

1 *Circulation: Cardiovascular Quality and Outcomes*

2 **International Classification of Diseases (ICD) Codes for Congenital Heart Defects (CHD) Have**
3 **Variable and Limited Accuracy for Detecting CHD Cases**

4 Authors: Lindsey C. Ivey, MPH¹, Fred H. Rodriguez III, MD^{1,2}, Haoming Shi, BS³, Cohen Chong, BS⁴, Joy
5 Chen, MPH⁵, Cheryl Raskind-Hood, MS, MPH⁴, Karrie F. Downing, MPH⁶, Sherry L. Farr, PhD⁶, Wendy M.
6 Book, MD^{1,4}

7 **Affiliations:**

8 ¹Division of Cardiology, Emory University School of Medicine, Division of Cardiology, Atlanta, GA

9 ²Children's Healthcare of Atlanta Cardiology, Atlanta, GA

10 ³Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA

11 ⁴Emory University Rollins School of Public Health, Atlanta, GA

12 ⁵PicnicHealth, San Francisco, CA

13 ⁶National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta,
14 GA

15

16 Corresponding Author:

17 Wendy M. Book

18 1365 Clifton Rd. NE, 2nd floor Cardiology

19 Atlanta, GA 30322

20 Email: wbook@emory.edu

21 Author email addresses: Lindsey C. Ivey, lindsey.carrie.ivey@emory.edu; Fred H. Rodriguez III,

22 fred.rodriguez@emory.edu; Haoming Shi, haoming.shi@emory.edu; Cohen Chong,

23 cohenchong3712@outlook.com; Joy Chen, joywchen@berkeley.edu; Cheryl Raskind-Hood,

24 craskin@emory.edu; Karrie F. Downing, yyx9@cdc.gov; Sherry L. Farr, bwa0@cdc.gov

25

26 **Keywords: Congenital heart defects, outcomes research, epidemiology**

27

28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50

Disclaimer: The findings and conclusion in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention(CDC). The analysis was replicated at CDC by co-author Karrie Downing.

Author CRediT statement

Lindsey C. Ivey: conceptualization, methodology, investigation, resources, data curation, writing original draft and review and editing, analysis, supervision, project administration

Fred H. Rodriguez III: conceptualization, methodology, investigation, data curation, review and editing, supervision, project administration

Haoming Shi: data curation, analysis, writing review and editing

Cohen Chong: data curation

Joy Chen: data curation

Cheryl Raskind-Hood: conceptualization, methodology, analysis, writing review and editing

Karrie Downing: conceptualization, methodology, review and editing, project administration

Sherry Farr: conceptualization, methodology, review and editing, project administration

Wendy M. Book: conceptualization, methodology, investigation, data curation, writing original draft and review and editing, preparation, supervision, project administration, funding acquisition

51 **ABSTRACT (word count 300)**

52 **Background:** Administrative data permits analysis of large cohorts but relies on International
53 Classification of Diseases, Ninth and Tenth Revision, Clinical Modification (ICD) codes that may
54 not reflect true congenital heart defects (CHD).

55 **Methods:** 1497 cases with at least one encounter between 1/1/2010 – 12/31/2019 in two healthcare
56 systems (one adult, one pediatric) identified by at least one of 87 ICD CHD codes were validated
57 through chart review for the presence of CHD and CHD anatomic group.

58 **Results:** Inter- and intra-observer reliability averaged > 95%. Positive predictive value (PPV) of ICD
59 codes for CHD was 68.1% (1020/1497) overall, 94.6% (123/130) for cases identified in both
60 healthcare systems, 95.8% (249/260) for severe codes, 52.6% (370/703) for shunt codes, 75.9%
61 (243/320) for valve codes, 73.5% (119/162) for shunt and valve codes, and 75.0% (39/52) for “Other
62 CHD” (7 ICD codes). PPV for cases with >1 unique CHD code was 85.4% (503/589) vs. 56.3%
63 (498/884) for one CHD code. Of cases with secundum atrial septal defect ICD codes 745.5/Q21.1 in
64 isolation, 30.9% (123/398) had a confirmed CHD. Patent foramen ovale was present in 66.2%
65 (316/477) of false positives (FP). The median number of unique CHD-coded encounters was higher
66 for true positives (TP) than FP (2.0; interquartile range [IQR]: 1.0-3.0 vs 1.0; IQR:1.0-1.0,
67 respectively, $p < 0.0001$). TP had younger mean age at first encounter with a CHD code than FP (22.4
68 years vs 26.3 years, $p = 0.0017$).

69 **Conclusion:** PPV of CHD ICD codes varies by characteristics for detection of CHD by ICD code
70 and anatomic grouping. While an ICD code for severe CHD and/or the presence of a case in more
71 than one data source, regardless of anatomic group, is associated with higher PPV for CHD, most TP
72 cases did not have these characteristics. The development of algorithms to improve accuracy may
73 improve administrative data for CHD surveillance.

74

75 INTRODUCTION

76 The prevalence of congenital heart defects (CHD) beyond the first year of life varies across
77 the United States depending on the data source, case definition, and methodology used by
78 investigators¹⁻³. Public health surveillance to determine disease prevalence and long-term outcomes
79 for this population often relies on large administrative and clinical datasets from various sources
80 including electronic medical records (eMR) and billing data from hospital systems and government
81 agencies⁴⁻⁶. The reliability of CHD case detection in eMR and administrative data depends on the
82 accuracy of International Classification of Diseases, Ninth and Tenth Revision, Clinical
83 Modification (ICD-9-CM & ICD-10-CM) diagnosis code documentation in patient records⁷⁻⁹.
84 Identifying CHD cases by ICD codes may be prone to errors from misspecification, upcoding, rule-
85 out codes, and other issues that contribute to false positive (FP) and false negative classifications⁷.
86 CHD studies often use ICD-9-CM and ICD-10-CM code groups, 745.xx – 747.xx (63 codes) and
87 Q20.x – Q26.x (73 codes) respectively, to define a CHD cohort^{2,10}. Some of the codes in this group,
88 however, represent gastrointestinal, renal, spinal, extremity, umbilical artery, and other non-cardiac
89 vascular anomalies. ICD-10-CM relocated many non-cardiac vascular anomalies into groups Q27.x
90 and Q28.x, but non-specific CHD codes remain within Q20.x – Q26.x. Additionally, some codes
91 intended to represent CHD lack diagnostic specificity and have resulted in poor positive predictive
92 values (PPV) for true CHD identification^{2,9,11-15}. For example, ICD-9-CM code 745.5, which
93 represents both secundum atrial septal defect (ASD) and patent foramen ovale (PFO) had a PPV of
94 only 23.7% in an adolescent and adult population^{2,11}. Exclusion of code 745.5 and CHD codes
95 categorized as “other” increased PPV of CHD codes over 10 percentage points to 86.5%,¹⁶ but at the
96 expense of losing an estimated 24% of true CHD cases with code 745.5 in an administrative data set.
97 Differences in coding practices between administrative and clinical databases can lead to
98 differences in case ascertainment, and subsequently, prevalence and outcome estimations^{17,18}. A

99 hierarchal algorithm of 17 CHD subgroups found that the c-statistic for the code hierarchy alone was
100 0.79 (0.77–0.80) and improved to 0.89 (0.88–0.90) after age, encounter type, and provider type were
101 included in the algorithm^{9, 19}. Therefore, algorithms and hierarchies^{2, 19-21} accounting for limitations
102 of administrative coding are often applied to restrict a cohort to those individuals most likely to have
103 the condition of interest, at the expense of unintentional exclusion of true cases from the dataset.

104 Due to limitations in administrative data, restrictive case definitions may underestimate
105 patients with CHD; however, unrestrictive case definitions that include the broader code groups
106 745.xx – 747.xx and Q20.x – Q26.x may overestimate patients with CHD. To understand limitations
107 of and improve the quality of administrative data for CHD surveillance, we evaluated the PPV of
108 select ICD-9-CM and ICD-10-CM CHD codes in two data sources and compared features associated
109 with both true positive (TP) and FP cases.

