

A Review of the Use of Large Datasets for Ascertainment of Cost Variation Across Institutions in Pediatric Medicine

James R. Shea, MD
Kimberly E. McHugh, MD, MSCR*

Address

*Medical University of South Carolina, 165 Ashley Avenue, MSC 915, Charleston, SC, 29425, USA
Email: mchughke@musc.edu

Published online: 17 October 2016
© Springer International Publishing AG 2016

This article is part of the Topical Collection on *Quality Improvement*

Keywords Datasets · Resource utilization · Cost variation

Opinion statement

Large datasets have been utilized frequently in the recent years to describe cost variation in pediatric medicine. Strengths of administrative datasets include their ability to provide cost data for a large nationwide sample in an economical manner. Limitations of these databases include the potential for coding errors and lack of clinical depth at the patient level. The linkage of administrative datasets with clinical registry or clinical trial data mitigates these shortcomings and provides a valuable method to accurately describe cost variation across institutions.

Introduction

The assessment of resource utilization is increasingly important in this era of rising health care costs. The value of health care has been defined as outcomes achieved relative to dollars spent [1], making methods to effectively measure the costs of our health care delivery of critical importance to considering value and quality among institutions [2].

Despite decades of conversation concerning quality of care, a wide variety of institutional practice habits persist.

As early as the 1990s, Wennberg unveiled a pattern of tremendous regional variation in medical practice in the USA [3]. His team detailed geographic variances in many areas ranging from physician supply to diagnostic testing to intervention rates and determined that different regions have a twofold variation in per capita Medicare spending. Of greater concern, the areas of increased expenditure did not confer improved outcomes for the patients [4]. Similar

examples seen in many individual medical subspecialties have led to a concerted turn toward value in the medical arena, resulting in the use of large datasets to examine institutional variation in both outcomes and costs in the practice of medicine. Table 1 details those databases referenced throughout this document.

Collaboration in pediatric clinical research

The challenges involved in the conduction of clinical research studies in children are well described. First, there are relatively few children with serious medical problems. For example, the American Cancer Society estimated 1.8 million new cancer cases in 2016, but only 10,380 involving patients aged 0–14 [5]. Secondly, there are unique ethical and regulatory restrictions that apply to the pediatric population [6]. These challenges have created a relative paucity of published data related to children compared to their adult counterparts [6].

The relative scarcity of potential participants makes creating a study that can be adequately powered to answer a given question particularly challenging. Campbell et al. reviewed randomized clinical trials in the *Archives of Diseases in Childhood* from 1982 to 1996 and reported that half of the studies had less than 40 enrollees [7]. To combat these challenges, some pediatric research collaborative networks developed. The Children's Oncology Group has been a model for decades in enrolling large numbers of member institutions to expand participant enrollment for trials in which single institution experience would be impossible to power. The Cystic Fibrosis Foundation formed a clinical trials network in 1998 that has served to generate trials for the development of novel therapies for cystic fibrosis [8]. The Pediatric Heart Network was established by the National Heart, Lung, and Blood Institute in 2001 and has been a robust source of collaboration for clinical trials in congenital and acquired pediatric heart disease since its inception.

Big data in pediatric medicine

The development of clinical trial networks has pushed forward multicenter evidence-based research in pediatrics. However, multicenter trials carry enormous logistical and financial challenges. They require significant resources to generate a trial that may be adequately powered to answer a single query.

The rise of electronic medical records has coincided with an increase in development of administrative databases. These data sets have the potential to aggregate hundreds of thousands of people, far outnumbering any feasible clinical trial. In addition, data entry does not require the time or expertise of clinical health care staff. These databases are relatively inexpensive and often efficient at answering a number of clinical questions.

