Databases for Congenital Heart Defect Public Health Studies Across the Lifespan

Tiffany J. Riehle-Colarusso, MD, MPH; Lisa Bergersen, MD, MPH; Craig S. Broberg, MD, MCR; Cynthia H. Cassell, PhD; Darryl T. Gray, MD, ScD; Scott D. Grosse, PhD; Jeffrey P. Jacobs, MD; Marshall L. Jacobs, MD; Russell S. Kirby, PhD, MS; Lazaros Kochilas, MD, MSCR; Asha Krishnaswamy, BEE, MS; Ariane Marelli, MD, MPH; Sara K. Pasquali, MD, MHS; Thalia Wood, MPH; Matthew E. Oster, MD, MPH for the Congenital Heart Public Health Consortium*

In a 2012 meeting at the Centers for Disease Control and Prevention (CDC), key experts and stakeholders identified public health knowledge gaps about congenital heart defects (CHDs), namely prevalence of CHDs across the life span, long-term outcomes of persons with CHDs, and health services delivery for persons with CHDs. These gaps, and strategies to address them, formed the basis of a CHD public health science agenda. The strategies included leveraging information in existing databases to examine the epidemiology, health outcomes, and health service utilization of the CHD population. Many databases with CHD data exist and are managed by hospitals, specialty organizations, partnerships, and public health and other governmental entities. Researchers may be familiar with some databases but not others. Anyone planning studies to address public health knowledge gaps may benefit from an understanding of this complex constellation of databases.

The Congenital Heart Public Health Consortium (CHPHC) was formed in 2009 as a collaboration of stakeholders with its mission to prevent CHDs and improve outcomes for affected individuals. The CPHHC created a database workgroup to increase awareness of opportunities to contribute to the public health science agenda for CHDs using existing databases. The workgroup, consisting of experts in various disciplines (cardiologists, surgeons, epidemiologists, health service researchers), identified databases located in Canada or the United States (US) with information on CHDs from 1990 onward. The goals of this article are to provide an overview of database types and to list examples of databases that may be used to address CHD public health knowledge gaps. IRB approval was not deemed necessary for this review.

Database characteristics that may be important to consider when designing a study to address CHD public health knowledge gaps can be grouped into 3 main areas: (1) population included, (2) data content, and (3) accessibility. The first area relates to aspects such as sample size, inclusion criteria, whether the database is population-based, and whether persons are followed for a period of time. The second relates to what variables are included (eg, type and amount of clinical detail, information on resource utilization, or financial information), data collection mechanisms and coding, and data timeliness, accuracy, and completeness. The last area involves obtaining access to use the data, which may be costly, time consuming, or restricted, and will vary depending on the database selected.

Using existing data is often more cost effective and reasonable than gathering new data; however, research is
limited to the data that are available, and there is often no perfect data set to answer a particular question. Features of particular databases vary in importance, depending on the research question. One database’s strength in answering a question may be a limitation for another question. For example, a database may be population-based but have limited clinical detail. This database may be good for an overall prevalence estimate but not as useful for analyzing treatment outcomes of a particular CHD phenotype. It is the role of the researcher to determine which characteristics are most important and to find the appropriate database that will best inform the particular research question. This article does not comment on the strengths or weaknesses of specific databases but, rather, presents general information and additional resources. Researchers may use this information to help determine the utility of existing databases for their particular CHD public health study.

Database Categories and Examples

We grouped examples of databases into categories based on type of data source (administrative healthcare, birth defect surveillance, clinical, survey, and vital records). We briefly describe each category below, with a discussion of strengths and limitations to consider when addressing public health knowledge gaps. We also determined whether identified example databases had individuals with only CHDs (cardiac-specific databases) or had individuals with many conditions, including CHDs (general databases). Examples of cardiac-specific and general databases in each of these categories are listed in Tables 1 through 4. Some databases have more than one type of data source and are therefore listed in Table 5 under a separate combined category heading (eg, Administrative and Clinical). The tables provide a brief description of the database, sponsoring organization, years of data, and a URL link for further information. An asterisk denotes cardiac-specific databases. Although basic information is provided on a variety of databases, researchers are encouraged to contact database hosts for further information to assess their utility. Also, because databases are constantly evolving, other databases not captured in these tables may be useful in addressing a particular question.

Administrative Healthcare Databases

Administrative healthcare databases are generally developed from facility records or health insurance claims for billing purposes and/or to document healthcare provided; they are typically not designed for research purposes. Most are not specific to CHDs but still are useful for research and public health investigations related to CHDs. We identified 13 administrative healthcare databases (1 of which is cardiac specific) (Table 1), 2 administrative/clinical databases, and 4 administrative/survey databases (Table 5).

Facility-based administrative healthcare databases include all patients at a certain institution, regardless of payer, and may be able to identify a person over multiple encounters. However, these databases do not have data on outside resources utilized by that individual. Facility-based databases usually include the nominal charges for the services provided, although the provision of hospital or aggregated department-specific cost-to-charge ratios allows the estimation of facility-perspective costs. On the other hand, claims-derived administrative healthcare databases may either overrepresent sicker patients or exclude those without access to care. Another limitation of US administrative healthcare databases is how data are coded. Typically, these databases use International Classification of Disease version 9 or 10 Clinical Modification (ICD-9-CM, ICD-10-CM) codes, which often lack sufficient detail to adequately characterize specific CHD phenotypes or procedures. Hence, researchers may be limited to investigating broad classes of CHDs or procedures. Administrative databases may also be difficult to access, because of restrictions and license fees, and to use, due to their size and need for strong programmers or computational power.

