Atherosclerosis 238 (2015) 336–343

Contents lists available at ScienceDirect

Atherosclerosis

journal homepage: www.elsevier.com/locate/atherosclerosis

Improving identification of familial hypercholesterolaemia in primary care: Derivation and validation of the familial hypercholesterolaemia case ascertainment tool (FAMCAT)

Stephen F. Weng [a], Joe Kai [a], H. Andrew Neil [b], Steve E. Humphries [c], Nadeem Qureshi [a, *]

[a] Division of Primary Care, School of Medicine, University of Nottingham, UK
[b] Wolfson College, University of Oxford, UK
[c] Centre for Cardiovascular Genetics, Institute of Cardiovascular Science, University College London, UK

ARTICLE INFO

Article history:
Received 19 September 2014
Received in revised form
18 November 2014
Accepted 9 December 2014
Available online 20 December 2014

Keywords:
Hypercholesterolaemia
Familial
Lipids
Primary care
Epidemiology

ABSTRACT

Objective: Heterozygous familial hypercholesterolaemia (FH) is a common autosomal dominant disorder. The vast majority of affected individuals remain undiagnosed, resulting in lost opportunities for preventing premature heart disease. Better use of routine primary care data offers an opportunity to enhance detection. We sought to develop a new predictive algorithm for improving identification of individuals in primary care who could be prioritised for further clinical assessment using established diagnostic criteria.

Methods: Data were analysed for 2,975,281 patients with total or LDL-cholesterol measurement from 1 Jan 1999 to 31 August 2013 using the Clinical Practice Research Datalink (CPRD). Included in this cohort study were 5050 documented cases of FH. Stepwise logistic regression was used to derive optimal multivariate prediction models. Model performance was assessed by its discriminatory accuracy (area under receiver operating curve [AUC]).

Results: The FH prediction model (FAMCAT), consisting of nine diagnostic variables, showed high discrimination (AUC 0.860, 95% CI 0.848–0.871) for distinguishing cases from non-cases. Sensitivity analysis demonstrated no significant drop in discrimination (AUC 0.858, 95% CI 0.845–0.869) after excluding secondary causes of hypercholesterolaemia. Removing family history variables reduced discrimination (AUC 0.820, 95% CI 0.807–0.834), while incorporating more comprehensive family history recording of myocardial infraction significantly improved discrimination (AUC 0.894, 95% CI 0.884–0.904).

Conclusion: This approach offers the opportunity to enhance detection of FH in primary care by identifying individuals with greatest probability of having the condition. Such cases can be prioritised for further clinical assessment, appropriate referral and treatment to prevent premature heart disease.

1. Introduction

Familial hypercholesterolaemia (FH) is the commonest autosomal dominant disorder, with between 1/200 to 1/500 individuals having the heterozygote form [1]. This genetic disorder is characterized by high serum cholesterol concentrations and is caused by mutations of the *LDLR* gene [1]. Without treatment, young adults aged 20 to 39 years with heterozygous FH are estimated to have nearly a 100-fold increase in mortality risk from CHD compared to unaffected adults [2,3]. Evidence indicates FH patients have up to a 37% reduction in CHD mortality following treatment with statins and improved life expectancy, emphasizing the major benefit of early identification and treatment [4]. If such patients are not recognized in primary care, they will be treated like other patients with common multifactorial causes for raised cholesterol and prescribed lower potency statins, or offered no medication at all if their global cardiovascular risk score is not elevated.

In the UK, the National Institute for Health and Care Excellence (NICE) recommends the Simon-Broome Register criteria [3] which includes cholesterol concentrations, clinical characteristics such as

* Corresponding author. Address: 1309 Tower, University Park, Nottingham NG7 2RD, UK.
E-mail address: nadeem.qureshi@nottingham.ac.uk (N. Qureshi).

http://dx.doi.org/10.1016/j.atherosclerosis.2014.12.034

presence of tendon xanthoma, and family history of premature heart disease and raised cholesterol. The Dutch Lipid Clinic criteria [5], recommended in Europe, is similar to the Simon-Broome criteria but also include arcus cornealis, peripheral vascular disease, and coronary artery disease. Despite established clinical assessment guidelines, the majority of FH cases remain undetected. Most European countries diagnose less than 20% of all estimated cases [1]. In the UK, only 12% to 15% of an estimated 120,000 heterozygous FH cases are diagnosed [6,7]. This may be due to several factors. Firstly, clinical assessment guidelines, developed from secondary care registers of FH patients, may have limited utility for the purposes of case-finding in primary care. For instance, in both Simon-Broome and Dutch Lipid Clinic criteria, a diagnosis of FH would require a comprehensive family history recording and genetic mutation testing. However, comprehensive family history (in particular age of onset of disease) and genetic testing are not routinely assessed or are undocumented in current primary care clinical systems [8,9]. Secondly, clinicians are underdiagnosing or under-recording recommended diagnostic characteristics such as tendon xanthoma, arcus cornealis and peripheral vascular disease [10–12]. To improve identification of FH in the general population, case-finding using a validated tool developed from data routinely available in primary care records may offer one potential approach. This could allow clinicians to prioritise individuals at greatest likelihood of having FH for further clinical assessment using established Simon-Broome or Dutch Lipid Clinical criteria. The aim of this study was to develop and validate a predictive modelling tool (FAMCAT) for identifying those patients with highest probability of FH in the general primary care population.

2. Methods

2.1. Data source

Data were obtained from the Clinical Practice Research Datalink (CPRD), a cohort of patients with prospectively collected data, derived from anonymized electronic medical records of more than 12 million patients from 681 UK general practices. CPRD records include demographic information, medical history, prescription details, clinical events, specialist referrals, hospital admissions, and laboratory test results. Approximately 8% of the UK population is currently included and the database is broadly representative of the UK population. Data undergo quality checks and practices are designated as meeting the CPRD quality criteria for research purposes and over 550 peer-reviewed studies using CPRD have been published [13]. Ethical approval was granted by the CPRD Independent Scientific Advisory Committee (protocol 13_167).

