

# Coronary Artery Disease Phenotype Detection in an Academic Hospital System Setting

Amy Joseph<sup>1</sup> Charles Mullett<sup>1,2</sup> Christa Lilly<sup>3</sup> Matthew Armistead<sup>2</sup> Harold J. Cox<sup>2</sup>  
 Michael Denney<sup>2</sup> Misha Varma<sup>1</sup> David Rich<sup>4</sup> Donald A. Adjeroh<sup>5</sup> Gianfranco Doretto<sup>5</sup>  
 William Neal<sup>1</sup> Lee A. Pyles<sup>1</sup>

<sup>1</sup>Department of Pediatrics, School of Medicine, West Virginia University, Morgantown, West Virginia, United States

<sup>2</sup>West Virginia Clinical and Translational Science Institute, West Virginia University, Morgantown, West Virginia, United States

<sup>3</sup>Department of Biostatistics, School of Public Health, West Virginia University, Morgantown, West Virginia, United States

<sup>4</sup>West Virginia University Hospital System, Morgantown, West Virginia, United States

<sup>5</sup>Lane Department of Computer Science and Electrical Engineering, Benjamin M. Statler College of Engineering and Mineral Resources, West Virginia University, Morgantown, West Virginia, United States

**Address for correspondence** Lee A. Pyles, MD, MS, Department of Pediatrics, West Virginia University School of Medicine, 1 Medical Center Dr., Box 9214, Morgantown, WV 26506, United States (e-mail: lpyles@hsc.wvu.edu).

Appl Clin Inform 2021;12:10–16.

## Abstract

### Keywords

- ▶ clinical phenotype
- ▶ clinical registry
- ▶ coronary artery disease phenotype
- ▶ accuracy
- ▶ problem list
- ▶ validation of algorithm
- ▶ knowledge management
- ▶ data validation and verification

**Background** The United States, and especially West Virginia, have a tremendous burden of coronary artery disease (CAD). Undiagnosed familial hypercholesterolemia (FH) is an important factor for CAD in the U.S. Identification of a CAD phenotype is an initial step to find families with FH.

**Objective** We hypothesized that a CAD phenotype detection algorithm that uses discrete data elements from electronic health records (EHRs) can be validated from EHR information housed in a data repository.

**Methods** We developed an algorithm to detect a CAD phenotype which searched through discrete data elements, such as diagnosis, problem lists, medical history, billing, and procedure (International Classification of Diseases [ICD]-9/10 and Current Procedural Terminology [CPT]) codes. The algorithm was applied to two cohorts of 500 patients, each with varying characteristics. The second (younger) cohort consisted of parents from a school child screening program. We then determined which patients had CAD by systematic, blinded review of EHRs. Following this, we revised the algorithm by refining the acceptable diagnoses and procedures. We ran the second algorithm on the same cohorts and determined the accuracy of the modification.

**Results** CAD phenotype Algorithm I was 89.6% accurate, 94.6% sensitive, and 85.6% specific for group 1. After revising the algorithm (denoted CAD Algorithm II) and applying it to the same groups 1 and 2, sensitivity 98.2%, specificity 87.8%, and accuracy 92.4; accuracy 93% for group 2. Group 1 F1 score was 92.4%. Specific ICD-10 and CPT codes such as “coronary angiography through a vein graft” were more useful than generic terms.

received  
 May 26, 2020  
 accepted after revision  
 October 9, 2020

© 2021. Thieme. All rights reserved.  
 Georg Thieme Verlag KG,  
 Rüdigerstraße 14,  
 70469 Stuttgart, Germany

DOI <https://doi.org/10.1055/s-0040-1721012>.  
 ISSN 1869-0327.

**Conclusion** We have created an algorithm, CAD Algorithm II, that detects CAD on a large scale with high accuracy and sensitivity (recall). It has proven useful among varied patient populations. Use of this algorithm can extend to monitor a registry of patients in an EHR and/or to identify a group such as those with likely FH.

