



Published in final edited form as:

*Birth Defects Res.* 2018 November 15; 110(19): 1395–1403. doi:10.1002/bdr2.1400.

## Population-based surveillance of congenital heart defects among adolescents and adults: surveillance methodology

Jill Glidewell<sup>1</sup>, Wendy Book<sup>2</sup>, Cheryl Raskind-Hood<sup>2</sup>, Carol Hogue<sup>2</sup>, Julie E. Dunn<sup>3</sup>, Michelle Gurvitz<sup>4</sup>, Al Ozonoff<sup>4,5</sup>, Claire McGarry<sup>6</sup>, Alissa Van Zutphen<sup>6,7</sup>, George Lui<sup>8</sup>, Karrie Downing<sup>1,9</sup>, Tiffany Riehle-Colarusso<sup>1</sup>

<sup>1</sup>Centers for Disease Control and Prevention (CDC), National Center on Birth Defects and Developmental Disabilities, Atlanta, Georgia

<sup>2</sup>Emory University, Atlanta, Georgia

<sup>3</sup>Massachusetts Department of Public Health, Boston, Massachusetts

<sup>4</sup>Boston Children's Hospital, Boston, Massachusetts

<sup>5</sup>Harvard Medical School, Boston, Massachusetts

<sup>6</sup>New York State Department of Health, Albany, New York

<sup>7</sup>University at Albany School of Public Health, Rensselaer, New York

<sup>8</sup>Stanford University School of Medicine, Stanford, California

<sup>9</sup>Oak Ridge Institute for Science and Education, Oak Ridge, Tennessee

### Abstract

**Background:** Improved treatment of congenital heart defects (CHDs) has increased survival of persons with CHDs; however, no U.S. population-based systems exist to assess prevalence, healthcare utilization, or longer-term outcomes among adolescents and adults with CHDs.

**Methods:** Novel approaches identified individuals aged 11–64 years who received healthcare with ICD-9-CM codes for CHDs at three sites: Emory University in Atlanta, Georgia (EU), Massachusetts Department of Public Health (MA), New York State Department of Health (NY) between January 1, 2008 (2009 for MA) and December 31, 2010. Case-finding sources included outpatient clinics; Medicaid and other claims data; and hospital inpatient, outpatient, and emergency visit data. Supplemental information came from state vital records (EU, MA), and birth defects registries (EU, NY). Demographics and diagnostic and procedural codes were linked, de-duplicated, and shared in a de-identified dataset. Cases were categorized into one of five mutually exclusive CHD severity groups; non-cardiac comorbidity codes were grouped into broad categories.

---

**Correspondence:** Jill Glidewell, Centers for Disease Control and Prevention 4770 Buford Hwy, MS E-86 Chamblee, GA 30341. [jill.glidewell@cdc.hhs.gov](mailto:jill.glidewell@cdc.hhs.gov).

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

#### SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Results:** 73,112 individuals with CHD codes in healthcare encounters were identified. Primary data source type varied: clinics (EU, NY for adolescents), claims (MA), hospital (NY for adults). There was a high rate of missing data for some variables and data varied in format and quality. Some diagnostic codes had poor specificity for CHD ascertainment.

**Conclusions:** To our knowledge, this is the first population-based, multi-site CHD surveillance among adolescents and adults in the U.S. Identification of people living with CHDs through healthcare encounters using multiple data sources was feasible, though data quality varied and linkage/de-duplication was labor-intensive.

## Keywords

birth defects; birth defects registry; congenital heart defects; surveillance

## 1 | INTRODUCTION

Congenital heart defects (CHDs) affect about 1% of U.S. births and are a leading cause of birth defect-associated infant mortality, morbidity, and healthcare costs (Arth et al., 2017; Botto, Correa, & Erickson, 2001; Hoffman & Kaplan, 2002; Rosano, Botto, Botting, & Mastroiacovo, 2000; Simeone et al., 2015; Yang et al., 2006; Yoon et al., 1997). Improved treatment of CHDs has resulted in many individuals, even those with severe types of CHDs, living into adolescence and adulthood. (Gilboa et al., 2016; Hoffman, Kaplan, & Liberthson, 2004; Marelli et al., 2014; Marelli, Mackie, Ionescu-Ittu, Rahme, & Pilote, 2007; Marelli, Therrien, Mackie, Ionescu-Ittu, & Pilote, 2009; Pillutla, Shetty, & Foster, 2009; Warnes et al., 2001) Birth defects surveillance systems in 42 states estimate the birth prevalence of many types of CHDs. However, there are no existing population-based systems in the U.S. to estimate prevalence beyond early childhood, or to assess the healthcare utilization and longer-term outcomes among adolescents and adults with CHDs. Based on extrapolated Canadian data, an estimated one million children and 1.4 million adults in the U.S. were living with a CHD in 2010, of whom approximately 12%, or 289,000 had a severe CHD (Gilboa et al., 2016).

The Centers for Disease Control and Prevention (CDC) funded three sites (Emory University [EU] in Atlanta, Georgia; Massachusetts Department of Public Health [MA]; and New York State Department of Health [NY]) for a surveillance project to better understand the prevalence, healthcare utilization, and longer-term outcomes of adolescents and adults with CHDs. This article describes the surveillance methodology. The participating sites had varying populations and data sources, completeness, and linkage methodologies. Project elements that were consistent among sites are described first, followed by unique site-specific methodology.

## 2 | OVERALL PROJECT METHODOLOGY

### 2.1 | Target population

Eligible cases resided in site-specific catchment areas and were aged 11 through 64 years at any time between January 1, 2008 and December 31, 2010 (January 1, 2009 and December 31, 2010 for MA). Catchment areas were statewide (MA) or county-specific (EU, NY).

