ESC European Society of Cardiology

European Heart Journal - Digital Health (2022) **00**, 1–9
https://doi.org/10.1093/ehjdh/ztac059

ORIGINAL ARTICLE

Electronic health record-based facilitation of familial hypercholesterolaemia detection sensitivity of different algorithms in genetically confirmed patients

Niekbachsh Mohammadnia[1,2,3], Ralph K. Akyea[4], Nadeem Qureshi[4], Willem A. Bax[1,†], and Jan H. Cornel[2,3,*†]

[1]Department of Internal Medicine, Northwest Clinics, Wilhelminalaan 12, 1815 JD, Alkmaar, The Netherlands; [2]Department of Cardiology, Northwest Clinics, Wilhelminalaan 12, 1815 JD, Alkmaar, The Netherlands; [3]Department of Cardiology, Radboudumc, Geert Grooteplein Zuid 10, 6525 GA, Nijmegen, The Netherlands; and [4]Primary Care Stratified Medicine (PRISM) Research Group, Centre for Academic Primary Care, School of Medicine, University of Nottingham, Applied Health Research Building, University Park, Nottingham NG7 2RD, UK

Received 6 June 2022; revised 8 August 2022; online publish-ahead-of-print 17 October 2022

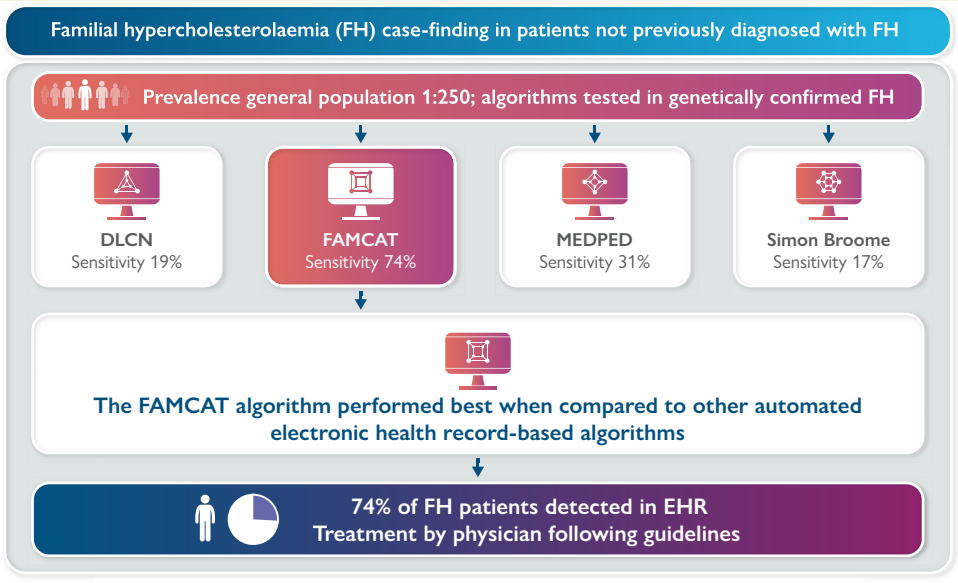
Aims Familial hypercholesterolaemia (FH) is a disorder of LDL cholesterol clearance, resulting in increased risk of cardiovascular disease. Recently, we developed a Dutch Lipid Clinic Network (DLCN) criteria-based algorithm to facilitate FH detection in electronic health records (EHRs). In this study, we investigated the sensitivity of this and other algorithms in a genetically confirmed FH population.

Methods and results All patients with a healthcare insurance-related coded diagnosis of 'primary dyslipidaemia' between 2018 and 2020 were assessed for genetically confirmed FH. Data were extracted at the time of genetic confirmation of FH (T1) and during the first visit in 2018–2020 (T2). We assessed the sensitivity of algorithms on T1 and T2 for DLCN ≥ 6 and compared with other algorithms [familial hypercholesterolaemia case ascertainment tool (FAMCAT), Make Early Diagnoses to Prevent Early Death (MEDPED), and Simon Broome (SB)] using EHR-coded data and using all available data (i.e. including non-coded free text). 208 patients with genetically confirmed FH were included. The sensitivity (95% CI) on T1 and T2 with EHR-coded data for DLCN ≥ 6 was 19% (14–25%) and 22% (17–28%), respectively. When using all available data, the sensitivity for DLCN ≥ 6 was 26% (20–32%) on T1 and 28% (22–34%) on T2. For FAMCAT, the sensitivity with EHR-coded data on T1 was 74% (67–79%) and 32% (26–39%) on T2, whilst sensitivity with all available data was 81% on T1 (75–86%) and 45% (39–52%) on T2. For Make Early Diagnoses to Prevent Early Death MEDPED and SB, using all available data, the sensitivity on T1 was 31% (25–37%) and 17% (13–23%), respectively.

Conclusions The FAMCAT algorithm had significantly better sensitivity than DLCN, MEDPED, and SB. FAMCAT has the best potential for FH case-finding using EHRs.

* Corresponding author. Tel: +31 72 548 4444, Fax: +31 72 548 2167, Email: j.h.cornel@nwz.nl
†Drs. Bax and Cornel contributed equally to this article.

