

How well can familial hypercholesterolemia be identified in an electronic health record database?

Katherine E Mues¹
Alina N Bogdanov²
Keri L Monda¹
Larisa Yedigarova³
Alexander Liede¹
Lee Kallenbach²

¹Center for Observational Research, Amgen Inc., Thousand Oaks, CA, USA; ²Practice Fusion, San Francisco, CA, USA; ³US Medical Affairs, Amgen Inc., Thousand Oaks, CA, USA

Background: Familial hypercholesterolemia (FH) is a condition characterized by high cholesterol levels and increased risk for coronary heart disease (CHD) that often goes undiagnosed. The Dutch Lipid Network Criteria (DLNC) are used to identify FH in clinical settings via physical examination, personal and family history of CHD, in addition to the presence of deleterious mutations of the *LDLR*, *ApoB*, and *PCSK9* genes. Agreement between clinical and genetic diagnosis of FH varies. While an ICD diagnosis code was not available for coding FH until 2016, Systematized Nomenclature of Medicine (SNOMED) clinical concept codes, including genetic diagnoses, for FH have been utilized in electronic health records (EHRs).

Objective: To evaluate the concordance of identifying FH via SNOMED and ICD-10 CM codes vs the DLNC in an EHR database.

Methods: Using the Practice Fusion EHR database, the sensitivity, specificity, positive predictive value (PPV), and negative predictive value were calculated comparing an FH cohort identified via SNOMED and ICD-10 CM codes to one identified via the DLNC.

Results: Among 907,616 patients with hypercholesterolemia, 2,180 were identified as FH via SNOMED code (zero were identified via ICD-10 CM), 259 had a DLNC score 6–8 (probable FH), and 45 had a DLNC score >8 (definite FH). Compared to DLNC score >8, the sensitivity, specificity, and PPV of the FH SNOMED code were 84.4%, 99.4%, and 6.4%, respectively. Compared to DLNC score ≥6, the sensitivity was 36.8% and the specificity was 99.5% with a PPV of 18.7%.

Conclusion: Compared to the clinical criteria for FH, identification of FH patients via SNOMED diagnosis codes had high sensitivity and specificity, but low PPV. The discordance of these two techniques in identifying FH patients speaks to the challenges in identifying FH patients in large electronic databases such as administrative claims and EHR.

Keywords: familial hypercholesterolemia, Dutch Lipid Network Criteria, electronic health record, SNOMED

Introduction

Familial hypercholesterolemia (FH) is a genetic disorder characterized by elevated cholesterol levels, specifically, high levels of low-density lipoprotein cholesterol (LDL-C) often resulting in premature coronary heart disease (CHD).¹ The most recent prevalence estimates of heterozygous FH in the USA suggest between 1 in 212 and 1 in 250 individuals may be affected.^{2,3} Historically, the prevalence of homozygous FH has been estimated to be 1 in 1,000,000, whereas new research has shown the prevalence may range from 1 in 160,000 to 1 in 300,000 individuals.^{1,4–7} Nordestgaard et al estimated that there are between 14 and 34 million individuals living with FH

Correspondence: Katherine E Mues
Amgen Inc., One Amgen Center Drive,
MS 24-2-A, Thousand Oaks 91320, CA,
USA
Tel +1 805 447 1864
Email kmues@amgen.com

worldwide.¹ Given the underutilization of genetic testing and the lack of awareness of FH in the clinical community, even the most current prevalence estimates may under-represent true prevalence of this condition in the general population. Thus, many individuals may not be getting the appropriate therapies to control their LDL-C.

Genetic variants leading to FH are present in the *LDLR*, *APOB*, or *PCSK9* genes, accounting for >90%, 5%, and 1% of FH cases, respectively.^{5,8-11} However, FH is most frequently diagnosed clinically based on elevated levels of LDL-C, premature CHD, family history of cardiovascular disease (CVD) and/or high LDL-C, and physical signs of elevated cholesterol.⁴ Although a positive genetic test for FH is conclusive evidence of the disease, FH cannot be ruled out in the result of a negative genetic test. In fact, ~20% of clinically evident FH patients do not appear to have a mutation identified in the FH-associated genes.¹² A recent study by Khera et al found that ~2% of individuals free of coronary artery disease (CAD) with LDL-C \geq 190 mg/dL (identified as FH patients based solely on LDL-C) tested positive for a mutation in one of the FH-associated genes.¹³ The same study also found the risk of CAD was higher among individuals with a genetic mutation for FH, even within LDL-C strata, suggesting that cumulative lifelong exposure to elevated LDL-C is associated with a higher CAD risk.

Concordance between the genetic and clinical diagnosis of FH varies widely due to inconsistencies in how FH is clinically diagnosed as well as lack of full understanding of the genetic basis of the disease.¹³⁻¹⁶ There are several existing clinical criteria used to diagnosis FH: 1) the Dutch Lipid Network Criteria (DLNC); 2) the Simon Broome criteria; and 3) the Make Early Diagnosis to Prevent Early Death (MEDPED) criteria.^{1,5,15,17} These three sets of criteria are similar in their diagnostic parameters and are generally viewed as interchangeable by clinicians. The DLNC is perhaps the most widely recognized score and uses five criteria (both clinical and genetic) to identify FH: family history of CVD, clinical history of premature CHD, physical examination for xanthomas or corneal arcus, very high LDL-C on repeated measurements, and a causative mutation detected by genetic testing.

Until the recent introduction of the ICD, 10th Revision (ICD-10 CM-CM) code for FH in October 2016 (ICD-10 CM E78.01), an ICD diagnosis code for FH did not exist in the USA, making it difficult to identify individuals with FH within large healthcare databases. However, Systematized Nomenclature of Medicine (SNOMED) clinical concept codes for FH do exist. SNOMED is a comprehensive and precise clinical coding system that was developed and supported by the International Health Terminology Standards

Organization and the US National Library of Medicine as a standard for electronic exchange of clinical health information.¹⁸ The Office of the National Coordinator for Health Information Technology has approved SNOMED as a vocabulary code set to ensure consistent clinical terminology between systems.¹⁹ As such, SNOMED codes are often implemented and mapped to other coding systems within EHR software applications. The SNOMED system includes a code for FH, as well as specific codes for heterozygous and homozygous FH, and FH associated with a mutation in a specific gene (*LDLR*, *APOB*, and *PCSK9*).