One example of multi-institutional facility-based databases is the Healthcare Cost and Utilization Project (HCUP) database developed and managed by the Agency for Healthcare Research and Quality (AHRQ) through a public-private partnership. The cornerstone of HCUP is facility-level inpatient and hospital outpatient discharge data that include diagnoses and procedure codes, admission source, discharge status, patient demographics, expected payment source, total billed hospital charges, estimated costs, length of stay, and specific hospital characteristics. Hospitals provide these data on all patients, including self-pay and uninsured patients, to state-level entities that create state-specific hospital discharge databases. Under Memoranda of Agreements, these entities
Table 1. Administrative Healthcare Database Examples in the United States and Canada With Data From 1990 Onward for Potential Use in Congenital Heart Defects Public Health Investigations

<table>
<thead>
<tr>
<th>Name</th>
<th>Brief Description</th>
<th>Sponsoring Organization</th>
<th>Data Years</th>
<th>URL (Accessed as of June 1, 2016)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kids’ Inpatient Database (KID)</td>
<td>Weighted sample of the SID data (see below) used to identify, track, &amp; analyze national trends in pediatric inpatient healthcare; sampling weights help provide national estimates</td>
<td>Agency for Healthcare Research and Quality (AHRQ)</td>
<td>Every 3 years; 1997–present</td>
<td><a href="http://www.hcup-us.ahrq.gov/kidoverview.jsp">http://www.hcup-us.ahrq.gov/kidoverview.jsp</a></td>
</tr>
<tr>
<td>Nationwide Emergency Department Sample (NEDS)</td>
<td>Sampled from the SID and SEDD data (see below), is the largest all-payer emergency department (ED) database in the US used to create estimates of ED care</td>
<td>Agency for Healthcare Research and Quality (AHRQ)</td>
<td>2006–2013</td>
<td><a href="http://www.hcup-us.ahrq.gov/nedsoverview.jsp">http://www.hcup-us.ahrq.gov/nedsoverview.jsp</a></td>
</tr>
<tr>
<td>Nationwide/National Inpatient Sample (NIS)</td>
<td>Weighted sample of discharges from US community hospitals, which is the largest publicly available all-payer inpatient healthcare database in the US; sampling weights help provide national estimates</td>
<td>Agency for Healthcare Research and Quality (AHRQ)</td>
<td>1988–present</td>
<td><a href="http://www.hcup-us.ahrq.gov/nisoverview.jsp">http://www.hcup-us.ahrq.gov/nisoverview.jsp</a></td>
</tr>
<tr>
<td>Nationwide Readmission Database (NRD)</td>
<td>Sampled from the SID data (see below), used to create estimates of national readmission rates for all payers and the uninsured</td>
<td>Agency for Healthcare Research and Quality (AHRQ)</td>
<td>2013</td>
<td><a href="http://www.hcup-us.ahrq.gov/nrdoverview.jsp">http://www.hcup-us.ahrq.gov/nrdoverview.jsp</a></td>
</tr>
<tr>
<td>State Ambulatory Surgical and Service Databases (SASD)</td>
<td>Encounter data for ambulatory surgery &amp; other outpatient services from hospital-owned facilities; capture of hospital-based outpatient diagnostic and/or pediatric cardiac catheterization is variable, as are data content &amp; years; some states have nonhospital outpatient data</td>
<td>Agency for Healthcare Research and Quality (AHRQ)</td>
<td>1997–present</td>
<td><a href="http://www.hcup-us.ahrq.gov/sasdovery.jsp">http://www.hcup-us.ahrq.gov/sasdovery.jsp</a></td>
</tr>
<tr>
<td>State Emergency Department Databases (SEDD)</td>
<td>Discharge data on all ED visits in a given state that do not result in an admission; ED visits resulting in admissions are captured in the SID</td>
<td>Agency for Healthcare Research and Quality (AHRQ)</td>
<td>1999–present</td>
<td><a href="http://www.hcup-us.ahrq.gov/seedoverview.jsp">http://www.hcup-us.ahrq.gov/seedoverview.jsp</a></td>
</tr>
<tr>
<td>State Inpatient Databases (SID)</td>
<td>Inpatient discharge data from participating states used to identify, track, &amp; analyze state trends in healthcare utilization, access, charges, quality, and outcomes</td>
<td>Agency for Healthcare Research and Quality (AHRQ)</td>
<td>1990–present</td>
<td><a href="http://www.hcup-us.ahrq.gov/sidoverview.jsp">http://www.hcup-us.ahrq.gov/sidoverview.jsp</a></td>
</tr>
</tbody>
</table>