Graphical Abstract



Keywords Familial hypercholesterolaemia • Algorithms • Electronic health records • Big data

Introduction

Familial hypercholesterolaemia (FH) is an autosomal dominant genetic disorder of LDL cholesterol (LDL-C) clearance resulting in a 10-fold increased risk of coronary heart disease (CHD). The prevalence of FH is estimated at 1:250–1:500 people worldwide, but detection rates below 5% are common in many countries.[1–3] When detected and treated adequately early in life, the high risk of CHD can be reduced significantly, comparable with the risk in people without FH.[4]

Several clinical algorithms have been developed to facilitate detection of FH, such as the Dutch Lipid Clinic Network (DLCN) criteria, the US Make Early Diagnoses to Prevent Early Death (MEDPED) programme, Simon Broome (SB), and the familial hypercholesterolaemia case ascertainment tool (FAMCAT).[2,5–10] Other studies used prediction modelling or machine learning algorithms with good diagnostic accuracy, but widespread implementation is lacking.[9–15] Currently, there is insufficient evidence on the best strategy to identify potential FH patients in population-based screening.[16,17] The European Society of Cardiology (ESC) guidelines on dyslipidaemia recommend the use of the DLCN criteria, a clinical scoring mechanism based on LDL-C levels, physical examination and patient-, and family history.[18] These criteria result in a cumulative score for an individual patient. A score of ≥6 points in the DLCN algorithm is considered an indication for genetic testing to confirm FH.[2] Lower cut-off has been shown to increase the sensitivity of DLCN for FH but decrease specificity.[19]

To increase detection of FH, we previously developed an automated, electronic health record (EHR)-based algorithm using the DLCN criteria, with automated correction of LDL-C levels for use of lipid-lowering therapy (LLT).[20] In DLCN, LDL-C is a strong discriminatory factor in the clinical diagnosis of FH.[21] In the same manner, the use of LLT may play a role in decreased physician alertness and underdetection of FH when applying the DLCN algorithm in an individual patient. Indeed, the previously described EHR-based DLCN algorithm depends largely on correction for LLT to facilitate the detection of FH.[20] Many algorithms include clinical features that are pathognomonic of FH, such as the presence of tendon xanthoma in FH. By contrast, FAMCAT was developed from a population-based perspective and may therefore perform differently in an EHR-based environment with substantial numbers of patients. Indeed, algorithms may perform differently when diagnosing an individual FH patient in comparison to identification of potential FH cases in a large population.[22]

In the present study, we investigated the sensitivity of the DLCN-based algorithm in a genetically confirmed FH population and compared it to the sensitivity of other algorithms to improve the facilitation of FH detection in automated EHR-based algorithms. This was investigated both when using coded EHR-accessible data and also when adding all available data to study the algorithm's full potential (including non-coded free text and non-coded data on classic phenotypical characteristics like tendon xanthoma and arcus cornealis).

Methods

Selection of study population

All patients (>18 years) with a healthcare insurance-related coded diagnosis of 'primary dyslipidaemia' in 2018–2020 were selected. Patients

were seen at the outpatient clinic by a physician specialized in lipidology. Patient files were manually assessed for genetically confirmed FH. Other types of dyslipidaemia were excluded (e.g. dysbetalipoproteinaemia and familial combined hyperlipidaemia). Patient data were studied at two different time points to assess the possible influence of changing therapeutic options [statins, ezetimibe, proprotein convertase subtilisin-kexin type-9 (PCSK9) inhibitors] and guidelines over time. Timepoint one (T_1) was defined as as close to the genetic confirmation date of FH as was available. If the patient was <18 years at T_1, but >18 years at T_2, the patient was also included. The moment of the patient's first follow-up visit in 2018–2020 was defined as time point two (T_2). Patients were excluded when information on T_1 was incomplete to perform further analyses. When T_1 was in 2018–2020 (i.e. within the timespan of T_2), patients were also included.

Description of the electronic health record system

HiX is an EHR developed by ChipSoft, Amsterdam, the Netherlands, used by Northwest Clinics, Alkmaar, the Netherlands since 2018 and in approximately 60% of hospitals in the Netherlands. HiX is classified as a medical device, class IIb, certified for ISO 13485:2016 complying with the European Medical Device Directive 93/42/EEC MDD. HiX EHR contains data of patients from Northwest Clinics and of 465.000 people residing in the Northwest Clinics Laboratory adherence, most of whom are cared for by general practitioners in the area. The EHR combines demographic data, medical history (preferably coded since 2018), coded current and previous medication, clinical notes, a full array of automatically downloaded laboratory, radiology, pathology and microbiology data, financial and logistic administration, and more.

Description of data used for analysis

We calculated the sensitivity of the algorithm for DLCN ≥ 6 (i.e. indication for genetic testing) both when using HiX EHR-accessible coded data and also when adding all available data (including free text) to demonstrate the full potential of the algorithm. We performed the same analysis for the other algorithms [the optimized FAMCAT (FAMCAT2), MEDPED and SB] and compared the sensitivity to that of the DLCN algorithm.[9,10] We specifically assessed EHR-recorded prevalence of classic phenotypical FH characteristics in this population, like tendon xanthoma and arcus cornealis, items often relevant to FH diagnostic algorithms. Free text data included all free text information on patient- and family history and physical examination. Only sufficiently specific algorithm-related information was taken into account when available in records, i.e. specific details of FH-related physical examination and family LDL-C levels and the specific age at which premature CHD had occurred in family members. Since the gold standard for this analysis was genetically confirmed FH, the criterion of a positive genetic test was not included for both DLCN and SB. An overview of data used by each of the algorithms and whether the data are coded in the HiX EHR is shown in *Table 1*. The study passed the Northwest Clinics Scientific Board and approval of the local ethical committee/Institutional Review Board (Northwest Clinics, Alkmaar, The Netherlands) for the present study was waived since the study does not fall within the scope of the Dutch Medical Research Involving Human Subjects Act (section 1.b WMO, 26 February 1998).

Dutch Lipid Clinic Network

DLCN scores were calculated after correction for LLT at the two different time points. Data on family-, and patient history as needed for DLCN were only included if the information was available before or at the time of diagnosis to simulate real-life practice. Correction for LLT was performed based on original papers and specified in the supplementary materials (see Supplementary material online, *Table S1*).[20,21,23,24]

FAMCAT

The optimized FAMCAT algorithm (sometimes called FAMCAT2) was applied.[9,10] In FAMCAT, all information on family history was included only if the information was available before or at the time of diagnosis to simulate real-life practice. Medical history considering hypertension, chronic kidney disease and diabetes mellitus were extracted separately for T_1 and T_2.[25] FAMCAT was applied using a threshold of 0.0047 as was found to be optimal in a UK population.[8,26]

Make Early Diagnoses to Prevent Early Death

MEDPED uses total cholesterol (TC) values, adjusted to the presence of a positive family history for FH. Family history is non-coded information in HiX EHR, and this information is not often recorded in those who are not suspected of FH. Therefore, sensitivity was only calculated for MEDPED using all available data (coded and non-coded data).