The recent introduction of the ICD-10 CM code for FH and the increased utilization of SNOMED codes provide an opportunity to more easily identify FH patients in EHR systems. However, it is unclear whether physicians are utilizing these codes while doing a clinical work-up for FH through criteria such as the DLNC. For example, providers may enter in clinical components of the DLNC without explicitly recording that the patient has FH; alternatively, providers may record that the patient has FH without explicitly recording the clinical components that comprise it. Thus, our objective was to evaluate the concordance of FH diagnosis codes with DLNC clinical criteria for FH within an EHR database.

Methods

Patient population

We used data from the Practice Fusion EHR database that consists of patient-level data from their proprietary cloud-based ambulatory EHR platform, currently in use at over 30,000 practices in all 50 US states. A majority of Practice Fusion practices are single provider or small group practices. Data are available for over 38 million unique patients starting in 2010, of which over 16 million are currently active on the platform. Data are updated daily and are made available for analysis in a Health Information Privacy Protection Act-compliant de-identified research database. Substantial data validation and quality checks are performed as part of this process. Amgen licensed the Practice Fusion EHR data for analytic use. Institutional review board approval was not necessary for this study, as we did not intervene on patients whose data are included in the Practice Fusion EHR.

SNOMED codes were introduced into the Practice Fusion EHR platform in 2013. When entering a diagnosis into a structured field, Practice Fusion EHR users search for clinically meaningful terms related to a disease condition, for example, “familial hypercholesterolemia”. From there, a drop-down list of matching diagnosis codes appears. A physician will then choose the code that best captures the patient’s diagnosis, which is then automatically mapped to all other correspond-

ing diagnosis codes. For example, a physician may type in “familial hypercholesterolemia” into the record. A drop-down list of diagnosis codes appears from which a physician chooses the ICD-10 CM code for FH. It is automatically mapped to the corresponding SNOMED and ICD-9 codes (Figure S1). As opposed to claims-based coding, the recording of diagnosis in an EHR is not related to billing and reimbursement.

For this analysis, four patient cohorts of interest were identified between January 1, 2013 and October 16, 2016. The hypercholesterolemia cohort was defined as the source population by the presence of at least one of the following ICD diagnosis codes for pure hypercholesterolemia: 272.0, E78.0. Within the hypercholesterolemia cohort, the FH SNOMED/ICD-10 CM cohort was defined by the presence of either a SNOMED or ICD-10 CM diagnosis code for FH (Table 1). Also within the hypercholesterolemia cohort, the FH DLNC cohort was defined via diagnoses and LDL-C lab test results that identify patients who meet the DLNC for FH. A patient receives a specific number of points for each DLN criterion that is present, which are then summed into an overall score. The score is categorized into the following groups: >8 definite FH; 6–8, probable FH; 3–5, possible FH; and <3, unlikely FH. See Table S1 for corresponding SNOMED, Logical Observation Identifiers Names and Codes, ICD-10 CM, and ICD-9 codes used to identify the DLNC.

Statistical analysis

Within each cohort, we calculated the distribution of patient demographics, number of patient visits, available follow-up time in the database, prescribed medications, LDL-C values, and other lab values (high-density lipoprotein cholesterol [HDL], triglycerides, total cholesterol, and LDL-C/HDL-C ratio). We evaluated the prescription of lipid lowering therapy (LLT) (statins, ezetimibe, and PCSK9 inhibitors) and other concomitant medications (insulin and other glucose lowering agents, as well as antihypertensives) at any point in the

patient’s record. We evaluated both the highest available and most recent LDL-C value in a patient’s record. For categorical variables, we calculated the number and percent of patients within each cohort with the variable of interest; for continuous variables, we calculated mean and SD within each cohort.

In this analysis, we compared a diagnosis of FH recorded via a SNOMED or ICD-10 CM code to the DLNC clinical diagnosis of FH. The sensitivity (Se), specificity (Sp), positive predictive value (PPV), negative predictive value (NPV), and the corresponding 95% CIs of identifying FH patients via SNOMED/ICD-10 CM codes vs several cut points of the DLNC score (DLNC >8, DLNC ≥6, and DLNC ≥3) were calculated. The calculation of Se, Sp, PPV, and NPV was limited to patients meeting at least one of the DLNC components. All analyses were conducted in IBM Netezza SQL.

Results

In the Practice Fusion database, 907,616 patients with hypercholesterolemia were identified between January 1, 2013 and October 16, 2016. Of these patients, 2,180 (0.24%) were identified as FH via an SNOMED code (there were no records of an ICD-10 CM code for FH), 96,392 (10.6%) had at least one criterion of the DLNC, 259 (0.03%) had a DLNC score 6–8 indicating probable FH, and 45 (0.005%) had a DLNC score >8 indicating definite FH (Table 2). Among patients identified as definite FH, 38 (84.4%) were also identified as FH via SNOMED codes, while among patients identified as probable FH, 74 (28.6%) were also identified as FH via SNOMED codes (Figure 1).