2.2. Study population

A cholesterol measurement is essential to establishing a suspected diagnosis of the FH. Thus, all included patients had at least one measurement of either total cholesterol or LDL-cholesterol between the (i) baseline date, 1 Jan 1999 or the earliest date which the CPRD practice started data after 1 Jan 1999 and (ii) end date, 31 August 2013 or the latest date the CPRD general practice finished contributing data prior to 31 August 2013. If follow-up was not completed, then the end date for the patient was specified as date of death, date of transfer, or date of final practice visit. For patients who were diagnosed with FH, the date of the diagnosis was specified as the end date. Patients aged less than 16 years were excluded as cholesterol thresholds for diagnosis and treatment of FH in children differ from adults [14]. Patients were also excluded if they had a prior FH diagnosis before the study entry date (1 Jan 1999).

2.3. Diagnostic variables

The diagnostic variables which were included in the analysis are recognized to be associated with FH (supplemental Box A). The variables of total cholesterol, LDL-cholesterol, family history of MI, and family history of raised cholesterol were based on Simon-Broome criteria [15] for identifying patients with possible FH. The cholesterol categories were assigned in line with *untreated* levels. If patients had both LDL-cholesterol and total cholesterol measured, the LDL-cholesterol was prioritized. If there were multiple cholesterol recordings, the highest cholesterol value was taken from each patient at any point between 1 Jan 1999 to 31 August 2013. Triglycerides (elevated levels are a negative indicator of FH [16]) were extracted for each patient during the time of the cholesterol measurement and categorized using established reference ranges [17] (elevated triglycerides classified as above 1.7 mmol/L). Cholesterol and triglyceride levels were checked for outlying observations ($\leq$0 mmol/L or >5 positive standard deviations [SD] from the mean) and data entry errors (non-numerical entries). Previous history of CHD <60 years may also result in a higher probability of being diagnosed with FH [5].

Family histories of MI and of raised cholesterol were included as potential diagnostic variables [3,14]. Although not explicitly included in previous criteria, family history of FH was also investigated. All the family history variables were dichotomized to either having a family history or not. If family history was not assessed, we assumed that there was no family history. Further categorization of family history to identify relative affected and age of onset of the condition was limited due to lack of recording [8].

Current diagnostic criteria [3,5,18] use untreated cholesterol levels to assess probable diagnosis of FH. However, individuals with raised cholesterols may be receiving lipid-lowering treatment. Hence, prescribing and potency of lipid-lowering treatment were included. If the most recent prescription ended within 30-days or overlapped with the date of the cholesterol measurement, the cholesterol level was considered treated. Otherwise, the cholesterol level was considered untreated. In categorizing patients as untreated, a 30-day washout period was utilized to account for any of the remaining effects of the lipid-lowering drugs when the drug treatment had been stopped [19]. The most recent recommendations for statin intensity in the UK NICE guidance for lipid modification [20] were utilized to classify potency based on a previous meta-analysis [21].

Current guidelines recommend that secondary causes of hypercholesterolaemia are negative indicators of FH [14,22]. Therefore, several key secondary conditions were analysed: liver disease (fatty liver disease, cirrhosis, chronic liver failure, alcoholic liver disease), diabetes mellitus (type I, type II), hypothyroidism (acquired and congenital), kidney disease (chronic kidney disease, renal impairment, acute renal failure), and nephrotic syndrome.

2.4. Outcome criteria

The case definition of FH was defined as all patients with a newly documented diagnosis of FH identified from patient records during the specified study period. Excluded from the analysis were patients with a previous diagnosis of other inherited lipid disorders. FH is specifically coded in UK general practice clinical systems by NHS Read Code “C3200”. An example of a typical UK general practice clinical system showing a diagnosis of FH can be found in supplemental Fig. B. This FH diagnosis must have occurred after the diagnostic variables, ensuring temporality between predictors and the outcome.

2.5. Statistical analysis

2.5.1. Derivation of the FAMCAT model

All of the analyses were completed in STATA 13 MP4. To develop the predictive model, a randomly selected 75% sample of the study population was used as the derivation cohort. As the outcome for diagnosis of FH was binary, stepwise logistic regression was used to develop a multivariate model for both men and women. Univariate analysis was first conducted to assess the association between each diagnostic variable and diagnosis of FH. A forward stepwise modelling approach was utilized to assess the impact of each additional predictor in the multivariate model. Predictors were included in the multivariate model the likelihood p-value was less than 5%. The final regression equation in the multivariate logistic models was used to determine the predicted probabilities of having FH.

2.5.2. Validation of the FAMCAT model

To validate the FAMCAT model, the predictive model was applied to all individuals within the validation cohort (remaining 25% random sample) to calculate each patients predicted probability of being identified with FH. Three comparison models were developed. The first comparator (Model 1) used only one diagnostic variable: total cholesterol > 7.5 mmol/L or LDL-cholesterol > 4.9 mmol/L. The second comparator (Model 2) was developed by incorporating variables indicative of a possible diagnosis of FH in the Simon-Broome criteria [3]. The third comparator (Model 3) was developed by adapting variables used in the Dutch Lipid Clinic [5] criteria. If specific variables of the Dutch Lipid Clinic criteria were not recorded in a patient's primary care record, then a null value was assumed. Discriminatory accuracy (*ability to distinguish between a case and non-case*) was assessed for all models by the area under the receiver operating curve (AUC) or Harrell's *c*-statistic; with higher values representing better discrimination. To generate confidence intervals for the *c*-statistic, a jack-knife procedure [23] was used to bootstrap standard errors.

2.5.3. Sensitivity analysis

In the first sensitivity analysis, we removed all secondary causes of raised cholesterol in the FAMCAT model and compared its discrimination to the primary analysis using the validation cohort. In the second sensitivity analysis using the validation cohort, we assessed the impact of improved family history recording of MI. To do this, we increased the proportion of positive family history cases for MI to 80.3% of FH cases and 9.3% of non-cases through random assignment of positive family histories to those who did not have a family history recording. These figures were based, firstly, on analysis of the Simon-Broome disease register which showed that 80.3% of all patients with possible FH had a positive family history of MI [3] (Neil HAW, personal communication, 25 Feb 2014), while the 9.3% figure was derived from previous analysis of medical coding for positive CHD family histories in primary care [8]. In the final sensitivity analysis, we evaluated the impact of family history in the FAMCAT model by excluding family histories of MI, FH, and raised cholesterol and compared its discrimination to the primary analysis using the validation cohort.

