Accepted Manuscript

BJGP OPEN

Identifying familial hypercholesterolaemia in primary care: Validation and optimisation of a clinical tool (FAMCAT)

Ralph Kwame Akyea, Nadeem Qureshi, Joe Kai, Simon de Lusignan, Julian Sherlock, Christopher McGee, Stephen Weng

DOI: https://doi.org/10.3399/bjgpopen20X101114

To access the most recent version of this article, please click the DOI URL in the line above.

Received 24 January 2020
Revised 07 February 2020
Accepted 10 February 2020

Identifying familial hypercholesterolaemia in primary care: Validation and optimisation of a clinical tool (FAMCAT)

Ralph K Akyea, MBChB MPH[1]; Nadeem Qureshi, MD[1]; Joe Kai, MD[1]; Simon de Lusignan, MD[2,3], Julian Sherlock[2], Christopher McGee[2,3], Stephen Weng, PhD[1]

Authors affiliations:

[1] Primary Care Stratified Medicine (PRISM), Division of Primary Care, University of Nottingham, Nottingham

[2] Department of Clincial and Experimental Medicine, University of Surrey, Guildford, UK

[3] Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC), London, UK

Correspondence to:

Dr Stephen Weng

Primary Care Stratified Medicine (PRISM), Division of Primary Care, Floor 13-15, The Tower Building, University Park Campus, University of Nottingham NG7 2RD UK

Stephen.Weng@nottingham.ac.uk

Word counts

Abstract: **203**

Main text: **2,694**

ABSTRACT

Background: Familial Hypercholesterolaemia (FH), an inherited lipid disorder causing premature heart disease, is severely underdiagnosed.

Aim: To evaluated the accuracy of a clinical tool (FAMCAT) for identifying FH in primary care.

Design and setting: Retrospective cohort study of 1,030,183 patients, from the UK Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC) database, aged over 16 years.

Method: The FAMCAT algorithm was compared to methods of FH detection recommended by national guidelines (Simon-Broome and Dutch Lipid Clinic Score diagnostic criteria and cholesterols levels >99th centile). Discrimination and calibration were assessed by area under the receiver operating curve (AUC) and comparing observed versus predicted cases.

Results: 1,707 patients had a diagnosis of FH. FAMCAT showed high levels of discrimination (AUC 0.844, 95% CI 0.834-0.854), performing significantly better than Simon-Broome criteria (AUC 0.730, 95% CI 0.719-0.741), Dutch Lipid Clinic Score (AUC 0.766, 95% CI 0.755-0.778), and screening cholesterols >99th centile (AUC 0.579, 95% CI 0.571-0.588). Inclusion of premature myocardial infarction and fitting cholesterol as a continuous variable improved the accuracy of FAMCAT (AUC 0.894, 95% CI, 0.885-0.903).

Conclusion: Better performance of the FAMCAT algorithm, compared to other approaches for case-finding of FH in primary care, has been confirmed in a separate population cohort.

Keywords: familial hypercholesterolaemia; case-finding; FAMCAT; validation; primary care; general practice

How this fits in

Many individuals with familial hypercholesterolaemia, an inherited lipid disorder, remain undiagnosed globally. This results in lost opportunities to identify and prevent many premature heart disease and premature death. This study evaluated the accuracy of a clinical tool (FAMCAT) in identifying FH in primary care.

In this study, FAMCAT has been confirmed to have a better predictive accuracy compared to other recommended approaches (Simone Broome criteria, Dutch Lipid Clinic Network Criteria and very elevated cholesterol alone) for case-finding FH in primary care.

BACKGROUND

Familial hypercholesterolaemia (FH) is a common inherited cause of raised cholesterol, affecting up to 320,000 adults in the UK and 834,000 adults in the US (1 in 250 prevalence for the adult general population).[1] Despite internationally recognised guidelines recommending clinicians actively identify individuals in primary care settings,[2–4] up to 80% of individuals with FH are still not identified,[3,5] leading to many avoidable heart attacks and early deaths. FH is a condition where preventive interventions to reduce premature cardiovascular disease, such as high intensity statins, are highly effective.[6,7]