Eligible cases were presumed to be alive as of January 1, 2010, and had a healthcare encounter between January 1, 2008 and December 31, 2010 (January 1, 2009 and December 31, 2010 for MA) which was captured in one of the data sources. Each site determined vital status by linking to state death certificates, the absence of which presumed, but did not verify, living status.

## 2.2 | Surveillance case definition

Cases were identified using *International Classification of Disease, Ninth Revision, Clinical Modification* (ICD-9-CM) CHD diagnostic codes 745.xx–747.xx, excluding: congenital heart block (746.86), pulmonary arteriovenous malformation (747.32), absent/hypoplastic umbilical artery (747.5), other anomalies of peripheral vascular system (747.6x), and other specified anomalies of circulatory system (747.8x). For data sources using other coding systems, the sites converted all codes to ICD-9-CM.

The remaining eligible CHD codes were categorized into five hierarchical groups similar to Marelli et al. (2007), integrating both hemodynamic severity and basic anatomy: severe, shunt plus valve, shunt, valve, and other CHDs (Table 1). Several codes were classified as “minor” CHD (Table 1). Severe CHD was defined as a CHD usually requiring surgical or catheter intervention within the first year of life. For CHDs with variable severity (e.g., aortic stenosis), case-specific severity could not be distinguished using administrative ICD-9-CM coding. All unique CHD codes were reported for each case. Cases were classified into one of the five mutually exclusive CHD groups according to a severity algorithm (Figure 1). Study clinicians determined a set of minor codes (Table 1) which were nonspecific (e.g., 746.9 unspecified defect of heart), usually associated with another CHD (e.g., 747.3 anomalies of pulmonary artery), or which may be an incidental or non-congenital finding (e.g., valve insufficiency). Cases with isolated minor codes were reported; otherwise, minor codes were not utilized in case classification (Figure 1).

## 2.3 | Surveillance data elements

Cases could have multiple healthcare encounters in multiple data sources. For each healthcare encounter, the following information was collected, when available: demographics, type of data source, coding system, encounter date, encounter type, length of stay (for hospitalizations), all ICD-9-CM and current procedural terminology (CPT) codes (diagnoses and procedures), and death-related variables. All data elements were defined by the collaborative group and standardized across sites.

ICD-9-CM codes unrelated to CHD diagnoses and all CPT codes were grouped into categories of comorbidities and procedures using the clinical classification software (CCS) tool, developed by the Agency for Healthcare Research and Quality (Healthcare Cost and Utilization Project, 2016). Non-cardiac diagnostic CCS categories were collapsed into 24 broader comorbidity groups for this project (Supporting Information Table 1). The codes in the cardiac- and vascular-related procedural CCS categories were collapsed into the following project-specific cardiac procedure categories: cardiac procedures and surgeries, diagnostic imaging, and vascular procedures (Supporting Information Table 2). Procedural codes not specified in these categories were considered non-cardiac procedures. Cases were

reported as having a non-cardiac procedure (e.g., knee replacement [yes/no]); the specific non-cardiac procedural codes were not reported.

## 2.4 | Data collection, cleaning, and linkage

Data use agreements were initiated and managed by sites for each data source. Data sources provided data to each site based on the case definition and variable list. After a case was identified with an eligible CHD code, data from all available encounters during the surveillance time period were extracted. Data varied in format and content, often requiring extensive standardization and verification of eligibility. Values were checked for permissible ranges, and data from all sources were used to supplement missing data. Multiple case encounters were de-duplicated and linked within and across data sources. Further linkage details are described in the site-specific methodologies.

## 2.5 | CHD code validation

Studies have shown that ICD-9-CM codes for CHDs lack specificity and have poor diagnostic accuracy. (Cronk et al., 2003; Frohnert et al., 2005; Strickland et al., 2008; Rodriguez et al., 2018) For example, 745.5 codes for both atrial septal defect (ASD), a true CHD, and patent foramen ovale (PFO), a normal newborn condition. The PFO, which may persist in 25% of adults, is usually clinically insignificant and asymptomatic, but potentially associated with strokes. The 745.4 codes for ventricular septal defect, a common congenital condition which may also be acquired in adults after a myocardial infarction. To investigate the validity of specific codes for identifying a true CHD, clinicians at participating sites reviewed medical records from a random sample of cases. Each site had different sample size and code selection (Table 2) based on number of cases with codes of interest, availability of medical records, and clinical review resources. Cases sampled had the following isolated codes potentially used for non-CHD conditions: 745.5, 745.4 in persons >40 years old, 746.85 (coronary artery anomaly), and 746.89/746.9 (other specified/unspecified heart anomaly).

# 3 | SITE-SPECIFIC METHODOLOGY

## 3.1 | Emory University

EU collected data statewide, but only reported data from five metropolitan Atlanta counties: Clayton, Cobb, DeKalb, Fulton, and Gwinnett for the final dataset, which was more saturated with healthcare institutions providing care to CHD patients. The source population was approximately 2,404,907 individuals and was racially diverse. Cases were identified from statewide Georgia (GA) Medicaid claims and from six clinical data sources: Emory Healthcare (EHC) (two large and multiple smaller hospitals, with affiliated outpatient multispecialty practices), St. Joseph's Hospital, a large, urban two-county safety net healthcare system (inpatient and outpatient) serving a large portion of socioeconomically disadvantaged people, a children's hospital system (includes three hospitals and affiliated outpatient multispecialty practices), and two pediatric cardiology private practices that collectively provide more than 90% of pediatric cardiology care in Georgia. GA Vital Records and CDC's Metropolitan Atlanta Congenital Defects Program (MACDP) provided supporting information on cases that were identified elsewhere. GA birth certificates

provided demographic information, and death certificates confirmed deaths, identified deaths not found in other sources, and provided additional death information. Data de-duplication and linkage was performed within and across data sources by deterministic matching of identifiers [name, date of birth, gender, Social Security Number (SSN), medical record number]. All discrepancies were individually reviewed and resolved. Linkages were continually updated as new data were obtained.