2.5.4. Calibration and reclassification

The FAMCAT model was assessed for calibration by comparing the proportion of predicted cases to the proportion of observed cases of FH in the validation cohort stratified by deciles (first decile representing the lowest probabilities) of predicted probabilities. In addition, we determined the impact of binary probability reclassification on sensitivity (*proportion of true positives identified*) and specificity (*proportion of true negatives identified*).

3. Results

3.1. Characteristics of the study population

There were 2,975,281 individuals in the CPRD study population from 1 Jan 1999 to 31 August 2013 with either a total cholesterol or LDL-cholesterol measurement during the study period. There were 3836 individuals (0.13%) in the total starting sample who were dropped from the analysis due to having outlying cholesterol measurements, data entry errors, or unspecified measurement units for the cholesterol readings. An additional 32 patients were excluded as the gender was not recorded. Thus, the complete cohort for the analysis comprised of 2,971,562 individuals. After excluding 246 patients with a diagnosis of FH before the 1 Jan 1999, 5050 cases with a documented diagnosis of FH were included in

Table 1
Clinical characteristics for men and women aged 16 or above in the derivation and validation cohorts. Values are means (standard deviations) unless stated otherwise.

Characteristics	Derivation cohort		Validation cohort	
	Men	Women	Men	Women
Total sample size	1,083,539 (48.6)	1,145,023 (51.4)	360,719 (48.6)	382,132 (51.4)
No (%) diagnosed with familial hypercholesterolaemia	1626 (0.2)	2152 (0.2)	535 (0.2)	737 (0.2)
Baseline age (years)	49 (15.9)	50 (17.4)	49 (15.8)	50 (17.4)
Age during cholesterol measurement (years)	56 (15.5)	58 (16.9)	56 (15.5)	58 (16.9)
No (%) with history of coronary heart disease <60 years	64,408 (5.9)	28,198 (2.5)	21,501 (5.9)	9481 (2.5)
Lipid Profile:				
Highest TC recorded (mmol/L)[a,b]	5.7 (1.21)	5.9 (1.3)	5.7 (1.2)	5.9 (1.3)
Highest LDL-cholesterol recorded (mmol/L)[c,d]	3.5 (1.0)	3.6 (1.1)	3.5 (1.0)	3.6 (1.1)
Triglycerides during cholesterol measurement (mmol/L)	1.9 (1.3)	1.5 (1.0)	1.9 (1.3)	1.5 (1.0)
Lipid-lowering drug usage during cholesterol measurement:				
No (%) prescribed fibrate, bile acid sequestrant, nicotinic acid	5712 (0.5)	4105 (0.4)	1863 (0.5)	1380 (0.4)
No (%) prescribed low potency statin[e]	21,064 (1.9)	16,735 (1.5)	6976 (1.9)	5651 (1.5)
No (%) prescribed medium potency statin[f]	70,161 (6.5)	55,154 (4.9)	23,488 (6.5)	18,469 (4.8)
No (%) prescribed high potency statin[g]	20,301 (1.9)	15,281 (1.3)	6905 (1.9)	5087 (1.3)
Family History:				
No (%) with family history of familial hypercholesterolaemia	5500 (0.5)	7485 (0.7)	1807 (0.5)	2547 (0.7)
No (%) with family history of raised cholesterol	3472 (0.3)	5324 (0.5)	1147 (0.3)	1727 (0.5)
No (%) with family history of myocardial infarction	34,493 (3.2)	37,103 (3.2)	11,520 (3.2)	12,340 (3.2)
Secondary causes of high cholesterol:				
No (%) diagnosed with liver disease	23,859 (2.2)	20,244 (1.8)	7956 (2.2)	6806 (1.8)
No (%) diagnosed with diabetes	157,070 (14.5)	128,695 (11.2)	52,270 (14.5)	43,120 (11.3)
No (%) diagnosed with hypothyroidism	29,939 (2.8)	115,114 (10.1)	10,028 (2.8)	38,528 (10.1)
No (%) diagnosed with kidney disease	111,817 (10.3)	149,641 (13.1)	37,386 (10.4)	50,052 (13.1)
No (%) diagnosed with nephrotic syndrome	1494 (0.1)	1101 (0.1)	475 (0.1)	396 (0.1)

[a] Median (Interquartile Range): Men = 5.6 (4.8–6.4); Women = 5.8 (5.0–6.7).
[b] Median (10th – 90th Percentile): Men = 5.6 (4.2–7.2); Women = 5.8 (4.3–7.5).
[c] Median (Interquartile Range): Men = 3.4 (2.7–4.1); Women = 3.5 (2.8–4.3).
[d] Median (10th – 90th Percentile): Men = 3.4 (2.2–4.8); Women = 3.5 (2.2–5.0).
[e] Fluvastatin or Pravastatin ≤ 40 mg/day; Simvastatin ≤ 10 mg/day.
[f] Fluvastatin or Pravastatin 80 mg/day; Simvastatin 20 mg/day or 40 mg/day; Atorvastatin ≤ 10 mg/day; Rosuvastatin 5 mg.
[g] Simvastatin 80 mg; Atorvastatin ≥ 20 mg/day; Rosuvastatin ≥ 10 mg/day.

Table 2
Multivariate logistic regression for mutually adjusted diagnostic variables of familial hypercholesterolaemia using the derivation cohort.

