Database Research for Fellows-in-Training and Early-Career Cardiologists

Tarek Alsaied, MD, Muhammad Shoaib Khan, MD, James Cnota, MD

I still remember meeting with 1 of my first-year cardiology fellows who was seeking a mentor and a research project. During that meeting, I felt his enthusiasm to have a well-designed research project that may affect the field of pediatric cardiology. I told him that although the time is limited in fellowship, there are always ways to do that. One way I recommended was using multi-institutional databases, which have become an important data resource in pediatric cardiology (1). We sought to provide a review of the potential research opportunities these databases may provide for fellows and early-career cardiologists.

Over the last 50 years, tremendous progress has been achieved in caring for children with congenital heart disease (CHD). Research in patients with CHD has never been easy, primarily due to its rarity and challenges with the complicated nature of the disease. Over the last 2 decades, the importance of collecting and reporting data for patients undergoing congenital heart surgery was recognized (1). This led many national organizations and congenital heart societies to start databases for the collection of outcomes data that eventually led to the creation of national benchmarks for congenital heart programs (2). The collection of data on a national scale addressed the problem of small sample sizes. The multi-institutional nature of the data helped in understanding institutional variations in treating different patients and adapting the most successful ones (1). In the past few years, there has been rapid growth in database research projects in CHD, due to the increased awareness of its usefulness and the massive evolution in information technology (2). Using databases has not only permitted tracking and comparison of outcomes, but has also led to quality improvement and meaningful collaboration between national and international centers (1,2).

Observations from database research may also generate hypotheses for further clinical trials (3). Database research is very valuable for fellows-in-training for multiple reasons. The availability of the raw data in a database may provide a time benefit and save you many steps in your research project. Databases also provide the sample size and statistical power that is very difficult to achieve with a single-center study (3). The data contributed by many centers result in generalizable results. For these reasons, approaching such databases with a good research question and sound analytic methodology will often lead to results that are accepted for presentation in scientific meetings—a goal for all fellows (4). Furthermore, as a good portion of the evidence-based practice in CHD comes from databases, this type of research will provide a good understanding of the source of evidence we have (5).

The most important aspect of having a successful research project using multi-institutional databases is finding a good mentor (4). As most institutions nowadays contribute to databases, you will likely be able to find a mentor at your institution with experience in databases. Having a mentor with some experience in database research is important, as your mentor will know what research questions a certain database will be able to answer and the limitations for the database you will be using. Your mentor also will guide you through the process of obtaining the data from the database. Having multiple mentors is helpful, and your mentors can be in different institutions (6). Even if your local mentors do not have experience in the database in which you are interested, they will be able to connect you with a mentor in a different department or institution who may help you in your project.
An important example of multi-institutional datasets in CHD is the pediatric heart network (PHN). The PHN was established in 2001 by the National Heart, Lung, and Blood Institute of the National Institutes of Health. It was created to help advance clinical research in CHD. Since it was established, the PHN has offered multiple educational and funding opportunities in clinical research for fellows-in-training and early-career pediatric cardiologists. Furthermore, the PHN has 3 large datasets for public use that are available free of charge (5).

Another important database registry is IMPACT (Improving Pediatric and Congenital Treatment), which is among the American College of Cardiology’s (ACC’s) National Cardiovascular Data Registries (1). The IMPACT registry collects pediatric cardiac catheterization data from >80 institutions across the United States (1,7). It has the potential to set the benchmarks in congenital interventional cardiology and supports powerful research studies (7). Interestingly, there were some abstracts presented as oral presentations by fellows-in-training at the 2015 ACC Scientific Sessions and later published in the Journal (8) from the IMPACT registry. A third success story is the NPC-QIC (National Pediatric Cardiology Quality Improvement Collaborative), which is the product of collaboration between multiple national societies, including the ACC and American Heart Association, to improve the care of children with CHD (9). The information from this database led to significant improvement in outcomes in children with single ventricle physiology (10). Multiple publications using this data were completed by early-career cardiologists and fellows-in-training (10).