Table 2 summarizes the baseline demographic and clinical characteristics among patients in the hypercholesterolemia cohort, FH SNOMED cohort, those meeting at least 1 of the DLN criterion, those with a DLNC score 6–8, and those with a DLNC score >8. Individuals identified as FH via SNOMED code and via the DLNC were younger and a higher proportion were female than those in the source hypercholesterolemia

Table 1 SNOMED and ICD-10 codes for familial hypercholesterolemia

Code	Code type	Description
31654005	SNOMED	Familial hypercholesterolemia (disorder)
398036000	SNOMED	Familial hypercholesterolemia (disorder)
238079002	SNOMED	Familial hypercholesterolemia – heterozygous (disorder)
238078005	SNOMED	Familial hypercholesterolemia – homozygous (disorder)
403831006	SNOMED	Familial hypercholesterolemia due to genetic defect of apolipoprotein B (disorder)
403829002	SNOMED	Familial hypercholesterolemia due to heterozygous low-density lipoprotein receptor mutation (disorder)
403830007	SNOMED	Familial hypercholesterolemia due to homozygous low-density lipoprotein receptor mutation (disorder)
E78.01	ICD-10	Familial hypercholesterolemia (effective from October 1, 2016)

Abbreviation: SNOMED, Systematized Nomenclature of Medicine.

Table 2 Baseline demographics and clinical characteristics among patients identified in the Practice Fusion EHR database (2013–2016)

Characteristics	Hypercholesterolemia	FH SNOMED	At least 1 DLN criterion	DLNC score 6–8	DLNC score >8
Total, N	N=907,616	N=2,180	N=96,392	N=259	N=45
Age (years), n (%)					
Mean (SD)	63.9 (15.4)	48.0 (20.9)	57.2 (13.2)	52.3 (12.0)	52.5 (14.5)
<18	9,652 (1.1)	288 (13.2)	530 (0.8)	5 (1.9)	1 (2.2)
18–39	49,727 (5.5)	392 (18.0)	7,498 (7.8)	49 (18.9)	7 (15.6)
40–64	381,168 (42.0)	997 (45.7)	62,060 (64.4)	155 (59.8)	27 (60.0)
65–74	234,175 (25.8)	319 (14.6)	16,690 (17.3)	33 (12.7)	8 (17.8)
75+	232,787 (25.6)	184 (8.4)	9,612 (10.0)	17 (6.6)	2 (4.4)
Female gender, n (%)	502,197 (55.3)	1,291 (59.2)	56,026 (58.1)	165 (63.7)	32 (71.1)
Ethnicity, n (%)					
Hispanic	131,168 (14.5)	287 (13.2)	12,753 (13.2)	30 (11.6)	3 (6.7)
Non-Hispanic	516,743 (56.9)	998 (45.8)	57,470 (59.6)	140 (54.1)	26 (57.8)
Missing/unknown	259,705 (28.6)	895 (41.1)	26,169 (27.1)	89 (34.4)	16 (35.6)
Race					
White	433,710 (47.8)	969 (44.4)	43,832 (45.5)	106 (40.9)	24 (53.3)
African American	84,599 (9.3)	105 (4.8)	11,545 (12.0)	36 (13.9)	3 (6.7)
Other	87,466 (9.6)	134 (6.1)	10,414 (10.8)	19 (7.3)	2 (4.4)
Missing/unknown	301,841 (33.3)	972 (44.6)	30,601 (31.7)	98 (37.8)	16 (35.6)
No of visits (mean)	10.2	9.5	13.6	13.6	21.0
>1 visit since 2013, n (%)	783,701 (86.3)	1,890 (86.7)	91,978 (95.4)	248 (95.8)	45 (100.0)
Comorbidities, n (%)					
Quan Enhanced Charlson comorbidity score ^a (mean)	0.96	0.61	0.92	0.96	1.15
Any malignancy	38,382 (4.2)	63 (2.9)	3,716 (3.3)	4 (1.5)	3 (6.7)
Cerebrovascular disease	46,972 (5.2)	73 (3.3)	8,001 (8.3)	17 (6.6)	4 (8.9)
Congestive heart failure	34,095 (3.8)	43 (2.0)	2,866 (3.0)	9 (3.5)	2 (4.4)
Diabetes with chronic complication ^b	40,309 (4.4)	64 (2.9)	3,273 (3.4)	21 (8.1)	1 (2.2)
Diabetes without chronic complication	247,919 (27.3)	288 (13.2)	21,285 (22.1)	42 (16.8)	10 (22.2)
Myocardial infarction	12,654 (1.4)	19 (0.9)	2,433 (2.5)	5 (1.9)	1 (2.2)
Peripheral vascular disease	50,268 (5.5)	70 (3.2)	6,784 (7.0)	15 (5.8)	6 (13.3)
Renal disease	50,117 (5.5)	69 (3.2)	4,976 (5.2)	23 (8.9)	2 (4.4)
Prescribed medications, n (%)					
Statin (any intensity)	563,821 (62.1)	1,055 (48.4)	61,137 (63.4)	178 (68.7)	24 (53.3)
Ezetimibe	35,910 (4.0)	177 (8.1)	4,967 (5.2)	35 (13.5)	7 (15.6)
PCSK9 inhibitor	639 (0.1)	105 (4.8)	275 (0.3)	9 (3.5)	7 (15.6)
Antihypertensive	421,260 (46.4)	535 (24.5)	40,146 (41.6)	83 (32.0)	13 (28.9)
Insulin	64,160 (7.1)	80 (3.7)	5,840 (6.1)	15 (5.8)	3 (6.7)

Notes: ^aThe Quan enhanced comorbidity score ranges from 0 to 37. ^bChronic complications of diabetes include renal manifestations, ophthalmic manifestations, neurological manifestations, and peripheral circulatory disorders.

Abbreviations: EHR, electronic health record; FH, familial hypercholesterolemia; SNOMED, Systematized Nomenclature of Medicine; DLNC, Dutch Lipid Network Criteria.

cohort. The average number of visits in the EHR during the study period ranged from 9.5 among the FH SNOMED cohort to 21.0 among those with DLNC score >8. The average Quan Enhanced Charlson comorbidity score was highest among those with DLNC score >8 and DLNC score 6–8, followed by the hypercholesterolemia cohort, those with at least 1 DLN criterion, and the FH SNOMED cohort. Prescription of statins ranged from 48.4% among those in the FH SNOMED cohort to 68.7% among those with a DLNC score 6–8 (Table 2).

Prescriptions for nonstatin LLTs were most frequent among those with a DLNC >8.