## Background and Significance

Electronic health records (EHRs) organize and collate medical information to a degree realized only in intensive research efforts a generation ago. Problem lists were made popular by Weed to organize medical thinking and evaluation.<sup>1</sup> Procedural terminology was standardized to facilitate medical billing. We now use these tools to enable case-finding for research, quality improvement, and preventive medicine. Definition of a clinical phenotype or grouping of patients from a collection of EHR records has been important for these operations.<sup>2</sup> Definition of a clinical population phenotype has typically required augmentation of the systematic cataloging of diagnoses and procedures with examination of clinical progress, admission, discharge, and procedure notes, often with the aid of artificial intelligence (AI) tools such as machine learning and natural language processing (NLP).<sup>3</sup> Clinical phenotypes have been defined for heart failure (HF), diabetes, hypertension, and other disease groups.<sup>4</sup> In this report, we describe a process to use the diagnosis and procedures encoded in random samples of EHRs to define a clinical phenotype for coronary artery disease (CAD). We hypothesized that an accurate CAD phenotype can be developed and validated from discrete data elements available in the West Virginia University Health System EHR.

Cardiovascular disease is the leading cause of mortality in the United States, particularly West Virginia, with the majority of deaths due to CAD.<sup>5</sup> In 2018, 18.2 million American adults (6.7%) were diagnosed with CAD.<sup>6</sup> We will attempt to answer the question of whether a CAD phenotype can be defined from diagnoses and procedures and we will utilize items from both clinical and administrative sources to define the phenotype. Pointedly, we will not employ advanced analytical techniques of AI to comb through information in clinical notes. The CAD phenotype is defined as presence of significant CAD as demonstrated by heart catheterization especially with stent or angioplasty, coronary artery bypass graft (CABG) placement or evaluation, and myocardial infarction. This project highlights the optimal operation of a mature EHR system from which researchers can synthesize diagnosis and procedure codes to define the CAD phenotype.<sup>7</sup>

The health informatics focused phenotype has been defined for a variety of diseases most notably HF. Aragam et al linked a clinical phenotype of nonischemic HF to genotype from the U.K. Biobank.<sup>8</sup> In this process they negated the group with ischemic HF as defined by diagnoses plus patient self-report. Kashyap et al noted sensitivity of only 47.5% and specificity of 96.7% for inpatient acute HF using International

Classification of Diseases (ICD)-9 codes for records from 2006 to 2014, highlighting the importance of improved usefulness in the current era for use of ICD plus Current Procedural Terminology (CPT) to create a phenotype.<sup>9</sup>

Ontologies such as SNOMED, ICD-9 and -10, and CPT codes have allowed use of big data techniques to initiate registries of patients of interest with various diseases such as diabetes, hypertension, and obesity.<sup>4,10-12</sup> These registries often involve development of a *clinical phenotype* that defines the patient population.<sup>2</sup> This phenotype derives from known clinical attributes, dependencies, or clinical course of progression or resolution of a disease process. This report describes validation of an algorithm to detect a phenotype derived from discrete data elements and their values and thus represents the maturation of functionality of an EHR.<sup>7</sup> Discrete data element items such as problem lists and procedure lists are analyzed rather than NLP analysis of the corpus of a clinical note.<sup>4,12,13</sup> The algorithm should be able to identify a phenotype from a varied population group, including a wide range of age groups and different disease cohorts.<sup>2</sup> By forming such an algorithm, large amounts of data can be extracted from EHRs and analyzed. In addition, the phenotype can provide a population for clinical trials.

## CAD Phenotype as a Building Block to Study Familial Hypercholesterolemia

Our interest in the CAD phenotype stems from a desire to identify families with familial hypercholesterolemia (FH) that is thought to be significantly underdiagnosed in the United States and Europe.<sup>14-17</sup> The FH diagnosis requires at a minimum identification of elevated low-density lipoprotein cholesterol (LDL-C) (or total or non-high-density lipoprotein cholesterol) plus evidence of early CAD.<sup>18</sup> Empiric diagnosis of FH as we propose requires a functional definition of CAD clinical phenotype to find family members with FH with either elevated or nondiagnostic LDL levels.<sup>19</sup> Thus, the analysis of the empiric relationship of elevated LDL and CAD that suggests FH requires specification of a population-based CAD phenotype. To detect FH, West Virginia pioneered the Coronary Artery Risk Detection in Appalachian Communities (CARDIAC) project in 1998, a multidimensional screening and research project that obtains fasting lipid profiles from over 60,000 fifth graders around the state.<sup>20-22</sup> The earlier a patient with FH can be detected and their lipid levels controlled, the greater the risk reduction for future coronary events. The recent Framingham report highlights the importance of optimum management of LDL in the 20- to 40-year-old age group to prevent CAD in the fifth or sixth decade.<sup>23</sup> Report of a Dutch series of FH families with childhood LDL control solidifies the thinking

that the atherosclerotic pathophysiology is dependent upon the LDL time integral that is cumulative from childhood, likely from birth.<sup>24,25</sup>