110 **METHODS**

111 This study utilized data from two Georgia-based statewide multi-hospital and outpatient
112 tertiary healthcare systems — a pediatric healthcare system (PHS), and an adult healthcare system
113 (AHS). Data were collected under a Cooperative Agreement between Emory University and Centers
114 for Disease Control and Prevention (CDC-RFA-DD19-1202B). Study approval by the Emory
115 University Institutional Review Board (IRB) was granted on August 26, 2020 (IRB#
116 STUDY00001030) and included a complete waiver of HIPAA authorization as well as a waiver of
117 informed consent.

118 **Case Selection for Chart Abstraction**

119 The cohort identified for case validation consisted of individuals born between 1/1/1955 and
120 12/31/2019, who were identified in at least one of the two clinical data sources with a healthcare
121 encounter between 1/1/2010–12/31/2019 and with at least one of 87 selected CHD ICD-9-CM and

122 ICD-10-CM codes noted in Table 1. Forty-nine ICD codes in the CHD code groups 745.xx – 747.xx
123 and Q20.x – Q26.x (25 ICD-9-CM and 24 ICD-10-CM) were excluded due to lack of specificity for
124 CHD. We hereafter referred to those meeting the case definition as “cases” even if determined not to
125 have a CHD after validation. Because patent ductus arteriosus (PDA) can naturally resolve in
126 infancy, cases with a PDA code associated only with encounters in the first three months of life were
127 excluded from selection, unless they had other included CHD codes. From a cohort of 15,504
128 individuals identified in the AHS meeting inclusion criteria, 750 cases at least 18 years of age as of
129 1/1/2010 were randomly selected. From a cohort of 36,399 in the PHS who met inclusion criteria,
130 another 750 cases were randomly selected for chart abstraction. Encounter level CHD codes were
131 grouped based on native anatomy into severe, shunt and valve, only shunt, only valve, and other
132 CHD² (Table 1).

133 **Research Electronic Data Capture (REDCap) Database Development**

134 Study data were collected and managed using Research Electronic Data Capture (REDCap),
135 a secure, web-based application designed to support data capture for research studies.^{22, 23} Only
136 demographics were entered into the REDCap database to enable abstractors to locate the eMRs for
137 each case, with abstractors blinded to ICD codes. The REDCap database was developed, tested, and
138 revised using an existing dataset from the same facilities, with clinician input.

139 **Abstractor Training, Process, and Quality Review**

140 Two board certified CHD clinicians (WMB and FHR) trained the abstractors through an
141 iterative process to identify the correct CHD diagnosis for a case. The clinicians reviewed cases
142 weekly for accuracy. Inter- and intra-observer reliability between chart abstractors and CHD
143 clinicians were measured in three formal training rounds for 102 test cases, in an iterative process
144 using the 2010–2019 dataset. The same cases were randomly assigned to each abstractor and

145 clinician. Clinician determination based on chart abstraction was considered the gold standard. For
146 the training cases, inter-observer and intra-observer reliability averaged 95%. For each case,
147 abstractors determined *CHD Yes/No/Unknown* and the anatomic group based on review of text
148 clinical documents of any available records, including healthcare system-associated health network
149 records and scanned outside records. *Unknown CHD* was defined as the absence of clinical records
150 sufficient to determine if a case has a CHD, such as no encounter notes in the eMR. *Not CHD* was
151 defined in the medical record documentation as the absence of heart defects recorded in imaging or
152 provider notes. PFO, and other non-CHD cardiac diagnoses were noted for all *Not CHD* and
153 reviewed by clinicians. During the validation process, a PDA in isolation in an infant less than 3
154 months of age was not considered a CHD unless the PDA persisted after this age. Similarly,
155 peripheral branch pulmonary stenosis was noted in pediatric cases, but not considered a CHD in
156 infancy in the absence of other associated heart defects. All *Not CHD* and *Unknown CHD* cases
157 were reviewed by at least three abstractors and two board certified CHD clinicians, to make a final
158 determination. For cases with *Unknown CHD* status that remained unknown after review, the cases
159 were replaced with other randomly selected cases.

160 **Chart Abstraction and Case Validation**

161 Abstractors and clinicians reviewed 1500 cases meeting criteria for case selection, not
162 including the training cases and 19 *Unknown CHD* cases that were replaced (three cases had to be
163 excluded due to being pulled using a code that should have been excluded in isolation, ICD-9-CM
164 747.3 and ICD-10-CM Q25.6), creating a final data set of 1497 cases. We categorized cases into six
165 anatomic groups and developed rules to best determine when multiple conflicting groups were
166 possible. The CHD diagnosis was based on review of healthcare clinical notes, imaging notes,
167 operative reports, and scanned clinical notes from outside each institution. Billing ICD codes that are

168 sometimes found in the eMR were not used to confirm a diagnosis. As cases could potentially have
169 records in both PHS and AHS, chart abstractors reviewed medical records from both healthcare
170 systems during the case validation process. Cases flagged to review with the clinicians included all
171 cases with atrioventricular canal defects (partial, complete, or transitional), transposition of the great
172 arteries (TGA) (either dextro-TGA or levo-TGA), cases classified as shunt + valve, *Unknown CHD*,
173 *Not CHD*, or those with conflicting diagnoses. Initially, clinicians reviewed all *Not CHD, Yes PFO*
174 cases. Once consistency and accuracy were assured, the clinicians no longer reviewed cases clearly
175 understood to be PFOs by the abstractor. *Unknown CHD* cases with no medical information were
176 linked back to the data source to verify the presence of a CHD ICD code in the medical record as a
177 quality check. Forty percent of cases (n=593) underwent adjudication by two clinicians.
178 Additionally, female cases between 11–50 years-old at any time in the surveillance window were
179 checked for the presence of a pregnancy in the eMR and recorded in REDCap due to the possibility
180 that a CHD code associated with a pregnant women’s chart might reflect CHD in the fetus.

181 **Intra- and Inter-observer Reliability**

182 To monitor intra-observer reliability, 5% of each abstractor’s cases were randomly duplicated
183 and reviewed. To monitor inter-observer reliability throughout the study, 10% of cases were
184 randomly duplicated and assigned to the clinicians, regardless of initial abstractor assignment. Intra-
185 observer reliability ranged from 95–100% and inter-observer reliability ranged from 88–100%
186 (Table 2).

187 **Statistical Analysis**

188 In addition to data source, sex, and anatomic severity groups, we assessed encounter
189 characteristics including the number of unique CHD ICD codes documented per case, number of
190 unique encounters with a CHD ICD code documented per case, number of cardiac encounters per

191 case, number of imaging encounters apart from a cardiac encounter per case, and number of non-
192 cardiac encounters per case. Cardiac encounters were defined as either cardiac encounter type, an
193 encounter with a cardiologist provider type, or an encounter with a National Provider Identifier
194 (NPI) number with a cardiac specialty. Imaging encounters were defined as Current Procedural
195 Terminology (CPT) codes for cardiac imaging (echocardiograms, cardiac CT scans, and cardiac
196 MRIs; eAppendix A) and did not include those that were also a cardiac encounter. Non-cardiac
197 encounters were those that did not fall within the cardiac or imaging encounter definition.
198 Additionally, age in years at first qualifying encounter (FQE, defined as the first encounter with a
199 CHD code in the surveillance period) was collected for each case.

200 Percentages by data source, sex, anatomic severity groups, and median encounter
201 characteristics were compared for patients with true CHD and patients without CHD (FP, i.e., cases
202 determined not to have a CHD in chart abstraction) as well as for FP with PFO and without PFO
203 using Pearson's chi-square test of independence for the categorical characteristics, and the non-
204 parametric two-sample Kolmogorov-Smirnov test for encounter characteristics (chosen due to
205 heavily skewed distributions). In addition, group differences in median age in years at the FQE were
206 analyzed via t-test. PPV was calculated as percent of TP: $(TP) / [TP + FP]$. Statistical significance
207 was defined as $p\text{-value} < 0.05$.