When pursuing database research in CHD, try to be aware of the limitations, which include variability in terminology and nomenclature between different institutions and organizations. Furthermore, administrative databases utilize data that are collected for hospital billing purposes, which may limit the validity and reliability of data. High-quality clinical databases are now available, but they may also lack the granular details of individual cases. To overcome these challenges, it is crucial to have a good study design and to be aware of the limitations and confounding factors in the dataset (11). From a public health standpoint, the cost of maintaining these databases adds financial burden to the health care system. Although no current mechanisms for efficient, real-time data are available, there is an ongoing effort to link multiple datasets to each other and to the electronic health record systems to decrease cost and increase efficiency (12). These costs are outweighed by the knowledge gained, and as today’s fellows become increasingly familiar with the resource, we believe the collaborative efforts needed to maintain the databases will further benefit patients and families affected by CHD.

REPRINT REQUESTS AND CORRESPONDENCE: Dr. Tarek Alsaied, Pediatric Cardiology, The Heart Institute, Cincinnati Children’s Hospital Medical Center, 3333 Burnet Avenue MLC, Cincinnati, Ohio 45229. E-mail: tarek.alsaied2@cchmc.org.

REFERENCES
RESPONSE: An Important Clinical Question, Rather Than a Database, Should Spark Research

Sara K. Pasquali, MD, MHS
Department of Pediatrics and Communicable Diseases, C.S. Mott Children’s Hospital, and Institute for Healthcare Policy and Innovation, University of Michigan, Ann Arbor, Michigan
E-mail: pasquali@med.umich.edu

Over the past 2 decades, the number and variety of data sources available to medical researchers has increased exponentially, mirroring similar data trends across many other fields. In pediatric cardiovascular medicine, there are now numerous datasets capturing information that may facilitate research, including national clinical registries, multicenter research datasets, administrative databases, the electronic health record, and emerging data sources containing genetic and biomarker information, data streams from monitoring systems, and longitudinal follow-up data, such as neurodevelopmental outcomes (1).

There are several key considerations regarding the use of these datasets by trainees and junior faculty for research purposes, in addition to the important points raised by Alsaied and colleagues. First, it is important to note that “database research” is not a type of research. Rather, outcomes research, health services and policy research, and comparative-effectiveness research are examples of types of research that, by their nature, may be facilitated by or require the use of large datasets. For example, it is not possible to study the national landscape of care delivery and outcomes without using a large multicenter dataset. However, there may be other research questions not well suited to the use of an existing dataset. It is best to develop a research project from the perspective of “How do I answer this important question?” (which may or may not involve the use of a certain database) versus “What analysis could I perform with this dataset?”

In addition, it is critical to understand the strengths and limitations of different datasets to gauge the types of questions that may or may not be answered. There can be wide differences in standards regarding data definitions, data entry, data audits, level of missing data, or capture of pertinent variables to the area of study. For example, as noted by Alsaied and colleagues, the limited granularity of International Classification of Diseases-9 (and 10) codes with regard to congenital heart diagnoses and procedures is important to recognize when considering the use of administrative datasets.

From a methodological standpoint, it is important to remember that correlation does not imply causation and that not all associations found in a large dataset are necessarily clinically meaningful. Analytic guidance from a mentor with expertise in working with large datasets is critical, as this often requires specific knowledge of advanced methodologies to deal with confounders, differences in case-mix across centers, center effects, and other factors.

Overall, strong mentorship is critically important. This may involve reaching out not only to others outside of one’s institution, as described by Alsaied and colleagues, but also to those outside of the field of cardiology. Although many cardiology programs now participate in these datasets, there is a different skillset involved in understanding how to submit data to a database versus the use of large datasets for research. As described previously, this includes knowledge of the overall strengths and weaknesses of the dataset and appropriate methodologic and analytic techniques, in addition to an understanding of the clinical question. My own mentorship team has included several individuals outside of cardiology, including a bariatric surgeon, infectious disease specialist, and health economist, who have all brought different and important expertise in various aspects of health services and policy research. For those interested in pursuing this type of research as a major part of their career, obtaining formal master’s-level training in addition to practical guidance from a group of expert mentors is often necessary to position yourself for success.

Finally, newer database management and linkage methodologies may further expand the types of research possible and the range of questions that may be answered (2). For example, integrating continuous data streams generated by monitoring systems with clinical outcomes data may enable better prediction and treatment of adverse events in intensive care settings, and leveraging existing registry data may hold the potential to more efficiently power clinical trials (3). These and other developments noted by Alsaied and colleagues make it an exciting time for early-career cardiologists to get involved in this area of research.
REFERENCES

