Administrative Data: Expanding the Infrastructure for Pediatric Research

Lilliam V. Ambroggio, PhD, MPH[1], and Samir S. Shah, MD, MSCE[1,2,3]

Research of pediatric medical conditions is challenging. There are far fewer affected children than adults for common diseases such as asthma and pneumonia, and studies of less common medical conditions cannot be conducted without involvement of multiple centers. Additionally, children who are eligible for inclusion in research studies are often too young to provide informed consent or even verbal assent, and investigators may have to obtain consent from multiple caregivers. Ethically, children are one of the most vulnerable research populations and additional considerations regarding their safety are taken into account during a study. Furthermore, measures of morbidity are usually more relevant for childhood illnesses as, unlike in adults, mortality is uncommon. However, morbidity can be difficult to quantify and many times cannot be compared directly across studies. All of these issues contribute to the high cost and complexity of conducting research in children.

Administrative databases offer a unique opportunity for researchers to assess healthcare at the population level while overcoming some of the challenges involved with research of pediatric conditions. Administrative data result from providing healthcare and reimbursing for services.[1] The primary aggregators of administrative data are the federal and state governments, private healthcare insurers, and healthcare networks. The content of administrative databases varies (**Table**) but often includes demographic characteristics, discharge diagnoses and procedure codes, and outcomes such as in-hospital mortality. Some administrative databases also include resource utilization information such as medications prescribed, radiologic imaging performed, and hospital charges and costs (usually derived from cost-to-charge ratios). Administrative databases permit identification of large cohorts of children retrospectively, making it far less expensive than identifying study cohorts prospectively. Administrative data include multiple years and multiple centers worth of data that represent healthcare in a real-world setting.[2] Statistical methods, such as propensity scores, applied to a large population sample in the analysis phase of a study can balance measured covariates within the cohort, ensuring that the treatment groups are comparable on average.[3] These methods are important in comparative effectiveness research, especially when a true randomized controlled trial is neither ethical nor feasible. Administrative data are particularly useful for the following types of study questions: (1) to determine which treatments or procedures are better[4]; (2) to estimate the incidence or prevalence of disease,[5] document trends of disease over time,[6] and hospital resource utilization for specific diseases[7]; (3) to identify associations between specific exposures and outcomes[8]; and (4) to provide insight into processes of care, errors of omission or commission, and appropriateness of care.[1]

Two articles recently published in *The Journal of Pediatrics* illustrate the use of administrative data in pediatric research. Puetz et al[5] used the Pediatric Health Information System (PHIS) to determine the prevalence, patterns and complications associated with administration of fresh frozen plasma (FFP). The PHIS database includes inpatient, observation, emergency department, and ambulatory surgery encounter information from more than 40 not-for-profit, free-standing children's hospitals throughout the US. Participating hospitals provide encounter data, including demographics, diagnoses, and procedures. Resource utilization data (eg, billing for medications, radiologic imaging, and laboratory testing) is also submitted to PHIS by most participating hospitals. Encrypted medical record numbers allow for identification of unique patients across multiple years within the same hospital.

The study by Puetz et al[5] was conducted in the context of recommendations to restrict FFP use as there is no proven benefit for many traditional indications, including correction of abnormal clotting tests, volume expansion, and prevention of intracranial hemorrhage in premature infants. The authors designed a retrospective cohort study that took advantage of 8 years of data from 40 children's hospitals that represented 3 252 149 admissions; 92 731 children (2.85%) received FFP. These numbers stand in stark contrast to the largest published pediatric randomized controlled trial, which included 776 children[9]; most studies of FFP have included fewer than 100 children.[5] Not surprisingly, FFP use varied across institutions, ranging from use in as few as 0.99% of admissions to use in as many as 5.84% of admissions. The study had 3 main findings. First, despite mounting evidence of lack of benefit of FFP for many clinical situations where FFP has traditionally been used, FFP use remained relatively constant over the 8-year study period.