208 RESULTS

209 Of the 1497 included validated cases with at least one of the 87 codes in ICD-9-CM code
210 group 745.xx – 747.xx and ICD-10-CM code group Q20.x – Q26.x (Table 1), 68.1% (n=1020) were
211 TP and 31.9% (n=477) were FP. Of the 1020 TP, 24.4% (n=249) had severe CHD, 11.7% (n=119)
212 had shunt and valve CHD, 36.3% (n=370) had shunt CHD, 23.8% (n=243) had valve CHD, and
213 3.8% (n=39) had “other CHD” or “other vascular” CHD. Among FP, 66.2% (316/477) were

214 confirmed to have a PFO. Of FP confirmed to have neither a CHD nor a PFO, 42.2% (68/161) had
215 other cardiac diagnoses, consisting of heritable and metabolic cardiomyopathies (n= 20), coronary
216 artery disease (n=14), non-congenital valvular heart disease (n=13), arrhythmias (n=11), heritable
217 aortopathies (n=5), pericardial disease (n=2), dextrocardia (n=1) and other cardiac diagnoses (n=2).

218 In Table 3, PPV of CHD ICD-9-CM and ICD-10-CM codes was 64.8% (481/742) in AHS
219 data, 66.6% (416/625) in PHS data, and 94.6% (123/130) for those individuals who had clinical
220 notes in both the AHS and PHS data. Males were as likely to be TP as females (p=0.2525).

221 Overall, the PPV for CHD among cases was 95.8% for severe codes (249/260), 73.5%
222 (119/162) for shunt and valve codes, 52.6% (370/703) for shunt codes, 75.9% (243/320) for valve
223 codes, and 75.0% (39/52) for “Other CHD” or “Other vascular” codes (Table 3). Just over a quarter
224 of the cohort, 26.9% (402/1497), had only 745.5 and/or Q21.1 (secundum ASD) shunt codes without
225 other CHD-related codes. Of these, 30.6% (123/402) were TP and 69.4% (279/402) were FP.
226 Overall, TP with only secundum ASD codes represented 12.1% (123/1020) of all TP.

227 The distribution of the number of unique CHD codes was significantly different for TP than
228 FP (Median: 2.0; interquartile range [IQR]: 1.0–3.0 vs. Median: 1.0; IQR:1.0–1.0, p<0.0001) (Table
229 3). The distribution of unique encounters with CHD codes was also significantly higher (p<0.0001)
230 for TP (3.0; IQR: 1.0–7.0) than FP (1.0; IQR:1.0–2.0). There was a significant difference between
231 the distribution of cardiac encounters for TP (2.0; IQR:1.0–5.0) and FP (1.0; IQR: 1.0–2.0),
232 p<0.0001. TP cases had a younger mean age at FQE than FP (22.4 years vs 26.3 years, p=0.0017).

233 Among FP, males were more likely to have a PFO than females (72.4% vs 60.7%,
234 p=0.0068). FP cases with codes in the shunt group, and specifically isolated 745.5 and/or Q21.1
235 codes, were more likely to have a PFO, whereas FP with codes in the valve group were less likely to
236 have a PFO (all p<0.0001) (Table 3). The median number of unique CHD ICD codes was similar

237 between FP cases with PFO and FP cases without PFO (both medians=1.0, $p=0.2662$). FP cases with
238 PFO did not have any differences in the distributions of unique encounters or non-cardiac
239 encounters. The mean age at FQE of FP cases with a PFO was lower than for FP cases without a
240 PFO (23.4 years vs 31.9 years, $p<0.0001$).

241 Table 4 shows PPV of the CHD-related ICD codes for individuals identified in AHS only
242 ($n=742$), PHS only ($n=602$), in both AHS and PHS ($n=129$), and among all cases with CHD coded
243 encounters in the surveillance period (i.e., AHS and/or PHS combined; $n=1473$). The presence of
244 more than one unique CHD ICD code resulted in higher PPV compared to those with only one
245 unique CHD code (85.4% vs. 56.3%, respectively; $p<0.0001$). When assessing PPV stratified by
246 date range, overall PPV differed slightly between 1/1/2010 to 9/30/2015 (ICD-9-CM coding era)
247 (75.3%; 579/769) and 10/1/2015 to 12/31/2019 (ICD-10-CM coding era) (70.9%; 742/1047; $p =$
248 0.0365). There was no statistically significant difference between cases that had CHD codes at
249 cardiac encounters (72.0%, 882/1225) and cases with codes at non-cardiac encounters (71.2%,
250 511/718) ($p=0.7056$). Cases with inpatient CHD codes had higher PPV (77.3%, 420/543) than cases
251 with only outpatient CHD codes (62.5%, 581/930) ($p<0.0001$); however, PPV improved to 80.4%
252 (333/414) when requiring at least 2 outpatient encounters with a CHD code separated by at least 30
253 days.

254 Of the 1497 cases, 71.2% (185/260) of cases with severe codes had severe CHD, 26.5%
255 (43/162) of cases with shunt and valve codes had shunt and valve CHD, 46.7% (328/703) of cases
256 with shunt codes had shunt CHD, 71.2% (230/320) of cases with valve codes had valve CHD, and
257 71.2% (37/52) with “Other CHD and Other vascular” codes had other CHD/other vascular CHD. For
258 those with shunt and valve codes, 32 of the 162 overall cases were confirmed to be shunt only, 40
259 were confirmed to be valve only, and 4 cases were neither shunt nor valve. Lastly, of the 46 women

260 with documented pregnancies during 2010–2019, 33 had a confirmed CHD resulting in a PPV of
261 71.7% for CHD ICD codes documented during pregnancy.

262 **DISCUSSION**

263 CHD represents a multitude of heart defects across a broad severity spectrum, so case
264 ascertainment presents a challenge in efforts to assess prevalence and outcomes of those living with
265 CHD. Administrative data can benefit CHD surveillance and research because it captures
266 information on a large scale from many patients across multiple care settings; however, the data have
267 variable and limited accuracy in detecting true CHD cases and may capture those with CHD-related
268 diagnosis codes who do not actually have CHD.

269 Study results demonstrated overall modest, but variable, PPV of the group of ICD-9-CM and
270 ICD-10-CM codes shown in Table 1 for detecting CHD cases after the exclusion of non-specific and
271 non-cardiovascular ICD codes from the case definition. Features associated with higher PPV include
272 identification of a case in more than one data source (PPV 94.6%), having CHD codes for severe
273 CHD anatomy (PPV 95.8%), and having CHD codes associated with an inpatient encounter (77.3%)
274 or 2 or more outpatient encounters (80.4%). However, some of these characteristics had relatively
275 low prevalence in the data; fewer than 10% of cases were found in both data sources, and those with
276 severe ICD codes made up only 17.4% of the dataset. Based on differences in distributions, TPs had
277 more unique CHD ICD codes, unique CHD-related encounters, CHD coded cardiac encounters, and
278 CHD coded non-cardiac encounters than FPs. A statistical difference in PPV was observed for cases
279 coded using ICD-9-CM vs. ICD-10-CM, suggesting a small loss in accuracy with the transition to
280 ICD-10-CM coding.

281 While PPV for CHD identified by ICD codes in administrative datasets could be increased by
282 constraining the dataset to inpatient encounters, multiple unique CHD codes, multiple encounters

283 with the CHD code, identification in more than one data source and/or to severe CHD codes, doing
284 so could lead to the exclusion of many cases with true CHD and the resulting sample would not be
285 representative of the larger CHD population. For example, requiring either an inpatient encounter
286 with a CHD code or at least two outpatient encounters with a CHD code separated by 30 days for
287 those without an inpatient encounter increased PPV from 68.0% to 80.4% in our sample, but resulted
288 in a loss of 33.1% of TP cases from the dataset. Furthermore, such restrictions can bias the sample
289 towards individuals with more severe CHD and/or more complicated clinical courses.