Simon broome

The SB algorithm consists of two components. The first component consists of TC or LDL-C values. The second component is either the presence of tendon xanthoma (in the patient or a relative) or a positive family history of myocardial infarction or raised cholesterol. It is not possible to use the SB algorithm without one of these components. The second component is non-coded in the HiX EHR and therefore sensitivity was only calculated using all available coded and non-coded data (including free text). LDL-C values were corrected for LLT in the same way as described for the DLCN algorithm.

Statistical analysis

Continuous variables were displayed as mean with standard deviation if normally distributed and as median with interquartile range (IQR) if non-normally distributed. Q–Q plots and histograms were used to determine normal distribution. Logarithmic transformation was performed to create a normal distribution for non-normally distributed data. Categorical variables were shown as count with percentages (%). Differences between T_1 and T_2 were calculated using the paired *t*-test for comparison of means, McNemar's test for categorical variables and the Wilcoxon signed-rank test for ordinal related variables. The sensitivity was calculated for MEDPED, SB, FAMCAT and DLCN (DLCN ≥6 and lower cut-off values of ≥5 and ≥4) on T_1 and T_2 with corresponding 95% confidence intervals. Statistical analyses were performed with IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp. *Figure 2* was made using R (The R Foundation for Statistical Computing, version 4.1.3).

Results

Selection of study population

In total, 721 patients with a visit in 2018, 2019 or 2020 with the healthcare insurance-related coded diagnosis 'primary dyslipidaemia' were identified. Of these, 398 (55%) were excluded because genetic testing did not reveal an FH-associated mutation or because genetic testing had not been performed. In total, 66 patients (9%) were excluded because the EHR mentioned a genetically confirmed FH, but further information on the exact mutation was missing. Forty-nine patients (7%) were excluded due to unavailable data on either T_1 or T_2 (*Figure 1*).

Table 1 Data used in the Dutch Lipid Clinic Network criteria, familial hypercholesterolaemia case ascertainment tool, Make Early Diagnoses to Prevent Early Death, and Simon Broome algorithms and data coded in the ChipSoft HiX electronic health record

	Dutch Lipid Clinic Network	Familial hypercholesterolaemia case ascertainment tool	Make Early Diagnoses to Prevent Early Death	Simon Broome	HiX electronic health record
Sex	No	Yes	No	No	Yes
Age	No	Yes	Yes	No	Yes
Use of lipid-lowering therapy	Yes[d]	Yes	No	Yes[d]	Yes
Laboratory results					
Total cholesterol	No	Yes	Yes	Yes	Yes
LDL cholesterol[a]	Yes	Yes	No	Yes	Yes
Triglycerides	No	Yes	No	No	Yes
Physical examination					
Tendon xanthoma	Yes	No	No	Yes	No
Arcus cornealis <45 years	Yes	No	No	No	No
Past medical history					
Premature coronary heart disease[b]	Yes	Yes	No	No	Yes[e]
Premature cerebral or peripheral vascular disease[b]	Yes	No	No	No	Yes[e]
Chronic kidney disease	No	Yes	No	No	Yes[e]
Diabetes mellitus Type 1 or 2	No	Yes	No	No	Yes[e]
Family history of					
Premature coronary heart disease[b]	Yes	Yes	No	Yes	No
Premature vascular disease[b]	Yes	No	No	No	No
Raised cholesterol[c]	Yes	Yes	No	Yes	No
Tendon xanthoma	Yes	No	No	Yes	No
Arcus cornealis	Yes	No	No	No	No
Familial hypercholesterolaemia	No	Yes	Yes	No	No

[a]Different LDL cholesterol cut-off values were used between the Dutch Lipid Clinic Network criteria, the familial hypercholesterolaemia case ascertainment tool and Simon Broome.
[b]In the Dutch Lipid Clinic Network criteria and the familial hypercholesterolaemia case ascertainment tool premature was defined before the age of 55 for males and before 60 years for females.
[c]Raised cholesterol was defined in the Dutch Lipid Clinic Network criteria as a first-degree family member with LDL-C > 95th percentile. In the familial hypercholesterolaemia case ascertainment tool, raised cholesterol was not specified with cut-off values.
[d]LDL-C levels were corrected for lipid-lowering therapy, but medication use is not used in the original Dutch Lipid Clinic Network criteria or the Simon Broome algorithm.
[e]Coded in HiX since 2018.
SB = Simon Broome

Baseline characteristics

The remaining 208 patients with genetically confirmed FH were included for further analysis. Fifty-nine percent of the population was female. T_1 was similar to T_2 in 35 patients (17%). As specified in the inclusion criteria, all patients were >18 years at T_2, but 23 patients (11%) were <18 years at the time of diagnosis (T_1). The median time between T_2 and T_1 was 3768 days (IQR: 813–5679). Baseline characteristics for T_1 and T_2 are shown in *Table 2*. The median uncorrected LDL-C levels were 5.0 mmoL/L (IQR: 3.9–6.6) on T_1 and 2.8 mmoL/L (IQR: 2.2–4.4) on T_2. After correction for use of LLT, median LDL-C levels were 6.5 mmol/L (IQR: 5.3–8.0) on T_1 and 6.3 mmoL/L (IQR: 5.3–8.1) on T_2 ($P = 0.47$). Tendon xanthomas were mentioned in free text in the EHR in four patients (1.9%) and arcus cornealis in three patients (1.4%).