### 3.2 | Massachusetts Department of Public Health

MA's catchment area was statewide with a racially diverse source population of approximately 4,818,851 individuals. The largest case-finding data source was the Massachusetts All-Payer Claims Database (MA APCD). MA APCD includes medical claims, member demographics, insurance information and provider information submitted by over 100 payers including commercial payers, third party administrators, and public programs (Medicare HMOs and MassHealth, Massachusetts' Medicaid program). The MA APCD does not include claims or information on individuals covered by Workers' Compensation, TRICARE or the Veterans Health Administration, the Federal Employees Health Benefit Plan, private insurers with fewer than 1,000 individuals covered, or uninsured individuals not enrolled in the Commonwealth's health safety net. Because of the high level of insurance coverage in MA (~96%), the MA APCD is estimated to include about 90% of state residents. (Massachusetts Health Insurance, 2013) MA APCD data were unavailable prior to 2009; therefore, MA collected 2009–2010 data for this project. Additional case finding sources were two large tertiary care referral centers serving the metropolitan Boston area and two smaller tertiary care centers in Central and Western Massachusetts. Identified cases were linked to MA electronic birth files, available from 1969 onward, and electronic death files. Records were de-duplicated within data source using a set of identifiers (last name, first name, date of birth, gender). Records were linked across data sources using FRIL's (Fine-Grained Records Integration and Linkage Tool), v. 2.1.5 (Jurczyk, Lu, Xiong, Cragan, & Correa, 2008).

### 3.3 | New York State Department of Health

NY's catchment area was 11 counties: Allegany, Bronx, Cattaraugus, Chautauqua, Erie, Genesee, Monroe, Niagara, Orleans, Westchester, and Wyoming. The source population was approximately 3,346,037 individuals, was racially diverse, and included both rural and metropolitan areas. A major case-finding and data source for cases of all ages in NY was the Statewide Planning and Research Cooperative System (SPARCS). SPARCS is a comprehensive, integrated information system available for healthcare resource planning, cost analysis, and surveillance of the state hospital services. Implemented by the New York State Department of Health (NYSDOH) in 1979, SPARCS receives hospital inpatient and outpatient discharge data from all facilities in NY. Adult cases (aged 20–64 years) were only identified through SPARCS; adolescent cases (aged 11–19 years) were also identified from seven pediatric cardiology healthcare facilities in and adjacent to the catchment area. These facilities provide a broad range of care for adolescents with CHDs, including cardiac surgeries and advanced treatment of complex CHDs. The NYSDOH's Congenital Malformations Registry (CMR), a state-mandated reporting system for children in NY with birth defects diagnosed under the age of 2 years, provided supporting information on

adolescent cases identified elsewhere. CMR data are routinely matched to birth and death certificate files using deterministic data linkage methods. NY was unable to link directly to vital statistics records. Only CMR data linked to death certificates and SPARCS discharge status were used to determine case vital status as of January 1, 2010. Records were de-duplicated within data source using all available identifiers (name, date of birth, address, SSN, medical record number, patient account number). The cases were then matched across data sources using FRIL v2.1.5 (Jurczyk et al., 2008) and SAS v9.3 (SAS Institute Inc., Cary, NC). Due to limited name information available in SPARCS data, the first two letters of the first name and the first two and last two letters of the last name were used for matching. A different matching weight was assigned for each variable, and matching was considered successful if 90% of information was matched, matching two data sources at a time. All variable and encounter data was combined with the unique linked and de-duplicated case.

### 3.4 | Surveillance dataset

Each site's Institutional Review Board approved sharing of de-identified data with the CDC. A Microsoft Access database was developed collaboratively with CDC and participating sites to house standardized data. De-identified, de-duplicated data, which combined and reconciled information from multiple data sources, were transmitted to CDC via a secure mechanism.

## 4 | RESULTS

Participating sites had varying populations and data sources, completeness, and linkage methodologies. However, all three sites linked and de-duplicated cases within and across multiple, existing datasets. A total of 73,112 unique individuals, aged 11–64 years, were identified who were presumed to be alive as of January 1, 2010, and who had an ICD-9-CM code for a CHD during healthcare encounters in the catchment areas (5,203 in EU, 62,605 in MA, 5,304 in NY). Overall, cases in MA were older (mean age = 37.1 years, median age = 38.0 years) compared to EU (mean = 30.5, median = 25.0) and NY (mean = 30.8, median = 22.0). Regarding case severity, shunts were the largest category at all sites. EU had a higher percentage of cases with “severe” CHDs and MA had a higher percentage of cases with “other” CHDs, relative to the other two sites (Figure 2). Subsequent papers will describe prevalence and characteristics of adolescents, adults, and pregnant women by surveillance site.

Each site had different amounts of cases found in each type of data source and across multiple data sources (Figure 3). The majority of EU and NY adolescent cases were found in clinics (75% and 81%, respectively) compared to only 10% of MA cases. NY did not use clinic data for adult case finding. Across all sites, <10% of cases were found in more than one case-finding data source. For example, of the 2,537 NY adolescent cases, only 42 (1.7%) were found in all three case-finding data sources. Conversely, 94% (EU), 93% (MA), 87% (NY adolescent) and 92% (NY adult) cases were found in a single data source (Figure 3).



Despite using multiple data sources for case-finding and supplemental information, sites had different amounts of missing demographic information (Table 2). Age, birth year, and gender were nearly always available; however, race was missing from 32% (EU), 48% (MA), and 20% (NY), body weight missing from approximately 90% of cases in all sites (Table 2). No data was available on employment status, education level, enrollment in special education, and receipt of social security income.