290 Similar to Rodriguez et al¹¹, the current analysis confirmed a low PPV of 30.9% for ICD-9-
291 CM and ICD-10-CM codes used for secundum ASD when occurring in the absence of other CHD
292 codes, and 84.6% of FP were likely assigned this code because they had a PFO. Therefore, CHD
293 studies may opt to exclude cases with isolated secundum ASD codes (745.5 or Q21.1) due to low
294 PPV²⁴, but they could be sacrificing more than 10% of true CHD cases from their sample. As of
295 10/01/2022, ASD and PFO have unique ICD-10-CM codes²⁵, making future analyses of
296 administrative data better able to identify true CHD cases and exclude those with PFO. However, as
297 PFO did not have a unique ICD-10-CM code until recently, this issue cannot be resolved by code
298 selection alone in studies including data prior to 10/01/2022, and thus will likely require algorithms
299 or clinician review to differentiate cases with secundum ASD from those with PFO. Understanding
300 features associated with TP cases captured with the secundum ASD codes and features associated
301 with FP cases captured with secundum ASD codes would aid in improving data quality of CHD
302 administrative datasets prior to 10/01/2022.

303 For the 33.8% of FP cases who did not have a PFO, we identified additional sources of
304 miscoding, which included the use of CHD codes for other cardiac conditions that may be heritable,
305 such Marfan or Brugada syndromes, heritable and metabolic cardiomyopathies, or non-congenital

306 valvular heart disease, but that are not considered structural congenital heart defects. FP cases in
307 women who are pregnant may result from CHD codes applied to the fetus rather than the pregnant
308 women, creating another potential source of miscoding.

309 We further identified variable and limited accuracy in detecting correct anatomic severity
310 group using ICD codes. Understanding features associated with TP versus FP cases and potential
311 sources of miscoding is an initial step in developing algorithms to detect CHD more accurately in
312 administrative records, prioritizing both PPV and sensitivity. Awareness of strengths and limitations
313 of administrative data in characterizing the CHD population will help interpret study results. Several
314 studies have found machine learning (ML) and natural language processing (NLP) to be promising
315 for detecting heart failure cases, lupus cases, and symptoms in hemodialysis patients²⁶⁻²⁹. ML and
316 NLP could potentially be applied to identify CHD cases more accurately in administrative datasets.³⁰

317 **Limitations and Strengths**

318 This ICD code validation study was strengthened by choosing specific ICD codes for the
319 case definition of CHD rather than entire code groups, the use of ICD-9-CM and ICD-10-CM codes,
320 high inter- and intra-observer reliability, and analysis of encounter-based features that are associated
321 with TP and FP. Conversely, this study was limited by the fact that administrative data from tertiary
322 referral center eMR are likely skewed towards more severe CHD given that individuals with severe
323 CHD are most likely to seek care and, as such, any of our estimates not stratified by severity may be
324 biased towards codes with a higher PPV. Tertiary referral centers may also code rare diseases with
325 greater accuracy regardless of case severity. Therefore, the applicability of findings to other data
326 sources, such as Medicaid data, may be limited. Additionally, ICD codes for CHD reflect native
327 anatomy and thus we were unable to validate use of these codes for post-operative anatomy.

328 **Conclusion**

329 Overall, ICD codes for CHD have modest PPV, which can be improved by requiring more
330 than one unique CHD code, multiple encounters and/or data sources, or excluding codes with low
331 PPV, but at the expense of lower sensitivity. Understanding the strengths and limitations of ICD
332 CHD codes will help interpret research findings derived from administrative sources. Development
333 of algorithms to detect CHD in administrative data may improve data quality.

334 **SOURCES OF FUNDING**

335 This work was funded by the Centers for Disease Control and Prevention (CDC), RFA DD-1902B.

336 **ACKNOWLEDGEMENTS**

337 This study was supported in part by National Center for Advancing Translational Sciences of the
338 National Institutes of Health under Award Number UL1TR002378 and the Medical Imaging,
339 Informatics, and Artificial Intelligence Core (MIIAI), which is supported by the Department of
340 Biomedical Informatics, Emory University School of Medicine. The content is solely the
341 responsibility of the authors and does not necessarily represent the official views of the National
342 Institutes of Health.

343 **DISCLOSURES**

344 The authors have no competing interests to disclose.

345

346

347 REFERENCES

- 348 1. Botto LD, Correa A, Erickson JD. Racial and temporal variations in the prevalence of heart defects.
349 *Pediatrics*. 2001;107:e32-e32
- 350 2. Glidewell J, Book W, Raskind-Hood C, Hogue C, Dunn JE, Gurvitz M, et al. Population-based
351 surveillance of congenital heart defects among adolescents and adults: Surveillance methodology.
352 *2018;110:1395-1403*
- 353 3. Hoffman Julien IE, Kaplan S. The incidence of congenital heart disease. *Journal of the American*
354 *College of Cardiology*. 2002;39:1890-1900
- 355 4. Virnig BA, McBean M. Administrative data for public health surveillance and planning. *Annual*
356 *Review of Public Health*. 2001;22:213-230
- 357 5. Egger JR, Lowe B, Johnson E, Durbin R, Griffin E, Dhakal S, et al. Evaluation of clinical and
358 administrative data to augment public health surveillance. *Online J Public Health Inform*. 2013;5:e30
- 359 6. Tomines A, Readhead H, Readhead A, Teutsch S. Applications of electronic health information in
360 public health: Uses, opportunities & barriers. *EGEMS (Wash DC)*. 2013;1:1019-1019
- 361 7. O'Malley KJ, Cook KF, Price MD, Wildes KR, Hurdle JF, Ashton CM. Measuring diagnoses: Icd code
362 accuracy. *Health Services Research*. 2005;40:1620-1639
- 363 8. Stausberg J, Lehmann N, Kaczmarek D, Stein M. Reliability of diagnoses coding with icd-10.
364 *International Journal of Medical Informatics*. 2008;77:50-57
- 365 9. Broberg C, McLarry J, Mitchell J, Winter C, Doberne J, Woods P, et al. Accuracy of administrative
366 data for detection and categorization of adult congenital heart disease patients from an electronic
367 medical record. *Pediatric Cardiology*. 2015;36:719-725
- 368 10. Duca LM, Pyle L, Khanna AD, Ong T, Kahn MG, DiGuseppi C, et al. Estimating the prevalence of
369 congenital heart disease among adolescents and adults in Colorado adjusted for incomplete case
370 ascertainment. *American Heart Journal*. 2020;221:95-105
- 371 11. Rodriguez III FH, Ephrem G, Gerardin JF, Raskind-Hood C, Hogue C, Book W. The 745.5 issue in code-
372 based, adult congenital heart disease population studies: Relevance to current and future icd-9-cm
373 and icd-10-cm studies. 2018;13:59-64
- 374 12. Strickland MJ, Riehle-Colarusso TJ, Jacobs JP, Reller MD, Mahle WT, Botto LD, et al. The importance
375 of nomenclature for congenital cardiac disease: Implications for research and evaluation. *Cardiology*
376 *in the young*. 2008;18 Suppl 2:92-100
- 377 13. Frohnert BK, Lussky RC, Alms MA, Mendelsohn NJ, Symonik DM, Falken MC. Validity of hospital
378 discharge data for identifying infants with cardiac defects. *Journal of Perinatology*. 2005;25:737-742
- 379 14. Cronk CE, Malloy ME, Pelech AN, Miller RE, Meyer SA, Cowell M, et al. Completeness of state
380 administrative databases for surveillance of congenital heart disease. *Birth Defects Research Part A:*
381 *Clinical and Molecular Teratology*. 2003;67:597-603
- 382 15. Steiner JM, Kirkpatrick JN, Heckbert SR, Habib A, Sibley J, Lober W, et al. Identification of adults with
383 congenital heart disease of moderate or great complexity from administrative data. *Congenital heart*
384 *disease*. 2018;13:65-71
- 385 16. Rodriguez FH, 3rd, Raskind-Hood CL, Hoffman T, Farr SL, Glidewell J, Li JS, et al. How well do icd-9-
386 cm codes predict true congenital heart defects? A centers for disease control and prevention-based
387 multisite validation project. *Journal of the American Heart Association*. 2022;11:e024911
- 388 17. Pasquali SK, Peterson ED, Jacobs JP, He X, Li JS, Jacobs ML, et al. Differential case ascertainment in
389 clinical registry versus administrative data and impact on outcomes assessment for pediatric cardiac
390 operations. *The Annals of Thoracic Surgery*. 2013;95:197-203
- 391 18. Jantzen DW, He X, Jacobs JP, Jacobs ML, Gaies MG, Hall M, et al. The impact of differential case
392 ascertainment in clinical registry versus administrative data on assessment of resource utilization in
393 pediatric heart surgery. *World Journal for Pediatric and Congenital Heart Surgery*. 2014;5:398-405