## Methods

A proposal was made to create an algorithm that would detect CAD using only data from the EHR discrete data element analysis. The records for each group of subjects were drawn from the WV Clinical Translational Science Institute Integrated Data Repository (WV CTSI IDR) that contained 1.6 million patient records as of September 2017.<sup>26</sup> This study was performed in compliance with the World Medical Association Declaration of Helsinki on Ethical Principles for Medical Research Involving Human Subjects and was reviewed by the West Virginia University Institutional Review Board. An American College of Cardiology nomenclature guideline paper was reviewed by three authors who then conducted an exercise to establish internal validity.<sup>27</sup> The guideline paper stressed unambiguous use of medical vocabularies to define clinical endpoints with a goal of improvement of clinical trials. We started by randomly selecting a set of 25 patients that were a mixture of CAD positive and CAD negative.<sup>27</sup> Each of our selection processes included use of a random number generation kernel. One author (M.V.) established the CAD status for each patient and this was reviewed in a blinded fashion by the first and senior authors. All three authors were found to be in complete agreement and further manual chart reviews proceeded.

Initially, we defined the parameters that would indicate CAD in EHRs. We primarily looked at diagnosis (billing) codes and procedure codes. Billing diagnosis codes are made through ICD-9 and -10 codes. The ICD-10 codes were cross-referenced back to ICD-9 except for Z95.1 in ICD-10; Presence of aortocoronary bypass graft (see [►Supplementary Appendix A](#) for ICD and CPT codes). The diagnoses and procedures could be found in the EHR problem list, visit diagnosis list, medical history, or billing information (provider or health system). Originally (see [►Supplementary Appendix A](#)), we included 410.x (acute myocardial infarction), 411.x (ischemic heart disease and coronary syndromes), 412.x (old myocardial infarction), 413.x (angina pectoris), and 414.x (coronary atherosclerosis). Procedure codes were assigned through CPT codes, which included codes for coronary angiography and catheterization, and CABG operative procedure. An algorithm (CAD Algorithm I) was constructed that included presence or absence of these discrete elements; no additive, exclusionary, or probabilistic operation was required. We tested it on a cohort of 500 patients drawn randomly from WV CTSI IDR that consisted of 250 patients assigned to have CAD and 250 without CAD from the criteria of Algorithm I, then

manually abstracted the charts and validated the diagnosis. Next, we drew another cohort of 500 patients that included patients that were known to be parents of children that had participated in the CARDIAC project, without regard for CAD status, including 191 with CAD determined by manual chart inspection including clinical notes and referral notes plus 309 without CAD. Following this, we evaluated variation between the manual data results and algorithm results. Standard demographics and descriptive analyses for the subjects were obtained. Relevance (sensitivity, specificity, positive predictive value, negative predictive value, and F-score) was calculated for the different groups and operations.<sup>28</sup> We decided to remove the 413.x ICD-9 codes, which primarily coded for angina, and added the 414.x codes (coronary atherosclerosis). We found that patients who presented to the emergency department with chest pain were frequently coded as angina pectoris regardless of the ultimately determined etiology and that term lacked any discriminatory value. This resulted in 40 false positive CAD patients in the first classification attempt, denoted “CAD Algorithm I” in [►Table 3](#). Patients who had angina due to CAD were identified from other codes. We also included diagnosis codes and past medical and surgical history in the algorithm to better identify CAD. After making the revisions, we ran the improved algorithm (CAD Algorithm II) for the same two sets of 500 patients that we had previously evaluated ([►Fig. 1](#)).