Current approaches to clinically predict FH-causing mutation in primary care use the Simon-Broome diagnostic criteria (SB), Dutch Lipid Clinic Network criteria (DLCN), Make Early Diagnosis to Prevent Early Deaths (MEDPED), or total cholesterol over the 99th percentile (> 7.5 mmol/L under the age of 30; > 9.0 mmol/L over the age of 30).[2,8] The Simon Broome criteria,[2] most commonly used in the UK, recommend that individuals with a total cholesterol concentration of more than 7·5 mmol/L and a family history of premature heart disease should be classified as having probable familial hypercholesterolaemia in primary care and should be referred for further lipid specialist assessment. Patients who then also meet specific clinical diagnostic criteria (example, tendon xanthoma), or diagnosis by genetic testing, are categorised as having definite familial hypercholesterolaemia. The DLCN criteria[9] use a points-based scoring system to classify possible, probable, or definite familial hypercholesterolaemia on the basis of differing LDL cholesterol thresholds, family history of premature vascular disease and raised cholesterol, personal history of premature vascular disease, clinical signs such as tendon xanthoma and arcus senilis, or mutation status. MEDPED criteria uses age-stratified total-cholesterol thresholds for both the general population and relatives, depending on degree of relation.[10] Identifying patients that fulfil these criteria in primary medical care settings usually leads to further specialist assessment but may be inefficient given only around 25% of referred patients may be subsequently confirmed to have FH.[11,12] The National Institute for Health and Care Excellence (NICE) recommended assessment against Simon Broome criteria or Dutch Lipid Clinic Network (DLCN) criteria to make a clinical diagnosis of FH in primary care settings.[2]

A case-finding algorithm, FAMCAT (Familial Hypercholesterolaemia Case Ascertainment Tool), has been previously derived and validated using data from almost 3 million primary care patients (including over 5,000 cases of FH) from 681 primary care centers in the Clinical Practice Research Datalink (CPRD) database.[13] The algorithm had a high predictive

accuracy to identify patients with documented FH in primary care – with an area under the receiver operating curve (AUC) of 0.86.[13] Area under the receiver operating curve (AUC) is an overall measure of the ability of a test to discriminate whether a specific condition is present or not present.[14] AUC value lies between 0.5 to 1 where 0.5 denotes a poor accuracy and 1 denotes a perfect accuracy. This study aimed to externally validate the FAMCAT algorithm in the UK's Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC) database, which is a separate database from the CPRD database from which the algorithm was originally derived.

METHODS

Study design and population

Primary care data were extracted from the Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC) database in the UK (Supplementary Method S1). The RCGP RSC sentinel system is the principal primary care public health surveillance data used by Public Health England for the UK National Health Service.[15,16] We undertook a retrospective cohort study in a large population of primary care patients. This comprised a randomly selected sample of adult patients registered for primary medical care from 1st January 1999 followed up until 31st January, 2017, who had at least one documented total or low-density lipoprotein (LDL) cholesterol measurement (necessary for establishing a suspected diagnosis). The cohort comprised all patients who were actively registered and contributing data and had visited their family medical practice up until the end date of when data were extracted. For patients who were diagnosed with FH, the date of diagnosis was specified as their ending date to ensure all predictors remained temporal to their diagnosis.

Patients less than aged 16 years were excluded from the analysis as cholesterol thresholds for diagnosis and treatment of FH in children differ from adults.[2] Patients were also excluded if they had a prior FH diagnosis before the study entry date (1st January 1999) or a diagnosis of other inherited lipid disorders.

The starting time point for database interrogation was consistent with the start date used when deriving the FAMCAT algorithm (Supplementary Table S1) using CPRD.[13]

Outcome

The primary outcome was defined as the incident diagnosis of FH, identified from a patient record, between 1st January 1999 and 31st January 2017. FH is specifically coded in UK primary electronic health records (EHRs) using the internationally recognised Read coding system. This diagnostic code is entered into primary care electronic records after specialist lipid assessment, based on clinical phenotype, and/or by genetic test.

Predictor variables

FAMCAT was developed as a multivariate logistic regression model, stratified by gender, to calculate an individual's probability of having FH.[13] Box 1 summarises all 10 predictors which were incorporated into FAMCAT. Age, cholesterol levels, and triglycerides were categorised. Statin potency was determined using classifications based on publication by Law et al,[17] incorporated in the most recent UK NICE Lipid Modification Guidelines (Supplementary Table S2).[18] Secondary causes of raised cholesterol, such as diabetes and chronic kidney disease, were included as predictor variables for lower probability of FH.

Validation of the FAMCAT algorithm with comparator models

The FAMCAT logistic regression equation developed in the CPRD database was applied directly to every patient in the cohort to calculate each patient's probability of having FH. This was done by applying the untransformed regression coefficients and constant term provided in Supplemetary Table S1. Descriptive characteristics of study population were provided: patient demographics and clinical characteristics. Patients with no data record for any clinical variables such as diabetes, chronic kidney disease, and prescribing of statins were considered either to not have the condition or not been prescribed the drug.