Continued
### Table 1. Continued

<table>
<thead>
<tr>
<th>Name</th>
<th>Brief Description</th>
<th>Sponsoring Organization</th>
<th>Data Years</th>
<th>URL (Accessed as of June 1, 2016)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrative—Others</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Healthcare Cost Institute (HCCI) Database</td>
<td>Collection of claims data by nonpartisan, nonprofit organization on over 50 million people with employer-sponsored insurance; annual reports published and data available to researchers to better understand determinants of US health care costs and utilization</td>
<td>Healthcare Cost Institute (HCCI)</td>
<td>2007–present</td>
<td><a href="http://www.healthcostinstitute.org/">http://www.healthcostinstitute.org/</a></td>
</tr>
<tr>
<td>MarketScan® Research Databases</td>
<td>Database system linking healthcare usage through commercial insurance, Medicaid, and Medicare to analyze a variety of outcomes</td>
<td>Truven Health Analytics</td>
<td>1995–present</td>
<td><a href="http://truenhealth.com/your-healthcare-focus/analytics-research/marketscan-research-databases">http://truenhealth.com/your-healthcare-focus/analytics-research/marketscan-research-databases</a></td>
</tr>
<tr>
<td>National Association of Children’s Hospitals (NACH)</td>
<td>Database of 95 children’s hospitals in the US with data to analyze inpatient populations, target quality improvement, enhance hospital utilization, &amp; support advocacy on behalf of children’s hospitals</td>
<td>National Children’s Hospital Association</td>
<td>2000–2012</td>
<td><a href="http://www.childrenshospitals.net/Am/Template.cfm?Section=Home3">http://www.childrenshospitals.net/Am/Template.cfm?Section=Home3</a></td>
</tr>
<tr>
<td>Pediatric Health Information System (PHIS)</td>
<td>Database of clinical and financial data from 49 tertiary-care pediatric hospitals in the US affiliated with the Children’s Hospital Association; data can be linked across encounters within the same hospital</td>
<td>Children’s Hospital Association</td>
<td>1992–present</td>
<td><a href="http://www.childrenshospitals.org/">http://www.childrenshospitals.org/</a></td>
</tr>
<tr>
<td>Quebec Congenital Heart Disease Database*</td>
<td>Database from 3 province-wide administrative databases, capturing demographics, diagnoses, procedures, and health services used throughout a patient’s life</td>
<td>McGill Adult Unit Congenital Heart Disease Excellence</td>
<td>1983–present</td>
<td>None</td>
</tr>
</tbody>
</table>

*Denotes cardiac-specific databases.
voluntarily share their files with AHRQ, and these files become part of HCUP. For 2013, the most current data year available, 48 states (accounting for 97% of the US population) participated in HCUP. The states decide which data elements are included in standardized State Inpatient Databases (SID) and whether AHRQ can release their files directly to users. For 2013, SID files for 28 states were available directly from AHRQ; files for the remaining states can potentially be obtained from the state-level organizations. Nationally representative databases based on aggregated SID data include the annual Nationwide/National Inpatient Sample (NIS) and the triennial Kids’ Inpatient Sample (KID). Other HCUP databases that capture CHD care are listed in Table 1. Copies of the HCUP databases can be purchased; aggregated data from select HCUP databases are freely available online at the HCUPnet site (http://hcupnet.ahrq.gov). Several health service research studies have used HCUP data to assess data on incidence, outcomes, facility costs, and factors related to hospitalization for individuals with CHDs.5-10

Health insurance claims databases include public insurers and proprietary insurance databases, such as Truven Health’s MarketScan® suite of databases. The MarketScan® research databases include commercial databases of employer-sponsored insurance, a Medicare database, and a Medicaid database representing claims from anonymized states that contract with Truven. MarketScan® data from 2005 were used to estimate health care use and costs for children with CHDs. Over 30 states have created, or are in the process of creating, all-payer claims databases (APCD) that combine claims from within their state from private and public payers. Some states have APCD data available on request, which could be useful in assessing resource utilization and healthcare costs for persons with CHD as well as surveillance of those with CHDs.

AHRQ has tools that states can use to improve quality of care for vulnerable populations. To help researchers answer specific health service questions, lists of databases with results for quality measures and databases from which measures could be calculated are available online (http://nhqrnet.ahrq.gov/inhqrdr/resources). This detailed compendium has information on over 100 databases and websites, including several listed in this article (eg, MarketScan®, HCUP, state APCDs, and Medical Expenditure Panel Survey [MEPS]), which can guide researchers to appropriate databases for a particular study question about CHDs.

### Birth Defects Surveillance

Surveillance of infants with birth defects is a core public health activity. Although the United States has no national birth defect...
### Table 3. Clinical Database Examples in the United States and Canada With Data From 1990 Onward for Potential Use in Congenital Heart Defects Public Health Investigations

<table>
<thead>
<tr>
<th>Name</th>
<th>Brief Description</th>
<th>Sponsoring Organization</th>
<th>Data Years</th>
<th>URL (Accessed as of June 1, 2016)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Congenital Cardiac Catheterization Outcomes Project (C3PO)*</td>
<td>Database of patient and procedural characteristics on catheterization procedures performed for congenital &amp; acquired heart disease in infants, children, and adults at 15 pediatric heart centers</td>
<td>Boston Children’s Hospital Cardiovascular Program</td>
<td>2007–present</td>
<td><a href="https://c3po-qi.chboston.org">https://c3po-qi.chboston.org</a></td>
</tr>
<tr>
<td>Congenital Cardiac Interventional Study Consortium (CCISC)*</td>
<td>Registry for demographic and procedural information on patients undergoing diagnostic &amp; interventional cardiac catheterizations for congenital heart disease (CHD)</td>
<td>Children’s Hospital of Michigan Foundation</td>
<td>2005–2010</td>
<td><a href="https://ccisc.med.wayne.edu/">https://ccisc.med.wayne.edu/</a></td>
</tr>
<tr>
<td>Congenital Evaluation, Reporting, and Tracking Endeavor (CONGENERATE)*</td>
<td>Database for providers of adults with CHD for multicenter collaboration, research, &amp; quality metric initiatives</td>
<td>McGill University, University of Sherbrooke, University of Montreal</td>
<td>2010–present</td>
<td><a href="http://www.congenerate.org/">http://www.congenerate.org/</a></td>
</tr>
<tr>
<td>Congenital Heart Surgeons Society (CHSS) Database*</td>
<td>Database for multi-institutional clinical studies evaluating surgical interventions for CHD. Goals: increase, correlate, &amp; disseminate knowledge of physiology, pathology and therapy</td>
<td>Congenital Heart Surgeons’ Society (CHSS)</td>
<td>1985–present</td>
<td><a href="http://www.chssdc.org/">http://www.chssdc.org/</a></td>
</tr>
<tr>
<td>IMProving Pediatric and Adult Congenital Treatments (IMPACT™) Registry*</td>
<td>Registry of demographics, management &amp; outcomes of pediatric and adult patients with CHD undergoing diagnostic &amp; intervention cardiac catheterizations and electrophysiology procedures at 55 sites; data for performance measurement, benchmarking, and quality improvement initiatives</td>
<td>American College of Cardiology/National Cardiovascular Data Registry</td>
<td>2010–present</td>
<td><a href="https://www.ncdr.com/webncdr/impact/home">https://www.ncdr.com/webncdr/impact/home</a></td>
</tr>
<tr>
<td>National Pediatric Cardiology Quality Improvement Collaborative*</td>
<td>Providers &amp; family network that collects data, conducts research, &amp; uses quality improvement science to improve outcomes; multicenter database to identify care variations &amp; best practices, and test hypotheses</td>
<td>Joint Council on Congenital Heart Disease</td>
<td>2006–present</td>
<td><a href="https://jchdqli.org/">https://jchdqli.org/</a></td>
</tr>
<tr>
<td>Organ Procurement Transplant Network Database</td>
<td>Database containing secure data on all wait lists, organ donation &amp; transplant events in the US; database can be queried online, and reports available</td>
<td>United Network for Organ Sharing (UNOS)</td>
<td>1987–present</td>
<td><a href="http://optn.transplant.hrsa.gov/">http://optn.transplant.hrsa.gov/</a></td>
</tr>
</tbody>
</table>