Results of site-specific case review for code validity are shown in Table 3. For a sample of cases identified with ICD-9-CM code 745.5, true CHD status varied from 24–59%. Thus, certain analyses from this project may report cases with isolated 745.5 separately due to the probability of misclassification. At EU, the majority of sampled cases with the other codes truly had a CHD. However, most cases sampled in NY with nonspecific or other specified codes (746.89/746.9) did not actually have a CHD (Table 3).

## 5 | DISCUSSION

This surveillance project used novel approaches to identify individuals who had healthcare encounters with ICD-9-CM codes for CHDs using multiple existing data sources. All sites were able to identify, link and de-duplicate cases within and across data sources, and collect demographics, clinical information, and healthcare utilization for characterization of adolescents and adults living with CHDs. Despite heterogeneity of surveillance methodology, there was similar distribution of case severity across the sites, which may be characteristic of the CHD population seeking healthcare or the entire CHD population. The case-finding methods used in MA identified many more individuals living with CHDs than the other sites. This may be result of differences in methodology, for example, coding practices; data sources and source population; healthcare access; geography; or other factors.

Data sources varied across sites in data quality, completeness, population, availability and accessibility. Even within a type (e.g., administrative claims) each site varied, depending on healthcare systems and policies. For example, APCD and SPARCS are not available in all states. Therefore, cases captured through healthcare encounters will vary based on data source availability in regions or states. There are advantages and disadvantages to all types of data sources. Advantages of clinical data are their detailed diagnostic, treatment, and clinical outcome information; limitations include capturing select cases, which may not be representative of the population, small sample size, and may not include healthcare utilization. Administrative claims data have large sample size and detailed healthcare utilization information; limitations include missing cases not receiving healthcare, coding systems that do not sufficiently specify types of CHDs or procedures, and potential inaccuracies in coding (Riehle-Colarusso et al., 2016).

There were several methodological strengths of this pilot surveillance project. First, this was the development of a case definition that utilized an existing hierarchical structure (Marelli et al., 2007) tailored to include the coding structure of the data sources. The case definition developed is specific to CHDs, excludes non-cardiac defects, and stratifies cases into mutually exclusive CHD severity categories. Another strength is the fact that the sites leveraged some data sources in which CHD diagnoses were already confirmed by clinical

experts in cardiology. Using data from these cardiac clinical care facilities increased the percentage of cases that were accurately coded. Finally, since 90% of cases were found in only one data source, it was clear that using multiple types of data sources increases the quantity and breadth of case ascertainment.

Data from this surveillance project should be viewed in light of several limitations. Persons with CHDs who did not have a healthcare encounter or have the CHD documented in their medical record during the target period were not identified. The likelihood of having a healthcare encounter could be affected by factors such as health insurance coverage, which differed considerably across the three sites. Second, limitations exist in using administrative ICD-9-CM codes for case inclusion. Although using clinical data sources might increase coding accuracy, variability in coding practices across the three geographic regions revealed differences in the number, choice, and consistency in code utilization. In addition, the use of ICD-9-CM codes may misclassify cases, for example, 745.5 is often used in adults to code for the normal anatomic variant, PFO, which would result in misclassification and an overestimation of cases with a CHD (Rodriguez et al., 2018), which would result in misclassification and an overestimation of cases with a CHD. Due to resource limitations, a uniform, extensive code validation across the three sites was not done.

There were several implementation challenges during this project. First, establishing partnerships and access to the data was time-consuming, taking nearly 18 months of the 3-year project period. Second, electronic health records were not uniformly implemented during the study period and data obtained varied in completeness and quality, leading to much time spent on cleaning for inclusion in the final dataset. Third, depending on the availability and completeness of matching variables, the linkage process was labor-intensive. Many demographic variables were missing, making de-duplicating individuals identified from multiple data sources challenging. Despite multiple data sources, basic variables such as race, height, and weight were missing at a significant rate at all sites. Finally, as national death records were not used to verify vital status, due to resource limitations, the absence of linkage to a state death record only presumes but does not verify that the individual was alive as of January 1, 2010.

This is the first step in gathering empiric data to understand the prevalence of CHDs beyond infancy in the United States, and in developing an approach to monitoring trends and regional variations. This surveillance project shows the importance of using multiple sources to conduct surveillance of individuals with CHDs. It is also important to note that individuals with an ICD-9-CM CHD-coded healthcare encounter are only a portion of the total population. During a predefined surveillance time period, eligible individuals with CHD may not seek health care or may not have a CHD recorded during a healthcare encounter. Having an estimate of the number and characteristics of adolescents and adults living with CHDs can provide important information to parents receiving a diagnosis of a CHD in their newborn child. This information may also be useful to public health professionals, policy-makers, and healthcare providers in planning resources and interventions for individuals living with CHDs.



## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

The authors would like to thank others involved in the project for their contributions: Bobby Lyles, Trenton Hoffman, Rusty Rodriguez, Marlene Anderka, Ami Bhatt, Charlotte Druschel, Marilyn Browne, Kristen Sommerhalter, and Daphne Hsu. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention. The authors have no conflicts of interest to declare.