Patients who had a DLNC score 6–8 and >8 had the highest mean LDL-C values (Table 3), and ~62% of patients with a DLNC score 6–8 had an LDL-C value of at least 330 mg/dL (data not shown). The mean LDL-C value among those in the FH SNOMED cohort and those meeting at least one DLN criterion was higher than among those in the hypercholesterolemia cohort (for both most recent and highest LDL-C

values) (Table 3). The mean levels of triglycerides, total cholesterol, and the LDL/HDL ratio were relatively higher among patients with DLNC score 6–8 and >8 compared to the other cohorts.

Among the full hypercholesterolemia cohort, 10.6% (N=96,392) had at least one DLN criterion compared with 27.4% (N=598) among the FH SNOMED cohort. Among

hypercholesterolemia patients with at least one of the DLN criterion, LDL-C between 155 and 189 mg/dL (58.0%) was the criterion most frequently met, followed by LDL-C 190–249 mg/dL (18.3%), clinical history of premature CAD (11.2%), and clinical history of premature cerebral or peripheral vascular disease (9.1%) (Figure 2). Among FH SNOMED patients with at least one of the DLN criterion, LDL-C 190–249 mg/dL was the most frequently met (33.9%), followed by LDL-C 155–189 mg/dL (27.9%), a positive genetic test (15.6%), clinical history of premature CAD (10.9%), LDL-C 250–239 mg/dL (9.7%), and family history of premature coronary and vascular disease or LDL-C above the 95th percentile (9.4%).

Among the 96,392 patients with at least one DLN criterion, 45 had a DLNC score >8, 304 had a DLNC score ≥6, 20,048 had a DLNC score ≥3, and 598 were identified as FH via an SNOMED code. Compared to the DLNC >8, the sensitivity of the FH SNOMED code was 84.4% and the specificity was 99.4%, with a PPV of 6.4% (Table 4). Compared to a DLNC ≥6, the sensitivity was 36.8% and the specificity was 99.5% with a PPV of 18.7%.

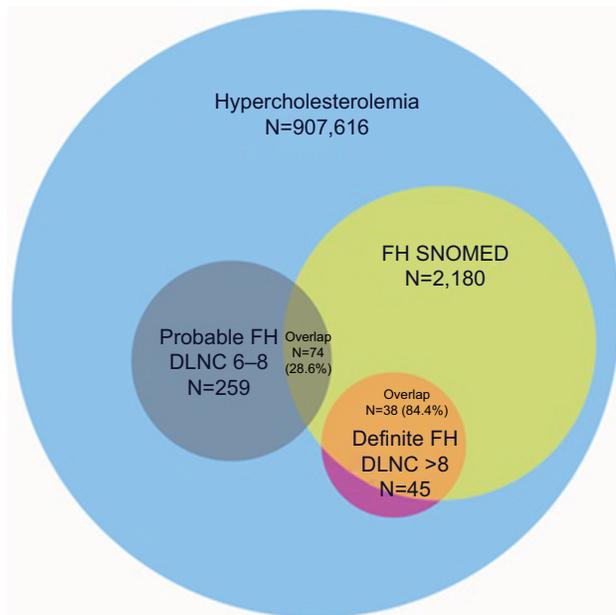


Figure 1 Overlap of study cohorts identified in the Practice Fusion EHR database, 2013–2016.

Abbreviations: EHR, electronic health record; FH, familial hypercholesterolemia; SNOMED, Systematized Nomenclature of Medicine; DLNC, Dutch Lipid Network Criteria.

Discussion

Compared to the clinical criteria for FH, the identification of FH patients via SNOMED diagnosis codes had high sensitivity and specificity, but low PPV. These results indicate that FH patients identified via SNOMED codes are unlikely to be identified using clinical criteria contained in their EHR, even at several cutoff levels of the DLNC. The discordance of these two techniques in identifying FH patients speaks to

Table 3 Lab values among patients identified in the Practice Fusion EHR database (2013–2016)

Lab	Hypercholesterolemia	FH SNOMED	At least 1 DLN criterion	DLNC score 6–8	DLNC score >8
Total	N=907,616	N=2,180	N=96,392	N=259	N=45
LDL-C: n (%) with measure	339,112 (37.4)	946 (43.4)	82,429 (85.5)	217 (83.8)	41 (91.1)
Most recent: mean (SD) (mg/dL)	111.1 (38.7)	139.2 (59.1)	147.4 (42.0)	242.6 (98.2)	186.9 (74.5)
Highest recorded: mean (SD) (mg/dL)	125.1 (40.3)	157.4 (63.2)	174.0 (32.6)	310.4 (65.2)	229.1 (69.0)
HDL-C: n (%) with measure	360,470 (39.7)	1,003 (46.0)	82,141 (85.2)	218 (84.2)	41 (91.1)
Mean (SD)	53.9 (16.7)	53.7 (16.9)	54.2 (15.5)	52.6 (18.2)	49.8 (13.0)
Triglycerides: n (%) with measure	296,922 (32.7)	855 (39.2)	71,722 (74.4)	186 (71.8)	36 (80.0)
Mean (SD)	140.0 (85.2)	146.0 (93.6)	146.1 (78.9)	163.7 (88.3)	204.9 (106.1)
Total cholesterol: n (%) with measure	381,397 (42.0)	1,077 (49.4)	83,407 (86.5)	226 (87.3)	41 (91.1)
Mean (SD)	192.7 (45.0)	222.6 (69.4)	229.4 (47.7)	324.8 (105.39)	274.1 (76.2)
LDL/HDL: n (%) with measure	48,397 (5.3)	137 (6.3)	12,211 (12.7)	30 (11.6)	6 (13.3)
Mean (SD)	2.2 (1.0)	3.0 (1.8)	2.8 (1.1)	4.8 (2.9)	3.7 (1.1)

Notes: *Based on the most recent lab value in the EHR.

Abbreviations: EHR, electronic health record; LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol; FH, familial hypercholesterolemia; SNOMED, Systematized Nomenclature of Medicine; DLNC, Dutch Lipid Network Criteria.