Performance of the risk prediction models was assessed by discrimination and calibration [19]. Specifically, discriminatory accuracy was assessed for all three models using the area under the receiver operating curve (AUC) or Harrell's *c*-statistics; with higher values representing better discrimination. To generate confidence intervals for the *c*-statistics, a jack-knife procedure[20] was used to estimate standard errors. We also compared the discrimination of FAMCAT against Simon-Broome diagnostic criteria,[2] Dutch Lipid Clinic Network Score,[9] and a simple classification of total cholesterol above 99th centile [2] (a new recommendation made by NICE guideline committee in the latest 2017 update) for

determining possible FH. Predictors included in the Simon-Broome and Dutch Lipid Clinic criteria were extracted using Read clinical classification codes and applied directly to the cohort.

Calibration was defined as how closely the predicted probability of FH agrees with the expected probability of FH. This was assessed by plotting the observed number of cases of FH against the expected number of cases of FH for each tenth of predicted probability to ensure 10 equally sized groups.[21]

Optimisation of the FAMCAT algorithm

To develop an optimised FAMCAT algorithm, the 10 predictors in the FAMCAT algorithm developed from the CPRD database were considered as *a priori* predictors. History of premature atherosclerotic CVD such as coronary heart disease and peripheral vascular disease (PVD) have been shown to be significantly associated with FH.[22] These conditions related to FH were explored as potential predictors hence included in the model and discriminatory performance of the model assessed using AUC. These new predictors included a personal history of premature myocardial infarction and history of peripheral vascular disease. Cholesterol level was built-in as a continuous variable for these optimised models, with an interaction term to specify whether the measurement was a total or LDL-cholesterol.

The study findings are reported in accordance with the Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) recommendations (Supplementary Table S3).

Patient Involvement

Involvement of patients and relevant advocate groups at all stages of our previous and current related research projects has proved invaluable in helping to further focus study design, output and dissemination on the needs of the public and the benefits that can be delivered for the community. FH patient representatives for this research project attended study steering meetings to advise on study conception and preparation, funding application, review of study protocols, and have contributed to interpretation, presentation and dissemination of the findings.

RESULTS

Study Population

From the 1,031,411 patients identified from RCGP database, 1,228 patients were excluded due to having other inherited lipid disorder or having all of their cholesterol measurements documented after a diagnosis of FH. The cohort of patients included in the analysis comprised 1,030,183 (52.1% female) eligible patients from 1 January 1999 to 1st September 2017. There were 649 men (0.13%) diagnosed with FH compared to 1,058 women (0.2%). The baseline age of the cohort was 56 years (SD, 15.3) for men and 57 years (SD, 16.7) for women. The mean highest total cholesterol was slightly higher in women at 5.8 mmol/L (SD 1.3) than in men (5.6 mmol/L; SD 1.2). Table 1 shows the full details of baseline characteristics for the entire cohort.

External validation

Discrimination

Table 2 shows the discrimination of FAMCAT algorithm compared to other clinical criteria. External validation of FAMCAT model in RCGP RSC database showed high level of discrimination (AUC 0.844, 95% CI 0.834 to 0.854). The performance of FAMCAT showed significantly better discrimination compared to Simon-Broome criteria (AUC 0.730, 95% CI 0.719 to 0.741) and Dutch Lipid Clinic Score (AUC 0.766, 95% CI 0.755 to 0.778). Figure 1 shows the receiver operating characteristics curves of the various models.

Calibration

The model showed good calibration across all deciles between observed and predicted cases, with slight under prediction of cases in the highest two deciles (Figure 2). There was an expected sharp increase in observed and predicted cases in the highest deciles of predicted probability where 414 cases were observed and 344.8 cases were predicted for the 9th decile, and 922 cases observed and 855.3 predicted for the 10th decile.

Sensitivity and specificity

A threshold corresponding to the top decile (10th) of predicted probability was used for case-finding in primary care setting[13] – a probability cut-off of $^{1}/_{250}$ or 0.004, the estimated prevalence of FH.[3] Using this cut-off, FAMCAT achieved a sensitivity of 77.5% (95% CI, 75.4% to 79.5%) and specificity of 81.1% (95% CI, 81.0% to 81.2%) with a corresponding

positive predictive value of 0.68% (95% CI, 0.64% to 0.71%) and a negative predictive value of 100%.

Optimised FAMCAT models

To optimise the FAMCAT model, cholesterol level was fitted as a continuous variable. Predictors considered to be related to FH (that is, personal history of premature MI and personal history of peripheral vascular disease) were included, with the risk factors/variables from FAMCAT algorithm serving as *a priori* predictors.

Fitting cholesterol as a continuous variable and including of personal history of premature MI and personal history of peripheral vascular disease, increased model discrimination by 5% (AUC 0.894, 95% CI 0.885 to 0.903) when compared to the validation model in the RCGP cohort. The optimised model showed good calibration across all deciles between observed and predicted cases (Figure 3). There was an expected sharp increase in observed and predicted cases in the highest decile of predicted probability where 1285 cases were observed and 1100 cases were predicted.