Diagnostic variables	Adjusted Odds Ratio [AOR] (95% confidence interval)	
	Men	Women
Highest TC or LDL recorded (mmol/L)		
Ideal (TC ≤ 5 OR LDL ≤ 3.3)	Ref	Ref
High (TC >5 to ≤6.5 OR LDL >3.3 to ≤4.1)	2.50 (2.03–3.08)	2.60 (2.13–3.18)
Very High (TC >6.5 to ≤7.5 OR LDL >4.1 to ≤4.9)	7.78 (6.28–9.64)	8.13 (6.61–9.99)
Extremely High (TC > 7.5 OR LDL > 4.9)	37.97 (30.99–46.52)	43.08 (35.43–52.40)
Age during cholesterol measurement (years)		
16–24	Ref	Ref
25–34	0.57 (0.42–0.76)	0.70 (0.54–0.92)
35–44	0.34 (0.26–0.45)	0.43 (0.33–0.55)
45–64	0.22 (0.17–0.29)	0.30 (0.23–0.38)
55–64	0.12 (0.09–0.17)	0.21 (0.16–0.26)
65–74	0.07 (0.05–0.10)	0.13 (0.10–0.17)
75–84	0.05 (0.03–0.08)	0.06 (0.05–0.09)
85 or above	0.05 (0.02–0.13)	0.04 (0.02–0.07)
Triglycerides during cholesterol measurement (mmol/L)		
Ideal (<1.7)	Ref	Ref
Borderline High (≥1.7 to <2.3)	0.94 (0.82–1.08)	0.96 (0.85–1.08)
High (≥2.3 to <5.6)	0.81 (0.71–0.92)	0.85 (0.76–0.96)
Very High (≥5.6)	0.72 (0.58–0.91)	0.58 (0.40–0.83)
Not Recorded	0.39 (0.32–0.50)	0.47 (0.39–0.57)
Lipid-lowering drug usage during cholesterol measurement		
No lipid-lowering drugs prescribed	Ref	Ref
Prescribed fibrate, bile acid sequestrant, or nicotinic acid	4.80 (3.39–6.79)	4.30 (2.98–6.18)
Prescribed low potency statins[a]	2.49 (1.72–3.58)	2.77 (2.05–3.74)
Prescribed medium potency statin[b]	4.47 (3.77–5.30)	3.51 (2.99–4.15)
Prescribed high potency statins[c]	10.64 (8.95–12.65)	6.31 (5.22–7.63)
Family history of familial hypercholesterolaemia		
No	Ref	Ref
Yes	10.99 (9.23–13.08)	8.21 (7.01–9.61)
Family history of myocardial infarction		
No	Ref	Ref
Yes	1.89 (1.58–2.27)	1.75 (1.49–2.06)
Family history of raised cholesterol		
No	Ref	Ref
Yes	3.22 (2.55–4.08)	3.23 (2.65–3.94)
Diagnosis of diabetes		
No	Ref	Ref
Yes	0.33 (0.27–0.41)	0.41 (0.34–0.49)
Diagnosis of kidney disease		
No	Ref	Ref
Yes	0.65 (0.51–0.84)	0.73 (0.62–0.85)

[a] Fluvastatin or Pravastatin ≤ 40 mg/day; Simvastatin ≤ 10 mg/day.
[b] Fluvastatin or Pravastatin 80 mg/day; Simvastatin 20 mg/day or 40 mg/day; Atorvastatin ≤ 10 mg/day; Rosuvastatin 5 mg.
[c] Simvastatin 80 mg; Atorvastatin ≥ 20 mg/day; Rosuvastatin ≥ 10 mg/day.

this cohort and used for analysis. To develop the FH probability model, 75% of the complete cohort (n = 2,228,562) was randomly sampled to become the derivation cohort, while the remaining 25% of the sample (n = 742,851) was assigned as the validation cohort.

Table 1 shows the descriptive characteristics of both the derivation and validation cohorts. Both cohorts showed similar frequencies, means, and medians for all clinical characteristics. In total, 10.8% of men and 8.1% of women were on lipid-lowering drugs at the time of the cholesterol measurement. The individuals who were on lipid-lowering drug treatment were mostly prescribed medium potency statins, with similar frequencies in both derivation and validation cohorts. The recording of a family history of FH and raised cholesterol was infrequent (0.3%–0.7%) although the recording of a family history of MI was higher in frequency (3.2%). Secondary causes of raised cholesterol showed similar frequencies in both cohorts.

3.2. Multivariate analysis

The multivariate predictive model was derived from the derivation cohort using the diagnostic indicator variables through stepwise logistic regression (univariate associations shown supplemental Tables C and D). Although personal history of premature MI is a recognised indicator of FH [5], the strength and significance of association excluded it from this algorithm. The optimal multivariate model retained nine diagnostic indicators as shown in Table 2. Total cholesterol > 7.5 mmol/L or LDL-cholesterol > 4.9 mmol/L were the strongest indicators of being diagnosed with FH (Males [M]: AOR 37.97, 95% CI 30.99–46.52; Females [F]: AOR 43.08, 95% CI 35.43–52.40). Having a family history of FH was strongly associated with diagnosis of FH (M: AOR 10.99, 95% CI 9.23–13.08; F: AOR 8.21, 95% CI 7.01–9.61). Having family histories of raised cholesterol (M: AOR 3.22, 95% CI 2.55–4.08; F: AOR 3.23, 95% CI 2.65–3.94) or MI (M: AOR 1.89, 95% CI 1.58–2.27; F: 1.75, 95% CI 1.49–2.06) were also significant predictors of FH. Another strong predictor of FH was being prescribed a high potency statin during the cholesterol assessment (M: AOR 10.64, 95% CI 8.95–12.65; F: AOR 6.31, 95% CI 5.22–7.63). For both males and females, age during the cholesterol measurement was indirectly associated with diagnosis. Elevated

Table 3
Model performance in the validation cohort for familial hypercholesterolaemia case identification.