- 394 19. Khan A, Ramsey K, Ballard C, Armstrong E, Burchill LJ, Menashe V, et al. Limited accuracy of
395 administrative data for the identification and classification of adult congenital heart disease.
396 2018;7:e007378
- 397 20. Downing KF, Oster ME, Olivari BS, Farr SL. Early-onset dementia among privately-insured adults with
398 and without congenital heart defects in the united states, 2015-2017. *International journal of*
399 *cardiology.* 2022;358:34-38
- 400 21. Downing KF, Tepper NK, Simeone RM, Ailes EC, Gurvitz M, Boulet SL, et al. Adverse pregnancy
401 conditions among privately insured women with and without congenital heart defects. *Circulation.*
402 *Cardiovascular quality and outcomes.* 2020;13:e006311
- 403 22. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture
404 (redcap)--a metadata-driven methodology and workflow process for providing translational research
405 informatics support. *Journal of biomedical informatics.* 2009;42:377-381
- 406 23. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The redcap consortium: Building
407 an international community of software platform partners. *Journal of biomedical informatics.*
408 2019;95:103208
- 409 24. Gurvitz M, Dunn Julie E, Bhatt A, Book Wendy M, Glidewell J, Hogue C, et al. Characteristics of adults
410 with congenital heart defects in the united states. *Journal of the American College of Cardiology.*
411 2020;76:175-182
- 412 25. Centers for Disease Control and Prevention. Comprehensive Listing ICD-10-CM Files [Internet].2022
413 [cited 2023 Jan 4]; Available from: [https://www.cdc.gov/nchs/icd/Comprehensive-Listing-of-ICD-10-](https://www.cdc.gov/nchs/icd/Comprehensive-Listing-of-ICD-10-CM-Files.htm)
414 [CM-Files.htm](https://www.cdc.gov/nchs/icd/Comprehensive-Listing-of-ICD-10-CM-Files.htm)
- 415 26. Chan L, Beers K, Yau AA, Chauhan K, Duffy Á, Chaudhary K, et al. Natural language processing of
416 electronic health records is superior to billing codes to identify symptom burden in hemodialysis
417 patients. *Kidney International.* 2020;97:383-392
- 418 27. Xu Y, Lee S, Martin E, D'Souza AG, Doktorchik CTA, Jiang J, et al. Enhancing icd-code-based case
419 definition for heart failure using electronic medical record data. *Journal of Cardiac Failure.*
420 2020;26:610-617
- 421 28. Jorge A, Castro VM, Barnado A, Gainer V, Hong C, Cai T, et al. Identifying lupus patients in electronic
422 health records: Development and validation of machine learning algorithms and application of rule-
423 based algorithms. *Seminars in Arthritis and Rheumatism.* 2019;49:84-90
- 424 29. Zhao Y, Fu S, Bielinski SJ, Decker PA, Chamberlain AM, Roger VL, et al. Natural language processing
425 and machine learning for identifying incident stroke from electronic health records: Algorithm
426 development and validation. *J Med Internet Res.* 2021;23:e22951
- 427 30. Diller G-P, Arvanitaki A, Opatowsky AR, Jenkins K, Moons P, Kempny A, et al. Lifespan perspective on
428 congenital heart disease research: Jacc state-of-the-art review. *Journal of the American College of*
429 *Cardiology.* 2021;77:2219-2235

430

431

432

433 **TABLES**

434 **Table 1: ICD-9-CM and ICD-10-CM CHD Code Inclusion Criteria for the Cohort from which**
 435 **the Random Sample was Selected**

ICD Diagnosis Code Label	ICD-9-CM	ICD-10-CM
Severe (15 ICD-9-CM codes and 14 ICD-10-CM codes)		
Hypoplastic Left Heart Syndrome (HLHS)	746.7	Q23.4
Tricuspid Atresia, stenosis, or absence	746.1	Q22.4
Hypoplastic right heart syndrome		Q22.6
Single ventricle, cor triloculare, double inlet left ventricle	745.3	Q20.4
Cor biloculare	745.7	
Pulmonary valve atresia or absence	746.01	Q22.0
Truncus Arteriosus, Common Truncus	745.0	Q20.0
Double outlet Right ventricle (DORV)	745.11	Q20.1
Double outlet Left ventricle (DOLV)		Q20.2
Tetralogy of Fallot	745.2	Q21.3
Transposition of the Great arteries (TGA), Complete TGA, dextro-TGA, TGA not otherwise specified, classical TGA ¹	745.1	Q20.3
	745.10	
	745.19	
Congenital Corrected transposition of great arteries (CCTGA), levo-TGA	745.12	Q20.5
Endocardial cushion defect; Atrioventricular septal defect; Complete atrioventricular canal defect (CAVCD) Endocardial cushion defect unspecified, Endocardial cushion defect, other ¹	745.6	Q21.2
	745.60	
	745.69	
Interrupted aortic arch	747.11	Q25.21
Total anomalous pulmonary venous return (TAPVR)	747.41	Q26.2
SHUNT AND VALVE²		
SHUNT (7 ICD-9-CM codes and 8 ICD-10-CM codes)³		
Ventricular septal defect (VSD)	745.4	Q21.0
Secundum atrial septal defect (ASD)	745.5	Q21.1
Primum atrial septal defect	745.61	
Other specified defects of septal closure, sinus venosus ASD, inferior sinus venosus ASD, superior sinus venosus ASD	745.8	Q21.8
Congenital malformation of cardiac septum, unspecified		Q21.9
Unspecified defect of septal closure	745.9	
Patent ductus arteriosus (PDA)	747.0	Q25.0
Aortopulmonary septal defect (AP window)		Q21.4
Partial anomalous pulmonary venous return (PAPVR)	747.42	Q26.3
Anomalous pulmonary venous connection, unspecified		Q26.4
VALVE (14 ICD-9-CM codes and 22 ICD-10-CM codes)⁴		
Anomalies of the pulmonary valve; Pulmonary valve anomaly unspecified ¹	746.0	Q22.3