FFP	Fresh frozen plasma
ICD-9-CM	*International Classification of Diseases, 9th Revision, Clinical Modification*
PHIS	Pediatric Health Information System

From the [1]Division of Hospital Medicine, and [2]Division of Infectious Diseases, Cincinnati Children's Hospital Medical Center, and [3]Department of Pediatrics, The University of Cincinnati College of Medicine, Cincinnati, OH

L.A. is funded by the National Research Service Award (T32HP10027-14). S.S. receives research support from the Children's Hospital Association for his role as an Executive Council member of the Pediatric Research in Inpatient Settings Network, PHIS+ (R01HS0986201; PI: R. Keren) from the Agency for Healthcare Research and Quality, and individual funding from the National Institutes of Health (R01HS0986201). The authors declare no conflicts of interest.

 http://dx.doi.org/10.1016/j.jpeds.2012.10.040

Table. Summary of databases commonly used in pediatric research

Variable	Kids' Inpatient Database (KID)	Nationwide Inpatient Sample (NIS)	National Ambulatory Medical Care Survey (NAMCS)	National Hospital Ambulatory Medical Care Survey (NHAMCS)	National Hospital Discharge Survey (NHDS)	PHIS	Discharge Abstract Database (DAD)
Sponsoring organization	Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality	Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality	National Center for Health Statistics, Centers for Disease Control and Prevention	National Center for Health Statistics, Centers for Disease Control and Prevention	National Center for Health Statistics, Centers for Disease Control and Prevention	Children's Hospital Association	Canadian Institute for Health Information
Visit types							
Outpatient (office-based)	…	…	X	X	…	…	…
Emergency room	X	X	…	X	X	X	X
Inpatient	X	X	…	…	X	X	X
First available dataset	1997	1988	1973	1992	1988	1992	1963
Frequency of data release	3 years	3 years	Yearly	Yearly	Yearly	Quarterly	Yearly
Number of ICD-9/ICD-10-CM discharge diagnosis codes (maximum)	15	25	3	3	7	41	25*
Number of ICD-9-CM/ICD-10-CM procedure codes (maximum)	15	15	…	…	…	41	20†
Medication data	…	…	X	X	…	X	…
Laboratory studies performed	…	…	X	X	…	X	…
Utilization of imaging studies	…	…	X	X	…	X	…

X, variable is present in the database; *ellipses* (…) variable is not available.
*ICD-10-CM codes for diseases, injuries, causes of death, injury and poisoning. These codes encompass the ICD-10-CM codes and are reported in place of ICD-10-CM codes in Canada.
†Canadian Classification of Health Interventions are the national standard for classifying health procedures in Canada and encompass ICD-10-CM codes.

Second, almost two-thirds of patients received FFP for indications without published evidence supporting its use. This does not imply that FFP transfusions are necessarily harmful to patients but rather that most are of unproven benefit. Third, venous or arterial thrombosis occurred in 15% of children receiving FFP. As the number of days receiving FFP per admission increased from 1 to ≥6, the rate of venous thrombosis increased from 8.0% to 20.0%, and the rate of arterial thrombosis increased from 3.4% to 12.6%. Although this association does not prove causation, these data have an important role in hypothesis generation. These findings, combined with recent data in adults demonstrating the harm of FFP in certain clinical situations, highlight the need for large, multicenter randomized trials examining the safety and efficacy of FFP use in children.

Administrative data can also substantively enhance primary data collection. Kirk et al linked data from 3361 5th-grade children participating in the 2003 Children's Lifestyle and School Performance Study with over 6 years of data from the Medical Services Insurance database and the Canadian Institute for Health Information Discharge Abstract Database in Nova Scotia[7]. The Medical Services Insurance database includes insurance records for any medical service (eg, hospital admissions, emergency room visits, and clinic visits) paid for by the Nova Scotia provincial health care system. The Canadian Institute for Health Information Discharge Abstract Database includes demographic, administrative and clinical data for hospital discharges across Canada.[10] Children's Lifestyle and School Performance Study is a population-based survey distributed to 5th grade children and their parents within Nova Scotia in 2003. This survey included a food frequency questionnaire, students' height and weight, information on sociodemographic factors, self-rated parental physical activity, parental diet, and frequency of the child's physical activity.

Kirk et al[7] sought to determine health care utilization among normal-weight, overweight, and obese fifth-grade children using the linked dataset. By age 14 years, obese children have 10% higher per capita physician costs than normal-weight children.[11] However, traditional obesity-related conditions such as hypertension did not contribute to the higher healthcare costs in obese children compared with their normal weight peers. Instead, overweight and obese children had significantly higher health care utilization for common pediatric conditions such as asthma and otitis media. This study highlights the need to focus on "non-traditional" consequences of obesity in children.

Studies that use administrative data must be interpreted in the context of limitations inherent in such databases. First, lack of detailed clinical information may limit inferences about individual patients. However, certain databases, such as PHIS, allow a participating hospital to re-identify their own patient population in order to augment their dataset with clinical information gleaned from the medical record. In addition, other databases such as Kids' Inpatient Database can be augmented with census tract information if deemed necessary to answer the research question at