**Continued**

DOI: 10.1161/JAHA.116.004148

Journal of the American Heart Association
<table>
<thead>
<tr>
<th>Name</th>
<th>Brief Description</th>
<th>Sponsoring Organization</th>
<th>Data Years</th>
<th>URL (Accessed as of June 1, 2016)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pediatric Cardiac Care Consortium (PCCC)*</td>
<td>Registry of cardiac catheterizations, surgeries, &amp; autopsies for infants, children, and adults with congenital or acquired heart disease from 57 pediatric cardiac centers; includes outcomes &amp; longitudinal patient tracking</td>
<td>University of Minnesota</td>
<td>1982–2011</td>
<td><a href="http://www.pcccoweb.com">http://www.pcccoweb.com</a></td>
</tr>
<tr>
<td>Pediatric Cardiac Critical Care Consortium (PC4)*</td>
<td>Consortium of pediatric cardiac critical care, cardiac surgery, &amp; cardiology that collects data on outcomes &amp; practice, provides performance feedback, and promotes improvement based on empirical analysis and collaborative learning</td>
<td>National Institutes of Health/ University of Michigan/ Participating Sites</td>
<td>2009–present</td>
<td><a href="http://pc4quality.org/">http://pc4quality.org/</a></td>
</tr>
<tr>
<td>Pediatric Heart Transplant Study (PHTS) Database*</td>
<td>International, prospective, event driven database for research in the field of pediatric heart transplantation. PHTS advances the science &amp; treatment of children during listing for and following heart transplantation</td>
<td>University of Alabama/ Birmingham School of Medicine</td>
<td>1993–present</td>
<td><a href="http://www.uab.edu/medicine/phts/">http://www.uab.edu/medicine/phts/</a></td>
</tr>
<tr>
<td>Society of Thoracic Surgeons Congenital Heart Surgery Database (STS-CHSD)*</td>
<td>Database for quality improvement, patient safety, and research which contains data on &gt;95% of pediatric cardiac operations in the US. Represents 120 United States pediatric cardiac surgery hospitals &amp; 3 in Canada</td>
<td>Society of Thoracic Surgeons</td>
<td>1994–present</td>
<td><a href="http://www.sts.org/sts-national-database/database-managers/congenital-heart-surgery-database">http://www.sts.org/sts-national-database/database-managers/congenital-heart-surgery-database</a></td>
</tr>
<tr>
<td>Western Canadian Children's Heart Network Database (WCCHN)*</td>
<td>Database containing data on all diagnosed pediatric heart disease and CHD cases for 5 Canadian sites and adult CHD cases in 1 site</td>
<td>Western Canadian Children's Heart Network</td>
<td>2006–present</td>
<td><a href="http://www.westernchildrenshearntnetwork.ca/">http://www.westernchildrenshearntnetwork.ca/</a></td>
</tr>
</tbody>
</table>

* Denotes cardiac-specific databases.
surveillance system, most states maintain their own surveillance programs, which can vary by which entity conducts the surveillance (eg, health department), objectives, case ascertainment method, age of children included, or defects included. Surveillance data can be used for epidemiologic investigations or health services research. We presented 3 examples of birth defect surveillance databases (Table 2) and 3 in the combined category entitled Birth Defects Surveillance/Survey (Table 5). It is beyond the scope of this article to list all birth defect programs. However, a list of programs with links can be found at the National Birth Defect Prevention Network (NBDPN) website (http://www.nbdpn.org/state_programs_and_related_lin.php). Researchers should contact specific birth defect surveillance programs to explore opportunities to analyze the state’s data.

The strengths of birth defects surveillance databases are that they usually include a comprehensive, population-based birth cohort of infants with birth defects. The NBDPN was formed to address issues of surveillance, research, and prevention among US birth defect programs. The NBDPN has created surveillance guidelines to help standardize data collection. Recently, the NBDPN developed data quality measures and trilevel performance criteria focused on data completeness, timeliness, and accuracy to assess strengths and weaknesses of programs. This information will be used to develop and implement national data quality standards for birth defects surveillance. Many programs also use chart review to validate diagnoses, obtain data from several data sources, or use modified ICD-9-CM or ICD-10-CM codes, which are more specific for birth defects; thus, the data quality may be quite high. However, surveillance databases have varied methodologies, rarely have resource utilization or financial details unless linked to databases with that information, and usually do not have detailed clinical data.