These types of data can be useful when directed at problems where clinical trials are underpowered or underfunded. For example, they may uncover subtle effects that cannot be elucidated with a limited sample size. Administrative datasets provide information on a large number of unselected patients across the country, often providing a more objective data sample than single institution studies, which may be biased toward publication of favorable outcomes. The regional and size variation present in institutions contributing to administrative datasets may also provide a sample more representative of nationwide

Table 1. Referenced datasets to ascertain resource utilization across institutions in pediatric medicine

Database	Overview	Pros	Cons	Access
NIS—National Inpatient Sample	Largest publicly available all-payer in-patient health care database in the USA with data on >7 million hospitalizations each year. Contains sample of discharges from all participating hospitals from 44 states. Data available from 1988.	Data representing more than 95 % of the US population. Large sample size. Charge information for all patients, regardless of payer.	Does not provide state and hospital identifiers, so cannot link to hospital-level data. Administrative database has limited clinical depth.	Available for purchase at: www.hcup-us.ahrq.gov/nisoverview.jsp Online data use agreement
KID—Kid's Inpatient Database	Largest publicly available all-payer pediatric in-patient care database in the USA. It contains data from approximately three million pediatric discharges each year from 44 states. Released every 3 years, since 1977.	Large sample size allows for analyses of rare conditions in pediatrics. Charge information for all patients, regardless of payer	Does not provide state and hospital identifiers, so cannot link to hospital-level data. Administrative database has limited clinical depth.	Available for purchase at https://www.hcup-us.ahrq.gov/kidoverview.jsp Online data use agreement
PHIS—Pediatric Health Information System	Clinical and resource utilization data for in-patient, ambulatory surgery, emergency department, and observation unit for more than 45 children's hospitals. Now incorporating laboratory and radiology data.	Contains granular resource utilization and charge data. Unblinded hospital ID to allow linkage to hospital-level data.	Administrative database has limited clinical depth.	Free to member institutions. More information at https://www.childrenshospitals.org/Programs-and-Services/Data-Analytics-and-Research/Pediatric-Analytic-Solutions/Pediatric-Health-Information-System
Inpatient Essentials (formerly Case Mix)	Contains in-patient clinical and resource utilization data for 90 children's hospitals.	Allows benchmark with peers on casemix index, readmissions, length of stay, mortality, and costs/charges.	Administrative database has limited clinical depth.	Free to member institutions. More information at https://www.childrenshospitals.org/Programs-and-Services/Data-Analytics-and-Research/Pediatric-Analytic-Solutions/Inpatient-Essentials

Table 1. (Continued)

Database	Overview	Pros	Cons	Access
STS CHSD—Society of Thoracic Surgeons Congenital Heart Surgery Database	Longitudinal follow-up of clinical outcomes for patients who undergo cardiothoracic surgery at participating centers.	Blinded hospital ID allows linkage to hospital-level data. Adjudicated clinical and operative data. Allows comparison to peers for quality assessment.	Limited variety of disease processes. No cost or charge data.	Access limited to Participants (member surgeons and their institutions). More information at http://www.sts.org/national-database

practice, when compared to single center reports or clinical trials performed at a network of large academic centers.

It is important to recognize that administrative datasets are not without their limitations. Often, these include International Classification of Diseases Ninth Revision, Clinical Modification (ICD-9) codes, which may not be nearly as detailed or current as coding designed for specific areas of medical subspecialties. In addition, these datasets lack detailed patient-level information, limiting the ability to risk adjust for baseline patient characteristics. Administrative databases also generally do not allow the follow-up of individual patients over time. Finally, they are more prone to coding errors and inaccurate case ascertainment than data entry performed by clinical staff [9–11].

Examples of the use of datasets to describe institutional variation in pediatrics

There has been a proliferation of published pediatric research using large datasets in the past 3 years. For example, Sun and colleagues utilized the US Nationwide Inpatient Sample (NIS) to identify 12,512 patients undergoing tonsillectomy in 2009, finding a median cost of \$4393 and mean cost of \$7525. They found the need of mechanical ventilation had a marked effect on the total encounter cost [12]. Meier et al. studied hospital costs for same-day pediatric adenotonsillectomy surgery within a multihospital network. They identified 26,602 cases from 18 hospitals from 1998 to 2012, ultimately showing significant variation in costs at different facilities (range \$1029–\$2385/case) [13].

Rice-Townsend and colleagues have performed several studies utilizing the Pediatric Health Information System (PHIS) database to examine practice variation and costs in pediatric surgery. Utilizing a cohort of 13,328 patients with appendicitis from 34 children's hospitals from 2010 to 2011, they concluded that median hospital costs differed fourfold for patients with uncomplicated disease, suggesting significant variation in practice [14]. This group also examined 2544 patients with intussusception using PHIS and found significant practice and cost variation among hospitals for this procedure as well [15].