Using the same threshold corresponding to the top decile (10^{th}) of predicted probability – a probability cut-off of $^{1}/_{250}$ or 0.004, the optimised FAMCAT model achieved the following: a sensitivity of 69.4% (95% CI, 67.2% to 71.6%) and specificity of 92.8% (92.8% to 92.9%) with a corresponding positive predictive value of 1.58% (95% CI, 1.49% to 1.67%) and a negative predictive value of 100%. The optimised FAMCAT model improved specificity by 14.4% from the standard FAMCAT model.

DISCUSSION

Summary

In this study, the FAMCAT algorithm has been validated in a separate cohort of over a million patients and has maintained high discriminatory accuracy. This algorithm also showed superior performance compared to recomended approaches in UK guidelines for case-finding. We have also demonstrated that the predictive accuracy of the FAMCAT algorithm can be further improved by incorporating personal history of premature myocardial

infarction and peripheral vascular disease and fitting cholesterol levels as a continuous variable.

Strengths and limitations

Our study has a number of strengths, especially the large, population-based sample and a long duration of follow-up, to validate and optimise an algorithm which identifies patients with the highest probability of existing FH. The RCGP RSC data is is nationally representative of the UK primary care patient population, and developed as a national disease and morbidity surveillance network. Given this purpose, disease coding and clinical measurements are better captured compared to other sources.[16] For instance, the proportion of patients with family history of MI recorded (5.7% in men; 6.7% in women) is higher compared to the UK's CPRD database (3.2% for both men and women).[13]

We acknowledge our study has limitations. The diagnosis of FH in the patient's electronic health records is based on the clinical phenotype, specifically those meeting clinical diagnostic criteria following specialist lipid assessment, which may or may not be confirmed by genetic testing. However, management of these patients to improve cardiovascular risk, will nevertheless be based on clinical phenotype. In UK national guidelines, the key role for genetic testing is to activate cascading testing to identify affected relatives by specialist care. The diagnosis is based on coded records rather than following an adjudication process which would not be feasible in such a large cohort of patients. The use of unadjudicated diagnosis coded in records by clinicians, is widely adopted in major clinical epidemiological research.[23,24].

Comparison with existing literature

FAMCAT is the first FH identification algorithm developed for use in primary care setting. Other tools, developed to improve identification of FH in primary care, have incorporated DLCN criteria. This includes, tool developed in the SEARCH Study [25], TARB-EX based on DLCN and correction for LDL-C [26], and the Caning Tool, an electronic extraction tool designed for primary care EHRs based on DLCN [27]. Our previous study [28] and current study show FAMCAT has significantly better predictive accuracy for clinical case-finding than any of these approaches including MEDPED, in very large primary care populations. The higher performance compared to recommended approaches is due to it being developed

directly from primary care EHRs. The nature of recording in routine EHRs has its limitations hence the application of very specific DLCN criteria developed outside primary care setting may not capture the distinict characteristics of individuals who may be at risk. Also, the use of blunt categorisations such as TC and family history, or lipid levels alone, capture too many individuals who do not have FH.

Implications for research and/or practice

The current, and previous,[28] external validations of FAMCAT in most general practice systems now show it can be confidently applied across UK primary care to identify people with possible FH. As with all available approaches, FAMCAT will not identify everyone with FH in the general population. Rather, it offers an accurate and practical approach to case-find those patients most likely to have FH, so they can be referred for specialist assessment and definitive genetic diagnosis (or its exclusion). Other methods such as child-parent screening or cascade testing in secondary care could further improve identification of FH. Our further research is exploring using machine-learning (ML) to identify FH in primary care; alongside similar work using secondary care data.[29]

For clinical practice, the FAMCAT algorithm has been integrated into some GP computer systems as an automated case-finding tool – https://www.nottingham.ac.uk/primis/tools/qi-tools/familial-hypercholesterolaemia.aspx. Although this is available for UK practice, the FAMCAT variables are all routinely recorded so the tool could be developed for wider use internationally. A web-based FAMCAT online risk calculator is also now available: https://prism-uon.shinyapps.io/FAMCAT/.

In conclusion, this study confirms FAMCAT performs better than other recommended approaches to case finding for FH using SB or DLCN criteria or very high cholesterol levels.. Use of FAMCAT in general practice will identify those patients with possible FH most likely to need referral for specialist diagnosis, and greater intervention to reduce risk of premature heart disease.

ADDITIONAL INFORMATION

Funding

The funding for this research was received from the National Institute for Health Research (NIHR) School for Primary Care Research (SPCR) (Project reference FR13-361). The views expressed are those of the authors and not necessarily those of the NIHR, the NHS, or the Department of Health and Social Care. Public Health England is the principal funder of RCGP RSC, for its surveillance activities.