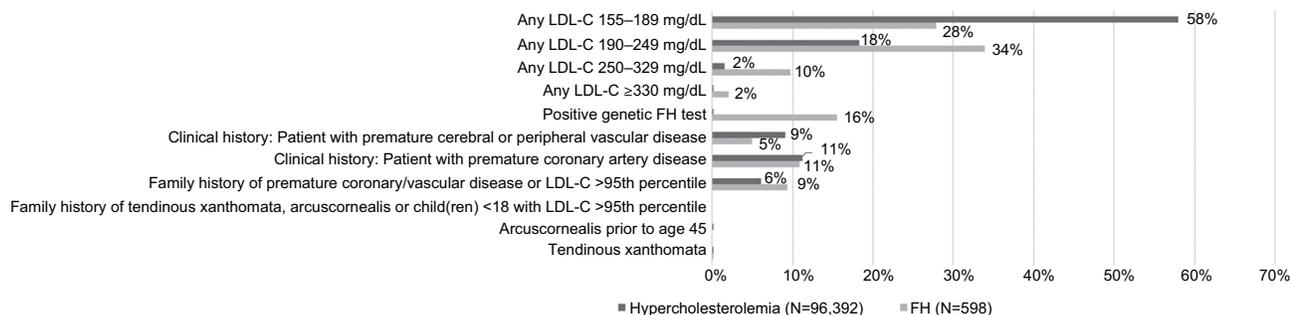


Figure 2 Distribution of the DLNC among subjects meeting at least one DLNC in the hypercholesterolemia cohort (N=96,392) and the FH SNOMED cohort (N=598). **Note:** Subjects can meet more than one criteria, yet must meet at least one to be included in the figure.

Abbreviations: LDL-C, low-density lipoprotein cholesterol; FH, familial hypercholesterolemia; SNOMED, Systematized Nomenclature of Medicine; DLNC, Dutch Lipid Network Criteria.

Table 4 Sensitivity, specificity, PPV, and NPV of ICD-10/SNOMED codes for FH vs the DLNC at varying score cutoffs (among subjects with at least one DLN criterion, N=96,392)

DLNC score	FH identification				Sensitivity	Specificity	PPV	NPV
	DLNC Y	DLNC N	DLNC Y	DLNC N				
	SNOMED Y	SNOMED Y	SNOMED N	SNOMED N				
>8	38	560	7	95,787	84.4%	99.4%	6.4%	100.0%
>6	112	486	192	95,602	36.8%	99.5%	18.7%	99.8%

Notes: DLNC Y indicates yes, the subject met the DLNC criteria for FH at the specified value in column 1. SNOMED Y indicates yes, the subject was identified as FH via a SNOMED code. DLNC N indicates no, the subject did not meet the DLNC criteria for FH at the specified value in column 1. SNOMED N indicates no, the subject was not identified as FH via a SNOMED code.

Abbreviations: PPV, positive predictive value; NPV, negative predictive value; FH, familial hypercholesterolemia; SNOMED, Systematized Nomenclature of Medicine; DLNC, Dutch Lipid Network Criteria.

the challenges in identifying FH patients in large electronic databases such as administrative claims and EHR.

If we consider clinical criteria as the gold standard of an FH diagnosis, standard coding such as SNOMED not only identifies a large percentage of “true FH” patients but also identifies patients who do not meet the gold standard of FH diagnosis. This may be a true misclassification of these patients via diagnosis codes or may speak to the lack of such clinical criteria being recorded in the structured portions of the medical record. Thus, the diagnosis codes may be accurate, but the gold standard may not be recorded as easily searchable content in an EHR, making identification of FH patients difficult. These findings may also indicate that clinicians are not fully aware of the clinical symptoms of FH (such as tendon xanthoma), and thus are not looking for and noting the clinical signs within the medical record, also contributing to the under-diagnosis of the condition.

Several other studies that have evaluated the application of clinical diagnostic systems in large databases have found similar challenges in identifying FH patients. Abul-Husn et al found that when applying the DLNC to a population of individuals with severe hypercholesterolemia, only 24% of

FH variant carriers were classified as either probable or definite FH, highlighting the rather low sensitivity of the DLNC when applied to an EHR database.²⁰ In the same study, of the 4,435 individuals with LDL-C >190 mg/dL in the EHR, only 112 (2.5%) had an FH genetic variant, again indicating the low capture of genetic variants for FH and potentially the underutilization of genetic testing in this population. In a general practitioner electronic healthcare database of patients in Australia, Vickery et al found that among active patients, 0.7% had an LDL-C value ≥5 mmol/L (~190 mg/dL), but only 0.013% had a diagnosis code for FH, again indicating the low diagnosis rates of FH. Among those with an LDL-C ≥5 mmol/L, 60% had been coded for a lipid disorder, 56% had record of a statin prescription, 40% had both, and 24% had neither.²¹

Although there was no ICD diagnosis code for FH until October 2016, SNOMED codes offer an alternate coding system to identify patients with FH in an electronic database. The SNOMED nomenclature is the most comprehensive healthcare terminology, an international standard for the electronic exchange of clinical health information supported by the US National Library of Medicine, which has

previously been incorporated into EHR-based algorithms to identify clinically relevant cardio- and cerebrovascular disease.^{19,22} Although practitioners may not be as familiar with SNOMED codes as other coding systems (eg, ICD-9/ICD-10 CM, HCPC), one of the strengths of the Practice Fusion data platform was the automatic mapping of physician diagnoses entered into the structured data fields into SNOMED codes. Thus, even if a physician was not familiar with the FH SNOMED codes, the Practice Fusion system automatically mapped each structured diagnosis of FH to an SNOMED code, allowing for efficient identification of FH patients within the database. Although we did use the ICD-10 CM code for identifying FH in this analysis, there were no cases identified in this manner, since the analysis was conducted through mid-October 2016, with limited (only 3 months) post-ICD-10 CM implementation. The hope is that, over time, this code will become more frequently used and awareness of FH will increase, to allow for more accurate diagnosis of FH within EHR systems.