	746.00	
Pulmonary valve stenosis (PS)	746.02	Q22.1
Pulmonary valve anomaly, other, pulmonary valve regurgitation	746.09	Q22.2
Ebstein anomaly of the tricuspid valve	746.2	Q22.5
Congenital malformations of tricuspid valve		Q22.8
Congenital malformation of tricuspid valve, unspecified		Q22.9
Aortic valve stenosis (AS)	746.3	Q23.0
Aortic insufficiency (congenital), bicuspid or unicuspid aortic valve	746.4	Q23.1
Other congenital malformations of aortic and mitral valves		Q23.8
Congenital malformation of aortic and mitral valves, and specified		Q23.9
Mitral stenosis or mitral valve abnormalities	746.5	Q23.2
Mitral insufficiency, cleft mitral valve	746.6	Q23.3
Subaortic stenosis, subaortic membrane	746.81	Q24.4
Infundibular or subvalvular pulmonary stenosis	746.83	Q24.3
Coarctation of the aorta ¹	747.1	Q25.1
	747.10	
Supravalvular aortic stenosis		Q25.3
Hypoplasia of the aorta		Q25.42
Atresia or stenosis of aorta	747.22	Q25.29
Absence and aplasia of aorta		Q25.41
Atresia of pulmonary artery		Q25.5
Pulmonary artery atresia, coarctation, or hypoplasia of the branch pulmonary arteries	747.31	Q25.71
Anomalies of the pulmonary artery, other	747.39	Q25.79
OTHER CHD (2 ICD-9-CM codes and 2 ICD-10-CM codes)⁵		
Cor triatriatum	746.82	Q24.2
Coronary artery anomaly, anomalous coronary artery, anomalous left coronary artery off the pulmonary artery (ALCAPA), anomalous right coronary artery off the pulmonary artery (ARCAPA)	746.85	Q24.5
OTHER VASCULAR (3 ICD-10-CM codes)⁶		
Double aortic arch		Q25.45
Right aortic arch ⁶ (component of vascular ring)		Q25.47
Anomalous origin of the subclavian artery ⁶ (component of vascular ring)		Q25.48

436 ¹ Some ICD-9-CM codes were adjusted over time to include an additional decimal place, but are considered the same code

437 ² Cases with no Severe codes and has both Shunt AND Valve codes.

438 ³ Cases with Shunt codes with or without any other group codes, but no Severe or Valve codes.

439 ⁴ Case with Valve codes with or without any other group codes, but no Severe or Shunt codes.

440 ⁵ Cases with only 'Other CHD' type group codes, AND no Severe, Shunt or Valve codes.

441 ⁶ Cases with 'Other Vascular' codes required either code Q25.45 or the combination of Q25.47 and Q25.48 to be included.

442 **Notes.** The following 25 ICD-9-CM codes were excluded from the case definition: 746.84 obstructive heart anomalies NEC, 746.86
 443 congenital heart block, 746.87 malposition of heart, 746.89 congenital heart anomaly NEC, 746.9 congenital heart anomaly NOS,
 444 747.20 congenital anomaly of aorta NOS, 747.21 anomalies of aortic arch, 747.29 congenital anomalies of aorta NEC, 747.3
 445 pulmonary artery anomaly, 747.32 pulmonary AV malformation, 747.39 other anomalies of the pulmonary artery and circulation,
 446 747.40 Great vein anomaly NOS, 747.49 great vein anomaly NEC, 747.5 and umbilical artery absence, 747.6 (end 1993) peripheral
 447 vascular anomaly NEC, 747.60 (begin 1993) unspecified peripheral vascular anomaly, 747.61 GI vessel anomaly, 747.62 renal vessel
 448 anomaly, 747.63 upper limb vessel anomaly, 747.64 lower limb vessel anomaly, 747.69 other specified peripheral vascular anomaly,

449 747.81 cerebrovascular anomaly, 747.82 spinal vessel anomaly, 747.83 persistent fetal circulation, 747.89 circulatory anomaly NEC,
450 747.9 circulatory anomaly NOS.
451 The following 24 ICD-10-CM codes were excluded from the case definition: Q20.6 isomerism of atrial appendages, Q20.8 other
452 congenitally malformations of cardiac chambers and connections, Q20.9 congenital malformation of cardiac chambers and
453 connections unspecified, Q24.0 malposition of heart or Apex, dextrocardia, Q24.1 levocardia, Q24.8 other specified congenital
454 malformations of the heart, Q24.9 unspecified defect of the heart, Q25.40 anomalies of aorta, unspecified, Q25.43 congenital
455 aneurysm of aorta, Q25.4 for congenital dilation of aorta, Q25.46 torturous aortic arch, Q25.47 right aortic arch, Q25.49 other
456 congenital malformations of aorta, Q25.6 branch pulmonary artery stenosis, Q25.72 congenital pulmonary arteriovenous
457 malformation, Q25.79 other congenital malformations of the pulmonary artery, Q25.8 other congenital malformations of other great
458 arteries, Q25.9 congenital malformation of great arteries unspecified, Q26.0 congenital stenosis vena cava, Q26.1 left superior vena
459 cava, Q26.5 anomalous portal venous connection, Q26.6 portal vein – hepatic artery fistula, Q26.8 other congenital malformations of
460 great veins, Q26.9 anomalies of great veins unspecified.
461 **Abbreviations:** CHD = congenital heart defect; ICD-9-CM = International Classification of Diseases 9th revision, Clinical Modification;
462 ICD-10-CM = International Classification of Diseases 10th revision, Clinical Modification.
463
464
465
466

467

468 **Table 2. Intra-observer and Inter-observer Reliability Results**
 469 **for Randomly Selected Cases from the 1500 Validated Dataset**

Observer	CHD Yes or No Agreement	Percent Agreement
Intra-Observer Reliability¹		
Abstractor1	25/25	100%
Abstractor2	20/21	95%
Abstractor3	21/22	95%
Abstractor4 ²	7/7	100%
Inter-Observer Reliability³		
Clinician1 ⁴ vs Abstractor1	23/25	92%
Clinician1 vs Abstractor2	27/27	100%
Clinician1 vs Abstractor3	21/24	88%
Clinician2 ⁵ vs Abstractor1	25/27	93%
Clinician2 vs Abstractor2	20/21	95%
Clinician2 vs Abstractor3	15/15	100%
Clinician2 vs Abstractor4	12/12	100%

470 ¹5% of each abstractor's cases were randomly duplicated and validated again by
 471 the same abstractor to assess intra-observer reliability.

472 ²Abstractor 4 only abstracted cases from PHS

473 ³10% of validated cases were randomly duplicated and validated by two different
 474 clinicians to assess inter-observer reliability.

475 ⁴Clinician 1 abstracted cases from AHS

476 ⁵Clinician 2 abstracted cases from PHS

477 AHS= Adult Healthcare System, PHS=Pediatric Healthcare System

478
 479

480

481

482

483

484

485 **Table 3. Characteristics of 1497 Validated Cases with at least one CHD ICD-9-CM and/or**
 486 **ICD-10-CM code(s)¹ and with at least one encounter between 1/1/2010 -12/31/2019 in an Adult**
 487 **Healthcare System and a Pediatric Healthcare System**
 488