Model	AUC *c*-statistic (95% confidence Interval)[a]
Primary Analysis	
Model 1: TC > 7.5 mmol/L or LDL-cholesterol > 4.9 mmol/L	0.556 (0.527–0.587)
[b]Model 2: TC > 7.5 mmol/L or LDL-cholesterol > 4.9 mmol/L + Family History MI	0.749 (0.735–0.763)
[c]Model 3: LDL categories + Family History + Clinical Assessment	0.737 (0.723–0.752)
Model 4: FAMCAT	0.860 (0.848–0.871)
Sensitivity Analysis	
FAMCAT excluding secondary disease causes[d]	0.858 (0.845–0.869)
FAMCAT with comprehensive family history of myocardial infarction[e]	0.894 (0.884–0.904)
FAMCAT excluding family history variables[f]	0.820 (0.807–0.834)

[a] Bootstrap standard errors using jack-knife procedure.
[b] Based on Simon-Broome criteria.
[c] Based on Dutch Lipid Clinic criteria.
[d] Excluded kidney disease and diabetes.
[e] Assumes 80.3% of familial hypercholesterolaemia cases and 9.3% of non-cases have positive family history of myocardial infarction.
[f] Excluded family history of myocardial infarction, raised cholesterol, and familial hypercholesterolaemia.

triglycerides ≥ 1.7 mmol/L were significantly less indicative of being diagnosed (M: AOR 0.72, 95% CI 0.58–0.91; F: AOR 0.58, 95% CI 0.40–0.83). Documented diagnoses of diabetes (M: AOR 0.33, 95% CI 0.27–0.41; F: AOR 0.41, 95% CI 0.34–0.49) or kidney disease (M: AOR 0.65, 95% CI 0.51 to 0.84; F: AOR 0.73, 95% CI 0.62–0.85) were also less indicative of a patient having FH.

3.3. Discrimination analysis

3.3.1. Primary analysis

Table 3 shows discriminatory accuracy according to the AUC *c*-statistic for each model. The FAMCAT model (Model 4) showed high model performance (AUC 0.860, 95% CI 0.848–0.871) with significantly improved discrimination when compared to Model 3, adapted from documented variables in the Dutch Lipid Clinic (AUC 0.737, 95% CI 0.723–0.752) and Model 2, adapted from variables in Simone-Broome (AUC 0.749, 95% CI 0.735–0.763). Model 1, which included only cholesterol showed poor discrimination (AUC 0.556, 95% CI 0.527–0.587). The improvement in discrimination for the FAMCAT compared to other models is shown in Fig. 1, a graph of the receiver operating curves.

As many of the variables present in the Dutch Lipid Clinic criteria could not be applied to the validation cohort, we also compared the FAMCAT to a simple gender and age adjusted log-linear LDL-cholesterol risk model. While this simple log-linear LDL model compared favourably to the model based on the Dutch Lipid Clinic criteria, it had poorer discrimination compared to the FAMCAT (Supplemental Fig. E).

3.3.2. Sensitivity analysis

In the first sensitivity analysis excluding secondary causes of raised cholesterol (diabetes and kidney disease) from the FAMCAT, there was no difference in discrimination (Fig. 2) from the primary analysis (AUC 0.858, 95% CI 0.845–0.869, Table 3). In the second sensitivity analysis, comprehensive family history recording of MI significantly improved discrimination (Fig. 2) compared to the primary analysis (AUC 0.894, 95% CI 0.884–0.904, Table 3). In the third sensitivity analysis, removing family histories of MI, FH, and raised cholesterol significantly decreased discrimination (Fig. 2) compared to the primary analysis (AUC 0.820, 95% CI 0.807–0.834, Table 3).

In the final sensitivity analysis, we found that expanding cholesterol categories to encompass an 'extremely high' total cholesterol threshold > 9.0 mmol/L or LDL-cholesterol > 6.5 mmol/L or incorporating total cholesterol and LDL-cholesterol as log-linear variables in the FAMCAT model resulted in no significant change in discrimination (Supplemental Fig. F).

3.4. Calibration and risk reclassification

The model showed accurate calibration across all deciles, with high levels of convergence between observed and predicted detection rates (supplemental Fig. G). There was also an expected sharp increase in observed and predicted cases in the highest decile of predicted probabilities where 829 cases were observed and 785 cases were predicted. To assess the impact of probability threshold reclassification, patients were stratified where the top decile of predicted probabilities defined as the 'high probability' and the remaining deciles defined as 'low probability'. The top decile corresponds to a predicted probability above 0.002, consistent with the estimated FH population prevalence of 1/500. Using this stratification, the FAMCAT achieved a sensitivity of 70% and a specificity of 88%.

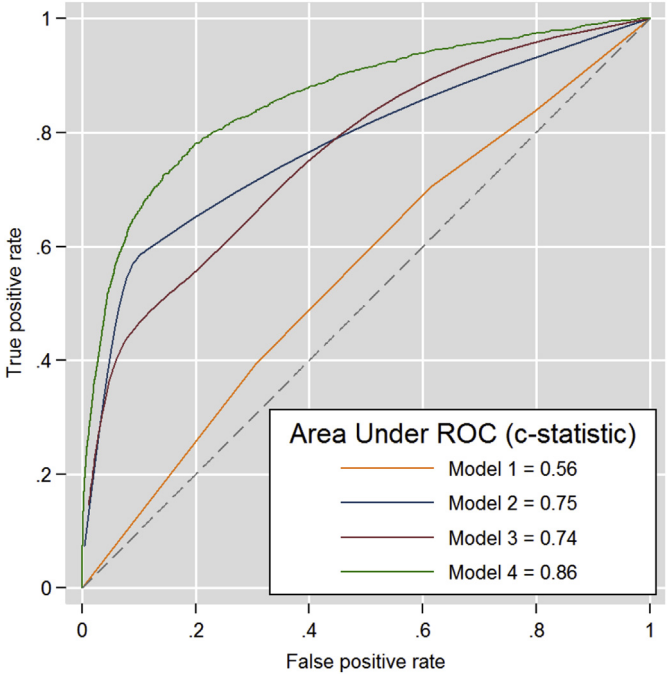

Fig. 1. Primary analysis: receiver operating curves derived from the validation cohort for FAMCAT variables in electronic health records compared to variables in established diagnostic criteria. Model 1: Total cholesterol > 7.5 mmol/L or LDL-cholesterol > 4.9 mmol/L Model 2: Total cholesterol > 7.5 mmol/L or LDL-cholesterol > 4.9 mmol/L + Family History MI (based on Simon-Broome) Model 3: LDL-cholesterol criteria + Family History + Clinical Assessment (based on Dutch Lipid Clinic) Model 4: FAMCAT model.