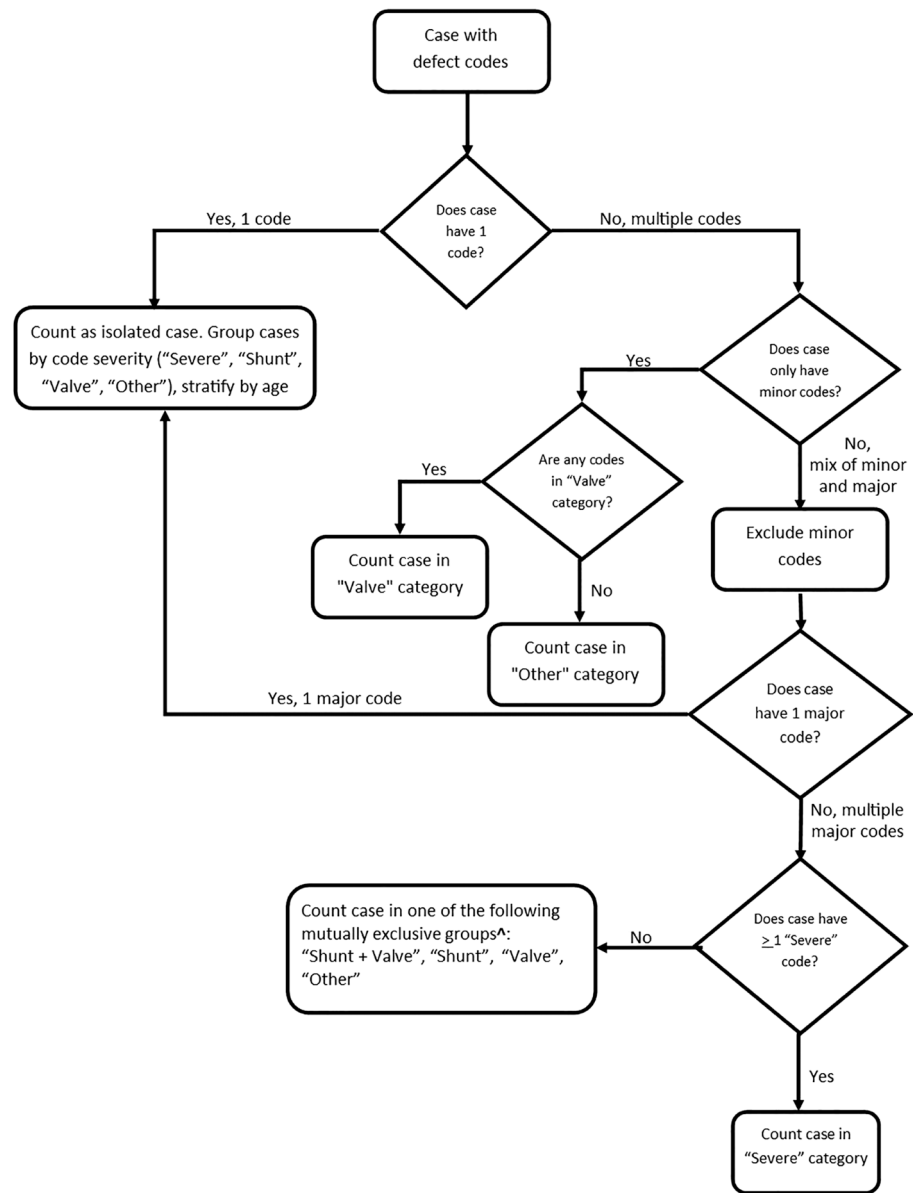
Funding information

Centers for Disease Control and Prevention, Grant/Award Number: CDC-RFA-DD12-1207