Table 4. Survey Database Examples in the United States and Canada With Data From 1990 Onward for Potential Use in Congenital Heart Defects Public Health Investigations

<table>
<thead>
<tr>
<th>Name</th>
<th>Brief Description</th>
<th>Sponsoring Organization</th>
<th>Data Years</th>
<th>URL (Accessed as of June 1, 2016)</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Community Survey (ACS)</td>
<td>Part of the Decennial Census Program, it is a nationwide continuous survey sent to a small percentage of US households to gather demographic, housing, social, and economic data and provide yearly reports</td>
<td>United States Decennial Census Program—Census Bureau</td>
<td>2005–present</td>
<td><a href="http://www.census.gov/programs-surveys/acs/">http://www.census.gov/programs-surveys/acs/</a></td>
</tr>
<tr>
<td>Decennial Census</td>
<td>Survey of all US households done every 10 years, consisting of short and long forms. As of 2010, only the short-form is done—the long form replaced by the ACS. Data are used for numerous purposes</td>
<td>United States Census Bureau</td>
<td>1790–present</td>
<td><a href="https://www.census.gov/">https://www.census.gov/</a></td>
</tr>
<tr>
<td>Medical Expenditure Panel Survey (MEPS)</td>
<td>Survey of households to estimate use of health services, cost, payment, &amp; availability; surveys have 3 components: core household, insurance/employer, and the medical provider</td>
<td>Agency for Healthcare Research and Quality (AHRQ)</td>
<td>1996–present</td>
<td><a href="http://meps.ahrq.gov/mepsweb/">http://meps.ahrq.gov/mepsweb/</a></td>
</tr>
<tr>
<td>National Health Interview Survey (NHIS)</td>
<td>Survey of households to estimate the amount, distribution, &amp; effects of illness &amp; disability in the US across demographics and socioeconomic status; updated questions on select topics; main source of health information on the US population</td>
<td>Centers for Disease Control and Prevention—National Center for Health Statistics</td>
<td>1957–present</td>
<td><a href="http://www.cdc.gov/nchs/nhis.htm">http://www.cdc.gov/nchs/nhis.htm</a></td>
</tr>
<tr>
<td>National Survey on Children with Special Health Care Needs (NS-CSHCN)</td>
<td>Random sample survey of households in all states to assess prevalence &amp; impact of special healthcare needs among children in the US; survey has core &amp; special topic areas such as CHDs</td>
<td>Centers for Disease Control and Prevention—National Center for Health Statistics—Maternal Child Health Bureau</td>
<td>2000, 2005, 2009</td>
<td><a href="http://www.cdc.gov/nchs/slats/cshcn.htm">http://www.cdc.gov/nchs/slats/cshcn.htm</a></td>
</tr>
</tbody>
</table>
on treatment course, unless it is related to the diagnosis of the CHD. Furthermore, due to Health Insurance Portability and Accountability Act (HIPAA) regulations, access to identifiable data is restricted and governed by the birth defects program. Birth defects surveillance databases, unless linked to other databases to provide information beyond infancy, are not

### Table 5. Combined Database Examples in the United States and Canada With Data From 1990 Onward for Potential Use in Congenital Heart Defects Public Health Investigations

<table>
<thead>
<tr>
<th>Name</th>
<th>Brief Description</th>
<th>Sponsoring Organization</th>
<th>Data Years</th>
<th>URL (Accessed as of June 1, 2016)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthcare Systems Research Network (HCSRN)</td>
<td>Collaboration of 18 integrated healthcare delivery systems implementing research findings in clinical practice; working over a broad scope of indicators, they aim to develop an extensive and usable database</td>
<td>None</td>
<td>2006–present</td>
<td><a href="http://www.hcsrn.org/en/">http://www.hcsrn.org/en/</a></td>
</tr>
<tr>
<td>Pediatric Health Information System Plus (PHIS+)</td>
<td>Database augmenting the existing PHIS (see Table 1) database by linking electronic laboratory and radiology reports from 6 of the 49 Children’s Hospital Association hospitals to conduct clinical comparative effectiveness research projects</td>
<td>Agency for Healthcare Research and Quality (AHRQ) and Children’s Hospital Association</td>
<td>2009–2012</td>
<td><a href="http://www.prisnetwork.org/research/phis_plus.html">http://www.prisnetwork.org/research/phis_plus.html</a></td>
</tr>
<tr>
<td>Administrative and Clinical</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>National Ambulatory Medical Care Survey (NAMCS)</td>
<td>National sample survey of nonfederal office-based physicians to provide data on ambulatory medical care services in the US</td>
<td>CDC*—National Center for Health Statistics</td>
<td>1973–present</td>
<td><a href="http://www.cdc.gov/nchs/ahcd/about_ahcd.htm">http://www.cdc.gov/nchs/ahcd/about_ahcd.htm</a></td>
</tr>
<tr>
<td>National Hospital Ambulatory Medical Care Survey (NHAMCS)</td>
<td>National sample survey of hospital emergency, outpatient, hospital-based, &amp; nonhospital ambulatory surgery centers; provides data on care at hospital-based ambulatory services and ambulatory surgical centers</td>
<td>CDC*—National Center for Health Statistics</td>
<td>1992–present</td>
<td><a href="http://www.cdc.gov/nchs/ahcd/about_ahcd.htm">http://www.cdc.gov/nchs/ahcd/about_ahcd.htm</a></td>
</tr>
<tr>
<td>National Hospital Care Survey (NHCS)</td>
<td>Survey combining data from NHAMCS, NHDS, and drug abuse network</td>
<td>CDC*—National Center for Health Statistics</td>
<td>2011–present</td>
<td><a href="http://www.cdc.gov/nchs/nhcs.htm">http://www.cdc.gov/nchs/nhcs.htm</a></td>
</tr>
<tr>
<td>Administrative and Survey</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birth Defect Surveillance and Survey</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birth Defects Study To Evaluate Pregnancy exposureS (BD-STEPs)</td>
<td>Multisite population-based, case-control study of 17 birth defects, building on findings from the NBDPS (see below)</td>
<td>Centers for Birth Defects Research and Prevention</td>
<td>2014–present</td>
<td><a href="http://bdsteps.org/">http://bdsteps.org/</a></td>
</tr>
<tr>
<td>National Birth Defects Prevention Study (NBDPS)</td>
<td>Multisite population-based, case-control study of 30 birth defects; includes maternal interview &amp; cheek cell specimens from family members; excludes syndromes &amp; chromosomal abnormalities</td>
<td>Centers for Birth Defects Research and Prevention</td>
<td>1997–2011</td>
<td><a href="http://www.nbdps.org/">http://www.nbdps.org/</a></td>
</tr>
<tr>
<td>Pregnancy Health Interview Study (Birth Defects Study)</td>
<td>Multisite case-control study of birth defects &amp; newborn health; focuses on environmental exposures (primarily medications) in pregnancy; includes maternal interview, medical record release; genetic specimens 1992–2008</td>
<td>Slone Epidemiology Center at Boston University</td>
<td>1979–present</td>
<td><a href="http://www.bu.edu/slone/research/studies/phis/">http://www.bu.edu/slone/research/studies/phis/</a></td>
</tr>
</tbody>
</table>

