

ORIGINAL RESEARCH

How Well Do *ICD-9-CM* Codes Predict True Congenital Heart Defects? A Centers for Disease Control and Prevention-Based Multisite Validation Project

Fred H. Rodriguez III , MD; Cheryl L. Raskind-Hood , MS, MPH; Trenton Hoffman , MS; Sherry L. Farr , PhD; Jill Glidewell, MSN, MPH; Jennifer S. Li , MD; Alfred D'ottavio , BSE, BEE; Lorenzo Botto, MD; Matthew R. Reeder , MPH; Daphne Hsu , MD; George K. Lui , MD; Anaclare M. Sullivan , MS; Wendy M. Book 

BACKGROUND: The Centers for Disease Control and Prevention's Surveillance of Congenital Heart Defects Across the Lifespan project uses large clinical and administrative databases at sites throughout the United States to understand population-based congenital heart defect (CHD) epidemiology and outcomes. These individual databases are also relied upon for accurate coding of CHD to estimate population prevalence.

METHODS AND RESULTS: This validation project assessed a sample of 774 cases from 4 surveillance sites to determine the positive predictive value (PPV) for identifying a true CHD case and classifying CHD anatomic group accurately based on 57 *International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)* codes. Chi-square tests assessed differences in PPV by CHD severity and age. Overall, PPV was 76.36% (591/774 [95% CI, 73.20–79.31]) for all sites and all CHD-related *ICD-9-CM* codes. Of patients with a code for complex CHD, 89.85% (177/197 [95% CI, 84.76–93.69]) had CHD; corresponding PPV estimates were 86.73% (170/196 [95% CI, 81.17–91.15]) for shunt, 82.99% (161/194 [95% CI, 76.95–87.99]) for valve, and 44.39% (83/187 [95% CI, 84.76–93.69]) for “Other” CHD anatomic group ($\chi^2=142.16$, $P<0.0001$). *ICD-9-CM* codes had higher PPVs for having CHD in the 3 younger age groups compared with those >64 years of age, ($\chi^2=4.23$, $P<0.0001$).

CONCLUSIONS: While CHD *ICD-9-CM* codes had acceptable PPV (86.54%) (508/587 [95% CI, 83.51–89.20]) for identifying whether a patient has CHD when excluding patients with *ICD-9-CM* codes for “Other” CHD and code 745.5, further evaluation and algorithm development may help inform and improve accurate identification of CHD in data sets across the CHD *ICD-9-CM* code groups.

Key Words: birth defects ■ congenital heart defects ■ epidemiology ■ surveillance ■ validation

Research and surveillance of patients with congenital heart defects (CHD) using administrative and clinical data both rely on the diagnostic accuracy of applying *International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)* and *International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM)* codes to detect CHD

cases. Administrative data, commonly referred to as claims data, are data created for the purpose of either the billing of health care encounters or record keeping for a health care system or an organization. *ICD* codes captured in administrative data may vary by medical practice, health care system, and region; thus, it is unknown how representative these *ICD* codes are for their intended

Correspondence to: Fred H. Rodriguez III, MD, 2835 Brandywine Rd, Suite 400, Atlanta, GA 30341. Email: fred.rodriguez@emory.edu

Supplemental Material is available at <https://www.ahajournals.org/doi/suppl/10.1161/JAHA.121.024911>

For Sources of Funding and Disclosures, see page 12.

© 2022 The Authors. Published on behalf of the American Heart Association, Inc., by Wiley. This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs License](#), which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

*JAH*A is available at: www.ahajournals.org/journal/jaha

CLINICAL PERSPECTIVE

What Is New?

- Four US sites associated with the Centers for Disease Control and Prevention's Surveillance of Congenital Heart Defects Across the Lifespan project validated a sample of 774 cases with *International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)* congenital heart disease (CHD) codes to assess true CHD cases.
- The majority of cases were associated with true CHD, though differences in positive predictive value (PPV) were noted based on anatomic complexity and ages of patients.
- Cases with complex CHD codes, multiple CHD codes, and age groups <65 years of age had greater PPV identifying true CHD.

What Are the Clinical Implications?

- ICD-9-CM* codes can identify patients with CHD in databases with high PPV for the complex code group, but lower PPV in certain patient groups, particularly those aged >65 years, and with "Other" CHD *ICD* code group.
- When attempting to identify cases with CHD, the presence of >1 CHD code increases the PPV for a true CHD, at the expense of sensitivity.
- Development of algorithms is needed to improve the identification of CHD cases in databases across anatomic code groups and age ranges.

Nonstandard Abbreviations and Acronyms

CDC	Centers for Disease Control and Prevention
eHR	electronic health record

disease state. In a recent study conducted by Khan et al. (2018), *ICD-9-CM* administrative codes extracted from the electronic health record (eHR) of patients with various types of CHD lesions seen at a large academic health care system were <50% accurate (48.7% [95% CI, 47%–51%]) at classifying those with a true CHD.¹ However, when only patients with moderate or complex CHD anatomy were included, the positive predictive value (PPV) of having CHD increased to 77.2%, (95% CI, 74%–81%). When other factors like younger age, adult CHD, provider type, and ECG, or echocardiogram were documented at the CHD-related encounter, the C-statistic was 0.89 (95% CI, 0.88–0.90).¹ Correctly and consistently applied definitions of CHD may increase the accuracy of CHD

prevalence, health care use, and health outcomes of individuals living with CHD using administrative health care data sets. However, prior studies have demonstrated that some *ICD-9-CM* codes may be associated with false positives and thus do not always reliably identify individuals who truly have CHD.^{1–5}

In studies of CHD using publicly available data sets like the National Inpatient Sample and the Kids' Inpatient Database or administrative data sources, a CHD case is typically defined by *ICD-9-CM* codes 745.xx to 747.xx for classification and more recently *ICD-10-CM* codes Q20 to Q28. While the range of these codes is broad and inclusive, this code group contains conditions that are not CHD, and thus, may include individuals who do not have CHD, creating misleading conclusions and misinformation. Furthermore, some codes in the CHD group may code for CHD, but may commonly be used incorrectly, as for "rule out" or normal variants. In particular, individuals with 1 CHD-related *ICD-9-CM* code —745.5 — are often misclassified as having CHD. Frequently found in large CHD administrative data sets and commonly included in CHD literature, the *ICD-9-CM* code 745.5 (hereafter referred to as "code 745.5") is used for both secundum atrial septal defect, a true CHD, and patent foramen ovale, a normal variant and not considered a CHD, seen in about 25% of the population.²

The current project aims to validate the extent to which CHD-related *ICD-9-CM* codes correctly identify CHD cases in administrative and clinical records by: (1) confirming patients as having a true CHD; and (2) classifying CHD anatomic grouping among those with true CHD. We hypothesized that both individual CHD codes and CHD anatomic groupings would have a high PPV (>80%) for CHD that would vary by anatomic group, but that coding errors would likely include patients who did not truly have a CHD.

METHODS

This analysis has been replicated by 2 independent analysts. Because of the sensitive nature of the data collected for this study, requests to access the data set from qualified researchers trained in human subject confidentiality protocols may be sent to the Centers for Disease Control and Prevention (CDC) at jill.glidewell@cdc.hhs.gov.

To improve upon CHD classification and as part of the multiyear CDC-sponsored project Surveillance of Congenital Heart Defects Across the Lifespan (CDC-RFA-DD15-1506),⁶ 4 sites reviewed the medical records of cases with CHD-related *ICD-9-CM* codes identified from administrative data sources to calculate the PPV of these codes (745.xx–747.xx) in correctly identifying a case with CHD. Cases were ascertained

by the presence of any CHD-related 745.xx to 747.xx *ICD-9-CM* code documented in a health care encounter between January 1, 2011 and December 31, 2013. Inclusion criteria defined a case having the presence of any included *ICD-9-CM* code at any encounter, regardless of how many encounters had CHD *ICD-9-CM* codes. The *ICD-9-CM* codes to define the *ICD* code-based anatomic groups were based on a hierarchy of codes, complex>shunt and valve>shunt or valve>“Other” CHD. The following codes were excluded as they were determined to be reflective of conditions other than true CHD: congenital heart block (746.86), absent/hypoplastic umbilical artery (747.5), pulmonary arteriovenous malformation (747.32), other anomalies of peripheral vascular system (747.6x), and other specified anomalies of circulatory system (747.8x); these codes were also excluded from the definition of CHD in the prior surveillance methods paper.^{5,6} Cases with code 745.5 without another included CHD *ICD-9-CM* code were also excluded based on previous studies.^{2,3} The Institutional Review Boards from Duke University in North Carolina (NC), Emory University in Georgia (GA), the New York State Department of Health (NY), and University of Utah (UT) approved an analysis of deidentified data to assess PPV of CHD-related *ICD-9-CM* codes.⁶ The requirement for informed consent was waived by each site’s respective Institutional Review Board. Eligible codes were classified into 1 of 5 CHD anatomic groups: complex, shunts, valves, shunts and valves, and “Other” CHD or non-specific defects^{5–8} (*Figure 1* and *Data S1*). Complex anatomy was based on native anatomy and defined as heart defects characterized by a recognized constellation of multiple specific defects which generally require intervention in the first year of life. Specific defects grouped as “complex” are defined in *Data S1*.^{5,6} A code-based hierarchy was developed such that the presence of a complex code designates the case as complex regardless of additional codes.⁶ In the absence of a complex code, the presence of both a shunt and valve code designated “shunt and valve” group inclusion. The absence of complex, shunt or valve codes and only “Other” CHD anatomic group codes designated the case as belonging to the “Other” CHD anatomic group (*Figure 1*).

Data and Procedures

Data from the 4 sites (GA, NC, NY, UT) over a 3-year project period, 2011 to 2013, were used for this validation. A total of 800 cases were planned for the validation study consisting of 200 cases from each of the 4 sites. Each site selected 50 cases from 4 mutually exclusive CHD anatomic groups as defined by the *ICD* code hierarchy (*Figure 1*) based on native anatomy: complex, shunt, valves, and “Other”, with each group further stratified by age: 1 to 10 years, 11 to 19 years, 20 to

64 years, >64 years (GA, NC, NY), or ages 11 to 19 years and 20 to 64 years (UT). Anatomic groups are described in *Figure 1* and *Data S1*. To ensure a comparable distribution of cases by age, the data set was stratified by age group and a proportion was selected, based on the age distribution of the larger cohort, into each of the 4 mutually exclusive anatomic groups (*Figure 2*). Only those data sources where medical charts were accessible for review were eligible for inclusion. Cases identified only in administrative data, but without clinical records to review, were excluded. Of the total cases from the larger CHD surveillance project, 69.4% of GA’s cases, 36.7% of NC’s cases, 32.5% of NY’s cases, and 97.6% of UT’s cases were eligible for medical chart review and selection for the validation project.

During medical chart review, clinical investigators at each site supervised the review of predetermined variables abstracted from eHRs and noted the presence/absence of a true CHD based on review of CHD anatomy as determined by: (1) cardiac imaging, (2) clinical diagnosis by an outpatient or inpatient encounter with a pediatric or adult CHD provider, (3) CHD surgery, and (4) autopsy report. Included in this review, clinical investigators evaluated: (1) CHD case (Yes/No), and (2) CHD anatomic group correct (Yes/No). All available information from the eHR (including any data before 2011 or after 2013) was also used to confirm or refute the presence of CHD for each selected case. During chart review, information on type of CHD recorded, number of unique CHD codes per case, diagnostic tests received (ie, echocardiograms, cardiac catheterizations, cardiac surgery), autopsy reports or clinic notes, as well as date of diagnosis and type of provider who made the CHD diagnosis was evaluated. Additionally, demographic information including age, sex, race, and ethnicity was abstracted. With respect to race, small sample size for the “Other” race category necessitated this group be combined with the “unknown” race group (“Other” race includes American Indian/Alaskan Native, Asian, native Hawaiian/Pacific Islander, and multi-racial).

The anatomic group of shunt and valve was excluded before case selection because it was assumed that if the case had codes for both of these anatomic groups, then the case was likely to be a true CHD case. Cases with code 745.5 in isolation or in combination with 746.89 or 746.9 were also excluded given the known poor PPV of these codes to represent a true CHD.² A total of 26 cases were excluded from analysis: 12 cases with code 745.5 in isolation or in combination with 746.89 or 746.9 among the sites that were found to be included in error after selection; 5 cases with both a shunt and valve diagnosis that were also erroneously selected for validation; 3 cases that did not have any clinical data to review; 3 cases that were inadvertently reviewed twice; and lastly, 3 cases whose only CHD code(s) were documented during

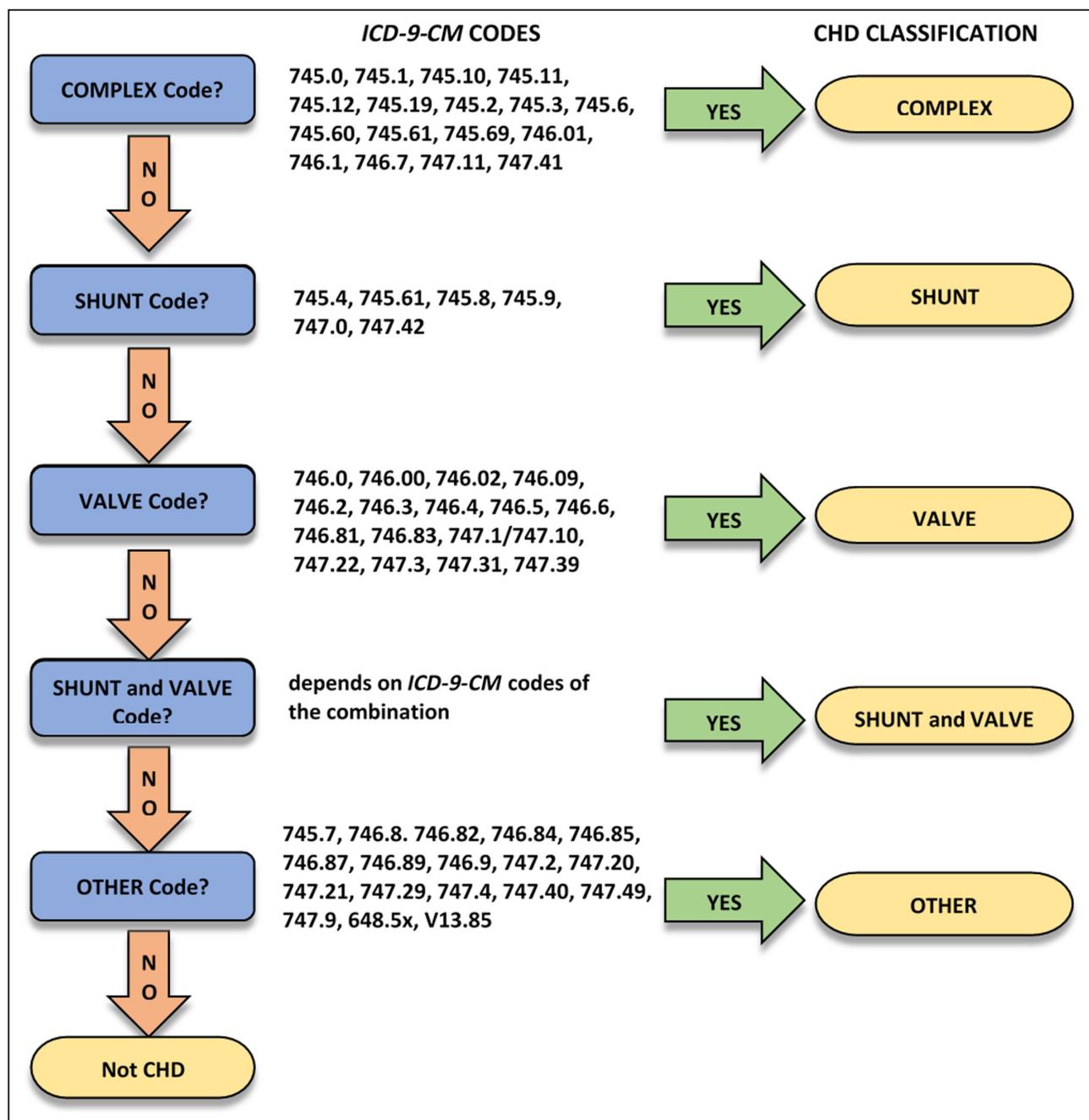


Figure 1. ICD-9-CM code-based hierarchy for congenital heart defect classification by native CHD anatomy group.

*Based on hierarchy reported in Ref. [6]. Individuals aged 1 to 64 years with documented congenital heart defects at health care encounters, 5 US surveillance sites, 2011 to 2013. CHD indicates congenital heart defect; and ICD-9-CM, International Classification of Diseases, Ninth Revision, Clinical Modification.

an encounter for a fetal echocardiogram defined by the associated Current Procedural Terminology code (Figure 2). CHD codes occurring during performance of a fetal echocardiogram were excluded based on unpublished data showing that only 2.9% (4 out of 138) of women who had a CHD code solely associated with a fetal echocardiogram encounter actually had a true CHD, whereas the majority of CHD diagnosis codes

documented during fetal echocardiogram encounters are intended for the fetus.

For GA, 200 cases, who resided in 1 of 5 metropolitan-Atlanta counties (Clayton, Cobb, DeKalb, Fulton, Gwinnett) and were seen at least once at 1 of 3 health care systems with records available, were randomly selected for review. For the 1- to 10- and 11- to 19-year-old groups, 13 cases for each CHD anatomic class were

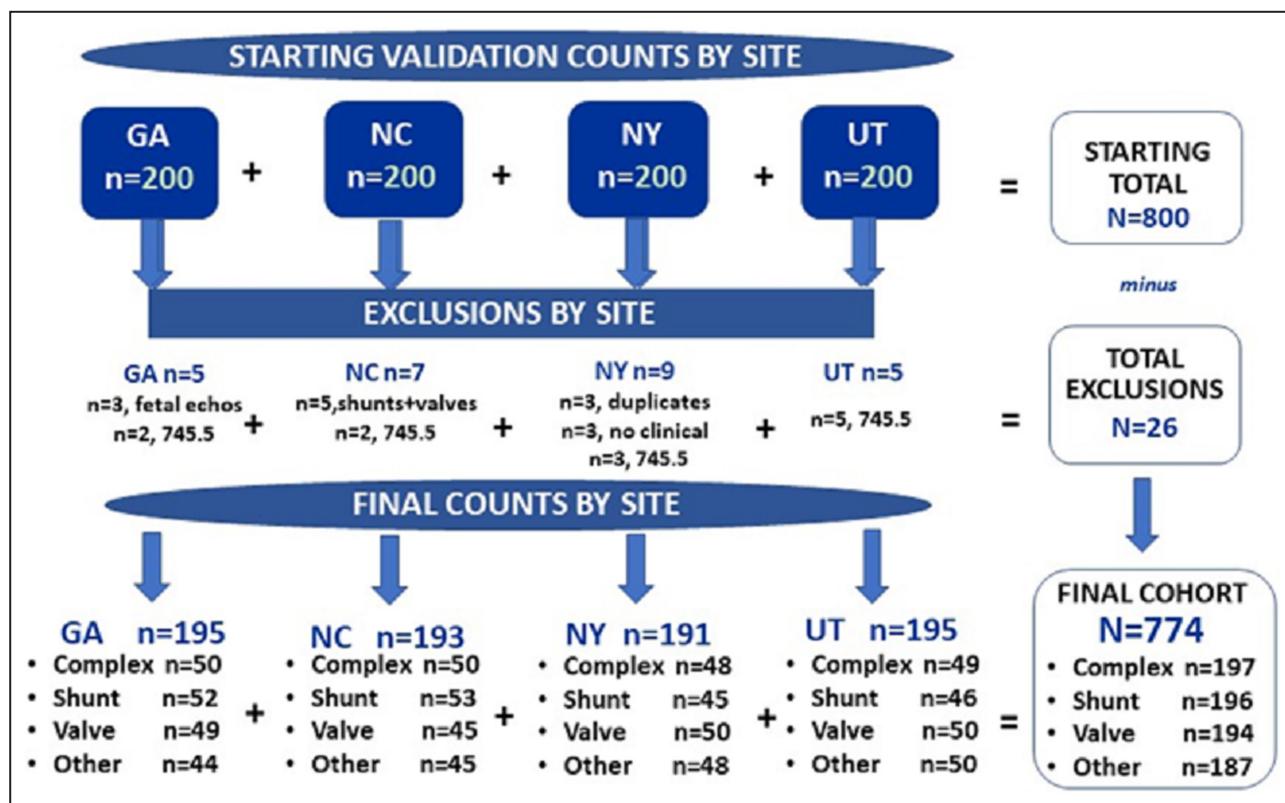


Figure 2. Cohort constructions and exclusions by site and congenital heart defect type.^{‡,§}

[‡]ICD-9-CM code 745.5 was omitted from the shunt group as it is used to indicate secundum atrial septal defect and patent foramen ovale, a normal variant. [§]“Other” congenital heart defect anatomic group consists of unspecific defects; congenital heart defect-related ICD-9-CM codes and their assigned CHD anatomic grouping are displayed in Data S1. GA indicates Emory University in Georgia; ICD-9-CM, International Classification of Diseases, Ninth Revision, Clinical Modification; NC, Duke University in North Carolina; NY, New York State Department of Health; and UT, Utah.

selected, and for the 20- to 64- and >64-year-old groups, 12 cases for each anatomic class were selected. A total of 195 validated GA cases were retained and contributed to the pooled analyses (Figure 2).

For NC, 200 cases, each with at least 2 encounters, were randomly selected from eHRs at one data source which captured patients with CHD statewide. Similar to GA, 13 cases each for the 1- to 10- and 11- to 19-year-old groups, and 12 cases each for the 20- to 64- and >64-year-old groups were selected for review. A total of 193 validated NC cases were included for pooled analyses with other sites’ data (Figure 2).

NY’s sample of 200 cases was composed of patients who resided in 1 of 11 counties (Allegany, Bronx, Cattaraugus, Chautauqua, Erie, Genesee, Monroe, Niagara, Orleans, Westchester, Wyoming) in NY and who had a health care encounter at 1 of 2 clinical data sources. NY, like GA and NC, randomly selected 13 cases for each CHD anatomic class for 1- to 10-year-olds and 11- to 19-year-olds, and 12 cases for the 20- to 64-year-olds and >64-year-old groups. A total of 191 NY cases were retained and contributed to pooled analyses (Figure 2).

In UT, data sources included eHRs from 2 health care systems. While 200 CHD cases were randomly selected and stratified by anatomic groupings with 50 in each category, these classes were further stratified by 2 age groupings, 11- to 19-year-olds and 20- to 64-year-olds, with 25 cases each. Although the UT site did not collect data on individuals aged <10 or >64 years, they contributed a total of 195 validated cases to the pooled, multisite data set (Figure 2).

Statistical Analysis

PPVs for CHD were calculated overall, and by anatomic group, site, and number of unique CHD codes associated with a case. In addition, separate PPV analyses were conducted by sex, race, ethnicity, age group, and anatomic group. For age-specific analyses, since UT did not contribute cases to the youngest age group, 1- to 10-year-olds, or the oldest age group category, >64-year-olds, UT was not included; age-specific analyses included GA, NC, and NY only. For age-specific analyses, PPV was first computed for age groups by sites, and then, also calculated omitting

the “Other” CHD anatomic group, followed by omitting the >64-year-old age group, and finally, by omitting both the “Other” CHD anatomic group and the >64-year-old age group. Analyses for CHD anatomic group proceeded similarly except all 4 sites were included. Lastly, PPVs for having CHD (Yes/No) were calculated for several individual CHD *ICD-9-CM* codes and calculated by number of unique CHD codes.

RESULTS

Table 1 shows number of cases by CHD anatomic groups, number of unique CHD-related *ICD-9-CM* codes (single/multiple) recorded in encounters, followed by demographic characteristics of the sample, overall and by site. Slightly over half (51.16%, n=396) of the 774 cases were female, and 53.49% (n=414) had a single unique CHD-related *ICD-9-CM* code recorded in the medical record. No significant differences were revealed for number of unique *ICD-9-CM* codes by site ($\chi^2=7.14$, $P=0.0675$) or sex by site ($\chi^2=2.77$, $P=0.4278$). No significant differences between age groups by site (GA, NC, and NY) in percent of individuals with one CHD-related *ICD-9-CM* code were observed ($\chi^2=4.04$, $P=0.6706$). Across and by site, however, White race was most prevalent accounting for 66.67% of the sample, followed by Black and other/unknown race, 16.93% and 16.41%, respectively. The contribution of Black cases varied by site, with GA providing almost 40% of the cases, while UT contributed <5% ($\chi^2=69.97$, $P<0.0001$). The majority of the sample were non-Hispanic (70.03%), and 18.86% of the sample’s ethnicity was unknown ($\chi^2=161.38$, $P<0.0001$) (**Table 1**).

The PPVs of having a CHD overall, by site, and by anatomic group with “Other” CHD anatomic group status (included/omitted) are seen in **Table 2**. PPV of having a CHD increased \approx 10 percentage points from 76.36% (591/774; [95% CI, 73.20–79.31]) to 86.54% (508/587 [95% CI, 83.51–89.20]) after omitting the “Other” CHD anatomic group ($\chi^2=22.28$, $P<0.0001$). This pattern was also observed within each site: the increase in PPV ranged from 7.4% for NC (79.79%; 154/193 [95% CI, 73.43–85.22]) to 87.16%; 129/148 [95% CI, 80.68–92.09] (ns)] to 14.6% for NY (70.68%; 135/191 [95% CI, 63.68–77.03] to 85.31%; 122/143 [95% CI, 78.43–90.67]); site-specific Chi-squares are for when “Other” CHD anatomic group was included versus omitted (**Table 2**). When all anatomic groups were combined regardless of whether “Other” CHD anatomic group was included or omitted, significant differences in PPV by site were not observed ($\chi^2=5.15$, $P=0.1612$ and $\chi^2=0.35$, $P=0.9509$, respectively) (**Table 2**). However, there was a significant difference in PPV for having a CHD by anatomic group. Of patients with a code for complex CHD, 89.85%

(177/197 [95% CI, 84.76–93.69]) had a CHD (ie, 10.15% (20/197; 95% CI, 6.31–15.24) did not have CHD); corresponding PPV estimates were 86.73% (170/196 [95% CI, 81.17–91.15]) for shunt, 82.99% (161/194 [95% CI, 76.95–87.99]) for valve, and 44.39% (83/187 [95% CI, 84.76–93.69]) for “Other” CHD anatomic group ($\chi^2=142.16$, $P<0.0001$) (**Table 2**). In Table **S1**, for the complex and shunt categories, all sites reported PPVs >84.00%; for the valve category, UT’s PPV was 76.0% (38/50 [95% CI, 61.83–86.94]) with other sites’ PPVs >80% (89.80%; 44/49 [95% CI, 77.77–96.60]) for GA, 84.44% (38/45 [95% CI, 70.54–93.51]) for NC, and 82.00% (41/50 [95% CI, 68.56–91.42]) for NY. In addition, the PPV of “Other” CHD anatomic group by site ranged from 27.08% (13/48 [95% CI, 15.28–41.85]) in NY to 55.56% (25/45 [95% CI, 40.00–70.36]) in NC (Table **S1**), and when “Other” CHD anatomic group was omitted, PPV for having a CHD did not differ significantly by anatomic group ($\chi^2=3.96$, $P=0.1383$) (**Table 2**). Lastly, when the PPVs of *ICD-9-CM* codes for correct CHD group assignment were assessed, among patients with \geq 1 complex anatomic codes, 74.62% (147/197 [95% CI, 67.94–80.54]) were confirmed to have complex CHD (ie, 25.4% had a shunt, valve, “Other” CHD anatomic grouping, or no CHD); corresponding PPV estimates for identifying a patient with CHD in the correct CHD group were 84.18% (165/196; 95% CI, 87.31–88.99) for shunt, 80.41% (156/194; 95% CI, 74.12–85.75) for valve, and 29.41% (55/187; 95% CI, 22.99–36.50) for “Other” ($\chi^2=168.02$, $P<0.0001$) (Table **S2**).

Table 3 shows PPVs of having a CHD by number of unique CHD-related *ICD-9-CM* codes overall and by anatomic group. Across all anatomic groups, when multiple unique codes were documented across 1 or multiple encounters, PPV was higher, 92.78% (334/360 [95% CI, 90.10–95.45]) compared with when a single CHD code was documented, 62.08% (257/414 [95% CI, 57.40–66.75]) ($\chi^2=100.53$, $P<0.0001$). This association was seen for anatomically complex CHD ($\chi^2=29.88$, $P<0.0001$), valve defects ($\chi^2=13.05$, $P<0.0001$), and the “Other” anatomic group ($\chi^2=5.54$, $P=0.019$). When >1 unique CHD-related *ICD-9-CM* code appeared in the patient’s eHR, PPV of having a CHD ranged from a low of 73.33% (11/15 [95% CI, 44.90–92.21]) in the “Other” CHD anatomic group to a high of 95.60% (152/159 [95% CI, 91.14–98.21]) in the complex group, compared with a range of 41.86% (72/172 [95% CI, 34.40–49.61]) in “Other” CHD anatomic group to 82.95% (73/88 [95% CI, 75.10–90.13]) in the shunt group for cases with a single CHD code.