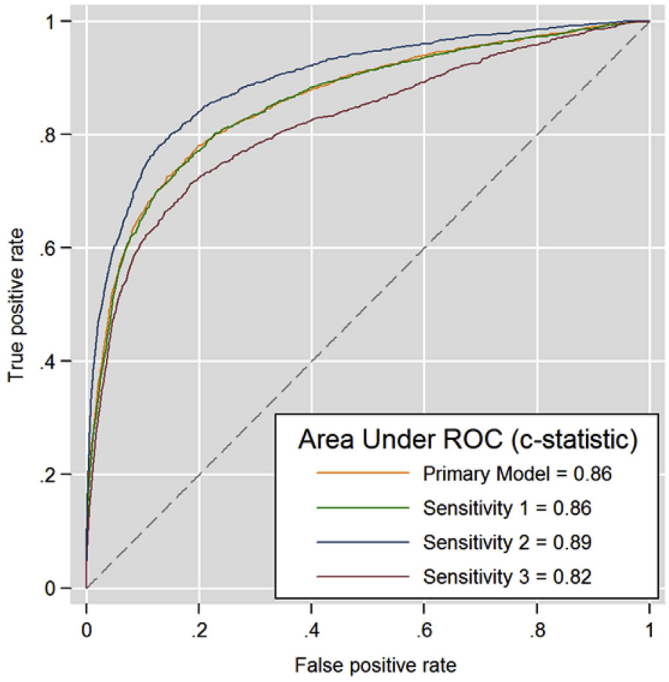

Fig. 2. Sensitivity analysis: receiver operating curves derived from the validation cohort for FAMCAT. Primary Model: FAMCAT model Sensitivity 1: FAMCAT model excluding secondary disease causes (kidney disease and diabetes) Sensitivity 2: FAMCAT model with complete recording of family history of myocardial infarction Sensitivity 3: FAMCAT model excluding all family history variables.

4. Discussion

We have derived and validated a new prediction tool for familial hypercholesterolaemia (FAMCAT), to enhance detection of individuals with possible FH. By identifying those in primary care with the highest probability of the condition, further clinical assessment of individuals can be proactively targeted, with referral for diagnosis and preventive care as appropriate, maximising efficient use of limited resources.

The majority of FH cases in the general population are not currently being diagnosed, and fail to benefit from preventive interventions to reduce their greatly elevated risk of premature CHD. FAMCAT offers the advantage of using coded variables routinely available in general practice electronic health records. In contrast, using all the variables in established diagnostic criteria for initial case ascertainment has poorer discriminatory accuracy because many of these variables are not recorded or are under-recorded in primary care. In addition, the importance of the family history was demonstrated, as removing family history variables (MI, FH, and raised cholesterol) from FAMCAT significantly reduced performance, whilst including more details on family history of MI improved prediction.

4.1. Clinical implications

In several countries there has been a drive to offer cardiovascular risk assessment to the adult population (UK NICE [22], European Society of Cardiology Joint Task Force [24]). In England, this has even been extended to a policy of offering universal cardiovascular screening to all adults aged 40 to 75. This expansion in cardiovascular screening is resulting in more individuals identified with cholesterol levels above the threshold at which further assessment for FH is recommended [1,3]. Assessing all these individuals for FH will lead to an inefficient use of limited primary care resources, with a large number of inappropriate referrals to specialist genetic and lipids services. On the other hand, not having any form of case-finding approach in primary care will continue to mean a large number of individuals with possible FH being missed.

An electronic health record database search in general practice, using the FAMCAT algorithm offers one possible solution to rationalise use of both primary care and specialist resources. A possible implementation pathway is illustrated in supplemental Fig. H. FAMCAT could be implemented as a toolkit, integrated within electronic health records, using routine primary care data extraction tools (e.g. PRIMIS CHART software [25]) to rank patients from highest to lowest probability of having FH.

Prior to this, a pre-defined threshold needs to be set for prioritising clinical assessment by General Practitioners to determine individuals at highest probability of having FH. If this threshold is unknown, we would suggest a risk-threshold of >1/500. The FAMCAT algorithm performs well at this threshold to identify individuals who should be assessed by primary care for possible FH using established diagnostic criteria. For instance, using this risk-threshold in an average English general practice of 6891 patients [26], nine of 13 estimated cases of FH will be identified. If this was implemented in all 8088 general practices in England, then 72,792 cases of possible FH will be identified. Patients who meet established diagnostic criteria could then be referred to secondary care, in line with current national guidelines [14]. Hence, we are recommending a stepwise approach which we anticipate will reduce the number of false-positive cases who will be referred to specialist services, compared to standard practice, due to the high sensitivity and specificity of the FAMCAT. Patients who have an identified mutation will provide the opportunity to instigate cascade screening in secondary care from confirmed cases, an approach that has been shown to be cost-effective [27,28]. The majority of patients in primary care would be successfully excluded from unnecessary referral, thereby saving healthcare resources. However, this is only one strategy for identifying FH cases. Patients with premature MI should ideally be assessed in secondary care for FH, but if overlooked, assessing such patients may come under the responsibility of primary care.

The study also highlights the importance of systematically collecting family histories in those individuals having cardiovascular risk assessment. We found family histories of MI, FH, or raised cholesterol were important predictive variables in the FAMCAT. Current recording and assessment of family history is known to be less than adequate in primary care [8]. However, by employing a more systematic approach, such as, asking patients attending primary care to complete a family history of CHD assessment instrument, higher proportions of individuals can be correctly identified [29]. This instrument could be a self-administered questionnaire or an online tool, completed in the practice or at home, with the patients forwarding the completed details to the responsible clinician, similar to the approach taken with familial cancer assessment [30].