In this analysis, only a small proportion of identified individuals with FH had evidence of a functional mutation on the *LDLR*, *APOB*, or *PCSK9* gene resulting from a genetic test. These results reiterate the low frequency of and differing opinions on the incremental value of FH genetic testing, even among a population of patients who have a diagnostic code for FH, suggesting that FH remains the first and foremost clinical diagnosis. Genetic testing is expensive and may not be covered by insurance plans.²³ Additionally, clinical guidelines do not recommend universal testing, but instead recommend cascade screening among first-degree relatives of an individual with FH, with universal screening limited to children <16 years with high levels of cholesterol.^{1,12} Finally, due to limitations in FH genetic testing, some individuals who carry a genetic mutation for FH may not be easily identifiable. This can happen for several reasons. First, a standard FH panel only covers a subset of all known variants, generally limited to those that contribute to a more severe disease phenotype. For example, although over 1,500 variants have been identified in the *LDLR* gene alone, not all of these are covered on a standard genetic array. Second, individuals that are compound heterozygotes (eg, two different mutations affecting two alleles) may display an FH phenotype without an easily identifiable genotype given that multiple low penetrance variants may be present.²⁴

This study has several limitations that should be considered when interpreting the results. First, the true gold standard of an FH diagnosis is unclear. Although we compare the SNOMED codes for FH to the DLNC, and calculate sensitivity and specificity accordingly, we must interpret these

findings with the understanding that clinical criteria are not the only method to diagnose FH; genetic testing may have been done and not necessarily recorded within the medical record. Despite this, patients identified as FH via the DLNC were slightly older, more female, had a higher disease burden, more statin and nonstatin LLT prescriptions, and had higher LDL-C values compared with FH patients identified via a SNOMED code. Given these findings, considering the DLNC as the gold standard in this study does not seem unreasonable. Second, the analysis was limited to only structured information captured by the EHR; FH diagnoses and results of genetic testing were not confirmed from unstructured data. We did not incorporate information captured in physician notes into the main analysis. Given that EHRs may still be relatively new to healthcare providers, it is possible that providers may prefer to record much of a patient's information via free text. Natural language processing is a lengthy and expensive process, limiting our ability to use the unstructured information included in the EHR, and more generally, speaks to the challenge of identifying FH patients in electronic databases. Third, this analysis was limited to information captured within the Practice Fusion EHR platform, which mostly covers primary care physicians, who may be less familiar with a condition such as FH. We do not have health information collected on patients who also sought care outside of Practice Fusion's network. Although the Practice Fusion network does capture some specialists such as cardiologists within community-based practices, it is possible that, compared to large health centers, these physicians may be less likely to diagnose and/or record a diagnosis of FH. Last, these results are not necessarily applicable to other EHR databases that may capture different types of health practitioners or may be related to an associated payer who may have specific requirements for FH diagnoses in order for appropriate treatments to be covered.

Comparing SNOMED diagnosis codes for FH to the DLNC clinical criteria, the sensitivity and specificity were high, whereas the positive predictive value was low. Thus, if considering the DLNC as the gold standard, the identification of FH patients in an EHR using a SNOMED code may result in the capture of many false positives. It is unclear if this is due to the true misclassification of FH via diagnosis codes or rather to the lack of recording of clinical symptoms of FH in the structured portions of an EHR (eg, not the clinical notes). Future work in the Practice Fusion database may include a comprehensive review of physician notes for identifying subjects with FH. Although EHRs provide an opportunity to extract physician notes with deeper clinical information, for research studies on large population-level databases,

natural language processing to extract information from the notes is expensive, time-consuming, and may be limited in its accuracy. Therefore, most research studies using EHR data on a large scale will most likely only be able to use structured information to identify subjects with FH, requiring improved techniques of identifying individuals with FH via structured data. Although the recent introduction of an ICD-10 CM code for FH in the USA will mitigate some of the challenges in identifying FH subjects in EHR and claims databases, the lag in data availability and need for historic data will result in continued challenges in research. However, until FH is more fully recognized by the clinical community, the addition of an ICD-10 CM code will not completely correct the under-diagnosis of FH.

Acknowledgment

Amgen, Inc. funded this study.

Author contributions

All authors contributed to data analysis, drafting and revising the article, gave final approval of the version to be published, and agree to be accountable for all aspects of the work.

Disclosure

KEM, KLM, LY, and AL are employees and stockholders of Amgen Inc. ANB and LK are employees of Practice Fusion, which received funding from Amgen to conduct the study. The authors report no other conflicts of interest in this work.