Table 4 shows PPVs of having a CHD by 4 age groups, 1- to 10-year-olds, 11- to 19-year-olds, 20- to 64-year-olds, and >64-year-olds, by 3 sites (GA, NC, and NY). Compared with age groups \leq 64 years whose PPVs were observed in the mid to lower 80% range

Table 1. Cases Identified With a Congenital Heart Defect *ICD-9-CM* Code: Number of Codes and Demographics, Overall and by Sites

	Overall	Sites*				
		GA	NC	NY	UT	
CHD anatomic groups						
Noncomplex						
Complex	197	50	50	48	49	
Shunt	196	52	53	45	46	
Valve	194	49	45	50	50	
“Other”†	187	44	45	48	50	
Included in analyses	774	195	193	191	195	
Excluded from data set	26	5	7	9	5	
No. of unique CHD-related <i>ICD-9-CM</i> codes for cases						
Single code	414 (53.49%)	114 (58.46%)	94 (48.70%)	111 (58.12%)	95 (48.72%)	
Multiple unique codes	360 (46.51%)	81 (41.54%)	99 (51.30%)	80 (41.88%)	100 (51.28%)	
Demographics						
Age group (in y)						
1–10	50 (8.64%)	18 (9.23%)	11 (5.70%)	21 (10.99)	...	
11–19	156 (26.94%)	52 (26.67%)	52 (26.94%)	52 (27.23%)	...	
20–64	208 (35.92%)	69 (35.38%)	75 (38.86%)	64 (33.51%)	...	
>64	165 (28.50%)	56 (28.72%)	55 (28.50%)	54 (28.27%)	...	
Sex						
Female	396 (51.16%)	105 (53.85%)	94 (48.70%)	91 (47.64%)	106 (54.36%)	
Male	378 (48.84%)	90 (46.15%)	99 (51.30%)	100 (52.36%)	89 (45.64%)	
Race						
Black	131 (16.93%)	51 (26.15%)	35 (18.13%)	43 (22.51%)	<11 (--)	
White	516 (66.67%)	102 (52.31%)	139 (72.02%)	111 (58.12%)	164 (84.10%)	
Other§ and unknown	127 (16.41%)	42 (21.54%)	19 (9.84%)	37 (19.37%)	29 (14.87%)	
Ethnicity						
Hispanic	86 (11.11%)	18 (9.23%)	<11 (--)	44 (23.04%)	17 (8.72%)	
Non-Hispanic	542 (70.03%)	132 (67.69%)	173 (89.64%)	142 (74.35%)	95 (48.72%)	
Unknown	146 (18.86%)	45 (23.08%)	13 (6.74%)	<11 (--)	83 (42.56%)	

Two numbers of unique CHD-related *ICD-9-CM* codes (single code, multiple unique codes); 4 age groups (1–10, 11–19, 20–64, >64) for GA, NC, and NY only; UT omitted as cohort does not include 1–10 or >64-year-olds; 2 sexes (male, female); 3 races (Black, White, Other/unknown); and 3 ethnicities (Hispanic, non-Hispanic, unknown). GA indicates Emory University in Georgia; *ICD-9-CM*, International Classification of Diseases, Ninth Revision, Clinical Modification; NC, Duke University in North Carolina; NY, New York State Department of Health; and UT, University of Utah.

*Site-specific percentages are column percentages; counts <11 not displayed.

†“Other” CHD anatomic group consists of unspecific defects; CHD-related *ICD-9-CM* codes and CHD anatomic groups are identified in Data S1.

‡ χ^2 applies to 5 Chi-square analyses, which includes 4 sites: GA, NC, NY, UT, except the χ^2 by age which omits UT; 2 number of unique CHD-related *ICD-9-CM* codes (single code, multiple unique codes); 4 age groups (1–10, 11–19, 20–64, >64) for GA, NC & NY only; UT omitted as cohort does not include 1 to 10 or >64-year-olds; 2 sexes (male, female); 3 races (Black, White, Other/unknown); and 3 ethnicities (Hispanic, non-Hispanic, unknown).

§“Other” race includes American Indian/Alaskan Native, Asian, native Hawaiian/Pacific Islander, and multiracial.

(86.00%; 43/50 [95% CI, 73.26–95.62]) for 1- to 10-year olds, 82.69% (129/156 [95% CI, 75.83–88.27]) for 11- to 19-year-olds, and 84.13% (175/208 [95% CI, 78.45–88.82]) for 20- to 64-year-olds), the PPV for those who were >64 years was significantly lower, 57.58% (95/165 [95% CI, 49.65–65.22]) ($\chi^2=45.23$, $P<0.0001$). When “Other” CHD anatomic group was omitted, the significant decreasing trend in PPV by age remained, with 95.24% (40/42 [95% CI, 83.84–99.42]) for 1- to 10-year-olds, 94.07% (111/118 [95% CI, 88.16–97.58]) for 11- to

19-year-olds, 93.90% (154/164 [95% CI, 89.07–97.04]) for 20- to 64-year olds, and 66.10% (78/118 [95% CI, 56.81–74.56]) for those >64 years ($\chi^2=58.82$, $P<0.0001$). Table 4 also reveals the overall PPV of having CHD increased from 76.34% (442/579 [95% CI, 72.66–79.74]) when all 4 age and CHD anatomic groups were included to 94.14% (305/324 [95% CI, 72.66–79.74]) after both “Other” CHD anatomic group and the >64-year-olds were omitted ($\chi^2=46.04$, $P<0.0001$). However, when the >64-year-olds were excluded regardless of whether or

Table 2. PPV* of ICD-9-CM CHD Codes for Having a CHD Overall, by Site and by CHD Anatomic Group, with “Other”† CHD Anatomic Group Included and Omitted

PPV* for having a CHD, overall				χ^2 ‡ P value	
“Other”‡ CHD included				22.28 $P<0.0001\$$	
“Other”‡ CHD omitted					
PPV for having a CHD by site					
	GA	NC	NY	UT	
“Other”‡ CHD included	78.46% 153/195 [72.02–84.01]	79.79% 154/193 [73.43–85.22]	70.68% 135/191 [63.68–77.03]	76.41% 149/195 [69.82–82.18]	5.15 $P=0.1612$
“Other”‡ CHD omitted	87.42% 132/151 [81.05–92.25]	87.16% 129/148 [80.68–92.09]	85.31% 122/143 [78.43–90.67]	86.21% 125/145 [79.50–91.37]	0.35 $P=0.9509$
PPV* for having a CHD by anatomic group				χ^2 ‡ P value	
Complex	Noncomplex			“Other”‡ CHD included	
	Shunt§	Valve	“Other”‡	“Other”‡ CHD omitted	
89.85%	86.73%	82.99%	44.39%	142.16 $P<0.0001 $	
177/197 [84.76–93.69]	170/196 [81.17–91.15]	161/194 [76.95–87.99]	83/187 [37.14–51.81]	3.96 $P=0.1383$	

ICD-9-CM indicates International Classification of Diseases, Ninth Revision, Clinical Modification; NC, Duke University in North Carolina; and PPV, positive predictive value.

*95% CI presented within brackets for positive predictive values.

† χ^2 - 5 Chi-square analyses that include CHD overall, by site and by anatomic group: Overall CHD status (2: Yes/No) comparing when “Other” CHD anatomic group was included and omitted; by site—separate χ^2 s conducted when “Other” CHD anatomic group included and omitted: CHD status (2: Yes/No) by site (4: GA, NC, NY, and UT); by anatomic group - separate χ^2 s conducted when: “Other” CHD anatomic group included: CHD status (2: Yes/No) by anatomic group (4: complex, shunt, valve, and “Other” CHD; and 2) “Other” CHD anatomic group omitted: CHD status (2: Yes/No) by anatomic group (3: complex, shunt, valve).

‡“Other” CHD anatomic group consists of unspecific defects; CHD-related ICD-9-CM codes and assigned CHD anatomic group are displayed in Data S1.

§ICD-9-CM code 745.5 was omitted from the shunt group as it is used to indicate secundum atrial septal defect and patent foramen ovale, a normal variant.

|| χ^2 analyses that revealed significance group differences at the $P<0.05$ level or better.

not the “Other” CHD anatomic group was retained in the analysis, no trend in PPV for having a CHD by age was observed (including “Other”: $X^2=0.34$, $P=0.8451$ and excluding “Other”: $X^2=0.11$, $P=0.9467$, respectively). In addition, PPV for having a CHD showed no significant group differences whether “Other” CHD anatomic group was included or omitted for sex ($X^2=1.16$, $P=0.2809$ and $X^2=1.48$, $P=0.2238$, respectively), for race ($X^2=5.62$, $P=0.0601$ and $X^2=0.35$, $P=0.8382$, respectively), or for ethnicity ($X^2=1.24$, $P=0.5381$ and $X^2=2.84$, $P=0.24$, respectively) (Table S3).

PPV of having a CHD for individual ICD-9-CM CHD codes with sufficient case numbers are presented in Table 5. In isolation, tetralogy of Fallot (745.2) had a PPV of 84.2% for having a CHD, ventricular septal defect (745.4) had a PPV of 89.1%, and patent ductus arteriosus had a PPV of 81.3%. However, unspecified anomaly of heart (746.9) had a PPV of 42.2% and other congenital anomaly of heart (746.89) had a PPV of 23.1% for having a CHD. PPV for each of these individual ICD-9-CM CHD-related codes were all >90.0% when multiple unique ICD-9-CM CHD-related codes were identified.

DISCUSSION

The primary aim of this validation project was to evaluate the PPV of ICD-9-CM codes and a code-based anatomic hierarchy for identifying CHD cases, and secondarily, to assess the PPV of these codes by anatomic groups. Overall, 76.36% of patients with ≥ 1 CHD-related ICD-9-CM codes had CHD. Across sites, of those with codes falling into the complex anatomic group, 86.00% to 92.00% had CHD, and only 27.08% to 55.56% of those with codes falling into the “Other” anatomic group had CHD. The PPV of ICD-9-CM codes for CHD was significantly higher for cases aged ≤ 64 years (83.82%) compared with those > 64 years (57.58%). Furthermore, for cases aged ≥ 64 years, PPV of having CHD by anatomic group revealed a low PPV of 36.17% (17/47 [95% CI, 22.67–49.91]) for the “Other” anatomic group and a high of 73.17% (30/41 [95% CI, 59.61–85.78]) for valve lesions. This is consistent with findings from a previous study assessing the accuracy of code 745.5 to predict true atrial septal defects, in which younger patients, ages 11 to 20 years, were

Table 3. PPV* of ICD-9-CM CHD Codes for Having a CHD By Number of Unique CHD Codes and Anatomic Group

	Single CHD code	Multiple unique CHD codes	χ^2 † P value
CHD anatomic group			
Overall	62.08% 257/414 [57.40–66.75]	92.78% 334/360 [90.10–95.45]	100.53 $P<0.0001$
Complex	65.79% 25/38 [48.65–80.37]	95.60% 152/159 [91.14–98.21]	29.88 $P<0.0001$
Shunt‡	82.95% 73/88 [75.10–90.13]	89.81% 97/108 [82.51–94.80]	1.98 $P=0.1590$
Valve	75.00% 87/116 [66.11–82.57]	94.87% 74/78 [87.39–98.59]	13.05 $P=0.0003$
“Other”§	41.86% 72/172 [34.40–49.61]	73.33% 11/15 [44.90–92.21]	5.54 $P=0.0186$

CHD indicates congenital heart defect; ICD-9-CM, International Classification of Diseases, Ninth Revision, Clinical Modification; and PPV, positive predictive value.

*95% CI presented within brackets [] for positive predictive values.

†ICD-9-CM code 745.5 was omitted from the shunt group as it is used to indicate secundum atrial septal defect and patent foramen ovale, a normal variant.

‡“Other” CHD anatomic group consists of unspecific defects; CHD-related ICD-9-CM codes and their assigned CHD anatomic grouping are displayed in Data S1.

more likely to have a true atrial septal defect (64.3%) compared with those aged 21 to 40 years (23.7%) or 41 to 64 years (19.0%) ($P<0.001$).² Using ICD codes in a “rule-out CHD” manner and incorrect coding for CHD may play a role in the low PPV for older CHD cases. Age-related changes in valve morphology and function may mistakenly be coded as CHD. The “Other” CHD category includes non-specific CHD codes such as “other congenital anomaly of the heart” that may be used inappropriately, for example, when ordering studies to investigate heart murmurs and symptoms of heart failure in older patients.

When examining the accuracy of ICD-9-CM-based native CHD anatomy group, about two thirds of cases across all sites were categorized into the correct anatomic group and the other one-third either did not have CHD or were categorized into the wrong group based on an incorrect code. The PPV of ICD-9-CM codes contained in eHR systems for identifying that a case truly has CHD of the indicated anatomic group ranged from 84.18% for shunt defect codes, excluding 745.5, to 29.41% for the “Other” CHD anatomic code group. Nonspecific codes in the “Other” CHD category may have a low PPV because they may have been used as “rule-out” codes when ordering diagnostic tests or were misused for an acquired heart condition. For example, if a patient has low blood oxygen saturations, then a “rule out CHD” code such as 746.9 (unspecified congenital anomaly of heart) may be used to assess

for CHD. A comparative PPV analysis for case classification revealed a lower PPV for single versus multiple code classification, 62.08% versus 92.78%, respectively (Table 3). While requiring multiple codes for case inclusion may likely avoid “rule out” code misclassification, it often comes at the expense of losing true CHD cases. In this study, relying on multiple codes alone would have resulted in reducing sensitivity by 33.2% (257/774) because 257 single-coded true CHD cases would be excluded. Additionally, reasons for the assignment of an incorrect anatomic group may also include coding errors such as when a code in the 745.xx to 747.xx ICD-9-CM group has a decimal dropped, leading to conversion from 745 to 745.0, a code for truncus arteriosus (745.0), rather than an intended code, for instance 745.4, ventricular septal defect. Misuse of CHD codes may occur in other ways. For example, this study included 3 cases that were coded as “other septal defect” rather than ventricular septal defect. These coding errors highlight that nonspecific CHD ICD-9-CM codes may actually represent specific and true CHD that could have been better captured by a more specific CHD code.

Of the 774 total cases, 414 had a single CHD code and 360 had multiple unique CHD codes. Patients with more than one unique CHD code were more likely to have CHD, and this increase in PPV was significant for the complex and valve anatomic groups. Certain ICD-9-CM codes that frequently occur in combination with other CHD codes had PPVs >90%. For instance, tetralogy of Fallot (745.2) and ventricular septal defect (745.4) all had high PPVs for CHD in combination with other CHD codes.

Prior studies have also examined whether complexity of disease may play a role in the accuracy of coding in administrative data. Steiner et al. (2017) found certain complex CHD codes, such as tetralogy of Fallot and truncus arteriosus, performed well.⁹ Khan et al. (2018) found that PPV of moderate or complex CHD codes for CHD was higher compared with simple shunt or valve defects or when coupled with other factors such as age, an encounter with an adult CHD provider, an echocardiogram or ECG compared with noncomplex CHD.¹ The current validation project found that the complex, shunt, and valve groups had higher PPV than the “Other” CHD group, supporting the conclusion that more specific diagnoses have a higher PPV for true CHD.

While this study did not examine sensitivity of administrative databases, other studies have demonstrated poor identification of CHD in state-specific administrative databases for pediatric patients.³ Cronk et al. (2003) investigated the sensitivity of ICD-9 codes for identifying individuals with CHD in 4 administrative databases in Wisconsin and found that only 57.9% of CHD cases identified at Children’s Hospital

Table 4. PPV* of ICD-9-CMCHD Codes† for Having a CHD, Overall and by Age Group, for Three Sites with “Other”‡ CHD Anatomic Group Included and Omitted

PPV* for having a CHD, for 4 age groups					χ^2 ‡, § P value
“Other”‡ CHD included					17.19 $P<0.0001$
“Other”‡ CHD omitted					
PPV* for having CHD by age groups					χ^2 ‡, § P value
	1–10y	11–19y	20–64y	>64y	All 4 age groups
“Other”‡ CHD included	86.00% 43/50 [73.26–95.62]	82.69% 129/156 [75.83–88.27]	84.13% 175/208 [78.45–88.82]	57.58% 95/165 [49.65–65.22]	45.23 $P<0.0001$
“Other”‡ CHD omitted	95.24% 40/42 [83.84–99.42]	94.07% 111/118 [88.16–97.58]	93.90% 154/164 [89.07–97.04]	66.10% 78/118 [56.81–74.56]	58.82 $P<0.0001$
PPV* for having A CHD, for 3 age groups					χ^2 ‡, ‡ P value
“Other”‡ CHD included					18.80 $P<0.0001$
“Other”‡ CHD omitted					

χ^2 for having a CHD when all 4 age groups and CHD anatomic groups were included (PPV=76.34%) compared with when >64-year-olds and “Other” CHD anatomic group were omitted (PPV=94.14%) was $\chi^2=46.04$, $P<0.0001$. CHD indicates congenital heart defect; ICD-9-CM, International Classification of Diseases, Ninth Revision, Clinical Modification; and PPV, positive predictive value.

*95% CI presented within brackets [] for positive predictive values.

†ICD-9-CM code 745.5 was omitted from the shunt group as it is used to indicate secundum atrial septal defect and patent foramen ovale, a normal variant.

‡“Other” CHD anatomic group consists of unspecific defects; CHD-related ICD-9-CM codes and assigned CHD anatomic group are displayed in Data S1.

§UT not included in these analyses as age groups 1–10 and >64 years were not reported.

of Wisconsin were identified by ICD-9 codes and/or a CHD checkbox indicative of the presence of CHD in any state database; a total of 216 cases (57.9%) of the 373 total cases were identified by a CHD ICD-9 code in at least one of the 4 state databases.¹⁰ Almost 62% (231 of 373) had a single CHD diagnosis and 91% (339 of 373) had 1 or 2 CHD diagnoses. Lack of reporting oversight and classification problems were thought to contribute to inadequate identification of CHD from such data sets.

Ideally, constructing an algorithm that utilizes data from administrative and eHRs will help to identify true CHD cases with improved accuracy and sensitivity. Machine learning is one possibility to improve both the PPV and sensitivity of CHD codes for CHD, and specific CHD type, based on ICD codes and other variables from clinical and administrative data sets, using the most predictive factors, while reducing the false negative rate. Restricting analyses to certain ICD codes and age categories may also improve the PPV. As a result of the current findings and previous research,^{2,11} researchers may consider excluding the

following codes from administrative data sets to improve the PPV for analyses seeking to study CHD:

- ICD-9-CM and anatomically equivalent ICD-10-CM codes that code for conditions other than heart defects, including congenital heart block (746.86, Q24.6), pulmonary arteriovenous malformation (747.32, Q25.72), absent/hypoplastic umbilical artery (747.5, Q27.0), other anomalies of peripheral vascular system (747.6x, Q27.9), and other specified anomalies of circulatory system (747.8x, Q28.8). These codes were also excluded in the surveillance methodology.^{5,6}
- The “Other” CHD anatomic group (as noted in Figure 1 and Data S1) and nonspecific ICD-9-CM codes and the equivalent ICD-10-CM codes including: other specified congenital anomalies of heart 746.8 (Q24.8), obstructive anomalies of heart, not elsewhere classified 746.84, coronary artery anomaly 746.85 (Q24.5), malposition of heart and cardiac apex 746.87 (Q24.0), other specified congenital anomalies of heart 746.89 (Q24.8), unspecified

Table 5. Positive Predictive Value of Specific CHD ICD-9-CM Codes for Having CHD by Number of Select Unique CHD Codes

ICD-9-CM code	Description of ICD-9-CM code	CHD anatomic group	Patients, n			PPV for having CHD		
			Single CHD code	Multiple unique CHD codes	Total # with code	Single CHD code	Multiple Unique CHD codes	Total % with code
745.2	Tetralogy of Fallot	Complex	19	53	72	84.2%	94.3%	91.7%
745.4	Ventricular septal defect	Shunt	64	93	157	89.1%	94.6%	92.4%
747.0	Patent ductus arteriosus	Shunt	16	36	52	81.3%	94.4%	90.4%
746.4	Bicuspid aortic valve and congenital aortic valve insufficiency	Valve	61	37	98	82.0%	97.3%	87.8%
746.02	Congenital pulmonary valve stenosis	Valve	10	37	47	100.0%	100.0%	100.0%
746.3	Congenital stenosis of aortic valve	Valve	16	24	40	87.5%	95.8%	92.5%
747.1	Coarctation of aorta	Valve	12	26	38	66.7%	100.0%	89.5%
746.9	Unspecified anomaly of heart	Other	45	89	134	42.2%	94.4%	76.9%
746.89	Other congenital anomaly of heart	Other	26	39	65	23.1%	92.3%	64.6%
V13.65	Personal history of corrected congenital Malformation of heart and circulatory system	Other	11	38	49	54.5%	92.1%	83.7%
746.85	Coronary artery anomaly	Other	26	12	38	61.5%	91.7%	71.1%
746.87	Malposition of heart and cardiac apex	Other	12	15	27	33.3%	93.3%	66.7%
747.29	Other anomaly of aorta	Other	14	12	26	35.7%	91.7%	61.5%

CHD indicates congenital heart defect; ICD-9-CM, International Classification of Diseases, Ninth Revision, Clinical Modification; and PPV, positive predictive value.

congenital anomaly of heart 746.9 (Q24.9), other congenital anomalies of aorta 747.2 (Q25.49), anomaly of aorta, unspecified 747.20 (Q25.40), anomalies of aortic arch 747.21, other anomalies of aorta 747.29 (Q25.49), congenital anomalies of great veins 747.4 (Q26.1), anomaly of great veins, unspecified 747.40 (Q26.9), other anomalies of great veins 747.49 (Q26.8), unspecified anomaly of circulatory system 747.9 (Q28.9, personal history of [corrected] congenital malformations of heart and circulatory system V13.65 (Z87.74)

- Patent foramen ovale / atrial septal defect 745.5 / Q21.1 alone or in combination with the “Other” CHD category.

Codes found in combination are likely to be more accurate than an isolated CHD code, though 53% (414/774) of the cases in this validation project had a single unique CHD code. Therefore, it is suggested to avoid excluding cases with single CHD codes from algorithms and analyses; doing so could result in missed cases and lack of generalizability. Additionally, severity of CHD is not indicated by the number of unique CHD codes applied to an individual. It is notable that

eliminating the ICD-9-CM codes above, or eliminating the >64-year-old group, will improve the PPV of CHD, but will also exclude a substantial number of true CHD cases. Certain codes are more likely to identify true CHD cases than others. Even though the vast majority of cases with complex and moderate CHD codes (80%–90%) have CHD, there remains about 1 in 4 or 5 cases who may either not have CHD or have a different severity type than what is coded. Sufficiently detailing how CHD is defined when using administrative data as well as understanding and documenting its limitations will improve the generalizability of the findings to the CHD population.

LIMITATIONS

The selected cases used in this analysis came from health care centers at locations where records could be reviewed. Thus, there was possible selection bias towards having true CHD, which might overestimate the PPV. However, using data from 2011 to 2013 may lead to an underestimation of PPV compared with more recent years as eHR use has become more standard. Coding

may vary by data source, year, and individuals who document the code (medical versus billing department staff), potentially limiting broad applicability to other data sets. Coding practices vary across both regions and medical centers, which is both a strength and limitation of our paper. Our data set used *ICD-9-CM* codes, which can be mapped to *ICD-10-CM* codes. However, differences in coding practices may vary between the *ICD-9-CM* and *ICD-10-CM* eras, thus making our results not directly applicable to *ICD-10-CM* based data sets. For the individual *ICD-9-CM* codes assessed, the reported PPV is for having CHD (Yes/No) and we were unable to examine PPV for whether the case had the specific CHD documented. We did not have access to false negatives cases in the data available and thus could not calculate sensitivities of CHD *ICD-9-CM* codes.

CONCLUSIONS

This validation study using data from 4 sites affiliated with the CDC's Surveillance of Congenital Heart Disease Across the Lifespan project revealed that *ICD-9-CM* codes accurately classify patients with true CHD in the majority of cases labeled as complex, shunt, and valve. The PPV of "Other" non-specific CHD *ICD-9-CM* codes was low and may not reflect true CHD. When >1 unique CHD code is associated with a case, the PPV for CHD increases. CHD codes had higher PPV in younger compared with older CHD cases. Further evaluation and algorithm development may help inform and improve the identification of CHD cases when administrative data sets are used.

ARTICLE INFORMATION

Received December 20, 2021; accepted May 17, 2022.

Affiliations

Division of Cardiology, Emory University School of Medicine and Emory University Rollins School of Public Health, Atlanta, GA (F.H.R., C.L.R., T.H., W.M.B.); Sibley Heart Center Cardiology, Atlanta, GA (F.H.R.); National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, GA (S.L.F., J.G.); Duke University School of Medicine, Durham, NC (J.S.L., A.D.); Division of Medical Genetics, Department of Pediatrics, University of Utah, Salt Lake City, UT (L.B., M.R.R.); Albert Einstein College of Medicine and The Children's Hospital at Montefiore, Bronx, NY (D.H.); Divisions of Cardiovascular Medicine and Pediatric Cardiology, Stanford University School of Medicine, Palo Alto, CA (G.K.L.); and New York State Department of Health, Bureau of Environmental & Occupational Epidemiology, Albany, NY (A.M.S.).

Acknowledgments

We thank the Pedigree and Population Resource of Huntsman Cancer Institute, University of Utah (funded in part by the Huntsman Cancer Foundation) for its role in the ongoing collection, maintenance, and support of the Utah Population Database. We also acknowledge partial support for the Utah Population Database through grant P30 CA2014 from the National Cancer Institute, University of Utah, and from the University of Utah's program in Personalized Health and Center for Clinical and Translational Science. We thank the University of Utah Center for Clinical and Translational Science (funded by National Institutes of Health Clinical

and Translational Science Awards), the Pedigree and Population Resource, University of Utah Information Technology Services and Biomedical Informatics Core for establishing the Master Subject Index between the Utah Population Database, the University of Utah Health Sciences Center and Intermountain Healthcare. This analysis has undergone replication by Anaclare Sullivan.

Sources of Funding

This work was supported by the CDC Cooperative Agreement, Surveillance of Congenital Heart Defects Across the Lifespan; Funding Opportunity Announcements #DD15-1506. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the CDC.

Disclosures

None.