*Centers for Disease Control and Prevention.

DOI: 10.1161/JAHA.116.004148
longitudinal. Although birth defects surveillance databases may not be able to address some clinical or outcomes questions, their strengths provide important information on the birth prevalence of CHDs.

One of the oldest birth defects surveillance programs is the Metropolitan Atlanta Congenital Defects Program (MACDP), maintained by the Centers for Disease Control and Prevention. Begun in 1967, MACDP collects information on birth defects in infants and children up to 6 years of age who were born to mothers residing in select metropolitan Atlanta counties. Cases are identified by trained abstractors who actively search newborn hospitals, pediatric hospitals, and other clinical sources, and cases are linked to vital records from the Georgia Department of Public Health. Records are reviewed, and those with a CHD diagnostic code are classified by physicians trained in pediatric cardiology, using standard clinical nomenclature derived from the Society of Thoracic Surgeons Congenital Heart Surgery Database (STS-CHSD). MACDP data on CHDs have been extensively analyzed, resulting in publications on trends in prevalence and survival, risk factors for CHDs, and a comparison of administrative and clinical coding for CHDs.

Birth defects surveillance programs monitor the CHD occurrence in their jurisdiction and contribute to CHD epidemiology. However, given the rarity of birth defects, there are often insufficient data in any one state to address some public health questions. The NBDPN also publishes pooled data from participating programs; in the 2012 annual report critical CHD surveillance data were highlighted, and the public health role in newborn screening for critical CHDs was discussed. There is also a data repository with data submitted by several states for infants with birth defects born 1999-2007, which has been used to study the association of race/ethnicity with birth defects, the survival of infants born with birth defects, and may be used to study other issues related to CHDs.

Clinical CHD Databases or Registries

Many databases with clinical information on persons with CHDs exist, including single- and multi-institutional databases as well as specialty care registries and research data sets. These databases vary in years of data collected, type of data, inclusion criteria, and purposes for utility. Research data sets may have uniquely different characteristics from clinical registries. Many clinical databases are designed to track patient outcomes, to improve quality of care, or for care benchmarking. However, since the early years of pediatric cardiac interventions, it was recognized that the experience of any single institution was limited, and collaboration between centers was necessary to have sufficient numbers to conduct meaningful outcomes analyses. In this article, we grouped examples of multi-institutional clinical data sets, specialty care registries, and research data sets in the “clinical” category. We identified 15 databases sourced primarily from clinical practice (13 cardiac-specific ones) (Table 3), and 2 administrative/clinical databases, sourced from a combination of large administrative healthcare databases combined with clinical practice data (Table 5).

The strength of clinical databases to address public health knowledge gaps lies in their detailed information on diagnosis, treatment, and clinical outcomes. Multi-institutional clinical databases usually amass a large sample size over time, with diversity in CHD phenotypes, patient characteristics, and geographic representation. Furthermore, clinical databases often use standard nomenclature and outcome measures, although the implementation of these standards may be inconsistent within or across institutions or databases, as recently documented. Clinical databases may also have information on comorbidities and noncardiac events, which is especially important for the older population. Clinical databases are useful, for example, when evaluating how clinical factors such as treatment or hospital course might influence the long-term outcomes of persons with a particular CHD phenotype. However, clinical databases may include only certain cohorts (eg, only persons with a specific diagnosis or undergoing a certain type of intervention), with little or no longitudinal follow-up of only limited outcome variables, may not be representative of the study population, and may not include resource utilization or financial data. Accessing the data may also require special approval or fee for access. These limitations may be important if a researcher is interested in an entire population or patient characteristics, which may not be consistently captured in clinical data (eg, birth information).

Efforts are ongoing to enhance and improve clinical databases for CHDs. The Multi-Societal Database Committee for Pediatric and Congenital Heart Disease was established in 2005 to provide infrastructure for collaboration among healthcare professionals interested in the outcomes of persons with CHDs. This committee is working to collaborate on use of common nomenclature, uniform core data set information, evaluation of case complexity, development of a mechanism for verifying case completeness and accuracy, and standardization of protocols for longitudinal follow-up of persons with CHDs. The outputs from this committee could help address not only questions related to treatment outcomes but public health questions as well.