Sensitivity of the different algorithms

The sensitivity on T_1 and T_2 with only EHR-coded data for DLCN ≥ 6 was 19% (95% CI: 14–25%) and 22% (95% CI: 17–28%), respectively (*Figure 2*). When using all available data, the sensitivity for DLCN ≥ 6 was 26% on T_1 (95% CI: 20–32%) and 28% on T_2 (95% CI: 22–34%). When using all available data on T_1, decreasing the cut-off for DLCN to a cut-off ≥5 and ≥4 resulted in a sensitivity of 51% (95% CI: 44–57%) and 55% (95% CI: 49–62%), respectively. When using all available data on T_2, the sensitivity was 54% (95% CI: 47–61%) for DLCN ≥4 and 49% (95% CI: 32–56%) for DLCN ≥ 5.

The FAMCAT algorithm had a sensitivity of 74% (95% CI: 67–79%) on T_1 and 32% on T_2 (95% CI: 26–39%) with only EHR-coded data. The sensitivity with all available data for

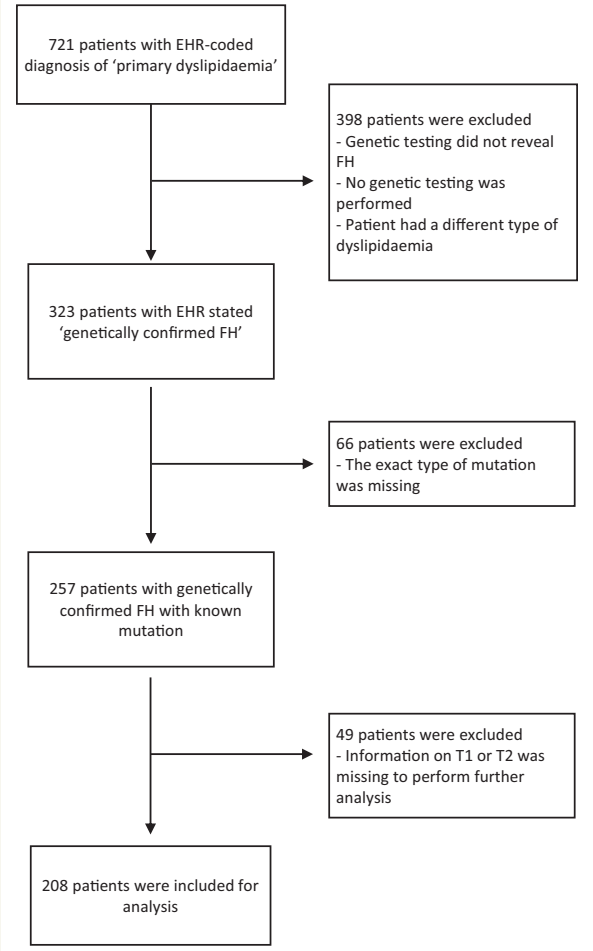


Figure 1 The selection process of the patients included in this study. EHR = electronic health record, FH = Familial hypercholesterolaemia.

FAMCAT was 81% on T_1 (95% CI: 75–86%) and 45% (95% CI: 39–52%) on T_2.

For MEDPED and SB, the sensitivity on T_1 was 31% (95% CI: 25–37%) and 17% (95% CI: 13–23%), respectively. Using all available data on T_2, sensitivity was 11% (95% CI: 7–15%) for MEDPED and 15% (95% CI: 11–21%) for SB. The sensitivity of FAMCAT on T_1 was higher than the sensitivity for DLCN ≥ 6 on T_1 and T_2 for both coded and with all available data and also when compared to the sensitivity of MEDPED and SB both on T_1 and T_2 (McNemar's test $P < 0.05$, *Figure 2*).

Discussion

The present data show that the sensitivity of an automated DLCN-based algorithm to facilitate detection of FH in electronic health records was relatively modest in a population with genetically confirmed FH. The FAMCAT algorithm was found to have significantly higher sensitivity than DLCN. Conclusions remained similar after the addition of all available data (including non-coded free text) to the EHR-coded data. Interestingly, EHR-recorded prevalence of classic phenotypical characteristics for FH, like tendon xanthoma and arcus cornealis, was low in this genetically confirmed FH population.

Previous validation of the DLCN algorithm in a population selected for FH phenotype (high LDL-C, TC and tendon xanthoma) observed a sensitivity for DLCN ≥ 6 of 67%, which may be different from the current study in which sensitivity for DLCN ≥ 6 was studied in a population with genetically proven FH.[19] A Western Australian study investigated the mutation spectrum of patients referred for genetic testing for FH. That study found a sensitivity of 29% for DLCN 6, 7, or 8 points (probable FH) and a sensitivity of 70% for DLCN > 8 points (definite FH) in 337 patients.[28] Recently, a study that validated FAMCAT and other algorithms in 260 genetically tested primary care patients of which 16 with an FH mutation, found a sensitivity for DLCN ≥ 6 of 35% with a specificity of 96%.[26] Importantly, high specificity would limit false-positive screening which may be particularly important when considering larger numbers of patients and cost associated with ensuing genetic confirmation. In the present study, decreasing the DLCN cut-off to lower values (≥5 and ≥4) increased sensitivity but has previously been shown to reduce specificity (i.e. increase false positives), possibly resulting in unnecessary genetic testing.[19]

FAMCAT was developed in a large population and uses 14 different variables, each with its coefficient resulting in a probability for FH.[9,10] An important difference from DLCN is a relatively lower reliance on recording family history, which may be difficult to extract automatically from an EHR, in particular in those individuals not suspected of FH. Although FAMCAT is less reliant on family history, discriminatory accuracy is improved with comprehensive family history.[8] In a validation study in genetically proven FH patients in the UK, FAMCAT showed a sensitivity of 69%, which was only slightly less than our findings for FAMCAT (with only coded data). In that study, specificity was 95%.[26] The present data may be influenced in part by a selection bias, created by including genetically confirmed FH patients only, where LDL-C, TC and triglycerides levels (three important variables in FAMCAT) are likely to be available more often than in the general population.