Supplemental Material

Data S1

Tables S1–S3

REFERENCES

- Khan A, Ramsey K, Ballard C, Armstrong E, Burchill LJ, Menashe V, Pantely G, Broberg CS. Limited accuracy of administrative data for the identification and classification of adult congenital heart disease. *J Am Heart Assoc.* 2018;7:e007378. doi: 10.1161/JAHA.117.007378
- Rodriguez FH III, Ephrem G, Gerardin JF, Raskind-Hood C, Hogue C, Book W. 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 Dis.* 2018;13:59–64. doi: 10.1111/chd.12563
- Agarwal S, Sud K, Menon V. Nationwide hospitalization trends in adult congenital heart disease across 2003–2012. *J Am Heart Assoc.* 2016;5:e002330. doi: 10.1161/JAHA.115.002330
- Broberg C, McLarry J, Mitchell J, Winter C, Doerne J, Woods P, Burchill L, Weiss J. Accuracy of administrative data for detection and categorization of adult congenital heart disease patients from an electronic medical record. *Pediatr Cardiol.* 2015;36:719–725. doi: 10.1007/s00246-014-1068-2
- Glidewell J, Book W, Raskind-Hood C, Hogue C, Dunn JE, Gurvitz M, Ozonoff A, McGarry C, Van Zutphen A, Lui G, et al. Population-based surveillance of congenital heart defects among adolescents and adults: surveillance methodology. *Birth Defects Res.* 2018;110:1395–1403. doi: 10.1002/bdr2.1400
- Glidewell JM, Farr SL, Book WM, Botto L, Li JS, Soim AS, Downing KF, Riehle-Colarusso T, D'Ottavio AA, Feldkamp ML, et al. Individuals aged 1–64 years with documented congenital heart defects at healthcare encounters, five U.S. surveillance sites, 2011–2013. *Am Heart J.* 2021;238:100–108. doi: 10.1016/j.ahj.2021.04.007
- Marelli AJ, Mackie AS, Ionescu-Ittu R, Rahme E, Pilote L. Congenital heart disease in the general population: changing prevalence and age distribution. *Circulation.* 2007;115:163–172. doi: 10.1161/CIRCULATIONAHA.106.627224
- Marelli AJ, Ionescu-Ittu R, Mackie AS, Guo L, Dendukuri N, Kaouache M. Lifetime prevalence of congenital heart disease in the general population from 2000 to 2010. *Circulation.* 2014;130:749–756. doi: 10.1161/CIRCULATIONAHA.113.008396
- Steiner JM, Kirkpatrick JN, Heckbert SR, Habib A, Sibley J, Lober W, Randall J, Curtis JR. Identification of adults with congenital heart disease of moderate or great complexity from administrative data. *Congenital Heart Dis.* 2017;13:65–71. doi: 10.1111/chd.12524
- Cronk CE, Malloy ME, Pelech AN, Miller RE, Meyer SA, Cowell M, McCarver DG. Completeness of state administrative databases for surveillance of congenital heart disease. *Birth Defects Res Part A: Clin Mol Teratol.* 2003;67:597–603. doi: 10.1002/bdra.10107
- Lui GK, McGarry C, Bhatt A, Book W, Riehle-Colarusso TJ, Dunn JE, Glidewell J, Gurvitz M, Hoffman T, Hogue C, et al. Surveillance of congenital heart defects among adolescents at three U.S. sites. *Am J Cardiol.* 2019;124:137–143. doi: 10.1016/j.amjcard.2019.03.044

Supplemental Materials

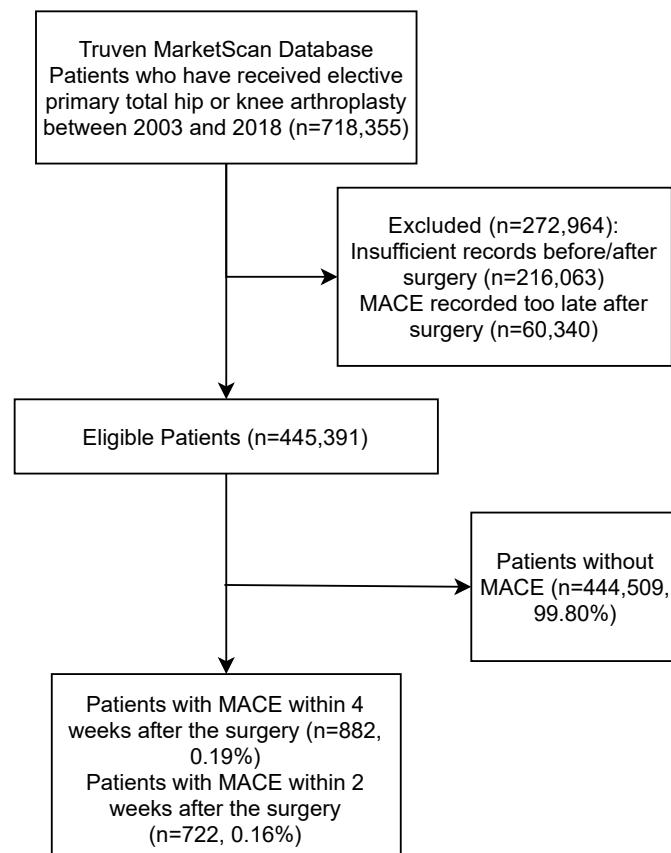


Fig. S1: CONSORT diagram conforming to the CONSORT-AI Extension guidelines stated at [https://doi.org/10.1016/S2589-7500\(20\)30218-1](https://doi.org/10.1016/S2589-7500(20)30218-1)

TABLE S1: Inclusion/Exclusion, Positive/Control Criteria & Cohort Definitions

	Definitions
Inclusion/Exclusion Criteria	<p>Age 45 - 95</p> <p>Has total hip/knee CPT codes (See Table S2) in medical history and length of history available before cardiac event spans ≥ 1 year</p> <p>Has a myocardial infarction or a cardiac arrest[‡] (See Table S4 for list of target codes used to identify cardiac event in diagnostic history) 4 weeks (2 weeks considered in secondary analysis) after surgery (positive cohort)</p> <p>Has 0.5 yr of medical history available after surgery (control)</p>
Positive & Control Cohorts	<p>Positive Cohort: At least one code for cardiac event (Table S4)</p> <p>Control Cohort: No code on cardiac event within 26 weeks of surgery</p>

TABLE S2: Current Procedural Terminology (CPT) codes for total hip/knee replacement used

CPT code	description
27130	Total Hip Replacement/Resurfacing
27132	Total Hip Replacement/Resurfacing
81.51	Total hip replacement
0SR9	Replacement: Hip Joint, Right
0SRB	Replacement: Hip Joint, Left
27442	Knee Total Replacement - (Arthroplasty)
27443	Knee Total Replacement - (Arthroplasty)
27445	Knee Total Replacement - (Arthroplasty)
27446	Knee Total Replacement - (Arthroplasty)
27447	Knee Total Replacement - (Arthroplasty)
81.54	Total knee replacement
0SRC	Replacement: Knee Joint, Right
0SRD	Replacement: Knee Joint, Left

TABLE S3: Codes used to determine RCRI

Description	Constituent Codes (*NDC: National Drug Code)
History of Heart Failure	ICD9 codes: 428.0, 428.1, 428.20, 428.21, 428.22, 428.23, 428.30, 428.31, 428.32, 428.33, 428.40, 428.41, 428.42, 428.43, 428.9, 428.2, 428.3, 428.4; ICD10 codes: I50.1, I50.20, I50.21, I50.22, I50.23, I50.30, I50.31, I50.32, I50.33, I50.40, I50.41, I50.42, I50.43, I50.810, I50.811, I50.812, I50.813, I50.814, I50.82, I50.83, I50.84, I50.89, I50.9, I50.2, I50.3, I50.4, I50.8.
History of Cerebrovascular Disease	ICD9 codes: 430, 431, 432.0, 432.1, 432.9, 433.00, 433.01, 433.10, 433.11, 433.20, 433.21, 433.30, 433.31, 433.80, 433.81, 433.90, 433.91, 434.00, 434.01, 434.10, 434.11, 434.90, 434.91, 435.0, 435.1, 435.2, 435.3, 435.8, 435.9, 436, 437.0, 437.1, 437.2, 437.3, 437.4, 437.5, 437.6, 437.7, 437.8, 437.9, 438.0, 438.10, 438.11, 438.12, 438.13, 438.14, 438.19, 438.20, 438.21, 438.22, 438.30, 438.31, 438.32, 438.40, 438.41, 438.42, 438.50, 438.51, 438.52, 438.53, 438.6, 438.7, 438.81, 438.82, 438.83, 438.84, 438.85, 438.89, 438.9, 432, 433, 434, 435, 437, 438, 433.0, 433.1, 433.2, 433.3, 433.8, 433.9, 434.0, 434.1, 434.9, 438.1, 438.2, 438.3, 438.4, 438.5, 438.8; ICD10 codes: I60.00, I60.01, I60.02, I60.10, I60.11, I60.12, I60.2, I60.30, I60.31, I60.32, I60.4, I60.50, I60.51, I60.52, I60.6, I60.7, I60.8, I60.9, I61.0, I61.1, I61.2, I61.3, I61.4, I61.5, I61.6, I61.8, I61.9, I62.00, I62.01, I62.02, I62.03, I62.1, I62.9, I63.00, I63.011, I63.012, I63.013, I63.019, I63.02, I63.031, I63.032, I63.033, I63.039, I63.09, I63.10, I63.111, I63.112, I63.113, I63.119, I63.12, I63.131, I63.132, I63.133, I63.139, I63.19, I63.20, I63.211, I63.212, I63.213, I63.219, I63.22, I63.231, I63.232, I63.233, I63.239, I63.29, I63.30, I63.311, I63.312, I63.313, I63.319, I63.321, I63.322, I63.323, I63.329, I63.331, I63.332, I63.333, I63.339, I63.341, I63.342, I63.343, I63.349, I63.39, I63.40, I63.411, I63.412, I63.413, I63.419, I63.421, I63.422, I63.423, I63.429, I63.431, I63.432, I63.433, I63.439, I63.441, I63.442, I63.443, I63.449, I63.49, I63.50, I63.511, I63.512, I63.513, I63.519, I63.521, I63.522, I63.523, I63.529, I63.531, I63.532, I63.533, I63.539, I63.541, I63.542, I63.543, I63.549, I63.59, I63.6, I63.81, I63.89, I63.9, I65.01, I65.02, I65.03, I65.09, I65.1, I65.21, I65.22, I65.23, I65.29, I65.8, I65.9, I66.01, I66.02, I66.03, I66.09, I66.11, I66.12, I66.13, I66.19, I66.21, I66.22, I66.23, I66.29, I66.3, I66.8, I66.9, I67.0, I67.1, I67.2, I67.3, I67.4, I67.5, I67.6, I67.7, I67.81, I67.82, I67.83, I67.841, I67.848, I67.850, I67.858, I67.89, I67.9, I68.0, I68.2, I68.8, I69.00, I69.010, I69.011, I69.012, I69.013, I69.014, I69.015, I69.019, I69.020, I69.021, I69.022, I69.023, I69.028, I69.031, I69.032, I69.033, I69.034, I69.039, I69.041, I69.042, I69.043, I69.044, I69.049, I69.051, I69.052, I69.053, I69.054, I69.059, I69.061, I69.062, I69.063, I69.064, I69.065, I69.069, I69.090, I69.091, I69.092, I69.093, I69.098, I69.10, I69.110, I69.111, I69.112, I69.113, I69.114, I69.115, I69.118, I69.119, I69.120, I69.121, I69.122, I69.123, I69.128, I69.131, I69.132, I69.133, I69.134, I69.139, I69.141, I69.142, I69.143, I69.144, I69.149, I69.151, I69.152, I69.153, I69.154, I69.159, I69.161, I69.162, I69.163, I69.164, I69.165, I69.169, I69.190, I69.191, I69.192, I69.193, I69.198, I69.20, I69.210, I69.211, I69.212, I69.213, I69.214, I69.215, I69.218, I69.219, I69.220, I69.221, I69.222, I69.223, I69.228, I69.231, I69.232, I69.233, I69.234, I69.239, I69.241, I69.242, I69.243, I69.244, I69.249, I69.251, I69.252, I69.253, I69.254, I69.259, I69.261, I69.262, I69.263, I69.264, I69.265, I69.269, I69.290, I69.291, I69.292, I69.293, I69.298, I69.30, I69.310, I69.311, I69.312, I69.313, I69.314, I69.315, I69.318, I69.319, I69.320, I69.321, I69.322, I69.323, I69.328, I69.331, I69.332, I69.333, I69.334, I69.339, I69.341, I69.342, I69.343, I69.344, I69.349, I69.351, I69.352, I69.353, I69.354, I69.359, I69.361, I69.362, I69.363, I69.364, I69.365, I69.369, I69.390, I69.391, I69.392, I69.393, I69.398, I69.80, I69.810, I69.811, I69.812, I69.813, I69.814, I69.815, I69.818, I69.819, I69.820, I69.821, I69.822, I69.823, I69.828, I69.831, I69.832, I69.833, I69.834, I69.839, I69.841, I69.842, I69.843, I69.844, I69.849, I69.851, I69.852, I69.853, I69.854, I69.859, I69.861, I69.862, I69.863, I69.864, I69.865, I69.869, I69.890, I69.891, I69.892, I69.893, I69.898, I69.90, I69.910, I69.911, I69.912, I69.913, I69.914, I69.915, I69.918, I69.919, I69.920, I69.921, I69.922, I69.923, I69.928, I69.931, I69.932, I69.933, I69.934, I69.939, I69.941, I69.942, I69.943, I69.944, I69.949, I69.951, I69.952, I69.953, I69.954, I69.959, I69.961, I69.962, I69.963, I69.964, I69.965, I69.969, I69.990, I69.991, I69.992, I69.993, I69.998, I60, I61, I62, I63, I65, I66, I67, I68, I69, I60.0, I60.1, I60.3, I60.5, I62.0, I63.0, I63.1, I63.2, I63.3, I63.4, I63.5, I63.8, I65.0, I65.2, I66.0, I66.1, I66.2, I67.8, I69.0, I69.1, I69.2, I69.3, I69.8, I69.9, I63.01, I63.03, I63.11, I63.13, I63.21, I63.23, I63.31, I63.32, I63.33, I63.34, I63.41, I63.42, I63.43, I63.44, I63.51, I63.52, I63.53, I63.54, I67.84, I67.85, I69.01, I69.02, I69.03, I69.04, I69.05, I69.06, I69.09, I69.11, I69.12, I69.13, I69.14, I69.15, I69.16, I69.19, I69.21, I69.22, I69.23, I69.24, I69.25, I69.26, I69.29, I69.31, I69.32, I69.33, I69.34, I69.35, I69.36, I69.39, I69.81, I69.82, I69.83, I69.84, I69.85, I69.86, I69.89, I69.91, I69.92, I69.93, I69.94, I69.95, I69.96, I69.99
History of Ischemic Heart Disease	ICD9 codes: 410.00, 410.01, 410.02, 410.10, 410.11, 410.12, 410.20, 410.21, 410.22, 410.30, 410.31, 410.32, 410.40, 410.41, 410.42, 410.50, 410.51, 410.52, 410.60, 410.61, 410.62, 410.70, 410.71, 410.72, 410.80, 410.81, 410.82, 410.90, 410.91, 410.92, 411.0, 411.1, 411.81, 411.89, 412, 413.0, 413.1, 413.9, 414.00, 414.01, 414.02, 414.03, 414.04, 414.05, 414.06, 414.07, 414.10, 414.11, 414.12, 414.19, 414.2, 414.3, 414.4, 414.8, 414.9, 410, 411, 413, 414, 410.0, 410.1, 410.2, 410.3, 410.4, 410.5, 410.6, 410.7, 410.8, 410.9, 411.8, 414.0, 414.1; ICD10 codes: I20.0, I20.1, I20.8, I20.9, I21.01, I21.02, I21.09, I21.11, I21.19, I21.21, I21.29, I21.3, I21.4, I21.9, I21.1, I21.21, I21.29, I22.1, I22.2, I22.8, I22.9, I23.0, I23.1, I23.2, I23.3, I23.4, I23.5, I23.6, I23.7, I23.8, I24.0, I24.1, I24.8, I24.9, I25.10, I25.11, I25.118, I25.119, I25.2, I25.3, I25.41, I25.42, I25.5, I25.6, I25.700, I25.701, I25.708, I25.709, I25.710, I25.711, I25.718, I25.719, I25.720, I25.721, I25.728, I25.729, I25.730, I25.731, I25.738, I25.739, I25.750, I25.751, I25.758, I25.759, I25.760, I25.761, I25.768, I25.769, I25.790, I25.791, I25.798, I25.799, I25.810, I25.811, I25.812, I25.82, I25.83, I25.84, I25.89, I25.9, I20, I21, I22, I23, I24, I25, I21.0, I21.1, I21.2, I21.A, I25.1, I25.4, I25.7, I25.8, I25.11, I25.70, I25.71, I25.72, I25.73, I25.75, I25.76, I25.79, I25.81
Pre-operative creatinine > 2 mg/dL / 176.8 µmol/L - Approximated by History of Chronic Kidney Disease	ICD9 codes: 585.3, 585.5, 585.6, 585.4; ICD10 codes: N18.30, N18.31, N18.32, N18.4, N18.5, N18.6, N18.3
Pre-operative treatment with Insulin	NDC* codes: 08881242112, 08881250305, 54868582400, 08881750023, 08881242120, 08881250313, 38396043277, 08881250321, 00169369619, 08881520665, 08881242138, 36652040218, 56151171101, 08881520673, 08496275501, 08881250354, 08496275511, 08881250362, 08808010055, 36652040276, 00002831101, 08080040028, 08080040029, 08080040030, 08396800100, 383968043377, 08881750130, 36652040318, 00002831517, 561511711201, 08881750155, 08881512597, 36652400801, 36652400802, 36652400803, 36652400804, 36652400805, 36652400806, 36652400807, 36652400808, 68258889903, 08881103025, 36652040376, 96295010494, 96295010495, 96295010496, 96295010497, 96295010498, 08881512647, 08396800200, 38396043477, 57515008218, 08881750239, 561511711301, 08881750254, 55948009710, 08881250545, 57515008258, 36652400901, 36652400902, 36652400903, 36652400904, 36652400905, 088080032010, 36652400906, 36652400907, 36652400908, 59060183302, 08881512738, 08881512746, 08396800300, 54569165101, 08881701166, 54569165102, 38396706339, 08881701174, 54274048310, 08290328888, 38396043577, 08222073150, 08881750338, 08222032195, 96295010629, 36652040518, 08881676624, 96295010643, 08881906005, 96295010645, 08881676632, 08881701216, 00002821001, 08881701224, 59060183402, 08881512811, 08080032110, 08222073198, 36652400576, 89134072200, 00002831501, 08396800400, 54569165200, 54569165202, 08881512852, 54274048410, 08881512860, 38396043677, 51927368100, 36652040618, 08881512878, 08881906104, 57515082180, 08080327114, 08080032210, 36652040676, 08881750510, 52297086578, 08881512944, 08287126003, 08287126004,

Continued on next page

Continued on next page

Continued on next page

Pre-operative treatment with Insulin (contd.)	NDC* codes: 08881533254, 08881566023, 08881500493, 08881500501, 00002834001, 08881820008, 00069072437, 08881516911, 00169030001, 0808020005, 08881713534, 08271352200, 08881516937, 08080200077, 08881533338, 08290324906, 00002751201, 08290324907, 08290324908, 08290324909, 08290324910, 08290324911, 08290324912, 08290841002, 08290841003, 96295104940, 96295104950, 76300022201, 96295104960, 24385098178, 55948022210, 96295104970, 00002841201, 00002871759, 96295104980, 98302013919, 98302013920, 98302013930, 98302013931, 08271352300, 98302013932, 98302013933, 98302013934, 98302013935, 98302013936, 98302013937, 98302013938, 98302013939, 98302013940, 98302013941, 08881050116, 08881050121, 08290841101, 08290841103, 08080323100, 08881050140, 00003183410, 00069005019, 24385098278, 08080323112, 00003244110, 08080020028, 08080020029, 08080020030, 00002821101, 00002831201, 08881533510, 08470205001, 08881050197, 76300071511, 76300071512, 76300071513, 76300071514, 98302014059, 08290841201, 08290841203, 08881640059, 68115083910, 24385098378, 55948022410, 08080528012, 55045362401, 08080823012, 24385098478, 00169022201, 08080823034, 08881713914, 87701074720, 87701074723, 87701074745, 08080823100, 08881075013, 08214352714, 08214352719, 55948022610, 55283061600, 00002751017, 08080200555, 08881075051, 08881247087, 08222094193, 86227080055, 08881075070, 08881247111, 38290846603, 08881247129, 08881247137, 38396040105, 86227080105, 38396040108, 00002751601, 54686136900, 08881533908, 08881026008, 08881026013, 08881501160, 00002879801, 08881501178, 64899093501, 08214352901, 08470352901, 38396040205, 38396040208, 38396040212, 00002879901, 08881501210, 38396040218, 08080127012, 00002831591, 54569137200, 08881501236, 08222094391, 08080200777, 64899093601, 08214353001, 08470353001, 38396040305, 38396040308, 38396040318, 38396040325, 52735038601, 61059052719, 08881501368, 38396081339, 64899093701, 08881501384, 08290841801, 08290841803, 00002951501, 38396040405, 08881501400, 38396040408, 38396040412, 38396040612, 00002879459, 00002811001, 52735038701, 38396040618, 08496313601, 08222094599, 76300072211, 76300072213, 76300072214, 00536991001, 76300072219, 76300072220, 00536991005, 38396040508, 38396040512, 38396040518, 877017474020, 38396040525, 08881608020, 08881608021, 08881608022, 08496305501, 08881608030, 08881608031, 08881608032, 08496305511, 00169244010, 08290309484, 00002844001, 08881501558, 51709007227, 54569383300, 54569383301, 54569383302, 51709007241, 08881903002, 38396040605, 38396040608, 08881903010, 08290309540, 08290309541, 08290309542, 08290309543, 08881501608, 08881608101, 08881608102, 08881608103, 08881608104, 08881608105, 08881608106, 08881608107, 08881608108, 08881608109, 08881608110, 08881608111, 08881608112, 08881608113, 38396040625, 49999099310, 52735038901, 08881305018, 08290309569, 08881608130, 08290309570, 08290309571, 08290309572, 08290309573, 08290309574, 08290309575, 08290309576, 08290309578, 08290309579, 08290309580, 08290309581, 08290309582, 08290309584, 08496305611, 08290309587, 08290309588, 08290309589, 08881608148, 08290309591, 08881608155, 08290309597, 08290309598, 08290309599, 08290309602, 08881608163, 00002831759, 08881501111, 54569383400, 54569383401, 08881608171, 08290309621, 08290309623, 08290309624, 08290309625, 08290309626, 08290309628, 08881608189, 08290309630, 08290309631, 08290309632, 08290309633, 08290309634, 08290309635, 08881600004, 08881608197, 08290309637, 08290309638, 08290309639, 08290309640, 08290309641, 08290309642, 08881608204, 08881608205, 08881608206, 08881608207, 08881608208, 08290309643, 08881608210, 08290309644, 08881608212, 08881608213, 08290309645, 00002873059, 08881305109, 08881608209, 08881608211, 38396040725, 52735039001, 08881305117, 54868137500, 87701747230, 08290309669, 08290309670, 08290309671, 08881600038, 89134052902, 00169244210, 87701747250, 082903096053, 08290309694, 08290309695, 54569383500, 54569383501, 54569383502, 87701747280, 58016478801, 38396040801, 38396040802, 38396040803, 38396040804, 38396040805, 38396040806, 38396040807, 38396040808, 38396040812, 08881608301, 08881608302, 08881608303, 08881608304, 08881608305, 08881608306, 08881608307, 08881608308, 08881608309, 08881608310, 08881608311, 08881305208, 08881608312, 08881608313, 38396040818, 38396040825, 52735039101, 08881501822, 08881600137, 08881600145, 08881051286, 08881600152, 08330131101, 08080062110, 08970213031, 00088250033, 00002864001, 00088250052, 08881051333, 38396040708, 08881600210, 38396040918, 52297086678, 38396040712, 52735039201, 08881600228, 08881600236, 62107006701, 96295106290, 08881600244, 38396040718, 87701747450, 08881600251, 08330131201, 08881501962, 38396040970, 54868588300, 00536991501, 08881600269, 08881501970, 08881600277, 08881600285, 763000515110, 763000515120, 08214353714, 08214353719, 38396041018, 52297086778, 763000515130, 763000515140, 52735039301, 08881600335, 62107006801, 08881600343, 00088250205, 08881600350, 08330131301, 38396041070, 08881600368, 00536991601, 96295106430, 08881600384, 08881600392, 38396041102, 96295106450, 08881846169, 08496281501, 38396041118, 52297086878, 08881600418, 00169007011, 08496281511, 52735039401, 08881600426, 383960411137, 08881600434, 38396041140, 62107006901, 08214509501, 38396041157, 08330131401, 08881240009, 08881600459, 08881240017, 00536991701, 08881051615, 08881600483, 08881051620, 08881600491, 38396041202, 08080218100, 08884533239, 99073071097, 99073071098, 99073071099, 99073071100, 52297086978, 08884533254, 08080128012, 38396041237, 38396041240, 62107006701, 08881600343, 08881600350, 08881600558, 00169351215, 08881600566, 00002751001, 00536991801, 08881600574, 08881600590, 38396041302, 08884533338, 08881600616, 08881600624, 38396041337, 38396041340, 08881535101, 08881600640, 38396041357, 08330131601, 08881600657, 0820090129, 61059053719, 08881600673, 38396082339, 76300712110, 76300712120, 08881600699, 08881600700, 76300712130, 088816006035, 8622707065035, 76300712140, 08881600723, 8622707065045, 088816006059, 08881600731, 38396041437, 86227065055, 38396041440, 088816006070, 89134061802, 08881600749, 0888160080, 38396041457, 08881600803, 08881600756, 00002841001, 38396000506, 08881600764, 08884533510, 08080300300, 08881600780, 00002861501, 86227065105, 08290309798, 08881600800, 08881912102, 08881600139, 08881600558, 0169351215, 08881600566, 00002751001, 00536991801, 08881600574, 08881600590, 38396041302, 08881600904, 54569630000, 54569630000, 08080201810, 08881600830, 0888160160, 00603699521, 38396041537, 38396041540, 08080030029, 08080030030, 08496011601, 38396041557, 38396000606, 08881600863, 08290843000, 08881600889, 08290843002, 08290843003, 00002751559, 87701445500, 08881600905, 08881609101, 08881600913, 54569630100, 00904395160, 08496290201, 00002850001, 00603699621, 38396041637, 38396041640, 08881609131, 89134062002, 08881609497, 08496011701, 08881609145, 38396041657, 08080325058, 08881600962, 38396000706, 54868277700, 58441112003, 54868605400, 08881600988, 08290843101, 08290843103, 08080325100, 08080325114, 00002811101, 00603699721, 00002877001, 08080300555, 08881601036, 08881601044, 08290490901, 08496011801, 08881601051, 00002851001, 58441112103, 00002751659, 08415001128, 08415001129, 08881601069, 25715669550, 08881560125, 08881601093, 08881560141, 08881601101, 08080620112, 38396041808, 08080622110, 00603699821, 08881601135, 08881609331, 86227090035, 08881560182, 54868359800, 08290491001, 89134052802, 08496011901, 08881609350, 86227090055, 088820091529, 08881609358, 08881609383, 08881609388, 08881609398, 08881609402, 08881609407, 08881609412, 08881609417, 08881609422, 08881609427, 08881609432, 08881609437, 08881609442, 08881609447, 08881609452, 08881609457, 08881609462, 08881609467, 08881609472, 08881609477, 08881609482, 08881609487, 08881609492, 08881609497, 08881609502, 08881609507, 08881609512, 08881609517, 08881609522, 08881609527, 08881609532, 08881609537, 08881609542, 08881609547, 08881609552, 08881609557, 08881609562, 08881609567, 08881609572, 08881609577, 08881609582, 08881609587, 08881609592, 08881609597, 08881609602, 08881609607, 08881609612, 08881609617, 08881609622, 08881609627, 08881609632, 08881609637, 08881609642, 08881609647, 08881609652, 08881609657, 08881609662, 08881609667, 08881609672, 08881609677, 08881609682, 08881609687, 08881609692, 08881609697, 08881609702, 08881609707, 08881609712, 08881609717, 08881609722, 08881609727, 08881609732, 08881609737, 08881609742, 08881609747, 08881609752, 08881609757, 08881609762, 08881609767, 08881609772, 08881609777, 08881609782, 08881609787, 08881609792, 08881609797, 08881609802, 08881609807, 08881609812, 08881609817, 08881609822, 08881609827, 08881609832, 08881609837, 08881609842, 08881609847, 08881609852, 08881609857, 08881609862, 08881609867, 08881609872, 08881609877, 08881609882, 08881609887, 08881609892, 08881609897, 08881609902, 08881609907, 08881609912, 08881609917, 08881609922, 08881609927, 08881609932, 08881609937, 08881609942, 08881609947, 08881609952, 08881609957, 08881609962, 08881609967, 08881609972, 08881609977, 08881609982, 08881609987, 08881609992, 08881609997, 08881610002, 08881610007, 08881610012, 08881610017, 08881610022, 08881610027, 08881610032, 08881610037, 08881610042, 08881610047, 08881610052, 08881610057, 08881610062, 08881610067, 08881610072, 08881610077, 08881610082, 08881610087, 08881610092, 08881610097, 08881610102, 08881610107, 08881610112, 08881610117, 08881610122, 08881610127, 08881610132, 08881610137, 08881610142, 08881610147, 08881610152, 08881610157, 08881610162, 08881610167, 08881610172, 08881610177, 08881610182, 08881610187, 08881610192, 08881610197, 08881610202, 08881610207, 08881610212, 08881610217, 08881610222, 08881610227, 08881610232, 08881610237, 08881610242, 08881610247, 08881610252, 08881610257, 08881610262, 08881610267, 08881610272, 08881610277, 08881610282, 08881610287, 08881610292, 08881610297, 08881610302, 08881610307, 08881610312, 08881610317, 08881610322, 08881610327, 08881610332, 08881610337, 08881610342, 08881610347, 08881610352, 08881610357, 08881610362, 08881610367, 08881610372, 08881610377, 08881610382, 08881610387, 08881610392, 08881610397, 08881610402, 08881610407, 08881610412, 08881610417, 08881610422, 08881610427, 08881610432, 08881610437, 08881610442, 08881610447, 08881610452, 08881610457, 08881610462, 08881610467, 08881610472, 08881610477, 08881610482, 08881610487, 08881610492, 08881610497, 08881610502, 08881610507, 08881610512, 08881610517, 08881610522, 08881610527, 08881610532, 08881610537, 08881610542, 08881610547, 08881610552, 08881610557, 08881610562, 08881610567, 08881610572, 08881610577, 08881610582, 08881610587, 08881610592, 08881610597, 08881610602, 08881610607, 08881610612, 08881610617, 08881610622, 08881610627, 08881610632, 08881610637, 08881610642, 08881610647, 08881610652, 08881610657, 08881610662, 08881610667, 08881610672, 08881610677, 08881610682, 08881610687, 08881610692, 08881610697, 08881610702, 08881610707, 08881610712, 08881610717, 08881610722, 08881610727, 08881610732, 08881610737, 08881610742, 08881610747, 08881610752, 08881610757, 08881610762, 08881610767, 08881610772, 08881610777, 088816