4.2. Strengths and limitations

The strengths of this study include its longitudinal design, large sample size and use of a data-driven approach from a general primary care population. Although FAMCAT is calibrated for the UK population, the methodology and implementation pathway are transferable to other populations with health systems using electronic health records. Previous attempts to develop a primary care electronic search strategy in one primary care centre [31] did not achieve the necessary sensitivity and specificity. With the advantage of large numbers of diagnosed FH patients in our study, we have developed an electronic search algorithm that has achieved high levels of sensitivity and specificity for routine use at the population-level. One of the core strengths of the FAMCAT model is its ability to incorporate important interactions between key diagnostic indicators. For instance, interpretation of univariate analysis incorrectly suggests that elevated triglycerides are a positive predictor of FH. However, in multivariate analysis, elevated triglycerides are correctly shown to be a negative predictor of FH. FAMCAT takes account of the differences in triglycerides across cholesterol levels between cases and non-cases (see supplemental File I).

The principal limitation of this study is that coding of FH in general practice records is typically based on a clinical diagnosis. This will include FH cases that are genetic mutation negative individuals with polygenic hypercholesterolaemia [32]. The proportion of individuals with a diagnosis confirmed by genetic mutation testing, which varies widely across the Europe [1], or referred to specialists are unknown, as this information is not routinely documented in UK primary care electronic health records. However, there is a low risk of miscoding the clinical diagnosis of FH in the electronic health records, as there is a specific diagnostic code for FH in the medical coding hierarchy in primary care clinical systems. Moreover, the clinical utility of the FAMCAT lies in its ability to apply available coded data in primary care records to identify patients with a high probability of having FH, who would then warrant clinical assessment using established diagnosis criteria. Subsequent referral would occur for those individuals who meet established diagnostic criteria for assessment in specialist care, including genetic testing to confirm diagnosis. Until molecular genetic testing of identified subjects using the FAMCAT algorithm is carried out, the mutation detection rate of the identified individuals is unknown, but previous research has shown that the diagnostic cut-offs for total and LDL-cholesterol used by the FAMCAT has a

mutation detection rate (LDL-R and APOB) from 35 to 40% [33]. A further limitation concerns inadequate family history assessment and documentation of clinical characteristics such as arcus cornealis and tendon xanthoma [8,12]. Such data limitation reduces the clinical utility of case-finding models based on variables present in the Simon-Broome and Dutch Lipid Clinic criteria. For instance, variables from the Dutch Lipid Clinic criteria can only be applied to 1.3% (n = 10,002) of the validation cohort in our study, with only 2.7% (n = 34) of 1272 FH cases having all elements of the criteria extracted from primary care records. Whilst the sensitivity analysis reinforces the importance of comprehensive family history as a predictor of FH, using incomplete recorded data in current primary care records is a common and established approach for developing risk algorithms from primary care databases [34]. As the quality of family history recording in primary care records improves, this can be incorporated into the algorithm through recalibration.

Finally, we have investigated the potential of including secondary causes of hypercholesterolaemia in the algorithm. However, we note it is the *uncontrolled* state of these secondary conditions that leads to changes in cholesterol levels. Due to limited recording and lack of further laboratory investigations (such as HbA1c for diabetes control, glomerular filtration rate (GFR) for renal function) at the time of the cholesterol measurement, information could not be obtained to determine the control of secondary conditions. Despite this, our analysis suggests this additional information would only have a marginal impact on ascertaining possible FH.

5. Conclusions

The FAMCAT algorithm performs well in a database derived from routinely-collected data in primary care, offering significant clinical utility for improving identification of potential FH for further targeted assessment in this setting. Currently, we are working in partnership with a primary care audit software developer to integrate the algorithm into UK General Practice computer systems, supported by a user-friendly interface. Our future research will look to further refine the model by external validation using an FH disease register with genetically-confirmed FH cases as the primary reference standard. Furthermore, evaluation of the clinical utility and cost-effectiveness of incorporating FAMCAT into primary care clinical practice is now needed.

Sources of funding

This work was supported by the Division of Primary, University of Nottingham and National Institute for Health Research School of Primary Care Research (NIHR-SPCR).

Conflict of interest

SEH holds a Chair funded by the British Heart Foundation (PG08/008) and the NIHR University College London Hospitals Biomedical Research Centre and is the Medical Director of StoreGene Ltd. SEH, HAW, and NQ were members of the National Institute for Health and Care Excellence (NICE) familial hypercholesterolaemia guideline development group (CG71). The remaining authors have no conflicts of interests to disclose.

Acknowledgements

The authors would like to thank Dr Boliang Guo for statistical and methodology advice and Dr Dermot Neely for advice on the aetiology and management of hypercholesterolaemia.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.atherosclerosis.2014.12.034.