## REFERENCES

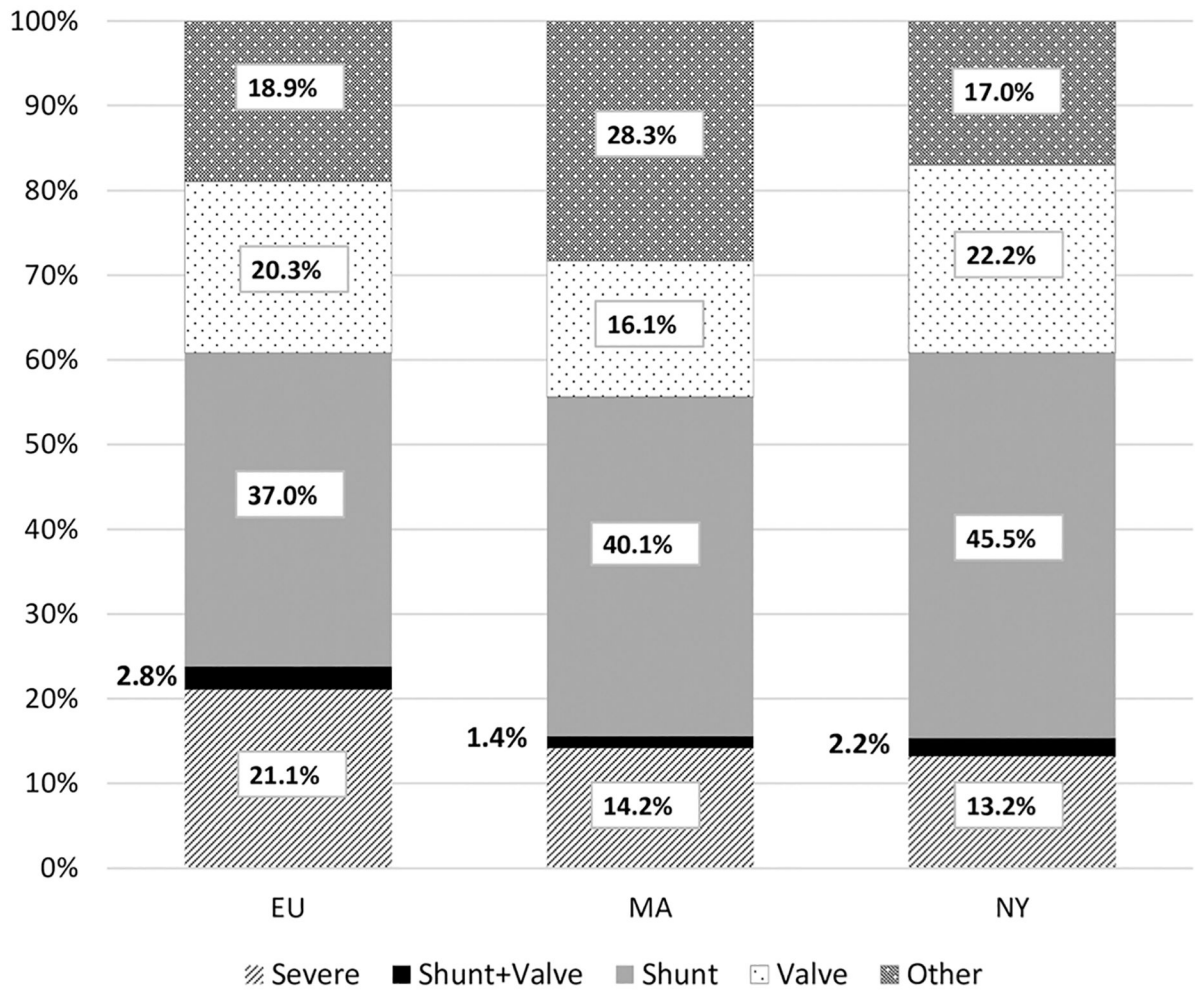
- Arth AC, Tinker SC, Simeone RM, Ailes EC, Cragan JD, & Grosse SD (2017). Inpatient hospitalization costs associated with birth defects among persons of all ages - United States, 2013. *MMWR. Morbidity and Mortality Weekly Report*, 66(2), 41–46. 10.15585/mmwr.mm6602a1 [PubMed: 28103210]
- Botto LD, Correa A, & Erickson JD (2001). Racial and temporal variations in the prevalence of heart defects. *Pediatrics*, 107(3), E32. [PubMed: 11230613]
- Cronk CE, Malloy ME, Pelech AN, Miller RE, Meyer SA, Cowell M, & McCarver DG (2003). Completeness of state administrative databases for surveillance of congenital heart disease. *Birth Defects Research. Part A, Clinical and Molecular Teratology*, 67(9), 597–603. 10.1002/bdra.10107 [PubMed: 14703780]
- Frohnert BK, Lussky RC, Alms MA, Mendelsohn NJ, Symonik DM, & Falken MC (2005). Validity of hospital discharge data for identifying infants with cardiac defects. *Journal of Perinatology*, 25(11), 737–742. 10.1038/sj.jp.7211382 [PubMed: 16163368]
- Gilboa SM, Devine OJ, Kucik JE, Oster ME, Riehle-Colarusso T, Nembhard WN, ... Marelli AJ (2016). Congenital heart defects in the United States: Estimating the magnitude of the affected population in 2010. *Circulation*, 134(2), 101–109. 10.1161/circulationaha.115.019307 [PubMed: 27382105]
- Cost Healthcare and Project Utilization. (2016). Agency for Healthcare Research and Quality. Clinical Classifications Software (CCS) for ICD-9-CM. Retrieved August 8, 2017, from <http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp>
- Hoffman JI, & Kaplan S (2002). The incidence of congenital heart disease. *Journal of the American College of Cardiology*, 39(12), 1890–1900. [PubMed: 12084585]
- Hoffman JI, Kaplan S, & Liberthson RR (2004). Prevalence of congenital heart disease. *American Heart Journal*, 147(3), 425–439. 10.1016/j.ahj.2003.05.003 [PubMed: 14999190]
- Jurczyk P, Lu JJ, Xiong L, Cragan JD, & Correa A (2008). Fine-grained record integration and linkage tool. *Birth Defects Research. Part A, Clinical and Molecular Teratology*, 82(11), 822–829. 10.1002/bdra.20521 [PubMed: 18985680]
- Marelli AJ, Ionescu-Ittu R, Mackie AS, Guo L, Dendukuri N, & Kaouache M (2014). Lifetime prevalence of congenital heart disease in the general population from 2000 to 2010. *Circulation*, 130(9), 749–756. 10.1161/circulationaha.113.008396 [PubMed: 24944314]
- Marelli AJ, Mackie AS, Ionescu-Ittu R, Rahme E, & Pilote L (2007). Congenital heart disease in the general population: Changing prevalence and age distribution. *Circulation*, 115(2), 163–172. 10.1161/circulationaha.106.627224 [PubMed: 17210844]
- Marelli AJ, Therrien J, Mackie AS, Ionescu-Ittu R, & Pilote L (2009). Planning the specialized care of adult congenital heart disease patients: From numbers to guidelines; an epidemiologic approach. *American Heart Journal*, 157(1), 1–8. 10.1016/j.ahj.2008.08.029 [PubMed: 19081390]