The PHIS database, run by the Children's Hospital Association (CHA), has also been employed by several other investigators to explore cost variation. Kharbanda et al. used the PHIS database to examine variation in resource utilization among pediatric emergency departments. Using diagnoses of asthma, gastroenteritis, or simple febrile seizure, they compared more than 250,000 Emergency Department (ED) visits at 21 institutions. Although practice and resource utilization varied across institutions, higher costs were not associated with lower rates of hospitalization or repeat ED visits [16]. In addition, Tieder et al. identified 24,890 admissions for diabetic ketoacidosis at 38 children's hospitals using the PHIS database. They compared resource utilization, length of stay, and readmission rates, finding a wide variance among institutional practices [17]. Finally, Brimley et al. evaluated the variation of costs and mortality among children with leukodystrophy in US children's hospitals using the PHIS database. They identified 122 patients and described a wide variety of costs at different institutions, although they did find a correlation between higher volumes of patients and increased cost efficiency [18].

Derrington and colleagues utilized the Massachusetts Pregnancy to Early Life Longitudinal Data System (PELL) to investigate the cost variation among different racial and ethnic backgrounds for children with Down syndrome. This was achieved with data collected on 504 children with Down syndrome and 468,000 in the control cohort. The study revealed higher costs in the birth hospitalization for both non-Hispanic black and Hispanic children when compared to non-Hispanic whites [19].

The Society of Thoracic Surgeons Congenital Heart Surgery Database (STS CHSD) is a voluntary registry that contains clinical outcome information on all congenital and pediatric cardiovascular operations performed at participating centers. Husain et al. accessed the STS CHSD to evaluate geographic variation among infants undergoing cardiac repair. They identified 23,379 patients in 94 centers with significant regional variation in all seven diagnostic groups examined [20]. The STS CHSD was also used by Jacobs et al. to describe variation in cost and outcomes among congenital heart centers [21]. They queried eight benchmark pediatric cardiac operations from performed from 2005 to 2009 and ultimately examined 18,375 index operations at 74 centers. Jacobs found significant interinstitutional variation in postoperative length of stay, which was most prominent for more complex operations.

Merging of clinical and administrative datasets

The linkage of clinical and administrative datasets has recently been utilized as a method to overcome the limited clinical details available in administrative datasets. Clinical registries and clinical trials contain detailed, adjudicated, clinical information on the patient level, whereas valuable cost data is only included in the administrative datasets. The method of linking datasets via indirect patient identifiers has been championed by Dr. Sara Pasquali in the field of congenital heart surgery. The linked clinical and administrative datasets allow for the examination of costs for congenital heart surgery, adjusted for important underlying preoperative risk factors [22].

For example, Pasquali et al. linked clinical data from the STS Congenital Heart Surgery Database with resource utilization data in PHIS to describe the cost variation for nine operations of differing complexity. A significant variation across centers in adjusted hospital costs per patient (up to ninefold) was observed for each operation. Differences in length of stay and complication rates explained 28 % of the between center variation and high-volume centers had lower costs for the most complex operations [23•].

Recently, the novel strategy of linking administrative datasets with clinical trial data has been explored in pediatric oncology and congenital heart surgery. The Children's Oncology Group (COG) merged clinical trial data from a phase III COG trial for patients with de novo acute myeloid leukemia at 43 centers with resource utilization data from PHIS using probabilistic matching. There was a 94 % success in linkage, and the standardized costs, blood product usage, and anti-infective exposures were described across centers [24].

Similar methods were employed to examine the impact of postoperative complications on costs for the Norwood operation. Detailed prospectively collected clinical information and postoperative complications data collected in the Pediatric Heart Network's Single Ventricle Reconstruction (SVR) Trial, a trial evaluating shunt types in patients with hypoplastic left heart syndrome, was linked at the patient level with cost data from the Case Mix administrative database. There was successful linkage of 98 % of eligible patient records, resulting in a study population of 334 patients. The adjusted hospital costs were demonstrated to increase with number of postoperative complications [25] and the hospital costs varied nearly fivefold across centers [26].