The SB algorithm had a sensitivity of 90% in the study of Damgaard *et al.*[19] Tendon xanthoma is one of the scoring criteria of SB and this was also an inclusion criterion for that study. In our study, the presence of tendon xanthoma was not often recorded: only four patients with tendon xanthoma were documented in the EHR. When incorporated in an algorithm aimed at automated EHR case-finding to facilitate FH recognition in patients not specifically suspected of FH, SB is therefore unlikely to be as sensitive as in the study by Damgaard *et al.*[19] In the study by Damgaard, MEDPED showed a sensitivity of 63–70% (depending on whether TC or LDL-C was used), which is much higher than in our study.[19] However, Damgaard *et al.* selected patients with a threshold TC or LDL-C value. MEDPED and SB are both quite dependent on family history. In MEDPED, family history is one of the three components used by the algorithm, and in SB it is not possible to determine 'possible FH' without information on family history. The often non-coded nature of family history in many electronic health record systems is therefore likely to result in limited value of the MEDPED and SB algorithms as an automated FH case-finding algorithm.

Table 2 Baseline characteristics

	T_1 ($n = 208$)	T_2 ($n = 208$)	*P*-value
Age (years), mean (SD)	39 (15)	49 (16)	<0.001
Sex			
Female, *n* (%)	122 (58.7)	122 (58.7)	—
Lipid levels			
Total cholesterol (mmol/L), median (IQR)	7.0 (5.6–8.7)	4.8 (4.0–6.2)	<0.001
LDL-C (mmol/L), median (IQR)	5.0 (3.9–6.6)	2.8 (2.2–4.4)	<0.001
LDL-C corrected for LLT (mmol/L), median (IQR)	6.5 (5.3–8.0)	6.3 (5.3–8.1)	0.47
Triglycerides (mmol/L), median (IQR)	1.1 (0.81–1.6)	1.1 (0.80–1.5)	0.11
HDL (mmol/L), median (IQR)	1.2 (1.1–1.6)	1.3 (1.1–1.6)	<0.001
Physical examination			
Tendon xanthoma[a], *n* (%)	4 (1.9)	—	—
Arcus cornealis[a], *n* (%)	3 (1.4)	—	—
Past medical history			
Premature coronary heart disease[a], *n* (%)	11 (5.3)	—	—
Premature cerebrovascular/peripheral disease[a], *n* (%)	4 (1.9)	—	—
Family history			
Premature coronary heart disease[a], *n* (%)	25 (12%)	—	—
First-degree relative with FH[a], *n* (%)	36 (17.3%)	—	—
Comorbidities			
Chronic kidney disease[b], *n* (%)	10 (4.8)	11 (5.3)	1.00
Diabetes mellitus[c], *n* (%)	3 (1.4)	10 (4.8)	0.02
Hypertension, *n* (%)	17 (8.2)	27 (13.0)	0.01
Medication use			
LLT use, *n* (%)	88 (42.3)	172 (82.7)	<0.001
Statin intensity[d]			<0.001
Low intensity, *n* (%)	5 (2.4)	1 (0.5)	—
Moderate intensity, *n* (%)	37 (17.8)	37 (17.8)	—
High intensity, *n* (%)	42 (20.2)	112 (53.8)	—
PCSK9 inhibitor usage, *n* (%)	2 (1.0)	32 (15.4)	<0.001
Ezetimibe, *n* (%)	13 (6.3)	109 (52.4)	<0.001
Type of FH mutation[a]			
LDL-receptor mutation, *n* (%)	177 (85)	—	—
ApoB mutation, *n* (%)	30 (14.5)	—	—
PCSK9 mutation, *n* (%)	1 (0.5)	—	—

[a]Variables were only extracted at T_1.
[b]Chronic kidney disease was defined according to the Kidney Disease: Improving Global Outcome (KDIGO) guidelines.[25]
[c]Diabetes mellitus was defined as the use or start of diabetes mellitus treatment.
[d]Statin intensity was classified by using the 2018 ACC/AHA Blood Cholesterol Guideline.[27]
Comparison for means and median (after logarithmic transformation) by paired *t*-test, for categorical variables McNemar's test, and for ordinal related variables, the Wilcoxon signed-rank test was used.
LLT = lipid-lowering therapy, PCSK9 = proprotein convertase subtilisin-kexin Type-9, FH = familial hypercholesterolaemia

This study illustrates that genotypical FH does not necessarily express as phenotypical FH. The average LDL-C levels were lower than perhaps expected and the recorded prevalence of classic phenotypical traits such as tendon xanthoma and arcus cornealis was <2%. Older studies estimate a prevalence ranging between 29% and 66% for tendon xanthoma.[29,30] This either implicates decreased recognition by physicians or that the prevalence is indeed lower, possibly due to earlier recognition and treatment nowadays. Even so, regardless of sometimes relatively low LDL-C levels, FH patients remain at an increased risk of CHD compared with someone without FH with the same LDL-C values.[31] This also illustrates that a screening strategy using only elevated LDL-C levels is likely to be less adequate in recognizing all at-risk FH patients. Indeed, the FAMCAT algorithm was found to perform better in this population-based automated FH-detection algorithm, possibly by combining 14 different generally available predictor variables.[10]

The strength of the present study is that we investigated algorithms in a large genetically confirmed FH population. We found in the current analysis that only a minority of the genetically confirmed FH population would have been recognized as DLCN ≥ 6 by the DLCN algorithm currently incorporated in the Chipsoft HiX EHR, which is available in 60% of hospitals in the Netherlands, emphasizing