Pre-operative treatment with Insulin	NDC* codes: 08290843801, 08881601689, 08290843803, 56151170201, 08881601697, 08881200292, 08881601705, 08881700010, 08881601713, 08881601721, 08881200318, 08881200326, 08496315601, 08881601747, 08881200342, 08881601762, 00536993001, 08881601770, 59060231404, 38396042490, 54569295100, 54569295101, 56151170301, 08881200383, 08881716503, 54868131100, 08881716511, 87701445930, 08881200433, 08881716529, 08881601846, 08881200441, 08881716537, 08881601853, 08881601861, 08881020233, 54868589900, 08881200466, 08881601879, 08290008410, 08290008411, 00182312285, 38396042590, 00182312288, 08290844001, 08290008430, 08290008431, 87701044593, 08881200508, 08222072191, 00904396160, 36652400001, 36652400002, 36652400003, 08881200516, 36652400004, 36652400005, 36652400006, 36652400007, 36652400008, 08290008465, 08290008466, 08881200573, 00839802306, 38396042690, 00069070737, 08881135068, 54868532700, 54868532701, 36652400101, 36652400102, 36652400103, 36652400104, 36652400105, 36652400106, 36652400107, 36652400108, 08881135084, 08214502901, 00002841101, 59060231704, 00839802406, 38396042790, 08080621100, 00002821601, 08881520178, 08080621112, 08881700408, 08881520186, 00002824001, 00002821517, 08881200714, 36652400202, 36652400203, 36652400205, 36652400206, 36652400207, 36652400208, 08222072399, 00169033301, 00002879359, 08214503001, 00169352815, 08881200755, 08881520251, 00839802506, 08290328203, 38396042890, 08080826012, 00003183715, 38396042912, 08881200805, 08290328233, 08496291501, 54868238001, 08881716917, 08496291511, 08881847993, 08290328278, 08290328279, 08290328280, 08290328281, 08290328282, 08290328283, 08290320096, 08290328289, 08290328290, 08290328291, 08881250016, 08881250024, 08881676012, 08290320109, 00839802606, 38396042990, 08881250032, 08290320119, 08881250040, 54868623100, 08080826112, 38396043012, 08214355719, 08881250057, 00002811201, 08881250065, 08222072597, 08881250073, 08881250081, 68258897701, 08881250099, 08881250107, 08881512258, 08881250115, 08881250123, 38396043090, 08881250131, 08290328410, 08290328411, 08290328412, 08290328418, 08881250149, 38396043112, 08290328430, 08290328431, 08881250164, 08290328438, 08290328440, 08881250172, 08881250180, 08881053577, 08290320271, 08290328465, 08290328466, 08290328468, 08881250198, 08290328471, 00002854001, 08881250206, 08881250214, 38396043177, 08881250222, 08881070001, 08080818100, 08080818101, 08881250230, 38396043190, 68115070905, 08080818112, 08080818113, 08881250248, 08881250255, 08080220112, 08881250263, 08881160157, 08881250271, 08881242088, 08881250289, 08881750007

TABLE S4: ICD codes for myocardial infarction used to identify positive cohort

ICD code	description
I46.8	Cardiac arrest due to other underlying condition
I21	ST elevation (STEMI) myocardial infarction involving left main coronary artery
410.72	Subendo infarct subseq
I21.01	ST elevation (STEMI) myocardial infarction involving left main coronary artery
410.01	AMI anterolateral init
I21.4	Non-ST elevation (NSTEMI) myocardial infarction
I21.A9	Other myocardial infarction type
I21.A1	Myocardial infarction type 2
410.61	True post infarct init
I21.3	ST elevation (STEMI) myocardial infarction of unspecified site
410.8	AMI NEC unspecified
410.42	AMI inferior wall subseq
427.5	Cardiac arrest
I46	Cardiac arrest due to underlying cardiac condition
I46.2	Cardiac arrest due to underlying cardiac condition
410	AMI anterolateral unspec
410.71	Subendo infarct initial
410.11	AMI anterior wall init
410.12	AMI anterior wall subseq
410.7	Subendo infarct unspec
I21.21	ST elevation (STEMI) myocardial infarction involving left circumflex coronary artery
410.4	AMI inferior wall unspec
410.21	AMI inferolateral init
410.82	AMI NEC subsequent
410.9	AMI NOS unspecified
410.2	AMI inferolateral unspec
I21.9	Acute myocardial infarction unspecified
I46.9	Cardiac arrest cause unspecified
410.1	AMI anterior wall unspec
410.02	AMI anterolateral subseq
410.51	AMI lateral NEC initial
410.52	AMI lateral NEC subseq
410.92	AMI NOS subsequent
I21.02	ST elevation (STEMI) myocardial infarction involving left anterior descending coronary artery
I21.19	ST elevation (STEMI) myocardial infarction involving other coronary artery of inferior wall
410.81	AMI NEC initial
410.41	AMI inferior wall init
410.31	AMI inferopost initial
410.62	True post infarct subseq
I21.09	ST elevation (STEMI) myocardial infarction involving other coronary artery of anterior wall
410.0	AMI anterolateral unspec
410.5	AMI lateral NEC unspec
410.6	True post infarct unspec
410.3	AMI inferopost unspec
410.91	AMI NOS initial
410.32	AMI inferopost subseq
I21.11	ST elevation (STEMI) myocardial infarction involving right coronary artery
410.22	AMI inferolateral subseq
I21.29	ST elevation (STEMI) myocardial infarction involving other sites

TABLE S5: Number of diagnostic codes encountered in dataset

gender	Number of codes	Number of unique codes
M	4879398	17554
F	7753318	19209
Total	12632716	36763

TABLE S6: CCoR phenotypes and maximum number of unique ICD codes defining CCoR phenotypes

CCoR phenotype	count of ICD codes in definition (Table S7)
Allergic	191
Cardiovascular	2017
CNS	765
Development	820
Digestive	1317
Endocrine	237
Frailty	557
Health-Services	501
Hematologic	429
Hypertension	80
Immune	1546
Infections-Bacterial	409
Infections-Fungal-and-Other	784
Infections-General	3612
Infections-Respiratory	712
Injuries	53265
Integumentary	1457
Metabolic	373
Musculoskeletal	7533
Neoplastic	3022
Ophthalmological	3401
Otic	856
PNS	394
Psychiatric	1478
Reproductive	2675
Respiratory	724

TABLE S7: Disease Categories With Detailed Set of ICD Codes Used in Definition. Not all infection and injury codes have been listed here.

	Description	Constituent ICD9 Codes
Allergic		477.2 493.81 T50.995A J67.2 495.6 T78.03x 372.14 J67 J67.0 M13.89 J30.1 995.63 995.65 558.3 T45.0X1A M13.859 716.27 D29.30 D29.1 L27.2 477.9 495.5 493.22 D69.2 T78.00x 287.33 995.60 J45.31 J45.51 D29.20 J67.7 T78.09x A D29.22 M13.80 J30.9 T78.08x 287.8 H10.45 B44.81 716.20 995.61 T78.05x A G93.92 693.1 493.90 T78.40x J45.20 493.82 J45.40 D69.42 495.7 J67.5 493.20 D69.49 J45.32 287.32 708.0 H65.119 995.64 D69.1 J45.21 D69.6 M13.819 716.23 495.4 995.67 287.1 T78.08x A T78.00x A 477.0 493.02 525.66 T78.02x A J67.1 D69.3 T78.04x T78.2xx A D29.4 716.25 T78.07x A 716.26 T78.07x M13.88 J67.3 495.9 J45.30 493.21 477 495.2 995.62 995.27 287.2 495.8 495 287.5 995.0 493 T78.05x L50.0 493.11 J45.902 D29.0 J45.990 287.9 J45 D29.21 J30.0 963.0 495.1 D29.32 L25.9 J44.9 J44.0 477.1 M13.879 493.01 J45.41 T50.995 J45.998 692.9 M13.849 995.66 D69.8 995.69 T78.04x A J30 495.3 M13.869 287.30 J45.991 J44.1 995.3 287.4 J45.52 287.0 381.06 716.21 J45.901 J67.4 287.39 493.91 373.32 287.31 T78.06x A J30.89 287 K08.55 K52.2 D29.31 J45.50 495.0 J67.6 D69.9 D29.8 T78.02x 716.24 477.8 381.05 D29.493.12 T78.03x A J67.9 716.22 T78.2xx J30.5 999.4 493.00 M13.829 T78.01x T78.06x 493.10 518.6 716.28 J30.2 H01.119 995.68 M13.839 D69.0 T78.09x 381.04 D29.9 T78.01x A 716.29 J30.81 J45.22 J45.42 T45.0X1 J45.909 D69.41 J67.8
Cardiovascular		I35.0 I48.0 I25.728 444.8 P29.38 I94 I63.212 402.00 I70.669 440.30 I89.9 I60.9 I20.1 413.9 I24.1 I80.3 415.1 I77.811 785.9 I69.319 I69.339 I82.5Z9 R04.1 429.6 G43.619 I82.503 I82.611 I70.512 I75.011 I69.834 I70.628 K64.9 I89.0 I21.0 492.42 447.5 442.8 I70.792 454.0 I70.318 150.83 I70.744 405.0 426.2 455.3 I82.B29 I12 415.433.8 169.320 I27.81 444.21 T0.735 182.602 167.894 I41.4 425.4 135.9 I70.693 I69.234 I65.23 427.1 T0.244 149.02 I82.91 P29.89 I70.719 I69.131 I36.8 160.8 160.11 442.8 I69.852 I75.029 438.22 I69.120 I70.641 I60.2 426.51 T10.302 417.9 I63.162 O4.2 R02.0 427.31 I25.781 105.8 I70.791 I89.246.50 I63.349 I49.49 444.89 I63.213 I83.12 I77.0 I82.433 I70.608 P29 I47.1 428 I70.348 I82.293 I82.532 I69.322 I63.311 I65.03 I82.291 I70.498 I97.791 I69.859 I70.644 I82.441 I63.413 I70.362 405.11 I62 I70.012 I80.292 I41.1 433 I70.532 I97.638 I87.091 I69.999 I45.01 169.399 I223.137.2 I87.183 I69.932 I70.120 288.182 I82.492 I68.281 I63.233 I70.319 I70.765 I70.341 I435.441.9 I40.0 I63.539 K64.4 I95.89 I69.028 I82.5Y9 I50.32 I70.731 I70.768 458.0 G43.601 I21.0 I41.71 I429.3 I69.398 I87.399 I44.0 I70.421 I73 I70.729 E86.0 I69.231 I28.8 I13.0 I70.568 I442.2 I63.431 I41.1 89.7 I70.709 I438.21 I438.5 I70 I50.84 I42.0 I70.212 I77.7 I40.8 P29.2 455.0 410.1 416.1 I24.8 433.10 I70.393 413.1 I70.561 I87.092 I83.218 I427.6 453.8 I70.534 I97.130 I97.821 I79.711 I63.543 I02.0 I405.01 I69.364 I07 I83.92 I69.928 I69.214 I70.418 346.63 I06.8 410.6 I31.2 I70.433 I75.013 I70.193 I70.431 I75.023 I82.403 I83.10 I25.730 I69.814 I50.812 I79.429 88.9 I69.833 I28.10 I77.6 I69.165 I60.52 459.81 I47.140 I10.10 I83.202 I82.419 I69.033 I59.4 404.0 I69.213 I441.7 I82.412 I69.261 I82.432 I417 I49.9 I69.315 I438.19 I78.19 I36.2 497.8 I70.536 I70.532 I69.220 I77.70 I63.0 I70.439 I70.643 I43.80 276.50 I45.93 I70.522 I69.121 I433.21 I24.2 464.6 I34.61 I61.175 202 I88.8 I55.785.1 I0.9 I75.279 I63.20 I45.4 I63.521 I70.769 I69.334 I72.4 150.31 438.50 434 I70.349 I48.2 435.1 I80.219 I60.169.115 T0.544 I41.0 440.0 I69.010 I66.12 445.8 I30.9 I27.1 I80.293 I432 I97.190 I69.864 I86.14 I41.0 I410.8 I438.9 I63.133 I25.89 I63.032 I34.8 I77.4 I82.422 I87.391 I41.0 I87.312 I440.4 I63.232 I87.009 R00.8 I69.012 I70.468 Q82.5 I21.3 I87.303 I25.6 I69.842 290.41 I455.8 I69.042 I82.232 I87.099 H69.314 I74.10 I77.812 I69.393 I42.5 I69.141 I70.491 I82.210 V78.550 I78.5 I69.351 I82.513 I25.785 I52 I77.819 I415.11 I80.3 I71.5 I82.422 I42.6 P29.30 I69.031 I70.222 I31.4 I82.603 I82.729 I437 I70.621 I70.509 I70.763 I95 I38 I433.3 I69.133 K50.4 456.6 I80.212 I70.223 I70.730 I70.229 I404.10 I83.023 I437.3 I426.13 I65.22 I75.012 180.202 I48.3 I77.83 I69.815 I75.022 I72.20 I70.711 I45.3 I65.9 I80.233 I41.0 I70.291 I40.2 I63.29 I404.11 I82.421 R04.89 I83.0 785.0 I438.0 I10 I70.469 I91 I69.163 I47.2 I50.1 414.4 I83.002 410.5 I70.243 I63.039 I70.249 I70.269 I80.211 I86.8 I70.499 I69.363 I40.02 I22.0 I69.211 I58.8 I83.028 I74.2 428.0 I70.545 I442 I443.89 I426.3 I69.362 I31 I70.398 I81 I69.810 I70.261 I443.22 I70.1 I70.33 I83.023 I69.232 I63.532 I410.2 I71.02 I70.331 I441.0 I69.262 I410.6 I65.29 I54.5 I45.3 I71.69.293 438.81 I69.110 I15 P65.21 I428.31 I453.89 I69.051 I67.7 I69.062 I69.111 I69.349 I454.1 I453.1 I63.39 I42.1 I66.03 I70.426 I8.8 I42.12 I45.8 I81.73 I70.634 I69.918 I95.1 I87.393 I410.0 I70.332 I82.549 I443.29 I77.5 440 I31.8 I21 I82.B23 I11.0 I433.01 I455.7 I403.00 I66.024 I438.14 I06.0 G45.0 I97.51 I69.112 I70.369 I82.5Y2 I45.2 I74.01 I415.19 I87.311 I63.22 I84 I20.703 I83.009 I87 I71 I43.3 I06.1 I45.6 I69.822 I199 I33.0 I70.218 I438.5 I60.10 I69.254 I415.1 I52 I69.065 I59 I410.62 I72.180 I8.196 453.76 437.2 435.3 I41.11 I83.019 I09.2 I83.934 I26.8 I63.323 I410.72 I43.52 I63.441 I83.011 I41.10 I82.B11 I41.0 I21.73 I438.41 I75.81 I404.9 I13.11 I66.09 I95.0 I83.022 I438.4 I82.433 I70.202 I70.342 I70.648 I23.3 I69.235 I437.9 I82.812 I82.439 I63.599 I79.120 I69.249 I455.6 I82.501 I82.2 A23 I74.09 I63.322 I70.722 I82.423 I87.9 I82.C21 I41.02 I42.9 I69.093 I404 I24 I455.9 I32 I70.798 I44.6 I40.042 I70.450 I80.322 I82.509 I441.7 I77.3 I69.993 I53 I83 I82.499 I69.034 I70.402 I60.01 I458.1 I438.82 I24.0 I456.8 I69.144 R01.2 I70.231 I65.1 I70.429 I70.612 I50.33 I82.612 I63.422 I57 I433.2 I69.221 I95.9 I79 I78 I97.11 I04 I69.963 I77.9 I82.410.52 I410.3 I70.298 438.20 I48.0 I50.82 I70.698 I70.533 I69.915 I69.052 I425.9 I69.243 I80.224 I427.32 I438.410 I69.092 I413.0 I290.4 I27.22 I69.042 I442.83 I453.86 I557 I453.75 I445.0 I23.2 438.1 I458.21 I70.303 I70.541 I70.201 I69.154 I41 429.1 I51.9 I69.244 I27.41 I72.6 I16.0 I290.4 I70.793 I82.182 I44.3 I51.5 I26.0 I93.225 I37.9 I66.13 I70.329 I69.191 I70.169 I69.338 I442.2 I36.1 I69.191 R01.1 I08.1 I70.133 I70.462 I82.401 I82.192 I63.519 K55.9 I82.412 I82.523 I33.410.7 I45.4 I69.333 I83.219 I69.198 I70.399 426.54 I25.710 I42.4 I42.2 I81.557.9 I80.231 I82.443 I80.0 I P29.12 I447.3 I70.245 I41.0 I70.439.1 I69.812 I50.23 I448.9 I43.4 I26.99 I42.7.5 I77.74 I453.6 I69.118 I35 I44.1.3 I82.C22 I08.8 I41.07 I28.0 I438.8 I97.42 I65.09 I37.0 I67.848 I70.449 I77.2 I41.8 I70.603 G40.5 I453.81 I44.5 I426.0 I06.2 I74.5 I42.7.9 I42.7.2 I70.213 I61.2 I44.9 I62.1 F01.51 I70.701 I56.8 I456.8 I65.8 I87.013 I01.8 I40.3 I453.83 I69.242 I42.4 I81.833 I69.161 I69.031 I69.133 I45.3 I70.121 I11.25.82 I87.339 I70.749 I83.892 I23.5 I97.611 I66.3 I63.449 I453.72 I88.0 I40.59 I69.049 I69.943 I70.748 I21.0 I42.8 32.47 I70.78 I40.3 I82.9 I82.162 I69.321 I69.321 I405.09 I453.40 I453.42 I40.8 I69.162 426.1 I69.192 I83.222 I87.033 I25.83 I74.3 I82.523 I458.9 I64.8 I438.83 I42.7 I70.521 I459 I83.221 I82.609 I82.703 I07.0 I97.110 I69.942 I437.6 I405.1 I60.6 I69.342 I70.728 405.99 I69.831 I69.30 I12.9 I43.8 I44.7.7 I41.1 069.828 I43.4 I91.0 I89.109 I23.4 I83.024 I47.0 I82.543 I50.810 I70.293 I82.211 I405.19 I86.2 I69.954 I70.409 I83.812 I97.622 437.0 I25.799 I403.01 I75.89 I70.201 I63.533 I78.1 I42.8 33 I82.422 I70.9 I410 402.414 I02.417 I82.169 I96.4 I70.62 I63.442 I47.2 I82.309 I63.339 I17.72 I70.363 I414.06 I404.92 I71.2 441.1 I82.522 I20.8 402.11 I82.819 I65.8 I36.9 I19.9 I78.5 I32.4 I69.344 I97.811 I63.113 I443.21 I69.020 I02.0 I70.328 I72.6 I62.0 I70.613 I447.72 I82.542 I78.51 I69.132 I83.015 I82.493 I67.4 I42.6 I42.5 I27.5 I80.2 I70.338 I63.10 I30.8 I79.4 I79.1 I79.4 I79.3 I79.4 I79.5 I79.6 I79.7 I79.8 I79.9 I79.10 I79.11 I79.12 I79.13 I79.14 I79.15 I79.16 I79.17 I79.18 I79.19 I79.20 I79.21 I79.22 I79.23 I79.24 I79.25 I79.26 I79.27 I79.28 I79.29 I79.30 I79.31 I79.32 I79.33 I79.34 I79.35 I79.36 I79.37 I79.38 I79.39 I79.40 I79.41 I79.42 I79.43 I79.44 I79.45 I79.46 I79.47 I79.48 I79.49 I79.50 I79.51 I79.52 I79.53 I79.54 I79.55 I79.56 I79.57 I79.58 I79.59 I79.60 I79.61 I79.62 I79.63 I79.64 I79.65 I79.66 I79.67 I79.68 I79.69 I79.70 I79.71 I79.72 I79.73 I79.74 I79.75 I79.76 I79.77 I79.78 I79.79 I79.80 I79.81 I79.82 I79.83 I79.84 I79.85 I79.86 I79.87 I79.88 I79.89 I79.90 I79.91 I79.92 I79.93 I79.94 I79.95 I79.96 I79.97 I79.98 I79.99 I79.10 I79.11 I79.12 I79.13 I79.14 I79.15 I79.16 I79.17 I79.18 I79.19 I79.20 I79.21 I79.22 I79.23 I79.24 I79.25 I79.26 I79.27 I79.28 I79.29 I79.30 I79.31 I79.32 I79.33 I79.34 I79.35 I79.36 I79.37 I79.38 I79.39 I79.40 I79.41 I79.42 I79.43 I79.44 I79.45 I79.46 I79.47 I79.48 I79.49 I79.50 I79.51 I79.52 I79.53 I79.54 I79.55 I79.56 I79.57 I79.58 I79.59 I79.60 I79.61 I79.62 I79.63 I79.64 I79.65 I79.66 I79.67 I79.68 I79.69 I79.70 I79.71 I79.72 I79.73 I79.74 I79.75 I79.76 I79.77 I79.78 I79.79 I79.80 I79.81 I79.82 I79.83 I79.84 I79.85 I79.86 I79.87 I79.88 I79.89 I79.90 I79.91 I79.92 I79.93 I79.94 I79.95 I79.96 I79.97 I79.98 I79.99 I79.10 I79.11 I79.12 I79.13 I79.14 I79.15 I79.16 I79.17 I79.18 I79.19 I79.20 I79.21 I79.22 I79.23 I79.24 I79.25 I79.26 I79.27 I79.28 I79.29 I79.30 I79.31 I79.32 I79.33 I79.34 I79.35 I79.36 I79.37 I79.38 I79.39 I79.40 I79.41 I79.42 I79.43 I79.44 I79.45 I79.46 I79.47 I79.48 I79.49 I79.50 I79.51 I79.52 I79.53 I79.54 I79.55 I79.56 I79.57 I79.58 I79.59 I79.60 I79.61 I79.62 I79.63 I79.64 I79.65 I79.66 I79.67 I79.68 I79.69 I79.70 I79.71 I79.72 I79.73 I79.74 I79.75 I79.76 I79.77 I79.78 I79.79 I79.80 I79.81 I79.82 I79.83 I79.84 I79.85 I79.86 I79.87 I79.88 I79.89 I79.90 I79.91 I79.92 I79.93 I79.94 I79.95 I79.96 I79.97 I79.98 I79.99 I79.10 I79.11 I79.12 I79.13 I79.14 I79.15 I79.16 I79.17 I79.18 I79.19 I79.20 I79.21 I79.22 I79.23 I79.24 I79.25 I79.26 I79.27 I79.28 I79.29 I79.30 I79.31 I79.32 I79.33 I79.34 I79.35 I79.36 I79.37 I79.38 I79.39 I79.40 I79.41 I79.42 I79.43 I79.44 I79.45 I79.46 I79.47 I79.48 I79.49 I79.50 I79.51 I79.52 I79.53 I79.54 I79.55 I79.56 I79.57 I79.58 I79.59 I79.60 I79.61 I79.62 I79.63 I79.64 I79.65 I79.66 I79.67 I79.68 I79.69 I79.70 I79.71 I79.72 I79.73 I79.74 I79.75 I79.76 I79.77 I79.78 I79.79 I79.80 I79.81 I79.82 I79.83 I79.84 I79.85 I79.86 I79.87 I79.88 I79.89 I79.90 I79.91 I79.92 I79.93 I79.94 I79.95 I79.96 I79.97 I79.98 I79.99 I79.10 I79.11 I79.12 I79.13 I79.14 I79.15 I79.16 I79.17 I79.18 I79.19 I79.20 I79.21 I79.22 I79.23 I79.24 I79.25 I79.26 I79.27 I79.28 I79.29 I79.30 I79.31 I79.32 I79.33 I79.34 I79.35 I79.36 I79.37 I79.38 I79.39 I79.40 I79.41 I79.42 I79.43 I79.44 I79.45 I79.46 I79.47 I79.48 I79.49 I79.50 I79.51 I79.52 I79.53 I79.54 I79.55 I79.56 I79.57 I79.58 I79.59 I79.60 I79.61 I79.62 I79.63 I79.64 I79.65 I79.66 I79.67 I79.68 I79.69 I79.70 I79.71 I79.72 I79.73 I79.74 I79.75 I79.76 I79.77 I79.78 I79.79 I79.80 I79.81 I79.82 I79.83 I79.84 I79.85 I79.86 I79.87 I79.88 I79.89 I79.90 I79.91 I79.92 I79.93 I79.94 I79.95 I79.96 I79.97 I79.98 I79.99 I79.10 I79.11 I79.12 I79.13 I79.14 I79.15 I79.16 I79.17 I79.18 I79.19 I79.20 I79.21 I79.22 I79.23 I79.24 I79.25 I79.26 I79.27 I79.28 I79.29 I79.30 I79.31 I79.32 I79.33 I79.34 I79.35 I79.36 I79.37 I79.38 I79.39 I79.40 I79.41 I79.42 I79.43 I79.44 I79.45 I79.46 I79.47 I79.48 I79.49 I79.50 I79.51 I79.52 I79.53 I79.54 I79.55 I79.56 I79.57 I79.58 I79.59 I79.60 I79.61 I79.62 I79.63 I79.64 I79.65 I79.66 I79.67 I79.68 I79.69 I79.70 I79.71 I79.72 I79.73 I79.74 I79.75 I79.76 I79.77 I79.78 I79.79 I79.80 I79.81 I79.82 I79.83 I79.84 I79.85 I79.86 I79.87 I79.88 I79.89 I79.90 I79.91 I79.92 I79.93 I79.94 I79.95 I79.96 I79.97 I79.98 I79.99 I79.10 I79.11 I79.12 I79.13 I79.14 I79.15 I79.16 I79.17 I79.18 I79.19 I79.20 I79.21 I79.22 I79.23 I79.24 I79.25 I79.26 I79.27 I79.28 I79.29 I79.30 I79.31 I79.32 I79.33 I79.34 I79.35 I79.36 I79.37 I79.38 I79.39 I79.40 I79.41 I79.42 I79.43 I79.44 I79.45 I79.46 I79.47 I79.48 I79.49 I79.50 I79.51 I79.52 I79.53 I79.54 I79.55 I79.56 I79.57 I79.58 I79.59 I79.60 I79.61 I79.62 I79.63 I79.64 I79.65 I79.66 I79.67 I79.68 I79.69 I79.70 I79.71 I79.72 I79.73 I79.74 I79.75 I79.76 I79.77 I79.78 I79.79 I79.80 I79.81 I79.82 I79.83 I79.84 I79.85 I79.86 I79.87 I79.88 I79.89 I79.90 I79.91 I79.92 I79.93 I79.94 I79.95 I79.96 I79.97 I79.98 I79.99 I79.10 I79.11 I79.12 I79.13 I79.14 I79.15 I79.16 I79.17 I79.18 I79.19 I79.20 I79.21 I79.22 I79.23 I79.24 I79.25 I79.26 I79.27 I79.28 I79.29 I79.30 I79.31 I79.32 I79.33 I79.34 I79.35 I79.36 I79.37 I79.38 I79.39 I79.40 I79.41 I79.42 I79.43 I79.44 I79.45 I79.46 I79.47 I79.48 I79.49 I79.50 I79.51 I79.52 I79.53 I79.54 I79.55 I79.56 I79.57 I79.58 I79.59 I79.60 I79.61 I79.62 I79.63 I79.64 I79.65 I79.66