Certain characteristics were also recorded for both sets of patients, for example, sex, age, body mass index (BMI), and maximum LDL-C levels found within the EHR. We wanted to determine if the algorithm would be accurate among two varied groups of patients, especially the younger group which consisted of parents of children who participated in the WV school-based cross-sectional CARDIAC program.

## Results

[►Table 1](#) shows demographic and laboratory features of the two groups of test subject. They were different in age, BMI, and LDL-C. The second group represented a pool of persons that likely wanted to know cardiometabolic risk factors for their children. [►Table 2](#) shows the different features of the ICD and CPT codes of CAD Algorithms I and II. Generic coronary angiography CPT and angina pectoris ICD codes were deleted to create CAD Algorithm II.

The first cohort consisted of 500 patients (group 1) that the algorithm detected as 250 patients with CAD and 250 patients without CAD. Blinded manual review of the EHRs for these 500 patients revealed 222 patients with CAD and 278 without ([►Table 3](#)). This was an 89.6% accurate correlation between the algorithm and manual validation. This resulted



**Fig. 1** Process of verifying algorithms. A separate “learning” group was not warranted because no artificial intelligence-based knowledge discovery was employed.

**Table 1** Patient characteristics from manual chart review

	Random cohort (Group 1) <i>n</i> = 500	Cohort with "cardiac parents" (Group 2) <i>n</i> = 500	<i>p</i> -Value
CAD positive	222	191	0.0465 <sup>a</sup>
CAD negative	278	309	
Female (%)	43.9%	45.2%	NS
Male (%)	55.9%	54.8%	
Age (minimum)	11	21	
Age (maximum)	96	86	
Age (average)	58.3	48.8	< 0.001 <sup>b</sup>
BMI (average)	30	31.5	0.005 <sup>b</sup>
LDL-C (average)	106.2	123.5	< 0.001 <sup>b</sup>

Abbreviations: BMI, body mass index; CAD, coronary artery disease; LDL-C, low-density lipoprotein cholesterol.

<sup>a</sup>Chi-square *p*-value.

<sup>b</sup>Two-tailed *t*-test *p*-value.

in a sensitivity of 94.6% and specificity of 85.6%. With the second set of 500 patients (group 2), the algorithm found 238 patients with CAD and 262 without CAD (→Table 3). On validating the data, 191 were CAD positive patients and 309 patients did not have CAD. This was 89% accurate (the combination of positive and negative predictive value), had a sensitivity of 97.9% and specificity of 83.5%. Then, after the algorithm was revised, creating CAD Algorithm II, the first cohort of 500 patients resulted in 252 patients that were diagnosed with CAD and 248 patients that did not have CAD (→Table 4). This gave a 92.4% accuracy, sensitivity of 98.2%, and specificity of 87.8%. For the second set of 500, the updated algorithm detected 222 patients with CAD and 278 without (→Table 4). This was 93% accurate, had a

sensitivity of 99%, and specificity of 89.3% (→Table 4). The second, younger group (group 2) had less CAD, as should be expected although the average LDL level was higher ( $p < 0.001$ ; →Table 1). Receiver operating characteristic curves are not determined in this methodology since dichotomous rather than probabilistic relationships are used to define the CAD status.

## Discussion

The development of EHRs has resulted in a large catalog of data to be mined, and efficient methods to extract information from this data need to be established. In our study, we have developed an algorithm that can detect CAD in a wide variety of patients. Our group 1 was 10 years older than group 2 who had children in CARDIAC in past years (up to 17 years previous). Considering our two different groups had different prevalences of CAD, the accuracy, sensitivity, and specificity were similar between both groups of patients, suggesting a robust algorithm. The most significant error cells consist of 34 false positive subjects in group 1 and 33 false positive subjects in group 2. These false positive determinations affect positive predictive value (precision) and accuracy. Performance measures were slightly stronger in group 2 which is the group of clinical interest that drives the investigation.