References

[1] B.G. Nordestgaard, M.J. Chapman, S.E. Humphries, 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. 34 (45) (2013) 3478–3490.
[2] Simon-Broome Register Group, Risk of fatal coronary heart disease in familial hypercholesterolaemia, BMJ 303 (6807) (1991) 893–896.
[3] Simon-Broome Register Group, Mortality in treated heterozygous familial hypercholesterolaemia: implications for clinical management, Atherosclerosis 142 (1) (1999) 105–112.
[4] A. Neil, J. Cooper, J. Betteridge, et al., Reductions in all-cause, cancer, and coronary mortality in statin-treated patients with heterozygous familial hypercholesterolaemia: a prospective registry study, Eur. Heart J. 29 (21) (2008) 2625–2633.
[5] Z. Reiner, A.L. Catapano, G. De Backer, et al., ESC/EAS guidelines for the management of dyslipidaemias: the task force for the management of dyslipidaemias of the European Society of Cardiology (ESC) and the European Atherosclerosis Society (EAS), Eur. Heart J. 32 (14) (2011) 1769–1818.
[6] R. Minhas, S.E. Humphries, N. Qureshi, H.A.W. Neil, Controversies in familial hypercholesterolaemia: recommendations of the NICE Guideline Development Group for the identification and management of familial hypercholesterolaemia, Heart Br. Card. Soc. 95 (7) (2009) 584–587.
[7] H.A. Neil, T. Hammond, R. Huxley, D.R. Matthews, S.E. Humphries, Extent of underdiagnosis of familial hypercholesterolaemia in routine practice: prospective registry study, BMJ 321 (7254) (2000) 148.
[8] P. Dhiman, J. Kai, L. Horsfall, K. Walters, N. Qureshi, Availability and quality of coronary heart disease family history in primary care medical records: implications for cardiovascular risk assessment, PLoS One 9 (1) (2014) e81998.
[9] M.A. Hoffman, The genome-enabled electronic medical record, J. Biomed. Informatics 40 (1) (2007) 44–46.
[10] T.R. Bates, J.R. Burnett, FMv Bockxmeer, S. Hamilton, L. Arnolda, G.F. Watts, Detection of familial hypercholesterolaemia: a major treatment gap in preventative cardiology, Heart Lung Circulation 17 (5) (2008) 411–413.
[11] A.T. Hirsch, M.H. Criqui, D. Treat-Jacobson, et al., Peripheral arterial disease detection, awareness, and treatment in primary care, JAMA 286 (11) (2001) 1317–1324.
[12] N. Qureshi, S.E. Humphries, M. Seed, P. Rowlands, R. Minhas, Identification and management of familial hypercholesterolaemia: what does it mean to primary care? Br. J. General Pract. 59 (567) (2009) 773–776.
[13] E. Herrett, S.L. Thomas, W.M. Schoonen, L. Smeeth, A.J. Hall, Validation and validity of diagnoses in the general practice research database: a systematic review, Br. J. Clin. Pharmacol. 69 (1) (2010) 4–14.
[14] National Institute for Health and Care Excellence, Identification and Management of Familial Hypercholesterolaemia, National Institute of Health and Care Excellence, London, 2008.
[15] D. Marks, M. Thorogood, H.A. Neil, S.E. Humphries, A review on the diagnosis, natural history, and treatment of familial hypercholesterolaemia, Atherosclerosis 168 (1) (2003) 1–14.
[16] G.D. Kolovou, P.M. Kostakou, K.K. Anagnostopoulou, Familial hypercholesterolemia and triglyceride metabolism, Int. J. Cardiol. 147 (3) (2011) 349–358.
[17] M. Miller, N.J. Stone, C. Ballantyne, et al., Triglycerides and cardiovascular disease: a scientific statement from the American Heart Association, Circulation 123 (20) (2011) 2292–2333.
[18] R.R. Williams, S.C. Hunt, M.C. Schumacher, et al., Diagnosing heterozygous familial hypercholesterolemia using new practical criteria validated by molecular genetics, Am. J. Cardiol. 72 (2) (1993) 171–176.
[19] N.J. Stone, Stopping statins, Circulation 110 (16) (2004) 2280–2282.
[20] National Institute for Health and Care Excellence, Draft for consultation: Lipid Modification – Cardiovascular Risk Assessment and the Modification of Blood Lipids for the Primary and Secondary Prevention of Cardiovascular Disease, National Institute of Health and Care Excellence, London, 2014.
[21] M.R. Law, N.J. Wald, A.R. Rudnicka, Quantifying effect of statins on low density lipoprotein cholesterol, ischaemic heart disease, and stroke: systematic review and meta-analysis, BMJ 326 (7404) (2003) 1423.
[22] National Institute for Health and Care Excellence, Cardiovascular Risk Assessment and the Modification of Blood Lipids for the Primary and Secondary Prevention of Cardiovascular Disease, National Institute for Health and Care Excellence, London, UK, 2008.
[23] R. Newson, Confidence intervals for rank statistics: Somers' D and extensions, Stata J. 6 (3) (2006) 309–334.
[24] The Fifth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice, European Guidelines on cardiovascular disease prevention in clinical practice, Eur. Heart J. 33 (13) (2012) 1635–1701.
[25] PRIMIS. Software Tools – CHART http://www.nottingham.ac.uk/primis/tools/chart/chart.aspx [accessed 10.08.14].

[26] Technical Steering Committee, General Practice Trends in the UK, Health and Social Care Information Centre, London, UK, 2013.
[27] P.J. Marang-van de Mheen, A.H.A. ten Asbroek, L. Bonneux, G.J. Bonsel, N.S. Klazinga, Cost-effectiveness of a family and DNA based screening programme on familial hypercholesterolaemia in The Netherlands, Eur. Heart J. 23 (24) (2002) 1922–1930.
[28] D. Marks, D. Wonderling, M. Thorogood, H. Lambert, S.E. Humphries, H.A.W. Neil, Cost effectiveness analysis of different approaches of screening for familial hypercholesterolaemia, BMJ 324 (7349) (2002) 1303.
[29] N. Qureshi, S. Armstrong, P. Dhiman, et al., Effect of adding systematic family history enquiry to cardiovascular disease risk assessment in primary care: a matched-pair, cluster randomized trial, Ann. Intern Med. 156 (4) (2012) 253–262.
[30] N. Qureshi, J.C. Carroll, B. Wilson, et al., The current state of cancer family history collection tools in primary care: a systematic review, Genet. Med. Official J. Am. Coll. Med. Genet. 11 (7) (2009) 495–506.
[31] J. Gray, A. Jaiyeola, M. Whiting, M. Modell, A.S. Wierzbicki, Identifying patients with familial hypercholesterolaemia in primary care: an informatics-based approach in one primary care centre, Heart Br. Card. Soc. 94 (6) (2008) 754–758.
[32] P.J. Talmud, S. Shah, R. Whittall, et al., Use of low-density lipoprotein cholesterol gene score to distinguish patients with polygenic and monogenic familial hypercholesterolaemia: a case-control study, Lancet 381 (9874) (2013) 1293–1301.
[33] M. Futema, R.A. Whittall, A. Kiley, et al., Analysis of the frequency and spectrum of mutations recognised to cause familial hypercholesterolaemia in routine clinical practice in a UK specialist hospital lipid clinic, Atherosclerosis 229 (1) (2013) 161–168.
[34] J. Hippisley-Cox, C. Coupland, P. Brindle, Derivation and validation of QStroke score for predicting risk of ischaemic stroke in primary care and comparison with other risk scores: a prospective open cohort study, BMJ 346 (2013) f2573.