Development	<p>533.60 K63.81 531.30 K13 K42.1 K52.81 K11.4 K68.11 R15.2 K94.13 K42.9 K52.3 R10.811 K10.823 K26.2 578 K22.719 K64.9 532 K29.81 K76.2 K57.53 K59.02 532.50 K80.42 K57.80 K80.34 K08.123 K54.90 K51.92 K56.601 K58.79 531.91 K76 K55.022 K55.30 K56.600 K80.133 K91.840 537.82 K54.5 K08.51 K45 527 K80.30 K67 K63.3 K22.2 538 787.5 K26.5 K25.2 K66 564 K80.13 K22.6 R13.14 K51.211 K39.00 K70.31 K00.4 K08.403 K13.29 K43.7 K70.51 R0.1 K51.89 740.9 K09.0 K59.1 K37 K57.40 533.91 K51.919 K05.216 K92.81 K13.79 R19.36 K51.1 K52.81 K51.1 K50.812 K51.2 K52.31 K51.2 K52.51 787.4 K40.4 K04.00 P77.3 K90.0 K22.3 K84 532.9 K64.4 K50.911 K69.7 K42.9 K9.3 Q25.1 K06.41 R19.00 K08.119 K27.9 K57.30 K01 K08.414 K08.530 K21 K43.1 K54.30 K08.60 K08.50 K30.3 K08.41 K56.12 K27.3 K56.699 K11.23 K50.118 R19.11 531.2 K13.1 K08.9 534.0 530.5 K11.9 527.8 K40.30 569.44 K62.3 K27.20 K08.433 K55.042 K70.30 K64.0 K94.01 K65 R10.2 533.31 K04.99 577.8 K94.23 K08.22 K41.40 K11.90 568.89 K41 K30 K08.52 K76.4 E16.9 K80.19 K02.7 K0.578.1 K27.6 K25.7 K76.8 J86.0 K86.9 K94.39 K10.5 323 P78.82 569.41 K32 K29.71 K04.3 K91.0 K58.8 K9.04 787 K75.17 K08.23 K95.89 K60.3 787.7 K26.9 K28.2 K14.1 K00 533.51 K08.494 K57.20 456.2 K04.90 K52.29 K56.0 K28.0 K13.23 530.82 K42.51 749.29 K9.31 K27.80 K31.9 K07.77 R14.2 R19.10 70 R11.2 R18.8 556.09 K28.4 78.4 727.821 K31.0 K55.21 K51.913 K38.8 R19.7 K46.0 K55.8 K0.51 K51.3 K51.31 K53.6 1 K85.80 K50.30 K56.50 537.84 K94.29 532.31 K53.4 30 K52.22 K40 R10.84 K73.2 K81.1 K80.1 66 K71.4 K53.0 8 K40.1 K59.8 K71 K58.9 K80.00 K50.814 527.2 787.23 K54 K08.431 532.40 569.85 531.31 564.01 537.83 569.87 R11.0 K14.3 537 K62.5 564.09 R14.3 K51.314 K22 K86.1 K91.5 K05.211 K71.29 K08.104 530.9 534.51 K74.1 21 P78.83 K14.3 K06.51 K18.6 568.2 K26 K51.311 K31.80 P78.1 787.1 K91.253.1 K09.4 11 K38.9 578.9 K03.89 K55.049 543.9 K51.514 K59.39 K41.10 K08.531 K03.81 K75.3 K51.411 K50.31 50.1 K22.9 K76 K40.823 K28.5 K29.01 K08.21 534.40 K31.80 530.2 K27.5 K51.512 K29.41 K22.4 K27.1 K02.61 R10.821 K51.518 K85.10 537.83 K73 K27.2 K58 558 536.8 541.0 K04.93 K06.0 K51.73 533.4 K28.9 K38.1 K93.51 K01.011 R10.826 K50.818 456.1 K51.912 K08.20 568.82 K61.0 579.2 K14.9 K82.1 K06.8 K08.50 R19.6 K51.218 565.0 K52.89 R13.19 K51.911 568.81 P76.1 K57.11 558.41 K08.424 K55.021 R10.812 K91.72 K50.013 K70.9 K21.519 K09.94 21 K05.312 K53.16 K09.1 777.6 R15.0 K29.20 562.13 K91.83 K56.691 K05.11 K31.1 K83 K60.0 K08.412 K55.032 K13.70 536.0 K08.101 K51.812 568.9 K44.0 K34.5 K50.964 K51.0 K12 50.42 P78.1 K05.230 532.30 K45.1 K01.73 K51.0 K56.9 K35.6 89 560.3 K80.70 557.2 K31.84 K08.499 K41.21 K78.83 K14.6 K08.111 K40.651 K18.6 568.2 K26 K51.311 K31.80 P78.1 787.1 K91.253.1 K09.4 11 K38.9 578.9 K03.89 K55.049 543.9 K85.31 K52.82 R19.02 K08.82 566 R10.30 556.51 K14.0 K85.22 531.90 K29.90 527.4 K94.1 10 K80.80 527.6 K08.409 K72.91 K49 R10.13 R19.05 532.6 K12.2 K41.01 K60.1 K29.00 K55.9 R15.9 K66.0 K08.124 K08.421 K02.63 K04.1 K26.0 K56.69 R18.0 K91.518.1 K21.87.6 K54.2 568.77 550.85 K12.0 K52.839 531.71 530.4 K90.2 K26.3 P76.0 K31.6 787.24 R19.32 K28.7 K03.29 532.3 533.90 534.10 579.21 K07.4 K22.1 K06.22 K09.199 K33.59 00 787.22 K86.3 R19.33 R19.12 K42.2 K03.3 K81.9 K50.914 K44.9 K85.00 787.02 568.8 P78.4 560.30 K13.5 530.89 K51.82 K51.213 K56.2 1 K30.6 K13.5 K66.90 K90.1 K1.74 K19.34 R13.0 K83.2 K06.021 K28 K02.51 K75.4 K64.8 531.51 K95 K11.6 K11.6 P78.89 540.0 K38.2 K55.062 K51.018 K56.9 493.3 K1.81 K51.0 K56.9 3 K91.1 K63.1 K53.1 5 51.4 40 533.0 777.9 K56.4 2 50.30 85 K05.20 K94.30 K43 P78.0 K51.50 K56.5 K05.5 534.31 K00.1 K56.2 R11.14 537.81 K53.9 K04.5 K29.80 K31.4 K96 R19.03 532.10 K60.5 R19.15 530.84 K55.031 K31.3 K08.401 K56.2 10 560.39 533.20 K12.33 R10.829 537.6 K05.342.1 K51.214 K82.3 532.21 K21.6 K66.8 K51.9 577.0 K29.31 R41.30 K76.1 R11.13 787.3 K56.5 30.7 R13.13 K57.52 K91.81 K53.2 K63.4 R10.10 K51.418 251.4 K02.62 K55.011 K62.82 K11.0 K91.81 K45.8 T0.741 K29.91 R0.822 K44.0 K57.00 787.29 K20.0 K00.8 K50.813 K14.5 K51.00 K56.49 456.21 K04.6 K77.0 P84 045 558.4 K52.0 569.42 R19.5 P17.7 K57.1 K51.813 K08.26 K41.40 K05.01 K85 K76.6 K05.313 K53.10 787.20 R10.819 K08.3 K25.1 K59.1 K08.89 K50.819 K68.19 K14.1 R11 K26.6 K51.019 K90.4 787.01 K20 R18.577 531.3 K65.9 K05.222 540.9 K49.3 K51.413 K53.70 K59.1 K40.41 K78.04 K91.858 K02.52 K53.7 K50.012 K38.3 K41.90 K51.819 K56.51 K57.32 K50.019 777.1 K80.43 K22.8 K71.51 K08.113 K98.194 K64.2 R19.01 K92.1 185.10 K76.0 K57.50 K80.60 K51.319 527.9 K71.81 K94. R13.10 K08.112 K03.6 K94.02 P76.8 K35.80 R10.83 530.81 K52.01 K31 K71.7 K29.61 527.1 569.1 K74.0 568.82 K25.1 K53.6 9 K52.32 577.9 K09.9 K57.1 K09.5 523.2 K08.409 K72.91 K49.0 777.3 537.8 533.41 K08.129 P76.2 K51.818 K57.41 530.83 K40 K50 K80.61 K41.91 K54.1 K59.03 K80.1 533.40 K02.3 R19.2 K90 K58.2 K94.20 K08.132 K27.0 K80.33 562.0 K26.1 K95.81 K08.129 P76.2 K51.818 K11.7 K40.01 K82.4 251.4 K35.3 K52.1 K52.2 K71.6 K21.0 K56.2 K71.3 K29.90 777.57 K1.563.0 560.81 K08.493 R19.35 K51.013 R15 787.1 777 533 537.8 533.41 K08.129 K13.3 K54.6 26.2 K7.2 K54.3 50.531 K08.419 K49.21 K70.1 K62.0 K65.4 R1.12 K51.3 K52.3 K13.2 K51.6 K54.5 K13.24 K53.11 K54.1 K01.5 K55.052 K50.119 K50.018 K94.22 K11.20 K13.21 K40.91 P77.1 K27.7 K04.70 K06 530.20 569.81 K66.1 K91.32 K50.811 K62.1 K21.32 K20.8 K08.439 K49.09 K24 558.9 K94.00 K56.609 533.71 564.0 K534.61 K51.413 K51.102 K80.46 K23.4 K03.41 K50.05 556.2 K02.55 K00.6 229.00 K6.001 K45.1 K55.1 K78 251 K31.9 K10.24 K08.491 K51.014 530.3 K31.819 K08.434 K80 K56.3 K51.519 532.7 K82.9 K70.40 532.70 532.560.32 527.5 K17 R16.2 560.8 527.0 K27.4 K72.11 K18 K80.62 K91.870 K08.402 540.1 530.0 K03.2 K91.850 K11.59 560.9 K13.22 K31.82 K08.131 K51 537.9 K80.41 K56.5 50.10 K08.139 K43.9 K91.89 K58.0 K21.72 K85.01 K51.21 K70.1 556.4 K40.9 K50.8 K03.7 K97 456.20 K56.60 534.41 K51.419 K60.02 K91.873 K05.311 K54.0 532.1 527.7 K90.54 K7.4 57.4 K7.4 57.4 K9.26 K9.50 K59.3 K08.121 K46 K85.02 K92.2 787.03 K81.2 P77.9 K09 K05.322 K35.3 K61.4 K09.8 K31.2 K43.2 K26.4 560.89 K80.12 K82.0 K45.0 K50.011 569.84 K41.11 K91.71 K51.511 K51.20 534.20 579.4 533.6 K08.191 536.3 K50.114 K11.21 K94.32 560.31 K75.1 K13.12 K56.7 K94.12 R16.0 K72.01 K28.6 777.52 K83.8 251.8 K90.81 K06.3 K14 K00.7 532.00 R10.814 K83.1 K08.56 787.99 K57 534.0 K80.18 K70.19 R19.8 K76.89 K08.24 R10.827 534.90 K06.2 K03.4 534.7</p>
Digestive	<p>242.11 E23.6 253 255.4 259.8 253.2 E27.0 242.90 253.0 250.62 250.72 250.02 P72 250.80 252.02 242.1 258.9 E05.40 E20.9 P72.8 250.3 242.2 P74.3 P72.2 259.2 775.6 P70.2 362.03 244.2 H05.239 250.42 250.10 E26.02 259.50 E23.</p>

Continued on next page

Continued on next page

	<p>041.4 017.10 B02.21 055 098.14 042.9 B09 011.80 569.5 B69 136.4 A19 078.0 057.8 A50.09 017.81 590.3 083.8 016.4 A01.38.3 003.8 B67 341.21 372.05 016.32 380.12 015.62 017.20 466.1 380.2 001.1 001 A52.10 B94.2 H70.009 B35.0 A92.8 017.44 017.82 123.1 115.00 016.71 B74.8 T79.A12S 081 016.60 011.8 B02.24 T50.A15A A05 B60.12 T50.A25D A04.9 B33.23 T50.A11A 682.6 098.41 B35.2 116.2 N37 018.92 040.89 A77.0 117.7 095.8 045.1 A07.4 A63 B08.1 018.01 070.3 B88.2 M90.869 B42.42 A85.1 005.9 041.11 015.3 375.30 A39.83 016.36 013.53 B96.81 011.93 488.11 041.04 A48.51 B26.1 A73 A08.32 B43.1 A18.4 B65.0 A41.02 T80.A19B 018.85 B16.9 006 B87.89 060.1 182.B29 324.9 P36.2 012.85 685 B50.0 027 A66.5 070.70 A66.1 B02.9 T50.A93A P35.1 A48 100 B17.9 590.1 058.1 079.3 130.1 B06.00 B42.89 112 017.9 B60.0 T50.A22D B74.4 B52.9 046.2 077.9 381.1 086.0 0483 056.01 A54.41 016.24 062.4 A83.9 K67 B97.11 449 085.0 007.8 013.62 A92.9 P36.39 T50.A94S 016.1 011.5 J05.10 466.0 A50.07 B00.82 015.61 B81.3 017.76 012.81 B08.01 015.04 771.3 A30.9 063.8 016.93 A52.06 B02.0 A22.2 059.22 041.03 008.69 011.23 381.0 A06.9 380.1 083.1 A28.2 A36.85 B46.0 013.14 070.52 B77.0 082.40 054.1 482.89 112.89 A02.0 T50.B12S 010.02 032.2 A18.5A B27.82 B73.00 008.5 047 036.42 T80.A10D M79.A21 B45.2 A08.39 681.0 111.0 T50.A25A B40.3 B48.3 375.31 094.1 017.33 014 J31.0 038.4 G00.3 B33.1 045.23 011.43 045.01 T50.B94 041.83 071 015.21 B43.0 016.53 048 094.8 472.1 016.04 010.94 073.7 A96.9 017.4 079.52 015.12 127.8 016.14 T80.A11S A00.1 097.9 A07 011.64 A66.3 038.49 039 100.81 137.3 136.3 382.2 B57 006.0 038.42 A51.45 B87 B95.7 771.1 H70.209 483.8 012.1 A36.9 010.01 063.9 J15.20 B97.89 420.91 B26.85 B45.8 081.2 074.22 A88.0 014.06 042.0 567.3 115.13 B55.1 016.54 016.35 B35.1 573.3 053.13 A37 015.75 079.2 J18.9 N11.8 008.45 056.05 A52.06 A54.83 A23.1 093.8 A53.0 424.9 A54.82 A40.0 H10.239 A48.52 A49.2 112.84 013.23 016.41 099.55 051.0 004.9 093.20 B14 B24 373.4 011.7 B79 098.39 M89.619 T50.B93 A80.9 091.1 A32 A17.82 031.0 M90.879 043.1 488.1 A51.9 J09.X3 A87.2 123.4 G00.9 A21.3 B40.0 B97.4 066.49 A02.24 038.0 B06.01 A48.8 017.06 730.76 421.1 098.5 094.8 020.9 001.9 115.9 015.66 A02.22 A50.03 B37.7 026.0 077.98 482.8 730.75 117.2 055.0 A93.7 A77.8 A04.5 A01.05 B02.32 003.21 A53.9 322.1 B48.0 B56.9 A44.1 098.50 015.70 B17.2 B67.99 H60.399 098.59 T50.A94 T50.B95D G03.9 A01.2 123 039.1 077.4 682.2 P35.2 T50.B15D B31 017.66 A80.39 S81.3 B78.81 B60.19 B32 038.12 074 321.2 131.9 012.8 A43.8 B57.42 131.8 015.92 482.8 070.7 T50.A13D 125 011.90 072.2 060.9 013.45 027.9 101 B30.9 A06.1 320.2 B97.5 484.8 A41.9 130.0 013.90 078.19 T50.B12 084.8 040.2 102.2 A25.0 B57.2 123.5 I31.2 J15.5 B27.81 099.59 B08.72 M86.9 005.2 094.82 H59.42 112.2 T79.A11 139.1 J85.1 003.9 008.43 003.0 A82.9 480.3 A08.2 083.2 123.9 070.43 078.4 078.1 B38.4 B40.1 B00.89 034.0 A51.5 012.01 B71 006.4 513 J17 074.3 A56.4 A18.12 A77.49 B95.8 132.9 B55.2 011.15 017 008.42 T50.B93D L03.039 B77.81 126.0 A56.09 077.3 B95.4 484.5 233.71 682.0 008.49 B17.11 010.05 A66.8 A85.2 B48.4 B96.5 A51.31 054.2 042.11.41 A52.8 132.2 A50.45 A04.7 A77.9 B46.5 006.5 115.04 009 A39.82 B37.42 011.0 B08.09 054.73 A44.8 012.84 P36 B85.0 383.21 059.0 040 126.6 T50.B95S 771.0 074.23 323.51 685.6 A26.0 382.01 B01.2 B88.4 A42.8 B20.7 B67.69 015.97 038.41 B94.1 421.0 021.1 098.86 A68 T50.A15A A49.02 T50.A21D 013.22 A32.7 B58.2 124 099.49 B67.32 079.83 121.0 A54.5 A52.3 J12.0 B87.3 070.1 B08.03 132.1 016.55 111.2 T50.A25 090.7 A36.81 A41.53 018.80 A81.01 B69.1 A21.1 098.82 I30.9 002 102.9 016.9 04 J04.30 A19.8 050.0 017.65 A18.18 A06.0 0 H04.429 P37.0 015.23 078 112.82 041.02 115.92 A41.52 B88.0 B96 022.2 A36.89 A52.9 B35.9 I31.8 A38.8 013.92 B2.2 B23 H66.90 015.94 M35.8 G06.1 391 B46.4 017.86 A02.25 A17.83 122 B90.2 B19.11 T50.A96A T80.A11D A43.0 A74.81 016.51 056.00 A18.39 472.2 102 099.5 067.9 H70.229 323.4 J09 X9 117.4 015.71 B80.0 B47.1 014.5 126.2 J21.8 484.7 I33.0 016.13 B95 016.95 054.71 043.2 422.93 045.10 128.8 B57.1 059.10 B57.49 091.4 038.2 016.61 021.56.71 B36.51 B78.91 B37.81 730.91 018.86 B65.3 372.0 B02.23 B39.5 A48.0 B18.8 482.0 381.10 005.3 017.6 464.50 A50.6 P37.8 T50.A96 A52.11 B67.39 H04.339 T50.A14A 039.3 421 T50.A24S B16.1 B51.0 I09.2 053.19 015.56 T79.A22A T80.A19 A52 G05.4 051.2 A44 B33 079 A67.1 380.10 T79.A12D B27 053.22 103.2 771.5 A37.80 B27.19 B82.B11 018.8 A22.0 A18.83 A40.0</p>
	<p>017.10 013.95 017.86 A17.83 J12.9 013.42 016.51 011.80 A15.1 017.83 017.42 B18.39 A19 010.91 J09 X9 017.81 015.71 017.0 013.30 A16.70 018.80 488.0 013.41 013.25 017.20 015.62 017.31 B25.0 J20.6 482.32 J21.8 017.02 484.7 016.13 012.31 016.95 017.44 017.82 016.71 013.61 016.60 T50.A15A A11.33 011.84 018.93 016.61 011.10 0485 015.50 018.92 018.86 J01.21 J12.3 017.07 482.0 018.01 046.50 011.31 J11.82 B42.42 010.06 J04.10 012.20 016.71 T79.A19S 013.81 013.33 J85.1 048.0 J01.11 048.5 A19.0 011.93 013.44 A18.4 017.22 T50.D20.9 A77.0 017.80 T80.A19D 011.91 018.85 013.06 016.92 013.05 016.75 019.0 017.89 J18.1 012.21 015.0 0482.82 011.42 0462 017.32 010.90 016.00 J01.11 017.22 010.90 016.00 016.92 013.05 016.75 019.0 017.89 J18.1 012.21 015.0 0482.82 011.42 0462 017.32 010.90 016.00 J01.11 017.22 010.90 016.00 016.92 013.05 016.75 019.0 017.89 J18.1 012.21 015.0 0482.82 011.42 0462 017.32 010.90 016.00 J01.11 017.22 010.90 016.00 016.92 013.05 016.75 019.0 017.89 J18.1 012.21 0</p>