An advantage of creating an algorithm without using AI is that the parameters that identify CAD are very clearly defined, and hence, unambiguous. Hence, there is a very high negative predictive value and patients without CAD are excluded. A disadvantage of not using NLP or other AI is that sometimes, evidence of CAD in a patient EHR may be written within a chart but may not have been clearly encoded or identified in the medical history. At times, a distant history of a CABG or coronary angiography at catheterization was generated prior to the patient record being added to EHRs. Another issue arose if the patient was

**Table 2** Features of CAD algorithms

	CAD Algorithm I	CAD Algorithm II
ICD-9/-10 codes		
410.x acute myocardial infarction	x	x
411.x ischemic heart disease and coronary syndromes	x	x
412.x old myocardial infarction	x	x
413.x angina pectoris	x	Deleted
414.x coronary atherosclerosis	x	x
CPT codes		
Coronary artery bypass graft procedure	x	x
Coronary angiography	x	Deleted CPT 93454, 93456, 93458, 93460,
Coronary angioplasty	x	x
Coronary stenting	x	x
Coronary angio through existing bypass graft		x

Abbreviations: CAD, coronary artery disease; CPT, Current Procedural Terminology; ICD, International Classification of Diseases.

Note: →Supplementary Appendix A provides detail of specific ICD-9 or ICD-10 codes and CPT codes.

**Table 3** Groups 1 and 2: CAD Algorithm I

Algorithm I results	Group 1 CAD true positive (manually validated)	CAD true negative (manually validated)	N	Group 2 CAD true positive (manually validated)	CAD true negative (manually validated)	N
CAD positive	210	40	250	187	51	238
CAD negative	12	238	250	4	258	262
Total	222	278	500	191	309	500
Positive predictive value/precision (%)	84			78.6		
Negative predictive value		95.2			98.5	
Sensitivity (%) / recall	94.6			97.9		
Specificity (%)		85.6%			83.5	
Accuracy (%) <sup>a</sup>	89.6			89.0		
F1 score	88.9			87.2		

Abbreviation: CAD, coronary artery disease.

Note: F1 score =  $(2 * \text{precision} * \text{recall}) / (\text{precision} + \text{recall})$ .

<sup>a</sup>Accuracy defined as true positive plus true negative divided by group totals.

being treated for CAD at a different hospital or with another provider. In that case, unless the CAD diagnosis was entered in the chart, then the patient would not be detected by the algorithm. These are known issues that have been considered in this evaluation and the noncommunicative outside patient arriving without records could also be missed with NLP or with our Algorithms I or II. Inclusion of self-report of CAD could help remedy this but the clinical experience suggests the inclusion of myriad symptoms as “heart attack” by patients.

The algorithms use parameters that include ICD-9 and CPT codes plus mapping to ICD-10 codes. Since the algorithms include constraints that are standard across the nation, it can be applied in other hospital systems. Therefore, this can help detect patients with CAD on a large scale among many hospitals, although it must be cautioned that the

coding can be dependent upon local custom, such as our observation that any adult chest pain in our hospital system was invariably coded “angina pectoris.”

While there are certain limitations with using our CAD Algorithm II, it is highly sensitive to detect CAD patients. Limitations include the problem of local custom in implementation of diagnostic and procedural coding and lack of standardized subject screening in clinical practice. Also, the algorithm detects EHR documentation of CAD rather than its actual presence or absence. Nonetheless, this work represents a validation of the promise of the EHR to revolutionize health care and an initial step toward use of data registries such as WV CTSI IDR to create a learning health system in an institution such as the West Virginia University Health System. The project helps to actualize the translational value of the WV CTSI Integrated Data Repository.

**Table 4** Groups 1 and 2: CAD Algorithm II

Algorithm II results	Group 1 CAD positive (validated)	Group 2 CAD negative (validated)	N	Group 1 CAD positive (validated)	Group 2 CAD negative (validated)	N
CAD positive	218	34	252	189	33	222
CAD negative	4	244	248	2	276	278
Total	222	278	500	191	309	500
Positive predictive value/precision (%)	86.5			85.1		
Negative predictive value (%)		98.4			99.3	
Sensitivity/Recall (%)	98.2			98.9		
Specificity (%)		87.8			89.3	
Accuracy (%) <sup>a</sup>	92.4			93		
F1 score	0.92			0.914		

Abbreviation: CAD, coronary artery disease.

Note: F1 score =  $(2 * \text{precision} * \text{recall}) / (\text{precision} + \text{recall})$ .

<sup>a</sup>Accuracy defined as true positive plus true negative divided by group totals.