One example of a large clinical database with geographical and diagnostic diversity is the STS-CHSD, founded in 1994 to support quality improvement in cardiothoracic surgery. As of December 31, 2015, STS-CHSD contains 394,980 operations reported from 124 pediatric and congenital heart surgery hospitals in the United States and 3 centers in...
Canada. With penetrance of over 95% in the United States, the data in STS-CHSD are representative of all US pediatric and congenital heart surgeries. Definitions of all terms and codes used in the STS-CHSD have been standardized and published, including the use of the International Pediatric and Congenital Cardiac Code (IPCCC). The STS-CHSD employs data quality measures and produces regular reports to better understand outcomes, provide benchmarks, and improve quality of care. Data from the STS-CHSD have also helped fill public health knowledge gaps. Application of the STS-CHSD nomenclature improved the quality of surveillance data for subsequent population-based analyses, eg, prevalence trends in CHDs, CHD survival, and receipt of special education by those with CHD. As with other clinical databases, aspects of the STS-CHSD may limit its utility to answer some public health questions (eg, access to care).

### Surveys

In surveys, individuals are usually sampled from a defined population and queried using a structured instrument (eg, telephone questionnaire) to generate information on a representative sample with respect to a target population of interest (eg, children <18 years of age). Data can be used to profile key issues in the population of individuals with CHDs to help set priorities for healthcare policy, develop programs, and improve services. The utility of survey data for answering CHD public health questions varies, depending on the survey design, sample composition and size, timeframe, and topics or questions included. In general, surveys that include persons with CHDs may be large overall (ie, a nationally representative sample) but may have a small number of total or specific CHD phenotypes, which may limit utility of the database. We identified several examples of databases with survey information that may be useful in public health studies of CHDs: 5 general survey databases (Table 4), 4 administrative/survey databases (Table 5), and 3 birth defect surveillance/survey databases (Table 5).

A strength of the identified surveys is that they ask the person or his or her proxy (eg, a parent) about a broad range of topics relevant to public health (ie, medical and nonmedical exposures, resource utilization, demographics, socioeconomic data, care coordination, continuity of care, barriers to care). Data important for understanding public health aspects of CHDs, such as self-reported information on quality of life or pregnancy exposures, may be available in survey data and not in other types of data sources. However, survey information is self-reported, often retrospective, and may have varying degrees of validity and recall bias. Data from surveys are typically cross-sectional—providing information about the population at one point in time—which may limit generalizability of research findings. Surveys typically lack identifiers that could otherwise be used for linking with other databases. Although the survey may be conducted repeatedly, it is usually on a different sample each time, as very few surveys recontact participants to obtain longitudinal data.

Two main sources of national population-based data are the Decennial Census and the American Community Survey (ACS). The Decennial Census has been conducted since 1790 as required by the US Constitution. Most households receive a short questionnaire, and prior to 2010 it was in every 6 households received a more detailed long questionnaire on socioeconomics. After 2000, the Census Bureau redesigned the census, and the socioeconomic questionnaire became the ACS. The ACS surveys households monthly and provides yearly information to communities in 1-, 3-, and 5-year reports. Data and tools to use the data from these surveys are publically available. The Census and ACS can be useful denominator and comparison data in studies of the CHD population. Furthermore, these data can be linked to other databases to study community-level factors influencing health and outcomes of persons with CHDs.

The National Survey of Children with Special Health Care Needs (NS-CSHCN) was a telephone survey sponsored by the federal Maternal and Child Health Bureau, designed to periodically sample the US population to identify children <18 years of age with special healthcare needs. trained interviewers asked the parent or guardian questions to identify all children in the household with special healthcare needs. It was administered 3 times between 2001 and 2010. In the 2009-2010 survey, CHDs were a specific condition prompt. Topics covered included child’s health and functional status, insurance coverage, access to healthcare, care coordination, and impact of health conditions on the child and the family. The survey is being integrated into the National Survey of Children’s Health but will still provide the same in-depth look at the lives of children with special healthcare needs. Survey strengths included that it was population-based and provided publicly available comparison data sets. It described the population of CSHCN and provided a snapshot of the impact of special healthcare needs. However, CHDs and treatment are not confirmed by a medical record source.

### Vital Records

The US vital records system is a federal-state partnership in which state vital records agencies receive federal funds for providing statistical data concerning vital events (live birth, death, and fetal death). Birth and death certificates enumerate all live births and deaths occurring in the United States and provide a comprehensive population-based cohort. Thus, vital records are important in CHD public health studies. Although all
states have vital records, data content varies slightly by state. The National Center for Health Statistics (NCHS) has promulgated national standard certificates that define the content and data elements. Researchers should contact the department of health in the particular states of interest to obtain information on available state-specific vital records databases.

Birth and death certificates contain protected personal identifiable information. However, NCHS has national, deidentified, publicly available data files (eg, birth, death, and period-linked birth-infant death data) useful for public health studies. For example, causes of death information from death certificates were used to describe annual CHD mortality in the United States by age, race, and sex. Period-linked birth-death data were used to identify racial differences in infant mortality due to birth defects such as CHDs. The NCHS also maintains the National Death Index (NDI), a restricted-access, centralized database of all state death records.