References

- Nordestgaard BG, Chapman MJ, Humphries SE, et al. 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):3478–3490.
- Buchholz EM, Rodday AM, Kolor K, Khoury MJ, de Ferranti SD. Prevalence and predictors of cholesterol screening, awareness, and statin treatment among US adults with familial hypercholesterolemia or other forms of severe dyslipidemia (1999–2014). *Circulation*. 2018;137(21):2218–2230.
- de Ferranti SD, Rodday AM, Mendelson MM, Wong JB, Leslie LK, Sheldrick RC. Prevalence of Familial hypercholesterolemia in the 1999 to 2012 United States National Health and Nutrition Examination Surveys (NHANES). *Circulation*. 2016;133(11):1067–1072.
- Goldstein JK, Hobbs HH, Brown MS. Familial hypercholesterolemia. In: Scriver CR, Beaudet AL, Sly WS, Valle D, editors. *The Metabolic & Molecular Bases of Inherited Diseases*. 8th ed. New York: McGraw-Hill; 2001:2863–2913.
- Austin MA, Hutter CM, Zimmern RL, Humphries SE. Genetic causes of monogenic heterozygous familial hypercholesterolemia: a HuGE prevalence review. *Am J Epidemiol*. 2004;160(5):407–420.
- Benn M, Watts GF, Tybjaerg-Hansen A, Nordestgaard BG. Familial hypercholesterolemia in the danish general population: prevalence, coronary artery disease, and cholesterol-lowering medication. *J Clin Endocrinol Metab*. 2012;97(11):3956–3964.
- Sjouke B, Hovingh GK, Kastelein JJP, Stefanutti C. Homozygous autosomal dominant hypercholesterolaemia. *Curr Opin Lipidol*. 2015;26(3):200–209.
- Brown MS, Goldstein JL, Arnold W, Perdue S. Expression of the familial hypercholesterolemia gene in heterozygotes: mechanism for a dominant disorder in man. *Science*. 1974;185(4145):61–63.
- Soria LF, Ludwig EH, Clarke HRG, Vegat GL, Grundyt SM, Mccarthy BJ. Association between a specific apolipoprotein B mutation and familial defective apolipoprotein B-100 (genetic disease/cholesterol metabolism). *Genetics*. 1989;86:587–591.
- Innerarity TL, Weisgraber KH, Arnold KS, et al. Familial defective apolipoprotein B-100: low density lipoproteins with abnormal receptor binding. *Proc Natl Acad Sci U S A*. 1987;84(19):6919–6923.
- Abifadel M, Varret M, Rabès JP, et al. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat Genet*. 2003;34(2):154–156.
- Goldberg AC, Hopkins PN, Toth PP, et al. Familial hypercholesterolemia: screening, diagnosis and management of pediatric and adult patients: clinical guidance from the National Lipid Association Expert Panel on Familial Hypercholesterolemia. *J Clin Lipidol*. 2011;5(3):133–140.
- Khera AV, Won HH, Peloso GM, et al. Diagnostic yield and clinical utility of sequencing familial hypercholesterolemia genes in patients with severe hypercholesterolemia. *J Am Coll Cardiol*. 2016;67(22):2578–2589.
- Civeira F International Panel on Management of Familial Hypercholesterolemia, International Panel on Management of Familial Hypercholesterolemia. Guidelines for the diagnosis and management of heterozygous familial hypercholesterolemia. *Atherosclerosis*. 2004;173(1):55–68.
- Williams RR, Hunt SC, Schumacher MC, et al. Diagnosing heterozygous familial hypercholesterolemia using new practical criteria validated by molecular genetics. *Am J Cardiol*. 1993;72(2):171–176.
- Palacios L, Grandoso L, Cuevas N, et al. Molecular characterization of familial hypercholesterolemia in Spain. *Atherosclerosis*. 2012;221(1):137–142.
- Haase A, Goldberg AC. Identification of people with heterozygous familial hypercholesterolemia. *Curr Opin Lipidol*. 2012;23(4):282–289.
- National Institutes of Health. U.S. National Library of Medicine. SNOMED CT. Available from: <https://www.nlm.nih.gov/healthit/snomedct/>. Accessed January 11, 2017.
- Health and Human Services Department. 2015 Edition Health Information Technology (Health IT) Certification Criteria, 2015 Edition Base Electronic Health Record (EHR) Definition, and ONC Health IT Certification Program Modifications; 2015. Available from: <https://www.federalregister.gov/documents/2015/10/16/2015-25597/2015-edition-health-information-technology-health-it-certification-criteria-2015-edition-base>. Accessed August 14, 2018.
- Abul-Husn NS, Manickam K, Jones LK, et al. Genetic identification of familial hypercholesterolemia within a single U.S. health care system. *Science*. 2016;354(6319):aaf7000.
- Vickery AW, Ryan J, Pang J, Garton-Smith J, Watts GF. Increasing the detection of familial hypercholesterolaemia using general practice electronic databases. *Heart Lung Circ*. 2017;26(5):450–454.
- Esteban S, Rodríguez Tablado M, Ricci RI, Terrasa S, Kopitowski K. A rule-based electronic phenotyping algorithm for detecting clinically relevant cardiovascular disease cases. *BMC Res Notes*. 2017;10(1):281.
- Singh S, Bittner V. Familial hypercholesterolemia-epidemiology, diagnosis, and screening. *Curr Atheroscler Rep*. 2015;17(2):482.
- Talmud PJ, Shah S, Whittall R, et al. Use of low-density lipoprotein cholesterol gene score to distinguish patients with polygenic and monogenic familial hypercholesterolaemia: a case-control study. *Lancet*. 2013;381(9874):1293–1301.

Supplementary materials

Table S1 Codes for identifying the Dutch lipid network criteria

Criteria	Code	Code type	Description
First-degree relative with known premature coronary and vascular disease	439724007	SNOMED	Family history of cardiovascular disease in first-degree male relative less than 55 years of age
	438825005	SNOMED	Family history of cardiovascular disease in first-degree female relative less than 65 years of age
	401067009	SNOMED	Family history of myocardial infarction in first-degree male relative less than 55 years (situation)
	401066000	SNOMED	Family history of myocardial infarction in first-degree female relative less than 65 years of age
	V17.3	ICD-9	Family history of ischemic heart disease
	V17.41	ICD-9	Family history of sudden cardiac death
	Z82.4	ICD-10 CM	Family history of ischemic heart disease and other diseases of the circulatory system
	Z82.41	ICD-10 CM	Family history of sudden cardiac death
	Z82.49	ICD-10 CM	Family history of ischemic heart disease and other diseases of the circulatory system
First-degree relative with tendinous xanthoma and/or arcus cornealis	699108005	SNOMED	Family history of tendinous xanthoma in first-degree relative (situation)
First-degree relative with known LDL-C level above the 95th percentile	Not able to evaluate in EHR		
Children aged less than 18 years with LDL-C level above the 95th percentile	Not able to evaluate in EHR		
Patient with premature* coronary artery disease	410.xx	ICD-9	Acute myocardial infarction
	411.xx	ICD-9	Other acute and subacute forms of ischemic heart disease
	412	ICD-9	Old myocardial infarction
	413.xx	ICD-9	Angina pectoris
	414.xx	ICD-9	Other forms of chronic ischemic heart disease
	I20.x	ICD-10 CM	Angina pectoris
	I21.xx	ICD-10 CM	STEMI and non-STEMI myocardial infarction
	I22.x	ICD-10 CM	Subsequent STEMI and non-STEMI myocardial infarction
	I23.x	ICD-10 CM	Certain current complications following ST elevation (STEMI) and non-ST elevation (NSTEMI) myocardial infarction (within the 28-day period)
Patient with premature* cerebral or peripheral vascular disease	I24.x	ICD-10 CM	Other acute ischemic heart diseases
	429.2	ICD-9	Cardiovascular disease unspecified
	430	ICD-9	Subarachnoid hemorrhage
	431	ICD-9	Intracerebral hemorrhage
	432.x	ICD-9	Other and unspecified intracranial hemorrhage
	433.xx	ICD-9	Occlusion and stenosis of precerebral arteries
	434.xx	ICD-9	Occlusion of cerebral arteries
	435.x	ICD-9	Transient cerebral ischemia
	436	ICD-9	Acute, but ill-defined cerebrovascular disease
	437.x	ICD-9	Other and ill-defined cerebrovascular disease
	438.xx	ICD-9	Late effects of cerebrovascular disease
	443.9	ICD-9	Peripheral vascular disease, unspecified
	I25.xxx	ICD-10 CM	Chronic ischemic heart disease
I60.xx	ICD-10 CM	Nontraumatic subarachnoid hemorrhage	