- Massachusetts Health Insurance and Employer Survey Chartbook: Updates for 2011. (2013). Retrieved September 20, 2017, 2017, from <http://chiamass.gov/assets/docs/r/pubs/13/mhischartpack-1-29-13.pdf>.
- Pillutla P, Shetty KD, & Foster E (2009). Mortality associated with adult congenital heart disease: Trends in the US population from 1979 to 2005. *American Heart Journal*, 158(5), 874–879. 10.1016/j.ahj.2009.08.014 [PubMed: 19853711]
- Riehle-Colarusso TJ, Bergensen L, Broberg CS, Cassell CH, Gray DT, Grosse SD, ... Oster ME (2016). Databases for congenital heart defect public health studies across the lifespan. *Journal of the American Heart Association*, 5(11), e004148 10.1161/JAHA.116.004148 [PubMed: 27912209]
- Rodriguez FH 3rd, Ephrem G, Gerardin JF, & Raskind-Hood C, Hogue C, Book W (2018). The 745.5 issue in code-based, adult congenital heart disease population studies: Relevance to current and future ICD-9-CM and ICD-10-CM studies. *Congenital Heart Disease*, 13(1), 59–64. [PubMed: 29266726]
- Rosano A, Botto LD, Botting B, & Mastroiacovo P (2000). Infant mortality and congenital anomalies from 1950 to 1994: An international perspective. *Journal of Epidemiology and Community Health*, 54(9), 660–666. [PubMed: 10942444]
- Simeone RM, Oster ME, Hobbs CA, Robbins JM, Collins RT, & Honein MA (2015). Population-based study of hospital costs for hospitalizations of infants, children, and adults with a congenital heart defect, Arkansas 2006 to 2011. *Birth Defects Research. Part A, Clinical and Molecular Teratology*, 103(9), 814–820. 10.1002/bdra.23379 [PubMed: 26069215]
- Strickland MJ, Riehle-Colarusso TJ, Jacobs JP, Reller MD, Mahle WT, Botto LD, ... Correa A (2008). The importance of nomenclature for congenital cardiac disease: Implications for research and evaluation. *Cardiology in the Young*, 18(Suppl 2), 92–100. 10.1017/S1047951108002515 [PubMed: 19063779]
- Warnes CA, Liberthson R, Danielson GK, Dore A, Harris L, Hoffman JI, ... Webb GD (2001). Task force 1: The changing profile of congenital heart disease in adult life. *Journal of the American College of Cardiology*, 37(5), 1170–1175. [PubMed: 11300418]
- Yang Q, Chen H, Correa A, Devine O, Mathews TJ, & Honein MA (2006). Racial differences in infant mortality attributable to birth defects in the United States, 1989–2002. *Birth Defects Research. Part A, Clinical and Molecular Teratology*, 76(10), 706–713. 10.1002/bdra.20308 [PubMed: 17022030]
- Yoon PW, Olney RS, Khoury MJ, Sappenfield WM, Chavez GF, & Taylor D (1997). Contribution of birth defects and genetic diseases to pediatric hospitalizations. A population-based study. *Archives of Pediatrics & Adolescent Medicine*, 151(11), 1096–1103. [PubMed: 9369870]

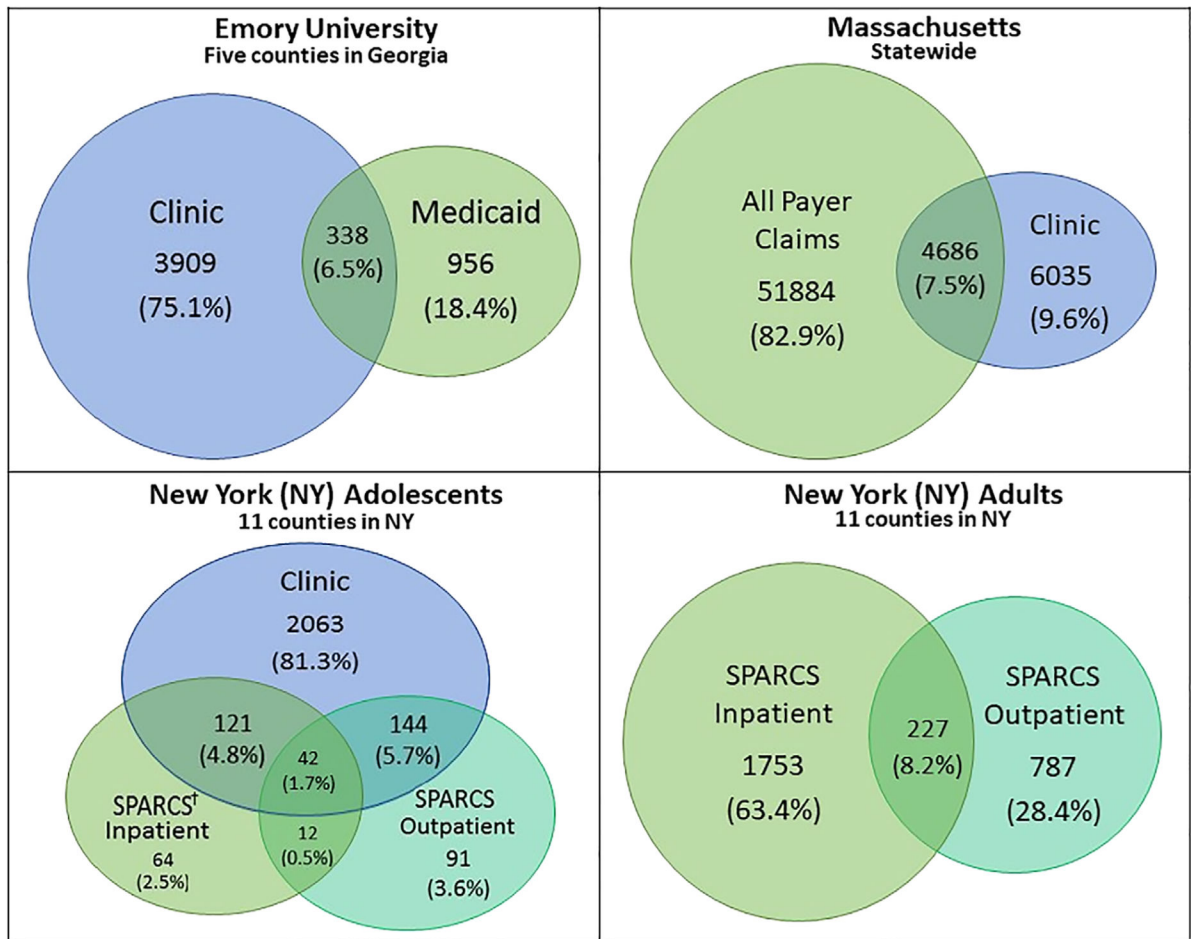


^ "Shunt+Valve" = codes in categories of Shunt + Valve +/- Other; "Shunt" = codes in categories of Shunt +/- Other; "Valve" = codes in categories of Valve +/- Other; "Other" = codes only in category of Other.