Ethical approval

Patients and practices are informed of this study and the option available to them to 'opt-out' of sharing data. All current research activities using pseudonymised data from the RCGP RSC network of general practices are listed on the RCGP RSC webpage (http://www.rcgp.org.uk/rsc) and practices are informed via the monthly newsletter.

Competing interest

NQ is a member of the most recent NICE Familial Hypercholesterolaemia & Lipid Modification Guideline Development Groups (CG71 & CG181). SW is a member of the Clinical Practice Research Datalink (CPRD) Independent Scientific Advisory Committee (ISAC), academic advisor to Quealth Ltd., and has received independent research grant funding from AMGEN. RKA currently holds an NIHR-SPCR funded studentship (2018-2021). The remaining authors have no competing interests.

Acknowledgements

We are grateful to Mr Phil Rowlands for his continued involvement and invaluable contributions to our on-going programme of research as a patient representative.

Patients for allowing their data to be used by RCGP RSC for surveillance and research. Practices who have agreed to be part of the RCGP RSC network and allow us to extract and used health data for surveillance and research. John Williams (Senior Clinical Research Fellow) and other members of the Clinical Informatics and Health Outcomes Research Group at University of Surrey. Apollo Medical Systems for data extraction. Collaboration with EMIS, TPP, and Micro-test CMR supplier for facilitating data extraction.

REFERENCES

1 de Ferranti SD, Rodday AM, Mendelson MM, *et al.* Prevalence of Familial Hypercholesterolemia in the 1999 to 2012 United States National Health and Nutrition Examination Surveys (NHANES). *Circulation* 2016;**133**:1067–72. doi:10.1161/CIRCULATIONAHA.115.018791

2 National Institute of Health and Care Excellence. Familial hypercholesterolaemia: identification and management. London: 2017.

3 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**:3478–90. doi:10.1093/eurheartj/eht273

4 Watts GF, Sullivan DR, Poplawski N, *et al.* Familial hypercholesterolaemia: A model of care for Australasia. *Atheroscler Suppl* 2011;**12**:221–63. doi:10.1016/j.atherosclerosissup.2011.06.001

5 Marks D, Thorogood M, Farrer JM, *et al.* Census of clinics providing specialist lipid services in the United Kingdom. *J Public Health (Bangkok)* 2004;**26**:353–4. doi:10.1093/pubmed/fdh176

6 Smilde TJ, van Wissen S, Wollersheim H, *et al.* Effect of aggressive versus conventional lipid lowering on atherosclerosis progression in familial hypercholesterolaemia (ASAP): a prospective, randomised, double-blind trial. *Lancet* 2001;**357**:577–81.

7 Alonso R, Mata P, De Andres R, *et al.* Sustained long-term improvement of arterial endothelial function in heterozygous familial hypercholesterolemia patients treated with simvastatin. *Atherosclerosis* 2001;**157**:423–9.

8 Chan DC, Pang J, Hooper AJ, *et al.* A Comparative Analysis of Phenotypic Predictors of Mutations in Familial Hypercholesterolemia. *J Clin Endocrinol Metab* 2018;**103**:1704–14. doi:10.1210/jc.2017-02622

9 Reiner Z, Catapano AL, De Backer G, *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* 2011;**32**:1769–818. doi:10.1093/eurheartj/ehr158

10 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**:171–6. doi:10.1016/0002-9149(93)90155-6

11 Qureshi N, Weng S, Tranter J, *et al.* Feasibility of improving identification of familial hypercholesterolaemia in general practice: intervention development study. *BMJ Open* 2016;**6**:e011734. doi:10.1136/bmjopen-2016-011734

12 Taylor A, Wang D, Patel K, *et al.* Mutation detection rate and spectrum in familial hypercholesterolaemia patients in the UK pilot cascade project. *Clin Genet* 2010;**77**:572–80. doi:10.1111/j.1399-0004.2009.01356.x

13 Weng SF, Kai J, Andrew Neil H, *et al.* Improving identification of familial hypercholesterolaemia in primary care: Derivation and validation of the familial hypercholesterolaemia case ascertainment tool (FAMCAT). *Atherosclerosis* 2015;**238**:336–43. doi:10.1016/j.atherosclerosis.2014.12.034

14 Hoo ZH, Candlish J, Teare D. What is an ROC curve? *Emerg Med J* 2017;**34**:357–9. doi:10.1136/emermed-2017-206735

15 de Lusignan S, Correa A, Smith GE, *et al.* RCGP Research and Surveillance Centre: 50 years' surveillance of influenza, infections, and respiratory conditions. *Br J Gen Pract* 2017;**67**:440–1. doi:10.3399/bjgp17X692645

16 Correa A, Hinton W, McGovern A, *et al.* Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) sentinel network: a cohort profile. *BMJ Open* 2016;**6**:e011092. doi:10.1136/bmjopen-2016-011092

17 Law MR, Wald NJ, Rudnicka AR. Quantifying effect of statins on low density lipoprotein cholesterol, ischaemic heart disease, and stroke: systematic review and meta-analysis. *BMJ* 2003;**326**:1423. doi:10.1136/bmj.326.7404.1423

18 Excellence NI for HC, National Institute for Health and Care Excellence. Cardiovascular disease: risk assessment and reduction, including lipid modification. London: : National Institute for Health and Care Excellence 2016.