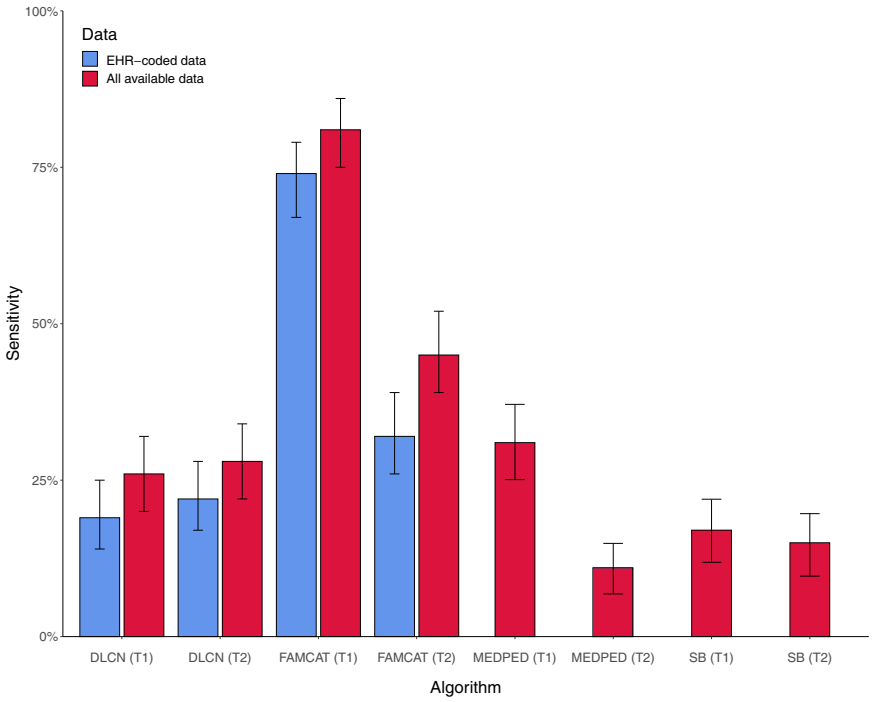


Figure 2 Sensitivity for all algorithms at T_1 and T_2. Sensitivity and 95% confidence intervals at T_1 and T_2 for Dutch Lipid Clinic Network ≥ 6 and familial hypercholesterolaemia case ascertainment tool (cut-off: 0.0047) for electronic health record-coded data in blue. Sensitivity and 95% confidence intervals for the Dutch Lipid Clinic Network criteria, the familial hypercholesterolaemia case ascertainment tool, Make Early Diagnoses to Prevent Early Death, and Simon Broome for all available data (i.e. coded and after also including non-coded-free text) are shown in red.

the need for improvement. Another strength is that, in the current study, data were analysed at two different time points to assess the possible influence of changing therapeutic options and guidelines over time. The observed lower sensitivity for FAMCAT on T_2, when compared with T_1, could have several reasons. First, this may be due to the relatively recent introduction and use of PCSK9 inhibitors. PCSK9 inhibitors are not incorporated in the FAMCAT algorithm.[8] Second, a different explanation could be the higher age of the population on T_2. The FAMCAT algorithm uses higher age as a negative predictor for the diagnosis for FH. Lastly, the FAMCAT algorithm is currently calibrated on a UK primary care population. The lower sensitivity on T_2 could also indicate that the FAMCAT algorithm is more useful for initial case-finding but not well enough calibrated to assess risk after specialist treatment is started (T_2). Therefore, although the FAMCAT algorithm performed best in the current setting, it may be optimized further.

Our study also has limitations. First, the present study focused on sensitivity whilst the specificity of algorithms is unknown. FAMCAT and DLCN showed high specificity in previous studies.[26] Second, some patients have typical phenotypical characteristics of FH, but genetic testing fails to detect a mutation, which could also relate to a yet unknown mutation of FH or polygenic hypercholesterolaemia. Detection of those patients, not included in the current study, may be equally important. Third, family history may be non-coded in an EHR, and other FH-related items may also be less accessible or available in EHRs, in particular in patients not suspected of FH. Furthermore, we must acknowledge that current findings apply to the use of an algorithm in case-finding in the general population as opposed to its use in an individual in whom algorithm performance to assist in the diagnosis of FH may be entirely different. Also, European General Data Protection Regulation (GDPR)-legislation prohibits the combination or use of patient data for purposes not entirely clear to the patient or its LDL-C ordering physician. The use of algorithms and its outcome or consequences should be known to the patient and the physician before an algorithm is applied, possibly limiting widespread use. An informative pop-up window when ordering LDL-C levels could assist in accommodating GDPR legislation. Lastly, the FAMCAT algorithm has not been recalibrated in a Dutch population and FAMCAT could have a different optimal cut-off value in a non-UK population. Therefore, further research should assess the discriminatory accuracy of the algorithms including both sensitivity and specificity across different populations. Perhaps most important, it is uncertain whether an automated algorithm results in actual subsequent genetic identification of FH in patients and their relatives and perhaps even in better outcome. This also requires further study.

EHRs contain large amount of data which, when combined and analysed adequately, can have important clinical consequences. Consistent coding is essential when EHR data is used for this purpose.[14,32] The use of big data analysis and machine learning are increasingly common in medical practice with promising results and are likely to increase the diagnostic yield in many clinical areas.[12,14,33–35]

In conclusion, current ESC guidelines recommend using the DLCN criteria for suspected cases, but the present data indicate

that FAMCAT performs better in an automated electronic health record-based algorithm, used to facilitate FH detection in large populations. FAMCAT also recognizes genotypical FH patients when classic but rare phenotypical characteristics are less accessible or available to the case-finding EHR algorithm. The performance of FAMCAT could be improved further by recalibrating the algorithm for populations with the use of PCSK9 inhibitors. Improvement of algorithms, embedded in Chipsoft HiX electronic health records, currently used in 60% of hospitals in the Netherlands, or in other electronic health records elsewhere, may result in improvement of FH identification.