CHARACTERISTICS	Validated Cases (N=1497)			Validated Cases without CHD (N=477)		
	Cases with CHD (Yes) n= 1020 (68.1%)	Cases without CHD (No) n= 477 (31.9%)	CHD Yes vs. CHD No	Cases without CHD, but with PFO n= 316 (66.2%)	Cases without CHD and PFO n= 161 (33.8%)	PFO vs. No PFO
CATEGORICAL VARIABLES[^]						
# Cases per Data Source¹						
<i>Adult Healthcare System (AHS)</i>	481 (64.8%)	261 (35.2%)	X²= 7.43 p=0.0064	152 (58.2%)	109 (41.8%)	X²=16.54 P<0.0001
<i>Pediatric healthcare System (PHS)</i>	416 (66.6%)	209 (33.4%)	X ² =1.23 p=0.2678	160 (76.6%)	49 (23.4%)	X²=17.64 P<0.0001
<i>Both AHS and PHS</i>	123 (94.6%)	<10 (--%)	--	<10 (--%)	<10 (--%)	--
Sex²						
<i>Male</i>	513 (69.5%)	225 (30.5%)	X ² =1.31 p=0.2525	163 (72.4%)	62 (27.6%)	X²=7.31 P=0.0068
<i>Female</i>	506 (66.8%)	252 (33.2%)	X ² =1.31 p=0.2525	153 (60.7%)	99 (39.3%)	X²=7.31 P=0.0068
CHD Anatomic Severity Groups³ as identified by ICD codes						
Severe group codes³	249 (95.8%)	11 (4.2%)	X²= 110.66 P<0.0001	<10 (--%)	10 (90.9%)	--
Shunt and Valve codes³	119 (73.5%)	43 (26.5%)	X ² =2.37 p=0.1238	39 (90.7%)	<10 (--%)	--
Shunt codes³	370 (52.6%)	333 (47.4%)	X²=146.76 P<0.0001	266 (79.9%)	67 (20.1%)	X²=91.68 P<0.0001
# (%) cases with only 745.5 and/or Q21.1⁴	123 (30.6%)	279 (69.4%)	X²= 356.72 p<0.0001	236 (84.6%)	43 (15.4%)	X²=101.11 P<0.0001
Valve codes³	243 (75.9%)	77 (24.1%)	X²=11.41 P=0.0007	10 (13.0%)	67 (87.0%)	X²=116.49 P<0.0001
Other CHD or Other Vascular codes³	39 (75.0%)	13 (25.0%)	X ² =1.17 p=0.2796	<10 (--%)	13 (100.0%)	--
CONTINUOUS VARIABLES^{^^}						
Median (IQR) of unique CHD ICD codes ⁵	2.0 (1.0-3.0)	1.0 (1.0-1.0)	D-stat =0.339 P<0.0001	1.0 (1.0-1.0)	1.0 (1.0-1.0)	D-stat =0.097 p=0.2662
Median (IQR) of unique encounters ⁶ with any CHD ICD code	3.0 (1.0-7.0)	1.0 (1.0-2.0)	D-stat=0.369, p<0.0001	1.0 (1.0-2.0)	1.0 (1.0-2.0)	D-stat=0.094 p=0.3255
Median (IQR) of cardiac ⁷ encounters per case with any CHD ICD code	2.0 (1.0-5.0)	1.0 (1.0-2.0)	D-stat =0.348 p<0.0001	1.0 (1.0-2.0)	1.0 (0.0-2.0)	D-stat =0.164 P=0.0113
Median (IQR) of imaging encounters ⁸ per case, but with no cardiac encounters	0.0 (0.0-1.0)	0.0 (0.0-1.0)	D-stat=0.028, P=0.9677	0.0 (0.0-0.0)	0.0 (0.0-1.0)	D-stat=0.186 p=0.0026
Median (IQR) of non-cardiac encounters ⁹ per case with any CHD ICD codes	0.0 (0.0-1.0)	0.0 (0.0-1.0)	D-stat =0.160 p<0.0001	0.0 (0.0-1.0)	0.0 (0.0-1.0)	D-stat =0.081 p=0.5468
Mean (SD) age ^{^^^} in years at FQE ¹⁰	22.4 (20.27)	26.3 (23.35)	t= 3.140 p=0.0017	23.4 (23.50)	31.9 (22.08)	t= 3.791 p=0.0002

489 ¹ Data sources are not mutually exclusive, and cases may be captured at either of the two healthcare systems or at both over the 10-year surveillance
 490 period.
 491 ² One case had unknown sex.

492 ³ CHD ICD-9-CM and ICD-10-CM codes and anatomic groups are defined in Appendix A.
493 ⁴ Includes only patients with ICD-9-CM code 745.5 and/or ICD-10-CM code Q21.2 without any other CHD ICD code in Appendix A, noted at any
494 encounter.
495 ⁵ Unique ICD-9-CM and ICD-10-CM codes defined as codes for a single CHD, as noted in Appendix A. Where two or more ICD-9-CM or ICD-10-CM
496 codes are assigned to a single CHD, codes are counted as one occurrence. Where both ICD codes are assigned to a single CHD, codes are counted as
497 a single occurrence.
498 ⁶ Unique encounters are defined as a unique encounter ID separated in time by a minimum of 24 hours.
499 ⁷ A cardiac encounter is an encounter that took place in pediatric cardiology, Sibley Heart Center, cardiothoracic surgery, cardiology, adult congenital
500 heart, heart failure, electrophysiology, interventional cardiology, structural cardiology, transplant cardiology, heart transplant, cardiac
501 catheterization, or had a provider NPI with a cardiac specialty.
502 ⁸ Imaging encounters are defined by Current Procedural Terminology (CPT) codes.
503 ⁹ Non-cardiac encounters are defined as encounters that do not have an encounter type noted in footnotes 7-8.
504 ¹⁰ Age at FQE is the age at first qualifying encounter, calculated at time from date of birth to first encounter w CHD ICD code.

505 **Notes.**
506 Statistically significant tests at p<0.05 are indicated in **bold**.
507 [^] For the categorical variables of data source, gender, and CHD anatomic severity groups, Pearson X² analyses were conducted; this test assessed
508 whether there was a statistically significant difference, at p<0.05, between expected and observed frequencies for 'Yes' CHD and 'No' CHD validated
509 cases, and for cases with CHD, but no PFO and those with neither CHD nor PFO.
510 ^{^^} For the continuous variables involving encounters as well as for # of unique ICD codes, *the non-parametric two-sample Kolmogorov-Smirnov tests*
511 were conducted; this test determined whether the distributions of encounters for 'Yes' CHD and 'No' CHD validated cases, and between cases with
512 CHD, but no PFO and cases with neither CHD nor PFO were statistically identical at the p<0.05.
513 ^{^^^} A t-test analysis was conducted to determine if there was a statistical mean difference in age in years at FQE, at p<0.05, between 'Yes' CHD and
514 'No' CHD validated cases, and for cases with CHD, but no PFO and those with neither CHD nor PFO.

515
516 **Abbreviations:** *ICD-9-CM* = International classification of diseases, 9th edition, clinical modification; *ICD-10-CM* = ICD 10th edition; *CHD* = congenital
517 heart defect; *PFO* = patent foramen ovale; *CPT* = Current Procedural Terminology; *FQE* = first qualifying encounter with an ICD-related CHD code.

518
519

520

521

522

523

524

525

526 **Table 4. Positive Predictive Value (PPV) of CHD ICD-9-CM and ICD-10-CM Codes by Defect**
 527 **Characteristics for Identifying CHD Cases, Validated through Chart Abstraction in Electronic**
 528 **Health Records in Two Tertiary Healthcare Systems, and Overall**

	Individuals with CHD ICD Codes in AHS only (N=742)	Individuals with CHD ICD Codes in PHS only (N=602)	Individuals with CHD ICD Codes in both AHS & PHS (N=129)	Individuals with CHD ICD Codes in AHS &/or PHS combined (N=1473) ¹
PPV FOR ANY CHD				
Any CHD codes ² at any encounter ³	64.8% (481/742)	66.1% (398/602)	94.6% (122/129)	68.0% (1001/1473)
Severe group codes ⁴	93.6% (131/140)	97.9% (47/48)	98.5% (64/65)	95.7% (242/253)
Shunt AND valve group codes ⁴	97.5% (39/40)	58.5% (55/94)	94.7% (18/19)	73.2% (112/153)
Only shunt group codes ⁴	36.1% (115/319)	64.7% (220/340)	83.9% (26/31)	52.3% (361/690)
Cases with only 745.5 and/or Q21.1 (subgroup of shunt)	25.9% (65/251)	36.0% (50/139)	100.0% (8/8)	30.9% (123/398)
Only valve group codes ⁴	83.1% (167/201)	60.0% (66/110)	100.0% (14/14)	76.0% (247/325)
Only Other CHD or Other vascular group codes ⁴	69.0% (29/42)	100.0% (10/10)	0% (0/0)	75.0% (39/52)
More than one unique CHD code ⁵	95.2% (217/228)	72.9% (191/262)	96.0% (95/99)	85.4% (503/589)
Only one unique CHD ICD code ⁵	51.4% (264/514)	60.9% (207/340)	90.0% (27/30)	56.3% (498/884)
Any CHD codes associated with encounters between 1/1/2010 and 9/30/2015 for a CHD	71.6% (303/423)	72.8% (177/243)	96.1% (99/103)	75.3% (579/769)
Any CHD codes associated with encounters between 10/1/2015 and 12/31/2019 for a CHD	67.4% (317/470)	68.8% (324/471)	95.3% (101/106)	70.9% (742/1047)
Any CHD codes ² associated with <i>cardiac</i> encounters ⁶	72.5% (371/512)	66.7% (392/588)	95.2% (119/125)	72.0% (882/1225)
Any CHD codes ³ associated with only <i>non-cardiac</i> encounters	66.3% (327/493)	72.9% (97/133)	94.6% (87/92)	71.2% (511/718)
Any CHD codes at <u>any</u> inpatient encounter ³ (Regardless of number)	73.3% (187/255)	74.9% (149/199)	94.4% (84/89)	77.3% (420/543)
Any CHD code at <u>any</u> outpatient encounter ³ (Regardless of number), no inpatient	60.4% (294/487)	61.8% (249/403)	95.0% (38/40)	62.5% (581/930)
Any CHD code in at least 2 outpatient encounters ³ separate for at least 30 days (subset of previous row)	77.1% (172/223)	82.7% (134/162)	93.1% (27/29)	80.4% (333/414)