## Clinical Relevance Statement

A capability to use diagnoses, procedures, and billing forms from providers or institutions and data elements of medical history to describe a group of patients with a diagnosis of coronary artery disease is demonstrated. This methodology can aid case-finding for management, quality improvement, and clinical research. The methodology demonstrates the maturity of the electronic health record of the West Virginia University Health System that contains sufficient discrete data elements to successfully identify CAD subjects.

## Multiple Choice Questions

- The accuracy of the algorithm attributes to which parameters?
  - 1 and 2
  - 2 and 4
  - None of the above
  - All of the above

**Correct Answer:** The correct answer is d, all of above: (TP plus TN) / (TP + FP + TN + FN).

- Heart cath for chest pain should always prompt a diagnosis of CAD.
  - True
  - False

**Correct Answer:** The correct answer is option b.

- CAD Algorithm II includes which group of patients:
  - Diagnostic heart cath with normal coronary angiography.
  - Open heart surgery for valve replacement.
  - Coronary angiography through existing vein graft.
  - TAVR: transcatheter aortic valve replacement.

**Correct Answer:** The correct answer is option c, coronary angiography through existing vein graft. Options a, b, and d do not diagnose coronary artery disease.

- A clinical phenotype represents: (choose 1)
  - The items in SNOMED that pertain to the clinical problem.
  - The clinical diagnoses that can be determined from GWAS for a subject.
  - A grouping of persons that exhibit the clinical symptom (s) or behavior of interest.
  - A group of persons that exhibit a genotype that reflects the clinical problem.

**Correct Answer:** The correct answer is option c. Option a is only an example of a classification system but it does not define a phenotype. Options b and d pertain to genotype, not phenotype.

## Protection of Human and Animal Subjects

This study was performed in compliance with the World Medical Association Declaration of Helsinki on Ethical Principles for Medical Research Involving Human Subjects and was reviewed by the West Virginia University Institutional Review Board.

## Funding

The project described was supported by the National Institute of General Medical Sciences, 2U54GM104942-02 and in part by funds from the National Science Foundation (NSF: # 1920920). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

## Conflict of Interest

None declared.

## References

- Wright A, Sittig DF, McGowan J, Ash JS, Weed LL. Bringing science to medicine: an interview with Larry Weed, inventor of the problem-oriented medical record. *J Am Med Inform Assoc* 2014;21(06):964-968
- Liao KP, Ananthakrishnan AN, Kumar V, et al. Methods to develop an electronic medical record phenotype algorithm to compare the risk of coronary artery disease across 3 chronic disease cohorts. *PLoS One* 2015;10(08):e0136651
- Popejoy LL, Khalilia MA, Popescu M, et al. Quantifying care coordination using natural language processing and domain-specific ontology. *J Am Med Inform Assoc* 2015;22(e1):e93-e103
- Teixeira PL, Wei WQ, Cronin RM, et al. Evaluating electronic health record data sources and algorithmic approaches to identify hypertensive individuals. *J Am Med Inform Assoc* 2017;24(01):162-171
- The Burden of Cardiovascular Disease in West Virginia. Published 2011. Available at: [http://www.wvdhhr.org/bph/hsc/pubs/other/burdenofcvd2010/cvh\\_burden\\_2010.pdf](http://www.wvdhhr.org/bph/hsc/pubs/other/burdenofcvd2010/cvh_burden_2010.pdf). Accessed January 29, 2017
- Benjamin EJ, Muntner P, Alonso A, et al; American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2019 update: a report from the American Heart Association. *Circulation* 2019;139(10):e56-e528
- Kennell TI Jr, Willig JH, Cimino JJ. Clinical informatics researcher's desiderata for the data content of the next generation electronic health record. *Appl Clin Inform* 2017;8(04):1159-1172
- Aragam KG, Chaffin M, Levinson RT, et al; GRADE Investigators. Phenotypic refinement of heart failure in a national biobank facilitates genetic discovery. *Circulation* 2018
- Kashyap R, Sarvottam K, Wilson GA, Jentzer JC, Seisa MO, Kashani KB. Derivation and validation of a computable phenotype for acute decompensated heart failure in hospitalized patients. *BMC Med Inform Decis Mak* 2020;20(01):85
- Rodrigues J, Schulz S, Rector A, et al. ICD-11 and SNOMED CT Common Ontology: Circulatory System. Copenhagen, Denmark: European Federation for Medical Informatics and IOS Press; 2014
- Lingren T, Thaker V, Brady C, et al. Developing an algorithm to detect early childhood obesity in two tertiary pediatric medical centers. *Appl Clin Inform* 2016;7(03):693-706
- Roldán-García MD, García-Godoy MJ, Aldana-Montes JF. Dione: an OWL representation of ICD-10-CM for classifying patients' diseases. *J Biomed Semantics* 2016;7(01):62