Lead author biography

Professor Jan Hein Cornel (Harlingen, The Netherlands, 1960) obtained his MD at the University of Groningen, the Netherlands in 1986. Subsequently, he specialized in the field of cardiology, and has been working as a cardiologist in Northwest Clinics (Alkmaar, the Netherlands) since 1995. He obtained his PhD with a thesis about myocardial viability at the Erasmus University in Rotterdam, the Netherlands in 1996. In 2020, he was appointed as professor at Radboudumc (Nijmegen, the Netherlands), where he continues his research on mechanisms in development of cardiovascular disease.

Supplementary material

Supplementary material is available at *European Heart Journal – Digital Health* online.

Funding

Funded, in part, by a Northwest Clinics Foreest Academy grant.

Conflict of interest: None declared.

Data availability

De-identified data that underlie the results in this article (text, table, figures, and appendices) will be accessible for researchers when provided with a methodologically sound proposal. Data sharing will be available starting 12 months after publication. Requests can be made via e-mail: j.h.cornel@nwz.nl.

References

1. Hu P, Dharmayat KI, Stevens CAT, Sharabiani MTA, Jones RS, Watts GF, Genest J, Ray KK, Vallejo-Vaz AJ. Prevalence of familial hypercholesterolemia among the general population and patients with atherosclerotic cardiovascular disease: a systematic review and meta-analysis. *Circulation* 2020;**141**:1742–1759.
2. Nordestgaard BG, Chapman MJ, Humphries SE, Ginsberg HN, Masana L, Descamps OS, Wiklund O, Hegele RA, Raal FJ, Defesche JC, Wiegman A, Santos RD, Watts GF, Parhofer KG, Hovingh GK, Kovanen PT, Boileau C, Averna M, Borén J, Bruckert E, Catapano AL, Kuivenhoven JA, Pajukanta P, Ray K, Stalenhoef AFH, Stroes E, Taskinen M-R, Tybjærg-Hansen A for the 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**:3478–3490.
3. Akioyamen LE, Genest J, Shan SD, Reel RL, Albaum JM, Chu A, Tu JV. Estimating the prevalence of heterozygous familial hypercholesterolaemia: a systematic review and meta-analysis. *BMJ Open* 2017;**7**:e016461.
4. Luirink IK, Wiegman A, Kusters DM, Hof MH, Groothoff JW, de Groot E, Kastelein JJP, Hutten BA. 20-Year Follow-up of statins in children with familial hypercholesterolemia. *N Engl J Med* 2019;**381**:1547–1556.
5. WHO Human Genetics Programme. *Familial hypercholesterolaemia (FH): report of a second WHO consultation, Geneva, 4 September 1998*. 1999. Geneva: World Health Organization. https://apps.who.int/iris/handle/10665/66346
6. Williams RR, Hunt SC, Schumacher MC, Hegele RA, Leppert MF, Ludwig EH, Hopkins PN. Diagnosing heterozygous familial hypercholesterolemia using new practical criteria validated by molecular genetics. *Am J Cardiol* 1993;**72**:171–176.
7. Risk of fatal coronary heart disease in familial hypercholesterolaemia. Scientific steering committee on behalf of the simon broome register group. *BMJ* 1991;**303**:893–896.
8. Weng SF, Kai J, Andrew Neil H, Humphries SE, Qureshi N. Improving identification of familial hypercholesterolaemia in primary care: derivation and validation of the familial hypercholesterolaemia case ascertainment tool (FAMCAT). *Atherosclerosis* 2015;**238**:336–343.
9. Akyea RK, Qureshi N, Kai J, de Lusignan S, Sherlock J, McGee C, Weng S. Evaluating a clinical tool (FAMCAT) for identifying familial hypercholesterolaemia in primary care: a retrospective cohort study. *BJGP Open* 2020;**4**:bjgpopen20X101114.
10. Weng S, Kai J, Akyea R, Qureshi N. Detection of familial hypercholesterolaemia: external validation of the FAMCAT clinical case-finding algorithm to identify patients in primary care. *Lancet Public Health* 2019;**4**:e256–e264.
11. Besseling J, Reitsma JB, Gaudet D, Brisson D, Kastelein JJ, Hovingh GK, Hutten BA. Selection of individuals for genetic testing for familial hypercholesterolaemia: development and external validation of a prediction model for the presence of a mutation causing familial hypercholesterolaemia. *Eur Heart J* 2017;**38**:565–573.
12. Myers KD, Knowles JW, Staszak D, Shapiro MD, Howard W, Yadava M, Zuzick D, Williamson L, Shah NH, Banda JM, Leader J, Cromwell WC, Trautman E, Murray MF, Baum SJ, Myers S, Gidding SS, Wilemon K, Rader DJ. Precision screening for familial hypercholesterolaemia: a machine learning study applied to electronic health encounter data. *Lancet Digit Health* 2019;**1**:e393–e402.
13. Pereira AC. A roadmap for familial hypercholesterolaemia control. *Lancet Digit Health* 2019;**1**:e376–e377.
14. Akyea RK, Qureshi N, Kai J, Weng SF. Performance and clinical utility of supervised machine-learning approaches in detecting familial hypercholesterolaemia in primary care. *NPJ Digit Med* 2020;**3**:142.
15. Banda JM, Sarraju A, Abbasi F, Parizo J, Pariani M, Ison H, Briskin E, Wand H, Dubois S, Jung K, Myers SA, Rader DJ, Leader JB, Murray MF, Myers KD, Wilemon K, Shah NH, Knowles JW. Finding missed cases of familial hypercholesterolemia in health systems using machine learning. *NPJ Digit Med* 2019;**2**:23.
16. Qureshi N, Da Silva MLR, Abdul-Hamid H, Weng SF, Kai J, Leonardi-Bee J. Strategies for screening for familial hypercholesterolaemia in primary care and other community settings. *Cochrane Database Syst Rev* 2021;**10**:Cd012985.
17. Silva L, Qureshi N, Abdul-Hamid H, Weng S, Kai J, Leonardi-Bee J. Systematic identification of familial hypercholesterolaemia in primary care-A systematic review. *J Pers Med* 2021;**11**:302.
18. Mach F, Baigent C, Catapano AL, Koskinas KC, Casula M, Badimon L, Chapman MJ, De Backer GG, Delgado V, Ference BA, Graham IM, Halliday A, Landmesser U, Mihaylova B, Pedersen TR, Riccardi G, Richter DJ, Sabatine MS, Taskinen MR, Tokgozoglu L, Wiklund O. 2019 ESC/EAS guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk. *Eur Heart J* 2020;**41**:111–188.
19. Damgaard D, Larsen ML, Nissen PH, Jensen JM, Jensen HK, Soerensen VR, Jensen LG, Faergeman O. The relationship of molecular genetic to clinical diagnosis of familial hypercholesterolemia in a danish population. *Atherosclerosis* 2005;**180**:155–160.
20. Pepplinkhuizen S, Ibrahim S, Vink R, Groot B, Stroes ESG, Bax WA, Cornel JH. Electronic health records to facilitate continuous detection of familial hypercholesterolemia. *Atherosclerosis* 2020;**310**:83–87.
21. Haralambos K, Whatley SD, Edwards R, Gingell R, Townsend D, Ashfield-Watt P, Lansberg P, Datta DB, McDowell IF. Clinical experience of scoring criteria for familial hypercholesterolaemia (FH) genetic testing in Wales. *Atherosclerosis* 2015;**240**:190–196.
22. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthc J* 2019;**6**:94–98.
23. Schwartz GG, Steg PG, Szarek M, Bhatt DL, Bittner VA, Diaz R, Edelberg JM, Goodman SG, Hanotin C, Harrington RA, Jukema JW, Lecorps G, Mahaffey KW, Moryusef A, Pordy R, Quintero K, Roe MT, Sasiela WJ, Tamby J-F, Tricoci P, White HD, Zeiher AM. Alirocumab and cardiovascular outcomes after acute coronary syndrome. *N Engl J Med* 2018;**379**:2097–2107.
24. Sabatine MS, Giugliano RP, Keech AC, Honarpour N, Wiviott SD, Murphy SA, Kuder JF, Wang H, Liu T, Wasserman SM, Sever PS, Pedersen TR. Evolocumab and clinical outcomes in patients with cardiovascular disease. *N Engl J Med* 2017;**376**:1713–1722.