Integumentary

Injuries

S06.9X4S T84.490D T63 S63 621D S82.232C S65.391S T85.614 S25.809D T85.398 S72.346 S83.011A S60.212A S66.526D T80.59X S52.232D S53.026A S82.115P
 S56.822 S86.391D T63.614D T50.992S S61.221S S19.9XKA 943.23 T23.179 S71.131 T80.410S S93.129 T40.0X2 T81.83XS S72.436R S63.433D T24.591D 832.01
 S72.91XK T24.031D S36.533A S56.196S T22.232A T24.332D S92.355K S59.129 S31.112 T78.3XX S66.120A T81.516 S21.442D S82.043P 803.50 S63.223S T53.0X4A
 S92.123A S36.020D S72.019A T63.511S S82.864R T65.831S S63.657D S14.157T T04 T84.620 S22.040R S56.892D S72.441A T56.1X1D T83.85X T53.7X3 S68.123A
 T47.1X6S T26.71XA S14.4XX S89.002D T53.1X1 S62.307S T65.213S S82.242M T22.741A S62.023D T46.6X1 S32.131K S52.515C T65.0X3S S46.012A S52.265C
 S92.221B S65.111A S06.1X1A S60.461 S61.549D S36.814.03 S24.151A S50.811A T62.1X1S S42.446P S52.264G T24.491A S52.042A S92.599K S06.4X9A S92.416D
 T20.411A T81.507D S12.600B S82.492P S21.351S T45.1X4S S32.000 T75.4XX T82.391 T43.623 S36.290 T54.3X1S S42.209K T56.3X3D S59.222 S99.829S
 S09.19XD S34.22XD S63.652A S94.45 S32.471A T40.3X2 S60.444D S82.146S S82.822K S50.10X T83.022S S56.128S S05.91XD S82.874Q S42.111D S25.402A
 T82.322 S62.614S S86.929S S90.464D S74.00XA T65.4X2 T61.784D S83.30X S32.456B T53.93X T24.501A S37.69XA S37.091D S63.022D S80.849S T46.991S
 S61.021 T49.0X5 S90.411D S30.826 T63.834 S64.494A S42.033D S32.05BD S42.191D S52.332C S31.110S S82.871M T43.4X5S S95.911S S81.20 S20.309A S01.01X
 S13.171D S72.366M S82.154B T79.5XXA T83.69XS S90.90X S00.252S S83.096S S01.90XA S72.031K S63.439A S06.365 T50.Z92 T25.139 S09.312A T50.Z11D
 S37.031 S82.426A S82.022B S95.809 S12.300K S35.338S S92.244D T23.799 S82.254 S62.312 T22.022S S01.118B S06.6X2D S60.940S S31.20X A T7.408D
 S09.312D S52.399F S82.251K T63.071D S82.464Q S52.243A S82.465Q S82.401N T49.8X6A 914.3 S55.212 S29.102K T86.23 S96.889 S22.050S S41.112D S50.11XA
 S52.272M S72.416C S82.876G S22.039D S42.144G S63.227A S66.518A T82.416S T85.511A T48.4X4 S82.034R T20.69X S82.62XP S42.112B S52.244 S92.031K
 S12.151K S42.216 S41.031A S15.202D T23.079D S39.91XS T19.4XXA S82.291A S92.491K S01.449S S52.361J T63.813S T63.713A S72.434J S82.016D S99.029B
 S22.002A S80.919D S83.231D S82.209H S90.933 T78.07XA T53.2X1S S20.461 T36.7X2 S92.78 T82.897A S72.365F T20.72X S63.065 S72.111B S21.06 S63.432D
 S65.911D S96.102A 944.6 S80.857A S83.001 S22.032D S87.93 S32.4D S15.9XK S41.031D S63.93X S82.154 S60.031 T23.469A S21.429A S52.282 T71.223
 S62.338G S82.443D S67.90X S82.111P 887.7 S15.121D 917.7 S66.516S S70.12X S23.143A 956.0 T39.8X3 S72.012D S90.01XD S61.205D S82.242S S82.843M
 T46.6X6D S39.002A S62.331B S09.312 T45.2X1S S82.843D S86.111S S20.02XA S42.335A S59.299A S20.112S S81.019 S32.120S S83.123A S92.331G S50.912
 S50.02XA S67.194 S02.630K S82.033P S66.002A S61.234S S61.217D 901.1 S10.12 S89.392P S60.342 S52.256M S60.471 S00.451D S53.091S S72.041D S59.101G
 923.03 T24.031A S01.81X S89.101K S81.842A T52.2X4 S92.323D S50.4X5S S82.234J T22.351D S76.311A S52.042R S82.872M S83.262S T55.1X2 T71.234A
 S82.156K T38.4X2D S14.154D S62.399P S63.217A T41.0X2 S92.611D S63.121D S52.225 S56.329D T23.091A S86.10 S54.011A S36.99XA S72.361A S52.234F S949.4
 S75.811S S81.822D S82.124P S72.442D S46.891 S92.326D T25.219A S72.425R S42.251B S66.120S S88.021S S27.69X D S52.513J S60.569D T63.832A S65.508S
 935.1 S82.113N S62.329D T83.39XA T63.823A T61.04XA T81.532 S42.035 S02.602A S52.609A S56.418D 806.31 S82.253Q S89.912S T21.31XS T37.5X6S S90.0XX
 S72.364D S39.002A S62.331B S09.312 T45.2X1S S82.843D S86.111S S20.02XA S42.335A S59.299A S20.112S S81.019 S32.120S S83.123A S92.331G S50.912
 S50.02XA S67.194 S02.630K S82.033P S66.002A S61.234S S61.217D 901.1 S10.12 S89.392P S60.342 S52.256M S60.471 S00.451D S53.091S S72.041D S59.101G
 923.03 T24.031A S01.81X S89.101K S81.842A T52.2X4 S92.323D S50.4X5S S82.234J T22.351D S76.311A S52.042R S82.872M S83.262S T55.1X2 T71.234A
 S82.156K T38.4X2D S14.154D S62.399P S63.217A T41.0X2 S92.611D S63.121D S52.225 S56.329D T23.091A S86.10 S54.011A S36.99XA S72.361A S52.234F S949.4
 S75.811S S81.822D S82.124P S72.442D S46.891 S92.326D T25.219A S72.425R S42.251B S66.120S S88.021S S27.69X D S52.513J S60.569D T63.832A S65.508S
 935.1 S82.113N S62.329D T83.39XA T63.823A T61.04XA T81.532 S42.035 S02.602A S52.609A S56.418D 806.31 S82.253Q S89.912S T21.31XS T37.5X6S S90.0XX
 S72.364D S39.002A S62.331B S09.312 T45.2X1S S82.843D S86.111S S20.02XA S42.335A S59.299A S20.112S S81.019 S32.120S S83.123A S92.331G S50.912
 S50.02XA S67.194 S02.630K S82.033P S66.002A S61.234S S61.217D 901.1 S10.12 S89.392P S60.342 S52.256M S60.471 S00.451D S53.091S S72.041D S59.101G
 923.03 T24.031A S01.81X S89.101K S81.842A T52.2X4 S92.323D S50.4X5S S82.234J T22.351D S76.311A S52.042R S82.872M S83.262S T55.1X2 T71.234A
 S82.156K T38.4X2D S14.154D S62.399P S63.217A T41.0X2 S92.611D S63.121D S52.225 S56.329D T23.091A S86.10 S54.011A S36.99XA S72.361A S52.234F S949.4
 S75.811S S81.822D S82.124P S72.442D S46.891 S92.326D T25.219A S72.425R S42.251B S66.120S S88.021S S27.69X D S52.513J S60.569D T63.832A S65.508S
 935.1 S82.113N S62.329D T83.39XA T63.823A T61.04XA T81.532 S42.035 S02.602A S52.609A S56.418D 806.31 S82.253Q S89.912S T21.31XS T37.5X6S S90.0XX
 S72.364D S39.002A S62.331B S09.312 T45.2X1S S82.843D S86.111S S20.02XA S42.335A S59.299A S20.112S S81.019 S32.120S S83.123A S92.331G S50.912
 S50.02XA S67.194 S02.630K S82.033P S66.002A S61.234S S61.217D 901.1 S10.12 S89.392P S60.342 S52.256M S60.471 S00.451D S53.091S S72.041D S59.101G
 923.03 T24.031A S01.81X S89.101K S81.842A T52.2X4 S92.323D S50.4X5S S82.234J T22.351D S76.311A S52.042R S82.872M S83.262S T55.1X2 T71.234A
 S82.156K T38.4X2D S14.154D S62.399P S63.217A T41.0X2 S92.611D S63.121D S52.225 S56.329D T23.091A S86.10 S54.011A S36.99XA S72.361A S52.234F S949.4
 S75.811S S81.822D S82.124P S72.442D S46.891 S92.326D T25.219A S72.425R S42.251B S66.120S S88.021S S27.69X D S52.513J S60.569D T63.832A S65.508S
 935.1 S82.113N S62.329D T83.39XA T63.823A T61.04XA T81.532 S42.035 S02.602A S52.609A S56.418D 806.31 S82.253Q S89.912S T21.31XS T37.5X6S S90.0XX
 S72.364D S39.002A S62.331B S09.312 T45.2X1S S82.843D S86.111S S20.02XA S42.335A S59.299A S20.112S S81.019 S32.120S S83.123A S92.331G S50.912
 S50.02XA S67.194 S02.630K S82.033P S66.002A S61.234S S61.217D 901.1 S10.12 S89.392P S60.342 S52.256M S60.471 S00.451D S53.091S S72.041D S59.101G
 923.03 T24.031A S01.81X S89.101K S81.842A T52.2X4 S92.323D S50.4X5S S82.234J T22.351D S76.311A S52.042R S82.872M S83.262S T55.1X2 T71.234A
 S82.156K T38.4X2D S14.154D S62.399P S63.217A T41.0X2 S92.611D S63.121D S52.225 S56.329D T23.091A S86.10 S54.011A S36.99XA S72.361A S52.234F S949.4
 S75.811S S81.822D S82.124P S72.442D S46.891 S92.326D T25.219A S72.425R S42.251B S66.120S S88.021S S27.69X D S52.513J S60.569D T63.832A S65.508S
 935.1 S82.113N S62.329D T83.39XA T63.823A T61.04XA T81.532 S42.035 S02.602A S52.609A S56.418D 806.31 S82.253Q S89.912S T21.31XS T37.5X6S S90.0XX
 S72.364D S39.002A S62.331B S09.312 T45.2X1S S82.843D S86.111S S20.02XA S42.335A S59.299A S20.112S S81.019 S32.120S S83.123A S92.331G S50.912
 S50.02XA S67.194 S02.630K S82.033P S66.002A S61.234S S61.217D 901.1 S10.12 S89.392P S60.342 S52.256M S60.471 S00.451D S53.091S S72.041D S59.101G
 923.03 T24.031A S01.81X S89.101K S81.842A T52.2X4 S92.323D S50.4X5S S82.234J T22.351D S76.311A S52.042R S82.872M S83.262S T55.1X2 T71.234A
 S82.156K T38.4X2D S14.154D S62.399P S63.217A T41.0X2 S92.611D S63.121D S52.225 S56.329D T23.091A S86.10 S54.011A S36.99XA S72.361A S52.234F S949.4
 S75.811S S81.822D S82.124P S72.442D S46.891 S92.326D T25.219A S72.425R S42.251B S66.120S S88.021S S27.69X D S52.513J S60.569D T63.832A S65.508S
 935.1 S82.113N S62.329D T83.39XA T63.823A T61.04XA T81.532 S42.035 S02.602A S52.609A S56.418D 806.31 S82.253Q S89.912S T21.31XS T37.5X6S S90.0XX
 S72.364D S39.002A S62.331B S09.312 T45.2X1S S82.843D S86.111S S20.02XA S42.335A S59.299A S20.112S S81.019 S32.120S S83.123A S92.331G S50.912
 S50.02XA S67.194 S02.630K S82.033P S66.002A S61.234S S61.217D 901.1 S10.12 S89.392P S60.342 S52.256M S60.471 S00.451D S53.091S S72.041D S59.101G
 923.03 T24.031A S01.81X S89.101K S81.842A T52.2X4 S92.323D S50.4X5S S82.234J T22.351D S76.311A S52.042R S82.872M S83.262S T55.1X2 T71.234A
 S82.156K T38.4X2D S14.154D S62.399P S63.217A T41.0X2 S92.611D S63.121D S52.225 S56.329D T23.091A S86.10 S54.011A S36.99XA S72.361A S52.234F S949.4
 S75.811S S81.822D S82.124P S72.442D S46.891 S92.326D T25.219A S72.425R S42.251B S66.120S S88.021S S27.69X D S52.513J S60.569D T63.832A S65.508S
 935.1 S82.113N S62.329D T83.39XA T63.823A T61.04XA T81.532 S42.035 S02.602A S52.609A S56.418D 806.31 S82.253Q S89.912S T21.31XS T37.5X6S S90.0XX
 S72.364D S39.002A S62.331B S09.312 T45.2X1S S82.843D S86.111S S20.02XA S42.335A S59.299A S20.112S S81.019 S32.120S S83.123A S92.331G S50.912
 S50.02XA S67.194 S02.630K S82.033P S66.002A S61.234S S61.217D 901.1 S10.12 S89.392P S60.342 S52.256M S60.471 S00.451D S53.091S S72.041D S59.101G
 923.03 T24.031A S01.81X S89.101K S81.842A T52.2X4 S92.323D S50.4X5S S82.234J T22.351D S76.311A S52.042R S82.872M S83.262S T55.1X2 T71.234A
 S82.156K T38.4X2D S14.154D S62.399P S63.217A T41.0X2 S92.611D S63.121D S52.225 S56.329D T23.091A S86.10 S54.011A S36.99XA S72.361A S52.234F S949.4
 S75.811S S81.822D S82.124P S72.442D S46.891 S92.326D T25.219A S72.425R S42.251B S66.120S S88.021S S27.69X D S52.513J S60.569D T63.832A S65.508S
 935.1 S82.113N S62.329D T83.39XA T63.823A T61.04XA T81.532 S42.035 S02.602A S52.609A S56.418D 806.31 S82.253Q S89.912S T21.31XS T37.5X6S S90.0XX
 S72.364D S39.002A S62.331B S09.312 T45.2X1S S82.843D S86.111S S20.02XA S42.335A S59.299A S20.112S S81.019 S32.120S S83.123A S92.331G S50.912
 S50.02XA S67.194 S02.630K S82.033P S66.002A S61.234S S61.217D 901.1 S10.12 S89.392P S60.342 S52.256M S60.471 S00.451D S53.091S S72.041D S59.101G
 923.03 T24.031A S01.81X S89.101K S81.842A T52.2X4 S92.323D S50.4X5S S82.234J T22.351D S76.311A S52.042R S82.872M S83.262S T55.1X2 T71.234A
 S82.156K T38.4X2D S14.154D S62.399P S63.217A T41.0X2 S92.611D S63.121D S52.225 S56.329D T23.091A S86.10 S54.011A S36.99XA S72.361A S52.234F S949.4
 S75.811S S81.822D S82.124P S72.442D S46.891 S92.326D T25.219A S72.425R S42.251B S66.120S S88.021S S27.69X D S52.513J S60.569D T63.832A S65.508S
 935.1 S82.113N S62.329D T83.39XA T63.823A T61.04XA T81.532 S42.035 S02.602A S52.609A S56.418D 806.31 S82.253Q S89.912S T21.31XS T37.5X6S S90.0XX
 S72.364D S39.002A S62.331B S09.312 T45.2X1S S82.843D S86.111S S20.02XA S42.335A S59.299A S20.112S S81.019 S32.120S S83.123A S92.331G S50.912
 S50.02XA S67.194 S02.630K S82.033P S66.002A S61.234S S61.217D 901.1 S10.12 S89.392P S60.342 S52.256M S60.471 S00.451D S53.091S S72.041D S59.101G
 923.03 T24.031A S01.81X S89.101K S81.842A T52.2X4 S92.323D S50.4X5S S82.234J T22.351D S76.311A S52.042R S82.872M S83.262S T55.1X2 T71.234A
 S82.156K T38.4X2D S14.154D S62.399P S63.217A T41.0X2 S92.611D S63.121D S52.225 S56.329D T23.091A S86.10 S54.011A S36.99XA S72.361A S52.234F S949.4
 S75.811S S81.822D S82.124P S72.442D S46.891 S92.326D T25.219A S72.425R S42.251B S66.120S S88.021S S27.69X D S52.513J S60.569D T63.832A S65.508S
 935.1 S82.113N S62.329D T83.39XA T63.823A T61.04XA T81.532 S42.035 S02.602A S52.609A S56.418D 806.31 S82.253Q S89.912S T21.31XS T37.5X6S S90.0XX
 S72.364D S39.002A S62.331B S09.312 T45.2X1S S82.843D S86.111S S20.02XA S42.335A S59.299A S20.112S S81.019 S32.120S S83.123A S92.331G S50.912
 S50.02XA S67.194 S02.630K S82.033P S66.002A S61.234S S61.217D 901.1 S10.12 S89.392P S60.342 S52.256M S60.471 S00.451D S53.091S S72.041D S59.101G
 923.03 T24.031A S01.81X S89.101K S81.842A T52.2X4 S92.323D S50.4X5S S82.234J T22.351D S76.311A S52.042R S82.872M S83.262S T55.1X2 T71.234A
 S82.156K T38.4X2D S14.154D S62.399P S63.217A T41.0X2 S92.611D S63.121D S52.225 S56.329D T23.091A S86.10 S54.011A S36.99XA S72.361A S52.234F S949.4
 S75.811S S81.822D S82.124P S72.442D S46.891 S92.326D T25.219A S72.425R S42.251B S66.120S S88.021S S27.69X D S52.513J S60.569D T63.832A S65.508S
 935.1 S82.113N S62.329D T83.39XA T63.823A T61.04XA T81.532 S42.035 S02.602A S52.609A S56.418D 806.31 S82.253Q S89.912S T21.31XS T37.5X6S S90.0XX
 S72.364D S39.002A S62.331B S09.312 T45.2X1S S82.843D S86.111S S20.02XA S42.335A S59.299A S20.112S S81.019 S32.120S S83.123A S92.331G S50.912
 S50.02XA S67.194 S02.630K S82.033P S66.002A S61.234S S61.217D 901.1 S10.12 S89.392P S60.342 S52.256M S60.471 S00.451D S53.091S S72.041D S59

Metabolic	<p>269 E83.30 270.5 P71.4 276.7 264.1 712.38 M1A.9xx0 274.82 E71.42 278.02 274.19 278.01 276.69 263.8 712.86 E72.8 268.1 275.40 274.03 712.28 M10.30 E66.2 783.1 E74.9 E71.41 270.8 712.19 E83.59 E87.70 M11.80 E50.7 E88.1 275.42 E67.0 260 M10.9 E46 274 E72.20 M11.88 276.3 712.84 E83.50 P19.0 264.3 E50.8 278.2 275.1 E50.1 M11.20 E83.81 R63.4 775.9 277.81 268 712.36 277.88 E83.40 E53.9 271.9 P71.3 E87.1 M11.879 270.7 E52 D84.1 273.9 266.2 264.4 265.1 274.8 712.1 M11.249 712.8 M11.849 277.82 E72.9 M11.269 712.16 E50.4 712.88 M10.40 M11.28 269.9 E88.9 264.9 268.0 E55.9 278.0 712.85 266.9 277.9 E75.21 712.3 712.89 712.27 712.97 277.8 P72.8 M11.859 278.3 264.0 278.03 G93.9 E53.0 E16.1 271.3 251.0 775.89 E70.0 P71.1 E51.11 P70.1 275.5 P70.4 277.7 272.3 712.15 E50.6 277.87 712.39 274.81 712.96 712.10 274.02 E87.8 712.90 E15 273.8 R63.0 271.2 E66.9 277.86 266.0 E53.8 E64.3 263.1 278 E87.0 272.5 E66.3 712.32 E67.1 262 712.92 M11.9 M11.29 712.31 272.1 P19.1 P71.8 M11.279 712.95 277.85 E65 E88.40 775.81 783.22 E61.4 783.2 263.2 261 E83.89 P71 M11.829 P74.0 E74.12 274.89 E70.21 E72.03 712.11 E50.2 M11.869 712.80 P19.9 E50.5 264.3 712.13 275.49 E16.2 E54 E67.8 712.83 276.8 278.4 775.7 712.37 E78.3 270.2 270.4 E53.1 E83.52 M83.9 269.2 269.1 783.21 712.33 M1A.00x E87.4 E87.5 P70.3 268.2 271.8 E87.3 M11.259 7E6.01 E87.2 266.1 P71.2 E88.01 E55.0 251.2 266 273.4 274.00 272.2 275.01 275.8 712.98 E56.8 E56.1 E74.21 E44.1 272.7 712.91 263.9 E88.81 E78.89 E78.9 712.34 269.3 274.01 263.0 P70.9 712.22 277.2 E74.39 263 E67.3 276.6 E88.09 E50.3 E72.10 R63.5 E43 P19 265 275.9 264.5 E66.01 E83.00 E80.0 E88.89 712.18 E83.118 275.4 E63.8 264 274.10 277.5 275.3 712.94 712.12 E44.0 E40 M11.819 276.2 274.0 E51.8 270.6 E79.8 712.23 276.9 272.4 P71.0 G93.89 712.2 712.87 E83.51 P70.2 271.1 276.4 E45 712.93 277.89 P84 P71.9 M11.887 E67.6 E41 712.17 265.0 E78.8 272.9 277.6 274.9 276.0 E56.9 E70.40 E83.110 M11.229 268.9 712.25 272 270 330.2 271 712.38 M11.839 264.2 275.09 712.30 276.1 R63.6 264.8 277.1 270.9 P70.7 712.21 278.8 272.6 712.9 275.03 M10.00 270.3 270.1 E88.3 E50.0 N20.0 277.8 P70.0 E63.9 E71.50 278.00 712.29 E74.4 274.11 E83.10 712.24 275.2 E78.1 P70.8 712.99 330.3 712.82 712.81 M11.219 M1A.9xx E50.9 251.1 E78.5 E78.6 272.8 P19.2 E71.318 264.7 E71.0 783.0 712.26 712.20 269.8 M1A.00x1 P72.9 265.2 M11.239 712.14 275.1</p>
Musculoskeletal	<p>553.9 M84.574D Q67.5 M84.442S M70.30 M25.761 M46.87 550.00 M60.231 M71.829 M85.622 M84.675A M84.753D M08.429 733.20 M84.472 M60.239 M84.346K 727.6 M12.332 735.1 M84.550K M62.241 M71.062 M66.219 M05.051 M24.275 552 M67.479 755.67 M89.322 M27.59 M25.842 M80.821S M84.753A M84.573P M54.02 718.80 M53.86 M85.58 Q65.00 733.43 717.41 M24.174 718.9 M85.012 M48.58XG M15.19 M45.12 M42.21 M20.21 M84.522K M16.10 M23.041 M66.211 M19 M13.822 M24.231 717.82 M90.871 M87.035 M48.57XD M96.679 M21.40 M32.12 M93.841 718.82 M46.96 M65.341 M11.10 M25.222 M67.279 M80.019S 719.99 M66.172 M84.462K M84.442K M96.639 M80.011S M08.231 K43.7 M08.849 M23.007 M12.212 M48.02 M84.619G M65.171 M67.229 M66.272 M60.819 719.90 M11.049 M84.553A M85.812 M89.312 719.70 M06.811 M12.449 M44.9 M24.174 718.9 M85.012 M48.58XG M15.19 M45.12 M42.21 M20.21 M84.522K M16.10 M23.041 M66.211 M19 M13.822 M24.231 717.82 M90.871 M87.035 M48.57XD M96.679 M21.40 M32.12 M93.841 718.82 M46.96 M65.341 M11.10 M25.222 M67.279 M80.019S 719.99 M66.172 M84.462K M84.442K M96.639 M80.011S M08.231 K43.7 M08.849 M23.007 M12.212 M48.02 M84.619G M65.171 M67.229 M66.272 M60.819 719.90 M11.049 M84.553A M85.812 M89.312 719.70 M06.811 M12.449 M44.9 M24.174 718.9 M85.012 M48.58XG M15.19 M45.12 M42.21 M20.21 M84.522K M16.10 M23.041 M66.211 M19 M13.822 M24.231 717.82 M90.871 M87.035 M48.57XD M96.679 M21.40 M32.12 M93.841 718.82 M46.96 M65.341 M11.10 M25.222 M67.279 M80.019S 719.99 M66.172 M84.462K M84.442K M96.639 M80.011S M08.231 K43.7 M08.849 M23.007 M12.212 M48.02 M84.619G M65.171 M67.229 M66.272 M60.819 719.90 M11.049 M84.553A M85.812 M89.312 719.70 M06.811 M12.449 M44.9 M24.174 718.9 M85.012 M48.58XG M15.19 M45.12 M42.21 M20.21 M84.522K M16.10 M23.041 M66.211 M19 M13.822 M24.231 717.82 M90.871 M87.035 M48.57XD M96.679 M21.40 M32.12 M93.841 718.82 M46.96 M65.341 M11.10 M25.222 M67.279 M80.019S 719.99 M66.172 M84.462K M84.442K M96.639 M80.011S M08.231 K43.7 M08.849 M23.007 M12.212 M48.02 M84.619G M65.171 M67.229 M66.272 M60.819 719.90 M11.049 M84.553A M85.812 M89.312 719.70 M06.811 M12.449 M44.9 M24.174 718.9 M85.012 M48.58XG M15.19 M45.12 M42.21 M20.21 M84.522K M16.10 M23.041 M66.211 M19 M13.822 M24.231 717.82 M90.871 M87.035 M48.57XD M96.679 M21.40 M32.12 M93.841 718.82 M46.96 M65.341 M11.10 M25.222 M67.279 M80.019S 719.99 M66.172 M84.462K M84.442K M96.639 M80.011S M08.231 K43.7 M08.849 M23.007 M12.212 M48.02 M84.619G M65.171 M67.229 M66.272 M60.819 719.90 M11.049 M84.553A M85.812 M89.312 719.70 M06.811 M12.449 M44.9 M24.174 718.9 M85.012 M48.58XG M15.19 M45.12 M42.21 M20.21 M84.522K M16.10 M23.041 M66.211 M19 M13.822 M24.231 717.82 M90.871 M87.035 M48.57XD M96.679 M21.40 M32.12 M93.841 718.82 M46.96 M65.341 M11.10 M25.222 M67.279 M80.019S 719.99 M66.172 M84.462K M84.442K M96.639 M80.011S M08.231 K43.7 M08.849 M23.007 M12.212 M48.02 M84.619G M65.171 M67.229 M66.272 M60.819 719.90 M11.049 M84.553A M85.812 M89.312 719.70 M06.811 M12.449 M44.9 M24.174 718.9 M85.012 M48.58XG M15.19 M45.12 M42.21 M20.21 M84.522K M16.10 M23.041 M66.211 M19 M13.822 M24.231 717.82 M90.871 M87.035 M48.57XD M96.679 M21.40 M32.12 M93.841 718.82 M46.96 M65.341 M11.10 M25.222 M67.279 M80.019S 719.99 M66.172 M84.462K M84.442K M96.639 M80.011S M08.231 K43.7 M08.849 M23.007 M12.212 M48.02 M84.619G M65.171 M67.229 M66.272 M60.819 719.90 M11.049 M84.553A M85.812 M89.312 719.70 M06.811 M12.449 M44.9 M24.174 718.9 M85.012 M48.58XG M15.19 M45.12 M42.21 M20.21 M84.522K M16.10 M23.041 M66.211 M19 M13.822 M24.231 717.82 M90.871 M87.035 M48.57XD M96.679 M21.40 M32.12 M93.841 718.82 M46.96 M65.341 M11.10 M25.222 M67.279 M80.019S 719.99 M66.172 M84.462K M84.442K M96.639 M80.011S M08.231 K43.7 M08.849 M23.007 M12.212 M48.02 M84.619G M65.171 M67.229 M66.272 M60.819 719.90 M11.049 M84.553A M85.812 M89.312 719.70 M06.811 M12.449 M44.9 M24.174 718.9 M85.012 M48.58XG M15.19 M45.12 M42.21 M20.21 M84.522K M16.10 M23.041 M66.211 M19 M13.822 M24.231 717.82 M90.871 M87.035 M48.57XD M96.679 M21.40 M32.12 M93.841 718.82 M46.96 M65.341 M11.10 M25.222 M67.279 M80.019S 719.99 M66.172 M84.462K M84.442K M96.639 M80.011S M08.231 K43.7 M08.849 M23.007 M12.212 M48.02 M84.619G M65.171 M67.229 M66.272 M60.819 719.90 M11.049 M84.553A M85.812 M89.312 719.70 M06.811 M12.449 M44.9 M24.174 718.9 M85.012 M48.58XG M15.19 M45.12 M42.21 M20.21 M84.522K M16.10 M23.041 M66.211 M19 M13.822 M24.231 717.82 M90.871 M87.035 M48.57XD M96.679 M21.40 M32.12 M93.841 718.82 M46.96 M65.341 M11.10 M25.222 M67.279 M80.019S 719.99 M66.172 M84.462K M84.442K M96.639 M80.011S M08.231 K43.7 M08.849 M23.007 M12.212 M48.02 M84.619G M65.171 M67.229 M66.272 M60.819 719.90 M11.049 M84.553A M85.812 M89.312 719.70 M06.811 M12.449 M44.9 M24.174 718.9 M85.012 M48.58XG M15.19 M45.12 M42.21 M20.21 M84.522K M16.10 M23.041 M66.211 M19 M13.822 M24.231 717.82 M90.871 M87.035 M48.57XD M96.679 M21.40 M32.12 M93.841 718.82 M46.96 M65.341 M11.10 M25.222 M67.279 M80.019S 719.99 M66.172 M84.462K M84.442K M96.639 M80.011S M08.231 K43.7 M08.849 M23.007 M12.212 M48.02 M84.619G M65.171 M67.229 M66.272 M60.819 719.90 M11.049 M84.553A M85.812 M89.312 719.70 M06.811 M12.449 M44.9 M24.174 718.9 M85.012 M48.58XG M15.19 M45.12 M42.21 M20.21 M84.522K M16.10 M23.041 M66.211 M19 M13.822 M24.231 717.82 M90.871 M87.035 M48.57XD M96.679 M21.40 M32.12 M93.841 718.82 M46.96 M65.341 M11.10 M25.222 M67.279 M80.019S 719.99 M66.172 M84.462K M84.442K M96.639 M80.011S M08.231 K43.7 M08.849 M23.007 M12.212 M48.02 M84.619G M65.171 M67.229 M66.272 M60.819 719.90 M11.049 M84.553A M85.812 M89.312 719.70 M06.811 M12.449 M44.9 M24.174 718.9 M85.012 M48.58XG M15.19 M45.12 M42.21 M20.21 M84.522K M16.10 M23.041 M66.211 M19 M13.822 M24.231 717.82 M90.871 M87.035 M48.57XD M96.679 M21.40 M32.12 M93.841 718.82 M46.96 M65.341 M11.10 M25.222 M67.279 M80.019S 719.99 M66.172 M84.462K M84.442K M96.639 M80.011S M08.231 K43.7 M08.849 M23.007 M12.212 M48.02 M84.619G M65.171 M67.229 M66.272 M60.819 719.90 M11.049 M84.553A M85.812 M89.312 719.70 M06.811 M12.449 M44.9 M24.174 718.9 M85.012 M48.58XG M15.19 M45.12 M42.21 M20.21 M84.522K M16.10 M23.041 M66.211 M19 M13.822 M24.231 717.82 M90.871 M87.035 M48.57XD M96.679 M21.40 M32.12 M93.841 718.82 M46.96 M65.341 M11.10 M25.222 M67.279 M80.019S 719.99 M66.172 M84.462K M84.442K M96.639 M80.011S M08.231 K43.7 M08.849 M23.007 M12.212 M48.02 M84.619G M65.171 M67.229 M66.272 M60.819 719.90 M11.049 M84.553A M85.812 M89.312 719.70 M06.811 M12.449 M44.9 M24.174 718.9 M85.012 M48.58XG M15.19 M45.12 M42.21 M20.21 M84.522K M16.10 M23.041 M66.211 M19 M13.822 M24.231 717.82 M90.871 M87.035 M48.57XD M96.679 M21.40 M32.12 M93.841 718.82 M46.96 M65.341 M11.10 M25.222 M67.279 M80.019S 719.99 M66.172 M84.462K M84.442K M96.639 M80.011S M08.231 K43.7 M08.849 M23.007 M12.212 M48.02 M84.619G M65.171 M67.229 M66.272 M60.819 719.90 M11.049 M84.553A M85.812 M89.312 719.70 M06.811 M12.449 M44.9 M24.174 718.9 M85.012 M48.58XG M15.19 M45.12 M42.21 M20.21 M84.522K M16.10 M23.041 M66.211 M19 M13.822 M24.231 717.82 M90.871 M87.035 M48.57XD M96.679 M21.40 M32.12 M93.841 718.82 M46.96 M65.341 M11.10 M25.222 M67.279 M80.019S 719.99 M66.172 M84.462K M84.442K M96.639 M80.011S M08.231 K43.7 M08.849 M23.007 M12.212 M48.02 M84.619G M65.171 M67.229 M66.272 M60.819 719.90 M11.049 M84.553</p>

Continued on next page

Continued on next page

Respiratory	J27 J63.0 J12.9 J37.0 P26.0 J39 R05 J69 J15.1 J67.0 516.0 P24.21 J95.859 770.3 502 J09 X9 J34.3 J52 770.81 J03.00 786.9 J9 J51.02 277.09 J20.6 J21.8 J33.8 P28.0 J95.850 J84.848 770.0 J67.7 P28.2 R06.7 J34.89 P27.8 478.30 J01.21 J12.3 518.3 G47.35 495.7 J11.82 J95.71 J64 J82 P28.11 J01.20 J15.8 J45.21 495.4 J39.9 J95.889 P23 P26.9 J95.811 J70.1 J73 J95.862 J66.1 J79.3 J20.9 J51.2 J41.8 J84.112 J03.518 518.1 J94.9 J47.18 J18.1 J34.81 J768.4 773.34 J22.9 J93 S22.9 R06 J68.8 J45.30 J51.9 327.29 J20.0 J63.5 J84.0 P91.63 770.15 R06.81 495 P19.0 J01.10 770.7 J95.863 S09.1 J02.0 J34.9 J84.02 J01.25 J12.16 J81.0 J94.0 478.3 J23 J10.81 J84.116 495.1 J50.7 J05.10 J26.1 J11.0 J06.0 J96.91 J33 R06.4 J45.991 516.3 327.70 J20.12 J49.1 J32.0 768.4 770.8 J01.00 R06.01 J01.94.1 J51.2 786.51 J38.02 J95.830 J38.01 J478.79 770.84 J68.3 J10.08 J84.115 J25 J01.41 M34.81 P27.0 J07 512.8 R06.00 J11.08 J46 R09.3 J49 J70.5 J95.4 J05.0 J95.02 J31.0 J95.821 P28.40 J91.8 768.71 J30.5 J38 J70.2 J63.4 J86.4 478.4 J10.00 J84.9 J96.20 J20.3 J30.2 J31 J01.01 J29 R07.9 J75 J15.7 327.22 J61 J93.1 J02.9 J47.8 75.75 P28.19 J45.22 J96.11 J67.8 J85.0 J35.0 J96.21 J21.1 J11.1 J35.01 P05 505 491.1 P28.10 J15.13 J38.00 J67 492.8 R06.2 770.88 J15.20 J58.5 P07.3 R07.3 J01.9 J51.62 J20.1 J22.1 J52.0 G47.30 H68.1 P23.0 J35.3 J35.3 J95.889 P10.6 J60.9 J95.860 J81 J43.8 J30.8 J68.0 J48.3 J78.4 34 J18.9 J94.2 J70.8 J35.9 J33.0 R07.1 P07.1 J70.5 401.0 J43.2 478.7 J4.75 J45.20 A70.4 J45.40 786.06 327.23 P28.89 R06.9 J39.0 J98.9 J47.0 J98.3 P22.1 J39.8 R06.83 J65 P24.10 R07.0 J37 R07.89 517 503 786.52 J01 491.20 J95.72 J28 J09.3 X70 P23.3 J01.80 J97 J12.89 J18.8 516 J67.3 495.9 J40 P24.1 E84.9 J10.01 500.0 J11.2 519.501 J93.0 J98.01 P22.49.8 J08 G47.51 518.8 327.20 J20.25 P78.06 J01 786.50 P24.40 J71 770.2 J18.843 J93.81 J51.0 495.3 J30.0 786.00 J92 J49.1 21 J51.9 J92 768.0 R06.89 J56 J01.11 770.5 P19.1 J70.3 516.3 J32 J41.1 J85.3 J20.8 J21.9 J10.89 J68.2 518.2 R06.3 P25.1 R06.5 J91 P25.2 G47.36 J45.901 J16.8 J84.8 P28.4 R07.82 J18.2 J34 J74 J30.89 J78.7 768.0 J11.83 J95.84 J59 478.5 J15.5 J48.17 J35.02 J18.0 768.72 J78.3 P51.92 J05.04 J63.6 P28.9 P19.9 J95.8 J51.9 491.1 J66.2 517.8 J95.822 J01.40 J48 519.09 J51.84 G47.39 J51 786.59 J27.03 J84.111 J86.0 P23.9 J34.0 J18.1 J49.1 R06.82 J32.2 J786.2 J96.2 J96.22 J71 J01.31 J63.3 J49.5 494.0 P24.81 J95.2 J42 J76 J95.09 J84.09 J76.9 J35.8 J01.91 G47.34 J06.9 J62.0 J96.10 J20.4 J40 491.9 R06.82 J32.2 J03.80 516.36 770.11 P23.4 J01.30 J84.8 J49.71 J98.4 J10.00 R06.03 J91.22 786.8 R07.491 J02.14 J66 P28.1 J15.211 J09.2 J93.9 J67.5 J70.4 J45.32 84.09 J70.6 J89.0 J01.81 P28.19 J51.8 J32.8 J3.3 J10.82 786.05 J47.8 32 J34.2 J84.8 J18.1 J98.8 327.26 J96.00 516.8 518.0 519.4 770.9 J38.3 R06.00 J09.3 J91 G47.37 327.2 J34.1 J12.0 J21.1 J01 P22.0 J12.2 J39.3 P25.3 786.02 J95.851 J31.1 P25.1 502 J15.4 J43.0 P23.1 495.2 478.31 P24.30 770.18 P24.31 J57 J63 J95.00 J69.0 J38.4 J09 277.0 J04.30 J95.5 J45.902 504 P24.1 J45 J03.02 P23.6 R06.1 519.0 786.1 J37.1 J15.29 J98.6 J44.9 J84.10 P27.9 J44.0 J96.00 J517.3 E84.0 J45.998 516.9 J86.9 471. J63.2 491.0 J68.1 786.07 J98.19 P25 J66.8 768.3 518.8 J94.5 J30.3 J72 J44.2 J15.9 J01.6 J16.8 J42.2 770.85 J66.0 J20.2 519.11 277.00 J47.9 P27.1 770.77 J91.62 P22.9 786.03 J84.117 J81.113 516.30 770.16 P23.2 495.0 J67.6 J39.3 J82 J98.5 J51.63 37 770.89 J88.7 768.6 J510 J40.4 J24.4 J32.24 J43.9 J77 J84.82 J20.8 510.9 770.13 P28.3 J12.87 478.32 478.70 J96 J38.6 J41.0 J24 J68.1 J11.89 J45.909 J03.81 J98.09 J51.89 P19.1 J61.77 70.10 J67.2 J15.0 J39.2 494.1 516.3 G47.33 327.22 J02.8 J62.6 E84.11 R09.89 J47 277.01 J95.831 J85 P23.5 J94.8 P24 J04.0 J69.8 517.2 R07.2 495.5 786.09 J20 J12.81 J45.31 J58 J45.51 J43.1 E84.19 P84 494 J55 J33.9 516.31 J30.9 P22.8 J98.59 J04.10 J38.1 J53 P25.8 P23.8 J62 J51.36 J84.842 J54 J92.0 J32.1 J76.9 J04.11 786.04 J68.9 J70.0 327.21 J90 P24.0 G47.33 327.2 J02.8 J62.6 E84.11 J71.10 J83.3 J32.3 J10.1 492.514 768.2 327.27 J95.61 J38.7 J16.0 J20.7 J47.8 0.471.1 R06.02 J15.6 J95.1 J02 B9.518 J15.4 J68.4 J96.01 J30.9 J35.1 516.35 J45.998 786.3 J35.2 D62.8 J91.1 J00.500 J93.12 770.82 J84.0 516.1 J45.41 J88 327.25 J84.114 P27 J20.3 471.9 J96.9 J11.80 J18.1 518.82 495.3 J39.1 J98.11 J10.2 J80 519.8 J45.52 J96.12 518.83 J05.11 R06.4 770.17 J69.1 J67.4 J84.841 J99.8 P26.2 J11 J63.1 J78 768.5 J85.2 J45.50 P26 P28.3 J15.9 J70.9 786.4 J36 J05 J95.861 J04.31 J67.9 J85 J04 478.71 471.0 518.81 E84.8 J98 J19 J21.0 J94 J43.9 770.14 516.34 J30.81 510.0 J45.42 J13 J26 768.73 R22.2 J96.92
-------------	--