Conclusions

The assessment of resource utilization is increasingly important in this era of rising health care costs and among efforts to accurately ascribe value in medicine. The use of large administrative datasets is commonly employed to measure healthcare costs. The studies identified above represent the result of a search of studies using large datasets to examine cost variation published in a 3-year span from 2012 to 2015. A wealth of outcomes and comparative cost analyses have been acquired across the medical spectrum by investigating readily available data in administrative datasets.

Although administrative datasets can efficiently provide cost information on a large number of patients across centers, administrative data does have its limitations. These shortcomings are particularly related to data entry errors, limited ability to track patients over time, and absence of detailed clinical information to control for patient-level risk factors which can clearly affect hospital stay and costs. It is the authors' belief that most of these limitations can be mitigated with the linkage of clinical and administrative databases, resulting in valuable datasets that could not be established individually. Identification of practice variation utilizing these linked clinical and administrative datasets can help direct initiatives to both improve outcomes and reduce costs across hospitals, improving the value of pediatric medicine.

Compliance with Ethical Standards

Conflict of Interest

James R. Shea declares that he has no conflict of interest.

Kimberly E. McHugh reports some research reported in this publication was conducted under a Scholar Award from the Pediatric Heart Network supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health under Award Number U10HL068270. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Human and Animal Rights and Informed Consent

This article does not contain any studies with human or animal subjects performed by any of the authors.

References and Recommended Reading

Papers of particular interest, published recently, have been highlighted as:

- Of importance

- Porter ME. What is value in health care? *The New England journal of medicine*. 2010;363(26):2477–81.
- Berwick DM. Quality of health care. Part 5: payment by capitation and the quality of care. *The New England journal of medicine*. 1996;335(16):1227–31.
- Wennberg JE. Understanding geographic variations in health care delivery. *The New England journal of medicine*. 1999;340(1):52–3.
- E SJaF. Reflections on Geographic Variations in US Health Care 2010 [cited 2010]. Available from: http://www.dartmouthatlas.org/downloads/press/Skinner_Fisher_DA_05_10.pdf
- Society AC. Cancer Facts & Figures 2016. American Cancer Society: Atlanta, Georgia; 2016.
- The ethical conduct of clinical research involving children: The National Academies Press; 2004.
- Campbell H, Surry SA, Royle EM. A review of randomised controlled trials published in *Archives of Disease in Childhood* from 1982–96. *Archives of disease in childhood*. 1998;79(2):192–7. Pubmed Central PMCID: Pmc1717649, Epub 1998/11/03. eng.
- Goss CH, Mayer-Hamblett N, Kronmal RA, Ramsey BW. The cystic fibrosis therapeutics development network (CF TDN): a paradigm of a clinical trials network for genetic and orphan diseases. *Advanced drug delivery reviews*. 2002;54(11):1505–28. Epub 2002/11/30. eng.
- Gutgesell HP, Hillman DG, McHugh KE, Dean P, Matherne GP. Use of an administrative database to determine clinical management and outcomes in congenital heart disease. *World journal for pediatric & congenital heart surgery*. 2011;2(4):593–6. Pubmed Central PMCID: 4004178.
- Pasquali SK, He X, Jacobs JP, Jacobs ML, Gaies MG, Shah SS, et al. Measuring hospital performance in congenital heart surgery: administrative versus clinical registry data. *The Annals of thoracic surgery*. 2015;99(3):932–8. Pubmed Central PMCID: 4707956.
- Jantzen DW, He X, Jacobs JP, Jacobs ML, Gaies MG, Hall M, et al. The impact of differential case ascertainment in clinical registry versus administrative data on assessment of resource utilization in pediatric heart surgery. *World journal for pediatric & congenital heart surgery*. 2014;5(3):398–405. Pubmed Central PMCID: 4275407.
- Sun GH, Auger KA, Aliu O, Patrick SW, DeMonner S, Davis MM. Variation in inpatient tonsillectomy costs within and between US hospitals attributable to post-operative complications. *Medical care*. 2013;51(12):1048–54. Epub 2013/08/24. eng.
- Meier JD, Zhang Y, Greene TH, Curtis JL, Srivastava R. Variation in pediatric outpatient adenotonsillectomy costs in a multihospital network. *The Laryngoscope*. 2015;125(5):1215–20. Epub 2014/11/05. eng.
- Rice-Townsend S, Barnes JN, Hall M, Baxter JL, Rangel SJ. Variation in practice and resource utilization associated with the diagnosis and management of appendicitis at freestanding children's hospitals: implications for value-based comparative analysis. *Annals of surgery*. 2014;259(6):1228–34. Epub 2013/10/08. eng.
- Rice-Townsend S, Chen C, Barnes JN, Rangel SJ. Variation in practice patterns and resource utilization surrounding management of intussusception at freestanding Children's Hospitals. *Journal of pediatric surgery*. 2013;48(1):104–10. Epub 2013/01/22. eng.
- Kharbanda AB, Hall M, Shah SS, Freedman SB, Mistry RD, Macias CG, et al. Variation in resource utilization across a national sample of pediatric emergency departments. *The Journal of pediatrics*. 2013;163(1):230–6. Epub 2013/01/22. eng.
- Tieder JS, McLeod L, Keren R, Luan X, Localio R, Mahant S, et al. Variation in resource use and readmission for diabetic ketoacidosis in children's hospitals. *Pediatrics*. 2013;132(2):229–36. Epub 2013/07/24. eng.
- Brimley CJ, Lopez J, van Haren K, Wilkes J, Sheng X, Nelson C, et al. National variation in costs and mortality for leukodystrophy patients in US children's hospitals. *Pediatric neurology*. 2013;49(3):156–62. e1. Epub 2013/08/21. eng.
- Derrington TM, Kotelchuck M, Plummer K, Cabral H, Lin AE, Belanoff C, et al. Racial/ethnic differences in hospital use and cost among a statewide population of children with Down syndrome. *Research in developmental disabilities*. 2013;34(10):3276–87. Pubmed Central PMCID: Pmc4453874, Epub 2013/07/31. eng.
- Husain SA, Pasquali SK, Jacobs JP, Hill KD, Kim S, Kane LC, et al. Congenital heart operations performed in the first year of life: does geographic variation exist? *The Annals of thoracic surgery*. 2014;98(3):912–8. Pubmed Central PMCID: Pmc4527868, Epub 2014/07/20. eng.
- Jacobs JP, O'Brien SM, Pasquali SK, Jacobs ML, Lacour-Gayet FG, Tchervenkov CI, et al. Variation in outcomes for benchmark operations: an analysis of the Society of Thoracic Surgeons Congenital Heart Surgery Database. *The Annals of thoracic surgery*. 2011;92(6):2184–91. Pubmed Central PMCID: Pmc3263755, discussion 91–2. Epub 2011/11/26. eng.
- Pasquali SK, Jacobs JP, Shook GJ, O'Brien SM, Hall M, Jacobs ML, et al. Linking clinical registry data with administrative data using indirect identifiers: implementation and validation in the congenital heart surgery population. *American heart journal*. 2010;160(6):1099–104. Pubmed Central PMCID: 3011979.
- Pasquali SK, Jacobs ML, He X, Shah SS, Peterson ED, Hall M, et al. Variation in congenital heart surgery costs

across hospitals. *Pediatrics*. 2014;133(3):e553–60. Pubmed Central PMCID: Pmc3934342, Epub 2014/02/26. eng.

This article by Dr Pasquali et al utilizes a validated technique to create data linkages using indirect identifiers between clinical registries and cost information. The linkage of detailed patient-level data with administrative cost data allows for adjustment of patient risk factors in determination of costs in a very complex and heterogenous population.

24. Aplenc R, Fisher BT, Huang YS, Li Y, Alonzo TA, Gerbing RB, et al. 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. *Pharmacoepidemiology and drug safety*. 2012;21 Suppl 2:37–43. Pubmed Central PMCID: 3359580.

25. McHugh KE, Pasquali SK, Hall MA, Scheurer MA. Impact of postoperative complications on hospital costs following the Norwood operation. *Cardiology in the young*. 2015;30:1–7. Epub 2015/12/31. Eng.

26. McHugh KE, Pasquali SK, Hall MA, Scheurer MA. Factors impacting variation in cost across centers for patients undergoing the Norwood operation. *J Am Coll Cardiol*. 2016;67(13_S):932. doi:10.1016/S0735-1097(16)30933-0.