25. KDIGO. Clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney Int* 2013;**3**:1–150.
26. Qureshi N, Akyea RK, Dutton B, Leonardi-Bee J, Humphries SE, Weng S, Kai J. Comparing the performance of the novel FAMCAT algorithms and established case-finding criteria for familial hypercholesterolaemia in primary care. *Open Heart* 2021; **8**:e001752.
27. Grundy SM, Stone NJ, Bailey AL, Beam C, Birtcher KK, Blumenthal RS, Braun LT, Ferranti Sd, Faiella-Tommasino J, Forman DE, Goldberg R, Heidenreich PA, Hlatky MA, Jones DW, Lloyd-Jones D, Lopez-Pajares N, Ndumele CE, Orringer CE, Peralta CA, Saseen JJ, Smith SC, Sperling L, Virani SS, Yeboah J. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the management of blood cholesterol: a report of the American college of cardiology/American heart association task force on clinical practice guidelines. *Circulation* 2019;**139**:e1082–e1143.
28. Hooper AJ, Nguyen LT, Burnett JR, Bates TR, Bell DA, Redgrave TG, Watts GF, van Bockxmeer FM. Genetic analysis of familial hypercholesterolaemia in western Australia. *Atherosclerosis* 2012;**224**:430–434.
29. Descamps OS, Leysen X, Van Leuven F, Heller FR. The use of achilles tendon ultrasonography for the diagnosis of familial hypercholesterolemia. *Atherosclerosis* 2001; **157**:514–518.
30. Bertolini S, Cantafora A, Averna M, Cortese C, Motti C, Martini S, Pes G, Postiglione A, Stefanutti C, Blotta I, Pisciotta L, Rolleri M, Langheim S, Ghisellini M, Rabbone I, Calandra S. Clinical expression of familial hypercholesterolemia in clusters of mutations of the LDL receptor gene that cause a receptor-defective or receptor-negative phenotype. *Arterioscler Thromb Vasc Bio* 2000;**20**:e41–e52.
31. Khera AV, Won H-H, Peloso GM, Lawson KS, Bartz TM, Deng X, Leeuwen EMv, Natarajan P, Emdin CA, Bick AG, Morrison AC, Brody JA, Gupta N, Nomura A, Kessler T, Duga S, Bis JC, Duijn CMv, Cupples LA, Psaty B, Rader DJ, Danesh J, Schunkert H, McPherson R, Farrall M, Watkins H, Lander E, Wilson JG, Correa A, Boerwinkle E, Merlini PA, Ardissino D, Saleheen D, Gabriel S, Kathiresan S. Diagnostic yield and clinical utility of sequencing familial hypercholesterolemia genes in patients with severe hypercholesterolemia. *J Am Coll Cardiol* 2016;**67**:2578–2589.
32. Alonso V, Santos JV, Pinto M, Ferreira J, Lema I, Lopes F, Freitas A. Health records as the basis of clinical coding: is the quality adequate? A qualitative study of medical coders' perceptions. *Health Inf Manag* 2020;**49**:28–37.
33. Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. *N Engl J Med* 2016;**375**:1216–1219.
34. Olivera P, Danese S, Jay N, Natoli G, Peyrin-Biroulet L. Big data in IBD: a look into the future. *Nat Rev Gastroenterol Hepatol* 2019;**16**:312–321.
35. Deo RC. Machine learning in medicine. *Circulation* 2015;**132**:1920–1930.