TABLE S8: Feature Definitions (Total number of features used: 380)

Feature name	Explanation	n _{features}
feature scores relative to phenotype score	Mean p-score of feature codes within sequence divided by general p-score of feature	26
feature scores relative to whole score	Mean p-score of feature codes within sequence divided by mean p-score of all codes in the record	26
aggregation score	aggregation of the p-scores in the record	13
high scores proportion	proportion of codes with very high p-scores among all codes in the record	1
low scores proportion	proportion of codes with very low p-scores among all codes in the record	1
dynamics of mean score	mean p-score of second half of the record divided by mean p-score of first half of the record	1
dynamics of geometric mean score	geometric mean p-score of second half of the record divided by mean p-score of first half of the record	1
dynamics of st.dev score	standard deviation of p-scores of second half of the record divided by standard deviation of p-scores of first half of the record	1
dynamics of score range	range of p-scores of second half of the record divided by range of p-scores of first half of the record	1
dynamics of score skew	skew of p-scores of second half of the record divided by skew of p-scores of first half of the record	1
aggregation relative to phn score	aggregation of all feature 's mean scores divided by corresponding general p-score of feature	9
aggregation relative to whole score	aggregation of all feature 's mean scores divided by mean p-score of all codes in the record	9
predicted risk from pfsa model	predicted risk from pfsa model	1
predicted risk from seq model	predicted risk from seq model	1
predicted risk from pscore model	predicted risk from pscore model	1
predicted risk from rare model	predicted risk from rare model	1
age at screening	Patient age at the moment of the screening	1
feature proportion	Ratio of number of weeks with the codes of a given phenotype to the total number of weeks in sequence	26
feature prevalence	Ratio of number of weeks with the codes of a given phenotype to the number of weeks with any diagnosis code recorded	26
feature first incident	Time interval from observation date to the first phenotype code, normalized by record length	26
feature last incident	Time interval from observation date to the last phenotype code, normalized by record length	26
feature mean position	Mean time position of phenotype codes in the record, normalized by record length	26
feature streak	Length of the longest uninterrupted subsequence of weeks with the codes of a given phenotype recorded	26
Max/Mean/Std/Range intermission	Maximum/Mean/Standard Deviation/Range of the lengths of subsequences of consequent weeks with codes	4
Max/Mean/Std cluster	Maximum/Mean/Standard Deviation of the lengths of subsequences of consequent weeks without codes	3
Max/Std/Range prevalence	Maximum/Standard Deviation/Range of the phenotype prevalences	3
Density of DX Record	Proportion of weeks in a record observed where at least one DX code was recorded	1
feature	Sequence Likelihood Defect for a given phenotype	26
feature neg llk	Negative LogLikelihood score for a given phenotype	26
feature pos llk	Positive LogLikelihood score for a given phenotype	26
feature llk ratio	Ratio of Positive to Negative LogLikelihood score for a given phenotype	26
Mean Δ	Mean Sequence Likelihood Defect	1
Std. deviation Δ	Range of Sequence Likelihood Defects	1
Range Δ	Standard Deviation of Sequence Likelihood Defects	1
Mean neg llk	Mean Negative LogLikelihood score	1
Range neg llk	Range of Negative LogLikelihood score	1
Std. deviation neg llk	Standard Deviation of Negative LogLikelihood score	1
Mean pos llk	Mean Positive LogLikelihood score	1
Range pos llk	Range of Positive LogLikelihood score	1
Std. deviation pos llk	Standard Deviation of Positive LogLikelihood score	1
Mean llk ratio	Mean LogLikelihood score ratio	1
Range llk ratio	Range of LogLikelihood score ratio	1
Std. deviation llk ratio	Standard Deviation of LogLikelihood score ratio	1
high scores proportion	proportion of codes with very high p-scores among all codes in the record	1
low scores proportion	proportion of codes with very low p-scores among all codes in the record	1

^{*} Δ : Sequence Likelihood Defect (See Methods)[†] neg llk: loglikelihood of observed sequence being generated by the model inferred from control (See Methods)[‡] pos llk: loglikelihood of observed sequence being generated by the model inferred from positive (See Methods)

TABLE S9: Proportion of 0's, 1's and 2's on average in trinary encodings with 95% CI

cohort	sex	proportion of 0	proportion of 1	proportion of 2
control	Female	0.879 ± 0.002	0.013 ± 0.001	0.106 ± 0.002
control	Male	0.891 ± 0.003	0.012 ± 0.001	0.095 ± 0.003
positive	Female	0.858 ± 0.040	0.015 ± 0.013	0.126 ± 0.038
positive	Male	0.873 ± 0.037	0.014 ± 0.012	0.111 ± 0.035
TOTAL	Female	0.879 ± 0.002	0.013 ± 0.001	0.106 ± 0.002
TOTAL	Male	0.891 ± 0.002	0.012 ± 0.001	0.095 ± 0.002

TABLE S10: Out-of-sample performance achieved (mean AUC) when training dataset is balanced (Note: performance degrades as we attempt to train with more balanced data, e.g., downsampling ratio of 1 is the case where we sample the control cohort to use only as many patients as in the positive cohort)

sex	downsampling ratio	all patients	age 65+ years	age < 65 years	frail
Female	1	0.755	0.715	0.694	0.732
Female	2	0.756	0.723	0.700	0.736
Female	5	0.768	0.735	0.727	0.752
Female	10	0.781	0.750	0.737	0.769
Female	20	0.772	0.750	0.728	0.759
Female	40	0.790	0.760	0.743	0.772
Male	1	0.724	0.665	0.690	0.721
Male	2	0.743	0.701	0.698	0.746
Male	5	0.754	0.708	0.722	0.761
Male	10	0.751	0.711	0.718	0.757
Male	20	0.759	0.718	0.725	0.759
Male	40	0.759	0.714	0.729	0.759

TABLE S11: Cohort Sizes

sex	cardiac event within week	n_{positive}	n_{control}	$n_{\text{high risk}}$	n
M	2	385	185528	146782	185913
M	4	464	185528	146782	185992
F	2	337	258981	204170	259318
F	4	418	258981	204170	259399
	Total	882	444509	350952	445391

TABLE S12: Out-of-sample predictive performance to predict MACE 4 weeks after surgery in sub-cohorts with pre-existing conditions

Pre-existing phenotype	Female CCoR	Female RCRI	Male CCoR	Male RCRI
Allergic	0.77	0.71	0.81	0.78
CNS	0.80	0.67	0.89	0.75
Cardiovascular	0.78	0.69	0.80	0.67
Development	0.77	0.83	0.86	0.68
Digestive	0.81	0.73	0.80	0.71
Endocrine	0.80	0.69	0.80	0.67
Frailty	0.78	0.69	0.85	0.73
Health Services	0.81	0.71	0.83	0.71
Hematologic	0.80	0.72	0.85	0.74
Hypertension	0.77	0.68	0.80	0.66
Immune	0.81	0.71	0.82	0.70
Infections Fungal and Other	0.80	0.76	0.84	0.68
Infections General	0.80	0.75	0.84	0.68
Infections Respiratory	0.76	0.64	0.83	0.63
Injuries	0.79	0.73	0.84	0.69
Integumentary	0.78	0.69	0.80	0.72
Metabolic	0.80	0.70	0.82	0.70
Musculoskeletal	0.81	0.71	0.83	0.70
Neoplastic	0.87	0.75	0.78	0.68
Ophthalmological	0.79	0.69	0.76	0.65
Otic	0.79	0.75	0.84	0.62
PNS	0.80	0.70	0.84	0.73
Psychiatric	0.85	0.74	0.89	0.72
Reproductive	0.80	0.70	0.82	0.79
Respiratory	0.81	0.70	0.83	0.67

TABLE S13: Out-of-sample predictive performance in sub-cohorts stratified by age

age	gender	auc CCoR	auc RCRI	n _{positive}	n _{control}
45 - 55	F	0.59	0.58	3	9056
45 - 55	M	0.89	0.79	5	7027
55 - 65	F	0.80	0.66	31	27256
55 - 65	M	0.78	0.63	39	20244
65 - 75	F	0.81	0.73	34	14235
65 - 75	M	0.73	0.58	31	9635
75 - 85	F	0.70	0.65	25	8515
75 - 85	M	0.79	0.71	36	5164
85 - 95	F	0.80	0.67	8	1578
85 - 95	M	0.75	0.48	12	847

TABLE S14: Out-of-sample* performance for predicting MACE with 4 weeks of Hip or Knee Arthroplasty (Primary Endpoint) at 99% Specificity: CCoR vs. RCRI**

sex	cohort	model	sensitivity	PPV	acc	LR+	LR-	AUC
Female	< 65	RCRI	0.01±0.02	0.008±0.000	0.987±0.006	0.04±0.1	1.00±0.02	0.639 ± 0.039
Female	< 65	CCoR	0.14±0.06	0.047±0.004	0.987±0.006	14.12±1.1	0.87±0.06	0.775 ± 0.035
Male	< 65	RCRI	0.07±0.03	0.025±0.003	0.987±0.000	7.42±0.8	0.94±0.03	0.682 ± 0.034
Male	< 65	CCoR	0.15±0.06	0.065±0.003	0.987±0.006	19.94±0.9	0.85±0.06	0.783 ± 0.030
Female	65+	RCRI	0.03±0.01	0.012±0.002	0.987±0.000	3.39±0.5	0.97±0.01	0.664 ± 0.028
Female	65+	CCoR	0.09±0.06	0.036±0.004	0.987±0.006	10.70±1.1	0.92±0.06	0.771 ± 0.025
Male	65+	RCRI	0.03±0.01	0.011±0.001	0.987±0.006	3.17±0.4	0.98±0.00	0.661 ± 0.026
Male	65+	CCoR	0.09±0.05	0.031±0.002	0.987±0.006	9.09±0.6	0.92±0.05	0.762 ± 0.023
Female	all patients	RCRI	0.05±0.01	0.016±0.003	0.987±0.000	4.67±0.8	0.96±0.01	0.688 ± 0.023
Female	all patients	CCoR	0.13±0.01	0.044±0.007	0.987±0.006	13.19±2.1	0.88±0.01	0.801 ± 0.019
Male	all patients	RCRI	0.05±0.01	0.019±0.002	0.987±0.000	5.44±0.7	0.95±0.01	0.705 ± 0.020
Male	all patients	CCoR	0.12±0.05	0.042±0.001	0.987±0.006	12.44±0.3	0.89±0.05	0.802 ± 0.018
Female	frail***	RCRI	0.03±0.00	0.009±0.002	0.987±0.006	2.59±0.6	0.98±0.00	0.670 ± 0.028
Female	frail	CCoR	0.11±0.03	0.036±0.002	0.987±0.000	10.36±0.7	0.90±0.03	0.791 ± 0.025
Male	frail	RCRI	0.06±0.04	0.020±0.002	0.987±0.006	5.91±0.7	0.95±0.04	0.727 ± 0.027
Male	frail	CCoR	0.15±0.03	0.050±0.002	0.987±0.000	15.02±0.7	0.86±0.03	0.810 ± 0.024
Female	high risk****	RCRI	0.02±0.00	0.008±0.001	0.987±0.006	2.19±0.3	0.99±0.00	0.581 ± 0.029
Female	high risk	CCoR	0.07±0.03	0.026±0.002	0.987±0.000	7.73±0.6	0.94±0.03	0.737 ± 0.026
Male	high risk	RCRI	0.03±0.01	0.010±0.001	0.987±0.006	2.91±0.3	0.98±0.00	0.617 ± 0.026
Male	high risk	CCoR	0.09±0.05	0.033±0.002	0.987±0.006	9.61±0.7	0.92±0.05	0.729 ± 0.024
Female	low risk†	CCoR	0.22±0.04	0.071±0.004	0.987±0.000	21.81±1.3	0.79±0.04	0.765 ± 0.036
Male	low risk	CCoR	0.11±0.04	0.042±0.003	0.987±0.000	5.96±0.9	0.90±0.04	0.766 ± 0.032

Abbreviations, AUC, area under the receiver operating characteristic curve; CCoR, Cardiac Co-Morbidity Risk Score; LR+, positive likelihood ratio; LR-, negative likelihood ratio; NPV, negative predictive value; PPV, positive predictive value; acc, accuracy; RCRI, Revised Cardiac Risk Index

*50% (n=445391) of cohort used for validation

**Because of insufficient availability of relevant laboratory data in the Truven dataset, presence of at least one diagnostic code for chronic kidney disease stage III or higher in the medical record in the year before the date of arthroplasty was used as a surrogate for the RCRI condition, serum creatinine concentration > 2.0mg/dL (to convert to micromoles per liter, multiply by 88.4).

***Frail subcategory was defined by codes specified in Table S7

****Low risk subcohort comprises patients with RCRI score 0. High risk patients have RCRI score > 0.

†No RCRI performance logged for low-risk patients, since their RCRI score is zero.

TABLE S15: Out-of-sample* performance for predicting MACE with 2 weeks of Hip or Knee Arthroplasty (Secondary Endpoint) at 99% Specificity: CCoR vs. RCRI**

sex	cohort	model	sensitivity	PPV	acc	LR+	LR-	AUC
Female	< 65	RCRI	0.01±0.02	0.009±0.000	0.987±0.006	0.04±0.1	1.00±0.02	0.647±0.044
Female	< 65	CCoR	0.11±0.01	0.073±0.023	0.987±0.006	22.87±8.5	0.90±0.01	0.787±0.039
Male	< 65	RCRI	0.09±0.02	0.032±0.009	0.987±0.000	9.33±2.8	0.92±0.02	0.688±0.037
Male	< 65	CCoR	0.14±0.04	0.056±0.004	0.987±0.000	16.83±1.3	0.87±0.04	0.797±0.033
Female	65+	RCRI	0.04±0.04	0.013±0.002	0.987±0.006	3.66±0.6	0.97±0.03	0.671±0.030
Female	65+	CCoR	0.10±0.03	0.041±0.002	0.987±0.000	12.08±0.7	0.91±0.04	0.787±0.027
Male	65+	RCRI	0.03±0.00	0.010±0.001	0.987±0.000	2.79±0.4	0.98±0.00	0.667±0.028
Male	65+	CCoR	0.09±0.03	0.032±0.002	0.987±0.000	9.32±0.6	0.92±0.03	0.780±0.025
Female	all patients	RCRI	0.05±0.01	0.018±0.003	0.987±0.000	5.17±0.9	0.96±0.01	0.692±0.025
Female	all patients	CCoR	0.14±0.06	0.048±0.001	0.987±0.006	14.48±0.3	0.87±0.06	0.809±0.021
Male	all patients	RCRI	0.05±0.01	0.018±0.002	0.987±0.000	5.15±0.7	0.96±0.01	0.710±0.022
Male	all patients	CCoR	0.14±0.02	0.047±0.007	0.987±0.000	13.98±2.1	0.87±0.02	0.813±0.019
Female	frail***	RCRI	0.03±0.01	0.012±0.003	0.987±0.000	3.43±0.9	0.98±0.01	0.676±0.032
Female	frail	CCoR	0.12±0.03	0.041±0.009	0.987±0.000	11.78±2.7	0.89±0.03	0.807±0.027
Male	frail	RCRI	0.06±0.01	0.020±0.003	0.987±0.000	5.82±0.9	0.95±0.01	0.736±0.029
Male	frail	CCoR	0.17±0.03	0.059±0.009	0.987±0.000	17.76±3.0	0.84±0.03	0.825±0.025
Female	high risk****	RCRI	0.02±0.03	0.008±0.001	0.987±0.006	2.35±0.4	0.99±0.02	0.584±0.032
Female	high risk	CCoR	0.08±0.00	0.029±0.006	0.987±0.006	8.45±1.9	0.93±0.01	0.742±0.028
Male	high risk	RCRI	0.03±0.03	0.009±0.001	0.987±0.006	2.69±0.4	0.98±0.02	0.628±0.028
Male	high risk	CCoR	0.09±0.02	0.033±0.007	0.987±0.000	9.66±2.3	0.92±0.02	0.737±0.026
Female	low risk [†]	CCoR	0.28±0.07	0.092±0.002	0.988±0.006	28.78±0.7	0.72±0.07	0.779±0.040
Male	low risk	CCoR	0.12±0.04	0.047±0.004	0.987±0.000	6.78±1.1	0.89±0.05	0.793±0.035

Abbreviations, AUC, area under the receiver operating characteristic curve; CCoR, Cardiac Co-Morbidity Risk Score; LR+, positive likelihood ratio; LR-, negative likelihood ratio; NPV, negative predictive value; PPV, positive predictive value; acc, accuracy; RCRI, Revised Cardiac Risk Index

*50% (n=445391) of cohort used for validation

**Because of insufficient availability of relevant laboratory data in the Truven dataset, presence of at least one diagnostic code for chronic kidney disease stage III or higher in the medical record in the year before the date of arthroplasty was used as a surrogate for the RCRI condition, serum creatinine concentration > 2.0mg/dL (to convert to micromoles per liter, multiply by 88.4).

***Frail subcategory was defined by codes specified in Table S7

****Low risk subcohort comprises patients with RCRI score 0. High risk patients have RCRI score > 0.

[†]No RCRI performance logged for low-risk patients, since their RCRI score is zero.

TABLE S16: Statistical significance of CCoR AUC > RCRI AUC (* denotes significance at 95% level, ** denotes significance at 99% level)

sex	model	weeks after surgery	RCRI auc	CCoR auc	p value	significance
Female	all patients	2	0.692	0.809	0.010	**
		4	0.688	0.801	0.007	**
	frail [†]	2	0.676	0.807	0.017	*
		4	0.670	0.791	0.014	*
	high risk [‡]	2	0.584	0.742	0.005	**
		4	0.581	0.737	0.003	**
	65+	2	0.671	0.787	0.020	*
		4	0.664	0.771	0.020	*
	< 65	2	0.647	0.787	0.036	*
		4	0.639	0.775	0.028	*
Male	all patients	2	0.710	0.813	0.010	*
		4	0.705	0.802	0.009	**
	frail	2	0.736	0.825	0.045	*
		4	0.727	0.810	0.042	*
	high risk	2	0.628	0.737	0.017	*
		4	0.617	0.729	0.011	*
	65+	2	0.667	0.780	0.017	*
		4	0.661	0.762	0.019	*
	< 65	2	0.688	0.797	0.047	*
		4	0.682	0.783	0.042	*

[†]Frail subcategory was defined by codes specified in Table S7

[‡]Low risk subcohort comprises patients with RCRI score 0. High risk patients have RCRI score > 0.

TABLE S17: Out-of-sample performance achieved using only PFSA component of the CCoR model (Note: the performance is significantly degraded, with all p-values < 0.01.)

sex	prediction horizon	AUC
Female	2 weeks	0.696 ± 0.082
Female	4 weeks	0.698 ± 0.074
Male	2 weeks	0.679 ± 0.074
Male	4 weeks	0.656 ± 0.067

SUPPLEMENTAL METHODS

A. Time-series Modeling of Diagnostic History

Individual diagnostic histories can have long-term memory [52], implying that the order, frequency, and comorbid interactions between diseases are important for assessing the future risk of our target phenotype. We analyze patient-specific diagnostic code sequences by first representing the medical history of each patient as a set of stochastic categorical time-series — one each for a specific group of related disorders — followed by the inference of stochastic models for these individual data streams. These inferred generators are from a special class of Hidden Markov Models (HMMs), referred to as Probabilistic Finite State Automata (PFSA) [53]. The inference algorithm we use is distinct from classical HMM learning, and has important advantages related to its ability to infer structure, and its sample complexity (See Supplementary text, Section VI). We infer a separate class of models for the positive and control cohorts, and then the problem reduces to determining the probability that the short diagnostic history from a new patient arises from the positive as opposed to the control category of the inferred models.

B. Inference & Event Periods

We train our predictive pipeline with all diagnostic codes that are recorded in the past 26 years from the point at which a prediction is made. This period from which we use data to train our pipeline is called the “inference window”. Our aim is to make predictions on the occurrence of the target diagnostic codes at 2 years from the end of the inference window. For patients in the control cohort, we make sure that no target code appears for 26 years after the end of the inference window. Additionally, when making predictions further into the future (upto 4 years, as described in the main text), we always make sure that the control group has no target codes for 1 year after the predicted time of diagnosis, i.e., if we are making a prediction of a diagnosis 4 years in future, then control group patients are chosen to have no diagnosis in at least next 5 years.

C. Step 1: Partitioning The Human Disease Spectrum

We begin by partitioning the human disease spectrum into 26 non-overlapping categories. Each category is defined by a set of diagnostic codes from the International Classification of Diseases, Ninth Revision (ICD9) (See Table SI-S7 for description of the categories used in this study). For this study, we ended up using 4879398 and 7753318 diagnostic codes for males and females respectively (17554 and 19209 unique codes) spanning both ICD9 and ICD10 protocols (using ICD10 General Equivalence Mappings (GEMS) [54] equivalents where necessary), from a total 445391 patients. Transforming the diagnostic histories to report only the broad categories reduces the number of distinct codes that the pipeline needs to handle, thus improving statistical power. Our categories largely align with the top-level ICD9 categories, with small adjustments, e.g. bringing all infections under one category irrespective of the pathogen or the target organ. We do not pre-select the phenotypes; we want our algorithm to seek out the important patterns without any manual curation of the input data. For each patient, the past medical history is a sequence $(t_1, x_1), \dots, (t_m, x_m)$, where t_i are timestamps and x_i are ICD9 codes diagnosed at time t_i . We map individual patient history to a three-alphabet categorical time series z^k corresponding to the disease category k , as follows. For each week i , we have:

$$z_i^k = \begin{cases} 0 & \text{if no diagnosis codes in week } i \\ 1 & \text{if there exists a diagnosis of category } k \text{ in week } i \\ 2 & \text{otherwise} \end{cases} \quad (1)$$

The time-series z^k is observed in the inference period. Thus, each patient is represented by 43 mapped trinary series.

D. Step 2: Model Inference & The Sequence Likelihood Defect Δ

The mapped series, disease-category, and perioperative cardiac event diagnosis-status are considered to be independent sample paths, and we want to explicitly model these systems as specialized HMMs (PFSAs). We model the positive and the control cohorts and each disease category separately, ending up with a total of 104 HMMs at the population level (26 categories, 2 perioperative cardiac event status categories: positive and control, and 2 sexes). Each of these inferred models is a PFSA; a directed graph with probability-weighted edges, and acts as an optimal generator of the stochastic process driving the sequential appearance of the three letters

(as defined by Eq. (1)) corresponding to disease category, and perioperative cardiac event status-type (See Section VI in the Supplementary text for background on PFSA inference).

To reliably infer the perioperative cardiac event status-type of a new patient, i.e, the likelihood of a diagnostic sequence being generated by the corresponding perioperative cardiac event status-type model, we generalize the notion of Kullbeck-Leibler (KL) divergence [55] between probability distributions to a divergence $\mathcal{D}_{KL}(G||H)$ between ergodic stationary categorical stochastic processes [56] G, H as:

$$\mathcal{D}_{KL}(G||H) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{x:|x|=n} p_G(x) \log \frac{p_G(x)}{p_H(x)} \quad (2)$$

where $|x|$ is the sequence length, and $p_G(x), p_H(x)$ are the probabilities of sequence x being generated by the processes G, H respectively. Defining the log-likelihood of x being generated by a process G as :

$$L(x, G) = -\frac{1}{|x|} \log p_G(x) \quad (3)$$

The cohort-type for an observed sequence x — which is actually generated by the hidden process G — can be formally inferred from observations based on the following provable relationships (See Supplementary Text Section VI, Theorem 6 and 7):

$$\lim_{|x| \rightarrow \infty} L(x, G) = \mathcal{H}(G) \quad (4a)$$

$$\lim_{|x| \rightarrow \infty} L(x, H) = \mathcal{H}(G) + \mathcal{D}_{KL}(G||H) \quad (4b)$$

where $\mathcal{H}(\cdot)$ is the entropy rate of a process [32]. Importantly, Eq. (4) shows that the computed likelihood has an additional non-negative contribution from the divergence term when we choose the incorrect generative process. Thus, if a patient is eventually going to be diagnosed with perioperative cardiac event, then we expect that the disease-specific mapped series corresponding to her diagnostic history be modeled by the PFSA in the positive cohort. Denoting the PFSA corresponding to disease category j for positive and control cohorts as G_+, G_0^j respectively, we can compute the *sequence likelihood defect* (SLD, Δ^j) as:

$$\Delta^j \triangleq L(G_0^j, x) - L(G_+, x) \rightarrow \mathcal{D}_{KL}(G_0^j||G_+) \quad (5)$$

With the inferred PFSA models and the individual diagnostic history, we estimate the SLD measure on the right-hand side of Eqn. (5). The higher this likelihood defect, the higher the similarity of diagnosis history to that of women with perioperative cardiac event.

E. Step 3: Risk Estimation Pipeline With Semi-supervised & Supervised Learning Modules

The risk estimation pipeline operates on patient specific information limited to the available diagnostic history in the inference period, and produces an estimate of the relative risk of perioperative cardiac event, with an associated confidence value. To learn the parameters and associated model structures of this pipeline, we transform the patient specific data to a set of engineered features, and the feature vectors realized on the positive and control sets are used to train a gradient-boosting classifier [57]. The complete list of 380 features used is provided in Table 6.