19 Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 2014;**35**:1925–31. doi:10.1093/eurheartj/ehu207

20 Newson R. Confidence intervals for rank statistics: Somers' D and extensions. *Stata J* 2006;**6**:309–34.

21 Hosmer DW, Lemeshow S, Sturdivant RX. *Applied Logistic Regression*. Wiley 2013.

22 Pérez de Isla L, Alonso R, Mata N, *et al.* Coronary Heart Disease, Peripheral Arterial Disease, and Stroke in Familial Hypercholesterolaemia. *Arterioscler Thromb Vasc Biol* 2016;**36**:2004–10. doi:10.1161/ATVBAHA.116.307514

23 Emanuelsson F, Nordestgaard BG, Benn M. Familial hypercholesterolemia and risk of peripheral arterial disease and chronic kidney disease. *J Clin Endocrinol Metab* 2018;**103**:4491–500. doi:10.1210/jc.2018-01058

24 Raymond Foley T, Singh GD, Kokkinidis DG, *et al.* High-Intensity Statin Therapy Is Associated With Improved Survival in Patients With Peripheral Artery Disease. *J Am Heart Assoc* 2017;**6**:e005699. doi:10.1161/JAHA.117.005699

25 Safarova MS, Liu H, Kullo IJ. Rapid identification of familial hypercholesterolemia from electronic health records: The SEARCH study. *J Clin Lipidol* 2016;**10**:1230–9. doi:10.1016/j.jacl.2016.08.001

26 Troeung L, Arnold-Reed D, Chan She Ping-Delfos W, *et al.* A new electronic screening tool for identifying risk of familial hypercholesterolaemia in general practice. *Heart* 2016;**102**:855–61. doi:10.1136/heartjnl-2015-308824

27 Vickery AW, Ryan J, Pang J, *et al.* Increasing the Detection of Familial Hypercholesterolaemia Using General Practice Electronic Databases. *Heart Lung Circ* 2017;**26**:450–4. doi:10.1016/j.hlc.2016.09.012

28 Weng S, Kai J, Akyea RK, *et al.* Detection of familial hypercholesterolaemia: external validation of a clinical case-finding approach in a retrospective primary care cohort study. *Lancet Public Heal* 2019;**4**:PE256-E264.

29 The Familial Hypercholesterolemia Foundation (FHF). Flag, Identify, Network, Deliver Familial Hypercholesterolemia (FIND FH). 2019.https://thefhfoundation.org/find-fh-initiative

FIGURE TITLES AND LEGENDS

Figure 1. Receiver operating curves derived from the external validation cohort (n = 1,030,183) for models of identifying familial hypercholesterolaemia in general practice (FAMCAT discrimination compared to recommended diagnostic criteria). Higher area under the curve (c-statistic) confers better discrimination

Figure 2. FAMCAT model calibration of observed versus predicted cases of familial hypercholesterolaemia in the external validation cohort by deciles of predicted probability

Figure 3. Calibration of observed versus predicted cases of Familial Hypercholesterolaemia (FH) in the external validation cohort by deciles of predicted probability using the optimised FAMCAT model

Box 1. Summary of predictor variables in FAMCAT

- Gender (male/female)
- Age in years (16-24; 25-34; 35-44; 45-54; 55-64; 65-74; 75-84)
- Highest cholesterol measurement recorded (mmol/L)
  - Ideal: TC* ≤ 5 OR LDL-C † ≤ 3.3
  - High: TC* > 5 to ≤ 6.5 OR LDL-C* > 3.3 to ≤ 4.1
  - Very High: TC* > 6.5 to ≤ 7.5 OR LDL-C* > 4.1 to ≤ 4.9
  - Extremely High: TC* > 7.5 OR LDL-C* > 4.9
- Triglycerides within one month of highest cholesterol measurement (mmol/L)
  - Idea: < 1.7
  - Borderline High: ≥ 1.7 to < 2.3
  - High: ≥ 2.3 to < 5.6
  - Very High: ≥ 5.6
  - Not assessed
- Lipid lowering drugs prescribed within one month of highest cholesterol measurement (none; fibrate/bile acid sequestrant/nicotinic acid; low potency statin; medium potency statin; high potency statin)
- Family history of familial hypercholesterolaemia (no; yes)
- Family history of myocardial infarction (no; yes)
- Family history of raised cholesterol (no; yes)
- Type I or Type II diabetes (no; yes)
- Chronic kidney disease (no; yes)