529 ¹ 24 cases did not have an encounter with a CHD code within the surveillance period.
530 ² ICD-9-CM and ICD-10-CM CHD codes, defined in Table 1.
531 ³ All included encounters occurred between 1/1/2010 and 12/31/2019.
532 ⁴ Anatomic CHD group by ICD codes, defined in Table 1.
533 ⁵ Unique ICD-9-CM and ICD-10-CM codes defined as codes for a single CHD, as noted in Table 1. Where two or more ICD-9-CM or ICD-
534 10-CM codes are assigned to a single CHD diagnosis, the codes are counted as one occurrence. Where both an ICD-9-CM code and
535 ICD-10-CM code are assigned to a single CHD diagnosis, the codes are counted as a single occurrence.
536 ⁶ Cardiac encounter = an encounter with a provider or encounter type listed as pediatric cardiology, Sibley Heart Center,
537 cardiothoracic surgery, cardiology, adult congenital heart, heart failure, electrophysiology, interventional cardiology, structural
538 cardiology, transplant cardiology, heart transplant, cardiac catheterization, or had a provider NPI with a cardiac specialty.
539 **Abbreviations:** CHD = congenital heart defect; AHS= Adult Healthcare System; PHS = Pediatric Healthcare System; PPV = positive
540 predictive value; ICD-9-CM = International classification of diseases, 9th edition, clinical modification; ICD-10-CM = International
541 classification of disease, 10th edition; ICD = International Classification of Diseases, and refers to both 9-CM and 10-CM.
542

543

544 eAppendix A

545

546 Cardiac Imaging CPT Codes

547

548 Congenital Echocardiography codes:

549 **93303** Transthoracic echocardiography for congenital cardiac anomalies; complete
550 **93304** Transthoracic echocardiography for congenital cardiac anomalies; follow-up or limited study
551 **93315** Transesophageal echocardiography for congenital cardiac anomalies; incl probe placement, image acquisition,
552 interpretation, and report
553 **93316** Transesophageal echocardiography for congenital cardiac anomalies; placement of transesophageal probe
554 only
555 **93317** Transesophageal echocardiography for congenital cardiac anomalies; image acquisition, interpretation &
556 report only
557

558 Non-Congenital Echocardiography codes:

559 **93306** Echocardiography, transthoracic, real-time with image documentation (2D), includes M-mode recording, when
560 performed, complete, with spectral Doppler echocardiography, & with color flow Doppler echocardiography
561 **93307** Echocardiography, transthoracic, real-time with image documentation (2D), includes M-mode recording, when
562 performed, complete, without spectral or color Doppler echocardiography
563 **93308** Echocardiography, transthoracic, real-time with image documentation (2D), includes M-mode recording, when
564 performed, follow-up or limited study
565 **93312** Echocardiography, transesophageal, real-time with image documentation (2d) (with/without m-mode
566 recording); incl probe placement, image acquisition, interpretation & report
567 **93313** Echocardiography, transesophageal, real-time with image documentation (2d) (with/without m-mode
568 recording); placement of transesophageal probe only
569 **93314** Echocardiography, transesophageal, real-time with image documentation (2d) (with/without m-mode
570 recording); image acquisition, interpretation & report only
571 **93318** Echocardiography, transesophageal (tee) for monitoring purposes, incl probe placement, real time 2-
572 dimensional image acquisition & interpretation leading to ongoing (continuous) assessment of (dynamically
573 changing) cardiac pumping function & to therapeutic measures on immediate time basis
574 **93320** Doppler echocardiography, pulsed wave and/or continuous wave with spectral display (list separately in
575 addition to codes for echocardiographic imaging); complete
576 **93321** Doppler echocardiography, pulsed wave and/or continuous wave with spectral display (list separately in
577 addition to codes for echocardiographic imaging); follow-up or limited study (list separately in addition to
578 codes for echocardiographic imaging)

- 579 **93325** Doppler echocardiography color flow velocity mapping (list separately in addition to codes for
580 echocardiography)
- 581 **C8925** Transesophageal echocardiography (tee) with/without contrast followed by with contrast, real time with
582 image documentation (2d) (with/without m-mode recording); incl probe placement, image acquisition,
583 interpretation & report
- 584 **C8926** Transesophageal echocardiography (tee) with/without contrast followed by with contrast, for congenital
585 cardiac anomalies; incl probe placement, image acquisition, interpretation & report
- 586 **C8927** Transesophageal echocardiography (tee) with/without contrast followed by with contrast, for monitoring
587 purposes, incl probe placement, real time 2-dimensional image acquisition & interpretation leading to ongoing
588 (continuous) assessment of (dynamically changing) cardiac pumping function & to therapeutic measures on an
589 immediate time basis
590
- 591 **Cardiac Magnetic Resonance Imaging (MRI) codes:**
- 592 **71555** Magnetic Resonance Angiography (MRA) of chest
- 593 **Cardiac MRI without gadolinium contrast**
- 594 **75557** Cardiac MRI for morphology & function without contrast
- 595 **75558** Cardiac MRI for morphology & function without contrast; with flow/velocity quantification
- 596 **75559** Cardiac MRI for morphology & function without contrast; with stress imaging
- 597 **75560** Cardiac MRI for morphology & function without contrast; with flow/velocity quantification & stress
598
- 599 **Cardiac MRI with gadolinium contrast**
- 600 **75561** Cardiac MRI for morphology & function without contrast followed by contrast and further sequences
- 601 **75562** Cardiac MRI for morphology & function without contrast followed by contrast and further sequences; with
602 flow/velocity quantification
- 603 **75563** Cardiac MRI for morphology & function without contrast followed by contrast and further sequences; with
604 stress imaging
- 605 **75564** Cardiac MRI for morphology & function without contrast followed by contrast and further sequences; with
606 flow/velocity quantification and stress
- 607 **Cardiac MRI Add-on Flow Code**
- 608 **75565** Cardiac MRI for velocity flow mapping
609
- 610 **Cardiac Computed Tomography (CT) codes:**
- 611 **71725** CT angiography of the chest (used for looking for extracardiac anatomy such as pulmonary arteries,
612 pulmonary veins, aortic arch, vascular rings)
- 613 **75571** Computed tomography, heart, without contrast, with quantitative evaluation of coronary calcium
- 614 **75572** Computed tomography, heart, with contrast, for evaluation of cardiac structure & morphology (incl 3D image
615 postprocessing, assessment of cardiac function & evaluation of venous structures, if performed)
- 616 **75573** Computed tomography, heart, with contrast, for evaluation of cardiac structure & morphology in setting of
617 congenital heart disease (incl 3D image postprocessing, assessment of LV cardiac function, RV structure &
618 function & evaluation of venous structures, if performed)
- 619 **75574** Computed tomographic angiography, heart, coronary arteries & bypass grafts (when present), with contrast,
620 incl 3D image postprocessing (incl evaluation of cardiac structure & morphology, assessment of cardiac function, &
621 evaluation of venous structures, if performed)