- 13 Wei WQ, Teixeira PL, Mo H, Cronin RM, Warner JL, Denny JC. Combining billing codes, clinical notes, and medications from electronic health records provides superior phenotyping performance. *J Am Med Inform Assoc* 2016;23(e1):e20–e27
- 14 Neal W, Knowles J, Wilemon K. Underutilization of cascade screening for familial hypercholesterolemia. *Clin Lipidol* 2014;9(03):291–293
- 15 Vinci SR, Rifas-Shiman SL, Cheng JK, Mannix RC, Gillman MW, de Ferranti SD. Cholesterol testing among children and adolescents during health visits. *JAMA* 2014;311(17):1804–1807
- 16 Ritchie SK, Murphy EC, Ice C, et al. Universal versus targeted blood cholesterol screening among youth: the CARDIAC project. *Pediatrics* 2010;126(02):260–265
- 17 Nordestgaard BG, Chapman MJ, Humphries SE, et al; European Atherosclerosis Society Consensus Panel. Familial hypercholesterolaemia is underdiagnosed and undertreated in the general population: guidance for clinicians to prevent coronary heart disease: consensus statement of the European Atherosclerosis Society. *Eur Heart J* 2013;34(45):3478–90a
- 18 Williams R, Schumacher M, Barlow G, et al. Documented need for more effective diagnosis and treatment of familial hypercholesterolemia according to data from 502 heterozygotes in Utah. *Am J Cardiol* 1993;72:18D–24D
- 19 Wald DS, Bestwick JP, Morris JK, Whyte K, Jenkins L, Wald NJ. Child-parent familial hypercholesterolemia screening in primary care. *N Engl J Med* 2016;375(17):1628–1637
- 20 Lilly CL, Gebremariam YD, Cottrell L, John C, Neal W. Trends in serum lipids among 5th grade CARDIAC participants, 2002-2012. *J Epidemiol Community Health* 2014;68(03):218–223
- 21 Pyles L, Elliott E, Neal W. Screening for hypercholesterolemia in children. *Curr Cardiol Rep* 2017;11:5
- 22 Elliott E, Lilly C, Murphy E, Pyles LA, Cottrell L, Neal WA. The Coronary Artery Risk Detection in Appalachian Communities (CARDIAC) project: an 18 year review. *Curr Pediatr Rev* 2017;13(04):265–276
- 23 Pletcher MJ, Vittinghoff E, Thanataveerat A, Bibbins-Domingo K, Moran AE. Young adult exposure to cardiovascular risk factors and risk of events later in life: the Framingham Offspring Study. *PLoS One* 2016;11(05):e0154288
- 24 Luirink IK, Wiegman A, Kusters DM, et al. 20-year follow-up of statins in children with familial hypercholesterolemia. *N Engl J Med* 2019;381(16):1547–1556
- 25 Wald DS, Kasturiratne A, Godoy A, et al. Child-parent screening for familial hypercholesterolemia. *J Pediatr* 2011;159(05):865–867
- 26 Denney MJ, Long DM, Armistead MG, Anderson JL, Conway BN. Validating the extract, transform, load process used to populate a large clinical research database. *Int J Med Inform* 2016;94:271–274
- 27 Hicks KA, Tcheng JE, Bozkurt B, et al. 2014 ACC/AHA key data elements and definitions for cardiovascular endpoint events in clinical trials: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Data Standards (Writing Committee to Develop Cardiovascular Endpoints Data Standards). *J Am Coll Cardiol* 2015;66(04):403–469
- 28 Akobeng AK. Understanding diagnostic tests 1: sensitivity, specificity and predictive values. *Acta Paediatr* 2007;96(03):338–341