† TC = total cholesterol; LDL-C = low densitiy lipoprotein cholesterol

Table 1. Clinical characteristics for the cohort of patients aged 16 years. Values are numbers and proportions unless stated otherwise

Characteristics	Men	Women
Total Sample Size	493400 (47.9)	536783 (52.1)
No (%) diagnosed with familial hypercholesterolaemia	649 (0.13)	1058 (0.2)
Baseline age years (SD)	56 (15.3)	57 (16.7)
No (%) with history of coronary heart disease < 60 years	15232 (3.1)	8203 (1.5)
Ethnicity		
No (%) White	319439 (64.7)	358612 (66.8)
No (%) Asian	26916 (5.5)	27749 (5.2)
No (%) Black	14181 (2.9)	17484 (3.3)
No (%) Mixed	3662 (0.7)	4465 (0.8)
No (%) Other	4151 (0.8)	4254 (0.8)
No (%) Unknown	125051 (25.3)	124219 (23.1)
Lipid Profile		
Highest TC recorded mmol/L (SD)	5.6 (1.2)	5.8 (1.3)
High LDL cholesterol recorded mmol/L (SD)	3.4 (1.0)	3.5 (1.1)
Triglycerides during cholesterol measurement mmol/L (SD)	1.6 (1.0)	1.4 (0.8)
Lipid-lowering drug usage at time of cholesterol measurement		
No (%) prescribed fibrate, bile acid sequestrant, nicotinic acid	1158 (0.2)	1491 (0.3)
No (%) prescribed low potency statin	6174 (1.3)	5521 (1.0)
No (%) prescribed medium potency statin	45510 (9.2)	38948 (7.3)
No (%) prescribed high potency statin	21860 (4.4)	17183 (3.2)
Family History		
No (%) with family history of FH	1136 (0.2)	1851 (0.3)
No (%) with family history of raised cholesterol	6698 (1.4)	10144 (1.9)
No (%) with family history of myocardial infarction	28213 (5.7)	36175 (6.7)
Secondary causes of high cholesterol at time of cholesterol measurement		
No (%) diagnosed with diabetes	84490 (17.1)	68978 (12.9)
No (%) diagnosed with chronic kidney disease	53866 (10.9)	71332 (13.3)

Asian includes Indian, Pakistani, Bangladeshi, Chinese and other Asians; FH – familial hypercholesterolaemia; LDL – low-density lipoprotein; SD – standard deviation; TC – total cholesterol;

Table 2. Model discrimination in the external validation cohort for identifying familial hypercholesterolaemia in general practice (n = 1,030,183)

Models	AUC (c-statistic)	Standard Error †	95% Confidence Interval
FAMCAT	0.844	0.005	0.834 – 0.854
Simon Broome Criteria ‡	0.730	0.006	0.719 – 0.741
Dutch Lipid Clinic Criteria §	0.766	0.006	0.755 – 0.778
Cholesterol above 99th centile ?	0.579	0.005	0.571 – 0.588

† Jack-knife procedure to estimate standard errors [20]

‡ Total cholesterol > 7.5 mmol/L or LDL-cholesterol > 4.9 mmol/L + family history of premature myocardial infarction [2]

§ Score based on LDL-cholesterol, family history, clinical history, and physical examination [9]

? The UK National Institute for Health and Care Excellence recommendation of screening for FH for cholesterol above 99th centile. That is, total cholesterol > 9.0 mmol/L or LDL-cholesterol > 6.6 mmol/L if age > 30 years; total cholesterol > 7.5 mmol/L or LDL-cholesterol > 4.9 mmol/L if age ≤ 30 years [2]