Although vital records data are useful, there are some limitations to consider. The quality of birth defects reporting on birth and fetal death certificates is generally poor and thus may influence the quality of a particular study. Researchers have identified limitations in ability to identify all decedents with a specific illness or health condition. The coding on birth or death records, or the checkboxes used on many birth/fetal death certificates, may not provide accurate or sufficient diagnostic details for some studies. Furthermore, birth and death certificates may use different coding systems. Death certificates have been coding underlying cause of death using ICD-10 since 1999, well ahead of clinical utilization of ICD-10-CM for billing purposes, which became official as of October 1, 2015. Resource utilization and cost/charge data are not presently reported in these documents. Finally, due to the personal identifying information, individual-level vital records are not easily accessible to general researchers and often must be linked at the health department or via the NDI.

Combining Databases Across Categories

Combining databases can maximize strengths and minimize limitations of individual databases to address issues in ways that may not be possible using a single database (Table 5). For example, linking data from a clinical database (STS-CHSD) with data from an administrative database (Pediatric Health Information System [PHIS]) has allowed multiple studies on healthcare utilization with robust clinical data to be conducted. Leveraging existing databases through linkage is also important to understand long-term and longitudinal outcomes for persons with CHDs. One example is the linkage of the Pediatric Cardiac Care Consortium (PCCC) with national registries. The PCCC contains data on patients who have undergone CHD interventions at 47 US centers between 1982 and 2011, with direct identifiers available for patients enrolled up to April 2003. The availability of direct identifiers allowed linkage of PCCC data with the NDI and the United Network for Organ Sharing (UNOS), thereby providing significant information regarding the long-term outcomes after palliative or corrective procedures. These linkages may address some of the individual database weaknesses regarding longer-term and longitudinal follow-up. Experts across disciplines agree that there needs to be a better mechanism for longitudinal follow-up of persons with CHDs across the life span. Longitudinal data can provide unique outcomes information. Restricted-access data files, such as NDI and the corresponding state-level records, may also be useful for other record-based linkage studies of persons with CHDs. Birth defects surveillance data have been linked to vital records to examine CHD prevalence and survival, and to longitudinal school records to investigate receipt of special education services among children with CHDs. Such population-based estimates are attainable only through linkage of multiple databases.

Throughout this article we have noted unique databases that span 2 database categories. However, databases from different categories have also been combined to form new stand-alone databases. One example of a database that spans 2 categories (ie, birth defect surveillance and surveys) is the National Birth Defects Prevention Study (NBDPS). The NBDPS is a multisite collaborative case-control study to evaluate potential genetic and environmental risk factors for major congenital malformations, including CHDs. Cases of CHDs are identified from birth defect surveillance data, and structured telephone interviews are conducted with mothers of cases and controls. Investigations using NBDPS data have contributed to understanding CHDs, including occurrence risk associated with maternal smoking, obesity, medication use, and descriptive epidemiologic studies of select CHDs. The strength of studies such as the NBDPS is that they are large, population-based, multicenter studies with standardized interview protocol, medical record review, and classification of CHDs. However, limitations exist, including potentially inaccurate or biased recall of exposures of interest due to self-report.

CDC recognized the possibilities for research and surveillance through linking data across various sources. In 2012, CDC awarded grants to the New York State Department of Health, Emory University in Atlanta, Georgia, and the Massachusetts Department of Public Health for a pilot study to develop population-based surveillance of adolescents and adults with CHDs. The grantees combined data within their states from a variety of data sources including birth defects surveillance data, Medicaid data, hospital discharge data, vital records, provider reports, and clinic billing data. As results are being analyzed from this pilot, a new collaborative study with 5 sites is expanding on this work.

Although examples of specific database combinations exist, a coordinated effort to use data for answering public
health questions concerning CHDs is lacking. The consolidation of heterogeneous datasets raises significant challenges related to confidentiality, governance, nomenclature and coding structure, and information technology capabilities. Even efforts at combining multi-institutional electronic health record (EHR) data on CHD have identified many obstacles. For example, there are inherent complexities of database interaction, such as nonstandard variable definitions or database structure. Furthermore, data are from disparate populations and different time points across the life span. Some represent a cross-section of the population, whereas others include only those patients seen in a specific healthcare setting or at the time of a specific event (such as surgery or cardiac intervention). Procedural data sets include far more clinical detail than administrative sources. The types of coding schemes used for each database vary, as well as the experience of the database manager or healthcare provider who selects the codes, both of which create inherent heterogeneity in the accuracy and granularity of the congenital diagnosis. Variables for accurate linkage between data sets may not be adequate, although this could be assisted through the use of a global unique identifier, as has been endorsed by the National Institutes of Health for other groups (https://ndar.nih.gov/tools_guid_tool.html). Furthermore, issues of HIPAA compliance may be raised because consent for data use in one database may not carry over to a conglomerate. To help address these challenges, in January 2015, the National Heart, Lung, and Blood Institute (NHLBI) convened a workgroup to develop a vision for an integrated data network for CHD research. The subsequent report summarizes the discussions and identifies critical elements as well as potential barriers for integrating CHD data.

Conclusion

There are numerous databases available to address public health knowledge gaps about CHDs across the life span. Databases can be grouped into broad categories with particular strengths and limitations. Understanding the relative characteristics of different databases is important for choosing the best data to answer a particular research question or to identify opportunities to maximize strengths and minimize limitations through database linkages.

Appendix

Participants in the Congenital Heart Public Health Consortium

Disclosures
None.

References
Databases for Congenital Heart Defects
Riehle-Colarusso et al


Key Words: congenital heart defects • databases • public health science
Databases for Congenital Heart Defect Public Health Studies Across the Lifespan

*J Am Heart Assoc.* 2016;5:e004148; originally published October 26, 2016;
doi: 10.1161/JAHA.116.004148

The *Journal of the American Heart Association* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Online ISSN: 2047-9980

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://jaha.ahajournals.org/content/5/11/e004148