We need two training sets: one to infer the models, and one to train the classifier with features derived from the inferred models. Thus, we do a random 3-way split of the set of unique patients into *feature-engineering* (25%), *training* (25%) and *test* (50%) sets. We use the feature-engineering set of ids first to infer our PFSA models (*unsupervised model inference in each category*), which then allows us to train the gradient-boosting classifier using the training set and PFSA models (*classical supervised learning*), and we finally execute out-of-sample validation on the test set. Fig. 1c in the main text shows the top 20 features ranked in order of their relative importance (relative loss in performance when dropped out of the analysis).

I. THRESHOLD SELECTION ON ROC CURVE

Once the ROC curve has been computed, we must choose a decision threshold to trade-off true positive rate and false positive rate. In situations where the number of negatives vastly outnumber the number of positives (which is the case in our problem), it is better to base this trade-off on a measure that is independent of the number of true negatives. The two popular measures considered in the literature are accuracy and the F1-score:

$$\text{accuracy} = \frac{t_p + t_n}{t_p + f_p + f_n + t_n} \quad (6)$$

$$F1 = \frac{2t_p}{2t_p + f_p + f_n} \quad (7)$$

The F1-score is the same as accuracy where the number of true negatives is the same as the number of true positives, thus partially correcting for the class imbalance.

The selection of the threshold may also be dictated by the current practice of ensuring high specificities in screening tests. Thus, a relevant clinically operating point is the one corresponding to 95% specificity, which is highlighted in Fig. 1a.

II. NOTE ON RECEIVER OPERATING CHARACTERISTICS (ROC) AND PRECISION-RECALL CURVES

The ROC curve is a plot between the False Positive rate (TPR) and the True Positive Rate (TPR), and the area under the ROC curve (AUC) is often used as a measure of classifier performance. For the sake of completeness, we introduce the relevant definitions:

In the following P denotes the total number of positive samples (number of patients who are eventually diagnosed), and N denotes the total number of negative samples (number of patients in the control group).

Definition 1. *True positive rate, true negative rate, false positive rate, positive predictive value (PPV), and prevalence (ρ) are defined as:*

$$TPR = \frac{t_p}{P} = \frac{t_p}{t_p + f_n} \quad (8)$$

$$TNR = \frac{t_n}{N} = \frac{t_n}{t_n + f_p} \quad (9)$$

$$FPR = 1 - TNR \quad (10)$$

$$PPV = \frac{t_p}{t_p + f_p} \quad (11)$$

$$\rho = \frac{P}{N + P} \quad (12)$$

where as before t_p, t_n, f_p, f_n are true positives, true negatives, false positives, and false negatives respectively.

Note that TPR is also referred to as **recall** or **sensitivity**, and PPV is also referred to as **precision**. True negative rate is also known as **specificity**.

A **precision-recall curve**, or a PPV-sensitivity curve is a plot between PPV and TPR.

Denoting sensitivity by s , and specificity by c , it follows that:

$$PPV = \frac{t_p/P}{t_p/P + (f_p/N)(N/P)} = \frac{TPR}{TPR + ((N - t_n)/N)(N/P)} \quad (13)$$

$$\Rightarrow PPV = \frac{s}{s + (1 - c)(\frac{1}{\rho} - 1)} \quad (14)$$

Thus, we note that for a fixed specificity and sensitivity, the PPV depends on prevalence. Indeed, it is clear from the above argument that PPV decreases with decreasing prevalence, and vice versa.

III. EFFECT OF CLASS IMBALANCE

ROC curves are generally assumed to be robust to class imbalance. Note that if we assume that patient outcomes are independent (which is well-justified in the case of a non-communicable condition, particularly in large databases), then t_p should scale linearly with the total number of positives P, implying:

$$TPR = \frac{t_p}{P} = \frac{t'_p}{P'} \quad (15)$$

implying that with different sizes of the set of positive samples (or negative samples), the ROC curve remains unchanged. In particular, note that even if the prevalence is very small (say 0.01%), we cannot cheat to boost the AUC by labeling all predictions as negative, or stating that risk is always zero: in that case, our P is very small, but our $t_p = 0$ strictly, implying that our $TPR = 0$, thus leading to a zero AUC. We can cheat to boost the accuracy (See the previous section), but not the AUC.

Note that while relative class sizes or imbalance does not affect the ROC (under the assumption that true positives and true negatives scale with the number of positives and negatives), very small absolute sample sizes might still result in poor performance of the model.

The precision-recall curves do get affected by class imbalance, or the prevalence, as shown by Eq (14). However, in diagnostic analysis, they are important since we are generally less interested in the number of true negatives; the ratio of false positives to the total number of positive recommendations by the algorithm is much more relevant, i.e., the PPV or the precision.

IV. GENERATING PFSA MODELS FROM SET OF INPUT STREAMS WITH VARIABLE INPUT LENGTHS

Our PFSA reconstruction algorithm [53] is distinct from standard HMM learning. We do not need to pre-specify structures, or the number of states in the algorithm, and all model parameters are inferred directly from data. Additionally, we can operate either with 1) a single input stream, or 2) a set of input streams of possibly varying lengths which are assumed to be different and independent sample paths from the unknown stochastic generator we are trying to infer. At an intuitive level, we use the input data to infer the length of histories one must remember to estimate the current state, and predict futures for the process being modeled. Thus, we do not step through the symbol streams with a pre-specified model structure, and avoid the need to have equal-length inputs. More details of the algorithm are provided in the next section.

The ability to model a set of input streams of varying lengths is particularly important, since medical histories of different patients are typically of different lengths.

V. PROBABILISTIC FINITE STATE AUTOMATA INFERENCE

A. Probabilistic Finite-State Automaton

Let Σ be a finite alphabet of symbols with size $|\Sigma|$. The set of sequences of length d over Σ is denoted by Σ^d . The set of finite but unbounded sequences over Σ is denoted by Σ^* , the Kleene star operation [58], i.e. $\Sigma^* = \bigcup_{d=0}^{\infty} \Sigma^d$. We use lower case Greek, for example σ or τ , for symbols in Σ , and lower case Latin, for example x or y , for sequences of symbols, i.e. $x = \sigma_1 \sigma_2 \dots \sigma_n$. We use $|x|$ to denote the length of x . The empty sequence is denoted by λ .

We denote the set of strictly infinite sequences over Σ by Σ^ω , and the set of strictly infinite sequences having x as prefix by $x\Sigma^\omega$. Let $\mathcal{S} = \{x\Sigma^\omega : x \in \Sigma^*\} \cup \{\emptyset\}$, we can verify that \mathcal{S} is a semiring [59] over Σ^ω . We use \mathcal{F} to denote the sigma algebra generated by \mathcal{S} .

Definition 2 (Stochastic Process over Σ). *A stochastic process over a finite alphabet Σ is a collection of Σ -valued random variables $\{X_t\}_{t \in \mathbb{N}}$ indexed by positive integers [56].*

We are specifically interested in processes in which the X_i s are not necessarily independently distributed.

Definition 3 (Sequence-Induced Measure and Derivative). *For a process \mathcal{P} , let $\Pr_{\mathcal{P}}(x)$ or simply $\Pr(x)$ denote the probability \mathcal{P} producing a sample path prefixed by x . The measure μ_x induced by a sequence $x \in \Sigma^*$ is the extension [59] to \mathcal{F} of the premeasure defined on the semiring \mathcal{S} given by*

$$\forall x, y \in \Sigma^*, \mu_x(y\Sigma^\omega) \triangleq \frac{\Pr(xy)}{\Pr(x)}, \text{ if } \Pr(x) > 0 \quad (16)$$

For any $d \in \mathbb{N}$, the d -th order derivative of a sequence x , written as ϕ_x^d , is defined to be the marginal distribution of μ_x on Σ^d , with the entry indexed by y denoted by $\phi_x^d(y)$. The first-order derivative is called the symbolic derivative and is denoted by ϕ_x for short.

Definition 4 (Probabilistic Nerode Equivalence and Causal States [60]). *For any pair of sequences $x, y \in \Sigma^*$, x is equivalent to y , written as $x \sim y$, if and only if either $\Pr(x) = \Pr(y) = 0$, or $\mu_x = \mu_y$. The equivalence class of a sequence x is denoted by $[x]$ and is called a causal state [61]. The cardinality of the set of causal states is called the probabilistic Nerode index, or the Nerode index for simplicity.*

We can see from the definition that causal states captures how the history of a process influences its future. Since the probabilistic Nerode equivalence is right invariant, it gives rise naturally to a automaton structure introduced below.

Definition 5 (Probabilistic Finite-State Automaton (PFSA)). *A PFSA G is defined by a quadruple $(Q, \Sigma, \delta, \tilde{\pi})$, where Q is a finite set, Σ is a finite alphabet, $\delta : Q \times \Sigma \rightarrow \Sigma$ is called the transition map, and $\tilde{\pi} : Q \rightarrow \mathbf{P}_\Sigma$, where*

\mathbf{P}_Σ is the space of probability distributions over Σ , is called the transition probability. The entry of $\tilde{\pi}(q)$ indexed by σ is denoted by $\tilde{\pi}(q, \sigma)$.

Definition 6 (Transition and Observation Matrices). The transition matrix Π is the $|Q| \times |Q|$ matrix with the entry indexed by q, q' , written as $\pi_{q,q'}$, satisfying

$$\pi_{q,q'} \triangleq \sum_{\{\sigma \in \Sigma | \delta(q, \sigma) = q'\}} \tilde{\pi}(q, \sigma) \quad (17)$$

and the observation matrix $\tilde{\Pi}$ is a $|Q| \times |\Sigma|$ matrix with the entry indexed by q, σ equaling $\tilde{\pi}(q, \sigma)$.

We note that both Π and $\tilde{\Pi}$ are stochastic, i.e. non-negative with rows summing up to 1.

Definition 7 (Extension of δ and $\tilde{\pi}$ to Σ^*). For any $x = \sigma_1 \dots \sigma_k$, $\delta(q, x)$ is defined recursively by

$$\delta(q, x) \triangleq \delta(\delta(q, \sigma_1 \dots \sigma_{k-1}), \sigma_k) \quad (18)$$

with $\delta(q, \lambda) = q$, and $\tilde{\pi}(q, x)$ is defined recursively by

$$\tilde{\pi}(q, x) \triangleq \prod_{i=1}^k \tilde{\pi}(\delta(q, \sigma_1 \dots \sigma_{i-1}), \sigma_i) \quad (19)$$

with $\tilde{\pi}(q, \lambda) = 1$.

Definition 8 (Strongly Connected PFSA). We say a PFSA is strongly connected if the underlying directed graph is strongly connected [62]. More precisely, a PFSA $G = (Q, \Sigma, \delta, \tilde{\pi})$ is strongly connected if for any pair of distinct states q and $q' \in Q$, there is an $x \in \Sigma^*$ such that $\delta(q, x) = q'$.

We assume all PFSA in the discussions in the sequel are strongly connected if not specified otherwise. For strongly connected PFSA G , there is a unique probability distribution over Q that satisfies $\mathbf{v}^T \Pi = \mathbf{v}^T$. This is the **stationary distribution** [63], [64] of G and is denoted as φ_G , or φ if G is understood.

Definition 9 (Γ -Expression). We can encode the information contained in δ and $\tilde{\pi}$ by a set of $|Q| \times |Q|$ matrices $\Gamma = \{\Gamma_\sigma | \sigma \in \Sigma\}$, where

$$\Gamma_\sigma|_{q,q'} \triangleq \begin{cases} \tilde{\pi}(q, \sigma) & \text{if } \delta(q, \sigma) = q', \\ 0 & \text{if otherwise.} \end{cases} \quad (20)$$

Γ_σ is called **event-specific transition matrix**, with the event being that σ is current the output. Γ_σ can also be extended to arbitrary $x \in \Sigma^*$ by defining $\Gamma_x = \prod_{i=1}^k \Gamma_{\sigma_i}$ with $\Gamma_\lambda = I$.

Definition 10 (Sequence-Induced Distribution on States). For a PFSA $G = (Q, \Sigma, \delta, \tilde{\pi})$ and a distribution φ_0 on Q , the **distribution on Q induced by a sequence x** is given by $\varphi_{G, \varphi_0}^T(x) = [\varphi_0^T \Gamma_x]$ with $\varphi_{G, \varphi_0}(\lambda) = \varphi_0$. The entry indexed by $q \in Q$ of the vector $\varphi_{G, \varphi_0}(x)$ is written as $\varphi_{G, \varphi_0}(x, q)$. When $\varphi_0 = \varphi_G$, the stationary distribution of G , we write $\varphi_{G, \varphi_0}(x)$ as $\varphi_G(x)$, or simply as $\varphi(x)$, if G is understood.

Definition 11 (Stochastic Process Generated by a PFSA). Let $G = (Q, \Sigma, \delta, \tilde{\pi})$ be a PFSA and let φ_0 be a distribution on Q , the Σ -valued stochastic process $\{X_t\}_{t \in \Sigma}$ generated by G and φ_0 satisfies that X_1 follows the distribution φ_0 and X_{t+1} follows the distribution $\varphi_{G, \varphi_0}(X_1 \dots X_t)$ for $t \in \mathbb{N}$.

For the rest of this paper, we will assume $\varphi_0 = \varphi_G$ if not specified otherwise. We can show that, when initialized with φ_G , the process generated by a PFSA G is stationary and ergodic. We also note the, for the process generate by G , we have $\phi_x = \varphi_G(x)^T \tilde{\Pi}$. Since $\varphi_G(\lambda) = \varphi_G$, the symbolic derivative of the empty sequence ϕ_λ is the stationary distribution on the symbols.

Definition 12 (Synchronizable PFSA and Synchronizing Sequence). A **synchronizing sequence** is a finite sequence that sends an arbitrary state of the PFSA to a fixed state [65]. To be more precise, let $G = (Q, \Sigma, \delta, \tilde{\pi})$ be a PFSA, we say a sequence $x \in \Sigma^*$ is a synchronizing sequence to a state $q \in Q$ if $\delta(q', x) = q$ for all $q' \in Q$. A PFSA is **synchronizable** if it has at least one synchronizing sequence. Given a sample path generated by a PFSA, we say the PFSA is **synchronized** if a synchronizing sequence transpires in the sample path.

Definition 13 (Equivalence and Irreducibility). Two PFSA G and H are **equivalent** if they generate the same stochastic process. A PFSA G is said to be **irreducible**, if there is not another PFSA with smaller state set that is equivalent to G .

Definition 14. Consider a PFSA G over state set Q . For a give $\varepsilon > 0$, we say a sequence x is a ε -synchronizing sequence to a state $q \in Q$ if

$$\|\varphi_G(x) - \mathbf{e}_q\|_\infty \leq \varepsilon. \quad (21)$$

Algorithm 1: GenESeSS

Data: A sequence x over alphabet Σ , $0 < \varepsilon < 1$

Result: State set Q , transition map δ , and transition probability $\tilde{\pi}$

```

/* Step One: Approximate  $\varepsilon$ -synchronizing sequence */
```

- 1 Let $L = \lceil \log_{|\Sigma|} 1/\varepsilon \rceil$;
- 2 Calculate the **derivative heap** D_ε^x equaling $\{\hat{\phi}_y^x : y \text{ is a sub-sequence of } x \text{ with } |y| \leq L\}$;
- 3 Let \mathcal{C} be the convex hull of D_ε^x ;
- 4 Select x_0 with $\hat{\phi}_{x_0}^x$ being a vertex of \mathcal{C} and has the highest frequency in x ;

```

/* Step Two: Identify transition structure */
```

- 5 Initialize $Q = \{q_0\}$;
- 6 Associate to q_0 the **sequence identifier** $x_{q_0}^{\text{id}} = x_0$ and the probability vector $d_{q_0} = \hat{\phi}_{x_0}^x$;
- 7 Let \tilde{Q} be the set of states that are just added and initialize it to be Q ;
- 8 **while** $\tilde{Q} \neq \emptyset$ **do**
- 9 Let $Q_{\text{new}} = \emptyset$ be the set of new states;
- 10 **for** $(q, \sigma) \in \tilde{Q} \times \Sigma$ **do**
- 11 Let $x = x_q^{\text{id}}$ and $d = \hat{\phi}_{x\sigma}^x$;
- 12 **if** $\|d - d_{q'}\|_\infty < \varepsilon$ **for some** $q' \in Q$ **then**
- 13 Let $\delta(q, \sigma) = q'$;
- 14 **else**
- 15 Let $Q_{\text{new}} = Q_{\text{new}} \cup \{q_{\text{new}}\}$ and $Q = Q \cup \{q_{\text{new}}\}$;
- 16 Associate to q_{new} the sequence identifier $x_{q_{\text{new}}}^{\text{id}} = x\sigma$ and the probability vector $d_{q_{\text{new}}} = d$;
- 17 Let $\delta(q, \sigma) = q_{\text{new}}$;
- 18 Let $\tilde{Q} = Q_{\text{new}}$;
- 19 Take a strongly connected subgraph of the labeled directed graph defined by Q and δ , and denote the vertex set of the subgraph again by Q ;

```

/* Step Three: Identify transition probability */
```

- 20 Initialize counter $N[q, \sigma]$ for each pair $(q, \sigma) \in Q \times \Sigma$;
- 21 Choose a random starting state $q \in Q$;
- 22 **for** $\sigma \in x$ **do**
- 23 Let $N[q, \sigma] = N[q, \sigma] + 1$;
- 24 Let $q = \delta(q, \sigma)$;
- 25 Let $\tilde{\pi}(q) = \llbracket (N[q, \sigma])_{\sigma \in \Sigma} \rrbracket$;
- 26 **return** $Q, \delta, \tilde{\pi}$;

While there exists PFSA that is not synchronizable, we can show that an irreducible PFSA always has an ε -synchronizing sequence for some state q for arbitrarily small $\varepsilon > 0$. Moreover, we can show that as length increases, sequences produced by PFSA become uniformly ε -synchronizing. These two are the underpinning properties for the inference algorithm of PFSA (See Alg. 1), because they imply that ϕ_x can be used to approximate $\tilde{\pi}(q)$ if x are properly prefixed and long enough.

Definition 15 (Joint ε -Synchronizing Sequence). *Let G and H be two PFSA over state sets Q_G and Q_H , respectively. For a fixed ε , a sequence x is said to be **jointly ε -synchronizing** to $(q, r) \in Q_G \times Q_H$ if x is ε -synchronizing to q and to r simultaneously. We define*

$$\Sigma_{\varepsilon, (q, r)}^d \triangleq \{x \in \Sigma^d : x \text{ jointly } \varepsilon\text{-synchronizing to } (q, r)\} \quad (22)$$

Definition 16 (Joint Pair of States). *Let G and H be two PFSA over state sets Q_G and Q_H , respectively. Define*

$$p_G(q, r) \triangleq \lim_{d \rightarrow \infty} p_G(\Sigma_{\varepsilon, (q, r)}^d) \quad (23)$$

*A pair of states $(q, r) \in Q_G \times Q_H$ is called a **G -joint pair** of states if $p_G(q, r) > 0$. We also define*

$$Q_c \triangleq \{(q, r) \in Q_G \times Q_H : (q, r) \text{ is a } G\text{-joint pair}\} \quad (24)$$

The inference algorithm for PFSA is called **GenESeSS** for Generator Extraction Using Self-similar Semantics. With an input sequence x and a hyperparameter ε , **GenESeSS** outputs a PFSA in the following three steps: 1) approximate an almost synchronizing sequence; 2) identify the transition structure of the PFSA; 3) calculate the transition probabilities of the PFSA. See Alg. 1 [53] for details.

Algorithm 2: Log-likelihood

Data: A PFSA $G = (\Sigma, Q, \delta, \tilde{\pi})$ and a sequence x over alphabet Σ
Result: Log-likelihood $L(x, G)$ of G generating x

- 1 Calculate the state transition matrix Π and observation $\tilde{\Pi}$;
- 2 Calculate the stationary distribution over states φ_G of G from Π ;
- 3 Calculate the stationary distribution of alphabet $\phi_\lambda^T = \varphi_G^T \tilde{\Pi}$;
- 4 Initialize p by φ_G and q by ϕ_λ ;
- 5 Let $L = 0$;
- 6 **for** i from 1 to $|x|$ **do**
- 7 Let σ be the i -th entry of x ;
- 8 Let $L = L - \log q|_\sigma$;
- 9 Let $p^T = [\![p^T \Gamma_\sigma]\!]$ where Γ_σ is defined in 9;
- 10 Let $q^T = p^T \tilde{\Pi}$;
- 11 **return** $L/|x|$;

VI. THEORETICAL DEVELOPMENT OF SEQUENCE LIKELIHOOD DEFECT

Definition 17 (Entropy Rate and KL Divergence). *By entropy rate of a PFSA, we mean the entropy rate of the stochastic process generated by the PFSA [32]. Similarly, by KL divergence of two PFSA, we mean the KL divergence between the two processes generated by them [66]. More precisely, we have*

$$\mathcal{H}(G) = - \lim_{d \rightarrow \infty} \frac{1}{d} \sum_{x \in \Sigma^d} p(x) \log p(x) \quad (25)$$

and the KL divergence

$$\mathcal{D}_{KL}(G \| H) = \lim_{d \rightarrow \infty} \frac{1}{d} \sum_{x \in \Sigma^d} p_G(x) \log \frac{p_G(x)}{p_H(x)} \quad (26)$$

whenever the limits exist.

Theorem 1 (Closed-form Formula for Entropy Rate and KL Divergence). *The entropy rate of a PFSA $G = (\Sigma, Q, \delta, \tilde{\pi})$ is given by*

$$\mathcal{H}(G) = \sum_{q \in Q} \varphi_G(q) \cdot h(\tilde{\pi}(q)) \quad (27)$$

where $h(v)$ is the based-2 entropy of the probability vector v .

Consider two PFSA $G = (Q_G, \Sigma, \delta_G, \tilde{\pi}_G)$ and $H = (Q_H, \Sigma, \delta_H, \tilde{\pi}_H)$ with μ_G being absolutely continuous with respect to μ_H . Let Q_c be the set of G -joint pairs of states, we have

$$\mathcal{D}_{KL}(G \| H) = \sum_{(q,r) \in Q_c} p_G(q, r) \mathcal{D}_{KL}(\tilde{\pi}_G(q) \| \tilde{\pi}_H(r)) \quad (28)$$

Definition 18 (Log-likelihood). *Let $x \in \Sigma^d$, the log-likelihood [32] of a PFSA G generating x is given by*

$$L(x, G) = -\frac{1}{d} \log p_G(x) \quad (29)$$

The calculation of log-likelihood is detailed in Alg. 2.

Theorem 2 (Convergence of log-likelihood). *Let G and H be two reduced PFSA, and let $x \in \Sigma^d$ be a sequence generated by G . Then we have*

$$L(x, H) \rightarrow \mathcal{H}(G) + \mathcal{D}_{KL}(G \| H) \quad (30)$$

in probability as $d \rightarrow \infty$.

Proof. We first notice that

$$\sum_{x \in \Sigma^d} p_G(x) \log \frac{p_G(x)}{p_H(x)} = \sum_{x \in \Sigma^{d-1}} \sum_{\sigma \in \Sigma} p_G(x) \varphi_G(x) \left. \tilde{\Pi}_G \right|_\sigma \log \frac{\varphi_G(x) \varphi_G(x) \left. \tilde{\Pi}_G \right|_\sigma}{p_H(x) \varphi_H(x) \left. \tilde{\Pi}_H \right|_\sigma} \quad (31)$$

$$= \sum_{x \in \Sigma^{d-1}} p_G(x) \log \frac{p_G(x)}{p_H(x)} + \underbrace{\sum_{x \in \Sigma^{d-1}} p_G(x) \sum_{\sigma \in \Sigma} \varphi_G(x) \left. \tilde{\Pi}_G \right|_\sigma \log \frac{\varphi_G(x) \left. \tilde{\Pi}_G \right|_\sigma}{\varphi_H(x) \left. \tilde{\Pi}_H \right|_\sigma}}_{D_d} \quad (32)$$

By induction, we have $\mathcal{D}_{KL}(G \parallel H) = \lim_{d \rightarrow \infty} \frac{1}{d} \sum_{i=1}^d D_i$, and hence by Cesàro summation theorem [67], we have $\mathcal{D}_{KL}(G \parallel H) = \lim_{d \rightarrow \infty} D_d$. Let $x = \sigma_1 \sigma_2 \dots \sigma_n$ be a sequence generated by G . Let $x^{[i-1]}$ is the truncation of x at the $(i-1)$ -th symbols, we have

$$-\frac{1}{n} \sum_{i=1}^n \log \varrho_H(x^{[i-1]}) \tilde{\Pi}_H|_{\sigma_i} = \underbrace{\frac{1}{n} \sum_{i=1}^n \log \frac{\varrho_G(x^{[i-1]}) \tilde{\Pi}_G|_{\sigma_i}}{\varrho_H(x^{[i-1]}) \tilde{\Pi}_H|_{\sigma_i}}}_{A_{x,n}} - \underbrace{\frac{1}{n} \sum_{i=1}^n \log \varrho_G(x^{[i-1]}) \tilde{\Pi}_G|_{\sigma_i}}_{B_{x,n}} \quad (33)$$

Since the stochastic process G generates is ergodic, we have

$$\lim_{n \rightarrow \infty} A_{x,n} = \lim_{d \rightarrow \infty} D_d = \mathcal{D}_{KL}(G \parallel H) \quad (34)$$

and $\lim_{n \rightarrow \infty} B_{x,n} = \mathcal{H}(G)$. \square

VII. PIPELINE OPTIMIZATION: HYPER-TRAINING, TRAINING, & VALIDATION

Our pipeline comprises a network of individually trained light gradient boosting machine (LGBM) [32] classifiers that focus on complementary aspects of the problem, and operate on different categories of input features as described next. Importantly, some of these features need to be generated non-trivially from the raw data, and these *feature generators* have parameters that need to be trained as well (or comprise models that need to be inferred). We call this inference of the feature-generators as **hyper-training**. Importantly, this is different from the more common notion of hyper-parameters. Hyper-parameters are one or more variables whose scalar values are commonly tuned by grid-search or via some meta-heuristics to optimize classifiers, whereas hyper-training produces generators of features, not simply a set of numbers.

Hyper-training & Training

Trinary Quantization of Medical Histories: The medical histories are mapped into trinary disease-phenotype-specific data-streams to enable generation of some of the features described below, as outlined in Section -C (Step 1).

Feature Categories: The features used in the pipeline maybe categorized as follows:

PFSA scores: The PFSA scores are computed on the basis of the inferred PFSA models as described in the previous sections. The generation of the PFSA models from the trinary data-streams is the first hyper-training step. These scores consist of the negative and positive log-likelihood of a phenotype-specific quantized medical history being generated by the PFSA models for the positive cohort and the control cohort of sex-stratified patients, and the corresponding sequence likelihood defects (See SI-Section -D). Recall that PFSA are specialized HMMs, and these measures encode the dynamics of the underlying processes, and are sensitive to the ordering, and frequency of the codes at the resolution of the disease phenotypes. Also, recall that diseases phenotypes are broad categories of diagnostic codes, and that we generate PFSA models for each category, and separately for the sexes and the positive cohort and the control cohorts).

Prevalence scores (p-scores): The p-scores focus on individual diagnostic codes, and we create a dictionary of the ratio of relative prevalence of each code (relative to the set of all codes present) in the positive category (for each sex) to the control category. This is the second hyper-training step. In the later steps of the pipeline, we use dictionary look ups to map codes to their p-scores, and also their aggregate measures such as mean, median, and variance to train a downstream LGBM.

Rare scores: These scores consist of a subset of p-scores which correspond to codes with particularly high and low relative prevalences ($p\text{-score} > 2$ or $< .5$). Thus, this feature category depends on the p-score dictionary generated in the second hyper-training step.

Sequence scores: Sequence scores are relatively straight-forward statsitical measures such as mean, median, variance, time since last occurrence etc.. on the trinary phenotype-specific sex-stratified histories. No hyper-training is required for the generation of the sequence features.

Thus we require three splits of the training dataset. The first split is used to carry out hyper-training of the PFSA models and the p-score dictionary. The second split is used to train the score-category specific LGBMs, one for each feature category. And the third split is used to train the final LGBM that takes inputs from the outputs of the four LGBMs in the previous layer.

Validation

In validation, or actual prediction of patient fate, we use the trinary mapping, generate the features using the PFSA models and the p-score dictionary, and calculate the raw-risk via the trained LGBM network. The relative score is then obtained by a choice of the operating point reflecting the specificity/sensitivity trade-off discussed before.