**FIGURE 1.**  
Congenital heart defect case classification algorithm



**FIGURE 2.**  
Percentage of cases by site and congenital heart defect severity group



**FIGURE 3.** Data sources and CHD case ascertainment by site. <sup>†</sup>Statewide Planning and Research Cooperative System

**TABLE 1**

Mutually exclusive congenital heart defect categories by ICD-9-CM<sup>a</sup> codes

Category	ICD-9-CM code	Code description
Severe <i>(case has a severe code, regardless of presence of shunt, valve, or other codes)</i>	745.0	Common truncus
	745.1	Transposition of the great arteries (TGA)
	745.10	Complete TGA (dextro-TGA), not otherwise specified (NOS) or classical
	745.11	Double outlet right ventricle, or incomplete TGA
	745.12	Corrected TGA (levo-TGA)
	745.19	TGA, other specified
	745.2	Tetralogy of Fallot
	745.3	Single ventricle, or cor triolculare
	745.6	Endocardial cushion defect
	745.60	Endocardial cushion defect unspecified
	745.61	Atrial septal defect (ASD) primum
	745.69	Endocardial cushion defect, other specified
	746.01	Pulmonary valve atresia or absence
	746.1	Tricuspid atresia, stenosis, or absence
	746.7	Hypoplastic left heart syndrome
	747.11	Interrupted aortic arch
	747.41	Total anomalous pulmonary venous return
Shunt + valve <i>(case has shunt AND valve codes)</i>	A combination of the shunt/valve codes below	A combination of the shunt/valve defects below
	745.4	Ventricular septal defect (VSD)
	745.5	ASD secundum or patent foramen ovale (PFO)
	745.8	Other specified defect of septal closure
	745.9	Unspecified defect of septal closure
	747.0	Patent ductus arteriosus (PDA)
	747.1/747.10	Coarctation of aorta
	746.0	Anomalies of pulmonary valve
Valve <i>(case has at least one valve code, no shunt or severe codes)</i>	746.00	Pulmonary valve anomaly, unspecified
	746.02	Pulmonary valve stenosis
	746.09	Pulmonary valve anomaly, other specified



Category	ICD-9-CM code	Code description
	746.2	Ebstein anomaly
	746.3	Aortic valve stenosis
	746.4 <sup>b</sup>	Aortic insufficiency or bicuspid/unicuspid aortic valve <sup>b</sup>
	746.5	Mitral stenosis or mitral valve abnormalities
	746.6 <sup>b</sup>	Mitral insufficiency <sup>b</sup>
	747.3 <sup>b</sup>	Anomalies of pulmonary artery <sup>b</sup>
	747.31	Pulmonary artery atresia, coarctation, or hypoplasia
	747.39	Anomalies of pulmonary artery, other specified
	745.7	Cor biloculare
	746.8	Other specified anomalies of heart
	746.81	Subaortic stenosis
	746.82	Cor triatriatum
	746.83	Infundibular or subvalvar pulmonary stenosis
	746.84	Obstructive anomalies of heart
	746.85	Coronary artery anomaly
	746.87	Malposition of heart or apex
	746.89 <sup>b</sup>	Other specified anomaly of heart (various types) <sup>b</sup>
	746.9 <sup>b</sup>	Unspecified defect of heart <sup>b</sup>
	747.2	Other anomalies of the aorta
	747.20	Anomalies of aorta, unspecified
	747.21	Anomaly of aortic arch
	747.22	Atresia or stenosis of aorta
	747.29	Other anomalies of aorta, other specified
	747.4	Anomalies of great veins
	747.40 <sup>b</sup>	Anomalies of great veins, unspecified <sup>b</sup>
	747.42	Partial anomalous venous return
	747.49	Other anomalies of great veins
	747.9 <sup>b</sup>	Unspecified anomalies of circulatory system <sup>b</sup>

Other only  
(case only has one or more codes in this category)

<sup>a</sup>International Classification of Disease, Ninth Revision, Clinical Modification.

<sup>b</sup>Codes considered to be minor. All other listed codes considered to be major.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**TABLE 2**

Demographic variables with percentage of data missing by site

Demographic variables	% data missing		
	Emory University	Massachusetts	New York
Age	0	0	0
Birth year	0	2.0	0
Gender	0	2.0	3.7
Race	32.1	48.0	19.8
Ethnicity	62.4	30.5	11.0
Height	69.4	91.6	100.0
Weight	87.5	90.8	90.5
Primary language	67.5	22.1	92.3
Marital status	28.2	96.6	87.0

**TABLE 3**

Code validation case reviews by ICD-9-CM<sup>a</sup> code and site

Code	Emory University		Massachusetts <sup>b</sup> (11–64 years)		New York <sup>b</sup> (11–64years)	
	N sampled	% true CHD	N sampled	% true CHD	N sampled	% true CHD
745.5 Isolated atrial septal defect or patent foramen ovale	11–21 years	81				
	22–40 years	32	123	42	27	59
	41–64 years	57				
	Total	170				
745.4 Isolated Ventricular septal defect	40–64 years	51	N/A	N/A	N/A	N/A
	Total	51				
746.85 Isolated Coronary artery anomaly	21–40 years	7				
	41–64 years	14	N/A	N/A	N/A	N/A
	Total	21				
746.9 Unspecified defect of heart	<40 years	12				
	40–64 years	9	N/A	N/A	20	35
	Total	21				
746.89 Isolated Other specified defect of heart	N/A	N/A	N/A	N/A	20	5
Total sampled cases		263	123		67	

<sup>a</sup>International Classification of Disease, Ninth Revision, Clinical Modification.

<sup>b</sup>MA and NY sampled cases across all age groups.