(Continued)

Table S1 (Continued)

Criteria	Code	Code type	Description
	I61.x	ICD-10 CM	Nontraumatic intracerebral hemorrhage
	I62.xx	ICD-10 CM	Other and unspecified nontraumatic intracranial hemorrhage
	I63.xxx	ICD-10 CM	Cerebral infarction
	I65.xx	ICD-10 CM	Occlusion and stenosis of precerebral arteries, not resulting in cerebral infarction
	I66.xx	ICD-10 CM	Occlusion and stenosis of cerebral arteries, not resulting in cerebral infarction
	I67.xxx	ICD-10 CM	Other cerebrovascular diseases
	I68.x	ICD-10 CM	Cerebrovascular disorders in diseases classified elsewhere
	I69.xxx	ICD-10 CM	Sequelae of cerebrovascular disease
	I73.9	ICD-10 CM	Peripheral vascular disease, unspecified
Elevated LDL-C	55440-2	LOINC	Cholesterol in LDL (real) (mass/volume) in serum or plasma by VAP
	2090-9	LOINC	Deprecated cholesterol in LDL (mass/volume) in serum or plasma
	2089-1	LOINC	Cholesterol in LDL (mass/volume) in serum or plasma
	35198-1	LOINC	Cholesterol in LDL (mass or moles/volume) in serum or plasma
	18262-6	LOINC	Cholesterol in LDL (mass/volume) in serum or plasma by direct assay
	13457-7	LOINC	Cholesterol in LDL (mass/volume) in serum or plasma by calculation
	18261-8	LOINC	Cholesterol in LDL (mass/volume) in serum or plasma ultracentrifugate
Arcus cornealis prior to age 45 years	231924000	SNOMED	Arcus of cornea
Tendon xanthoma	69880002	SNOMED	Xanthoma tendinosum
Functional mutation in the <i>apo B</i> gene	403831006	SNOMED	Familial hypercholesterolemia due to genetic defect of apolipoprotein B (disorder)
Functional mutation in the <i>LDLR</i> gene	403829002	SNOMED	Familial hypercholesterolemia due to heterozygous low-density lipoprotein receptor mutation (disorder)
	403830007	SNOMED	Familial hypercholesterolemia due to homozygous low-density lipoprotein receptor mutation (disorder)

Abbreviations: SNOMED, Systematized Nomenclature of Medicine; LOINC, Logical Observation Identifiers Names and Codes; EHR, electronic health record; LDL-C, low-density lipoprotein cholesterol; VAP, vertical auto profile.

The figure consists of two side-by-side screenshots of the Practice Fusion EHR platform's diagnosis section, connected by a black arrow pointing from left to right.

Left Screenshot: Search Results

DIAGNOSIS*
 Search results | My Dx list (0)
 E78.01 Familial hypercholesterolemia
Z83.42 Family history of familial hypercholesterolemia
 E78.01 Familial Combined Hypercholesterolemia
 E78.01 Familial homozygous hypercholesterolemia
 E78.01 Familial hypercholesterolemia - heterozygous
 E78.2 Familial hypercholesterolemia with hyperlipemia
 E78.01 Familial hypercholesterolemia due to heterozygous LDL receptor mutation
 E78.01 Familial hypercholesterolemia due to homozygous LDL receptor mutation
 Add custom diagnosis: "familial hypercholesterolemia"

Buttons: Cancel, Save

Right Screenshot: Diagnosis Details

E78.01 Familial hypercholesterolemia Edit
 Associate diagnosis with encounter 06/14/2018

ACUITY Clear
 Chronic
 Acute

START DATE
 MM/DD/YYYY Today

STOP DATE
 MM/DD/YYYY Today

MEDICATIONS ASSOCIATED WITH DIAGNOSIS Add

MAPPED CODES Hide

TYPE	CODE	DESCRIPTION
ICD-9	272.0	Pure hypercholesterolemia
SNOMED CT	398036000	Familial hypercholesterolemia

COMMENT Add
 ENCOUNTER COMMENT Add

RESOURCES
 Patient education materials

Buttons: Cancel, Add another, Save

Figure S1 Diagnosis section of the Practice Fusion EHR platform (example only: the left panel displays a physician's ability to type in a diagnosis and the menu it populates; the right panel displays the mapping to diagnosis codes that occurs when a physician choose a specific diagnosis from the drop-down menu).

Abbreviation: EHR, electronic health record.

Clinical Epidemiology

Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification,

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>

systematic reviews, risk and safety of medical interventions, epidemiology and biostatistical methods, and evaluation of guidelines, translational medicine, health policies and economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Dovepress