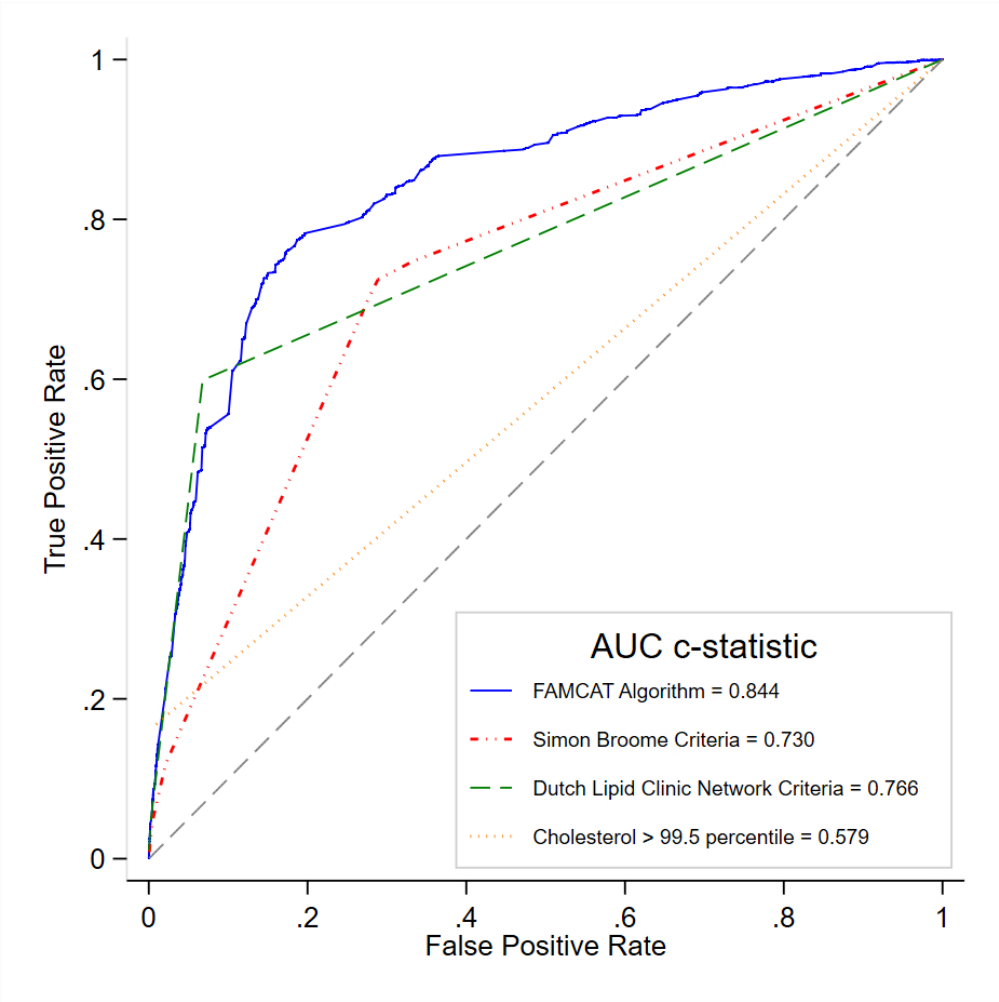


Figure 1. Receiver operating curves derived from the external validation cohort (n = 1,030,183) for models of identifying familial hypercholesterolaemia in general practice (FAMCAT discrimination compared to recommended diagnostic criteria). Higher area under the curve (c-statistic) confers better discrimination

78x78mm (300 x 300 DPI)

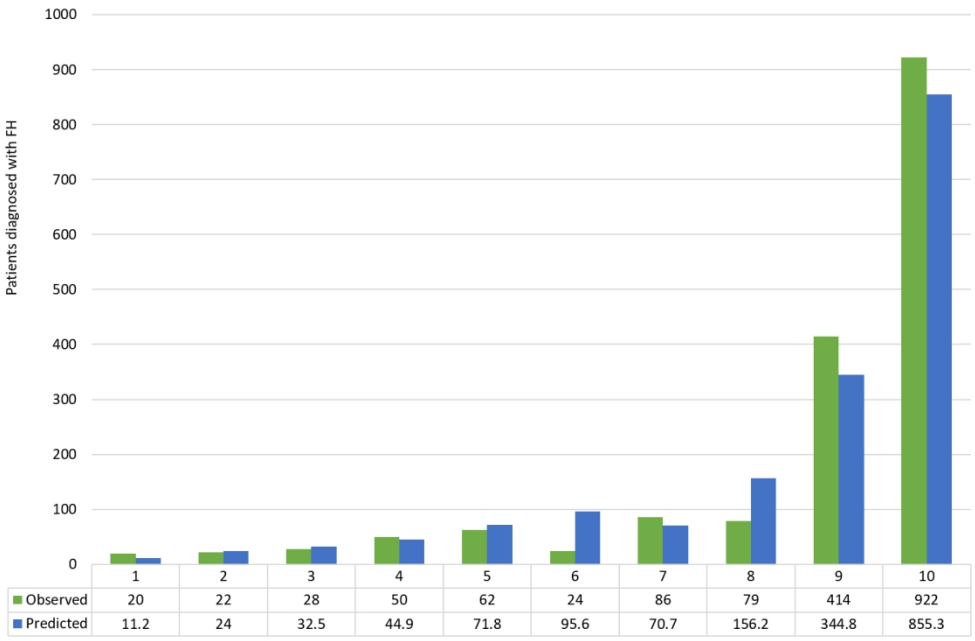
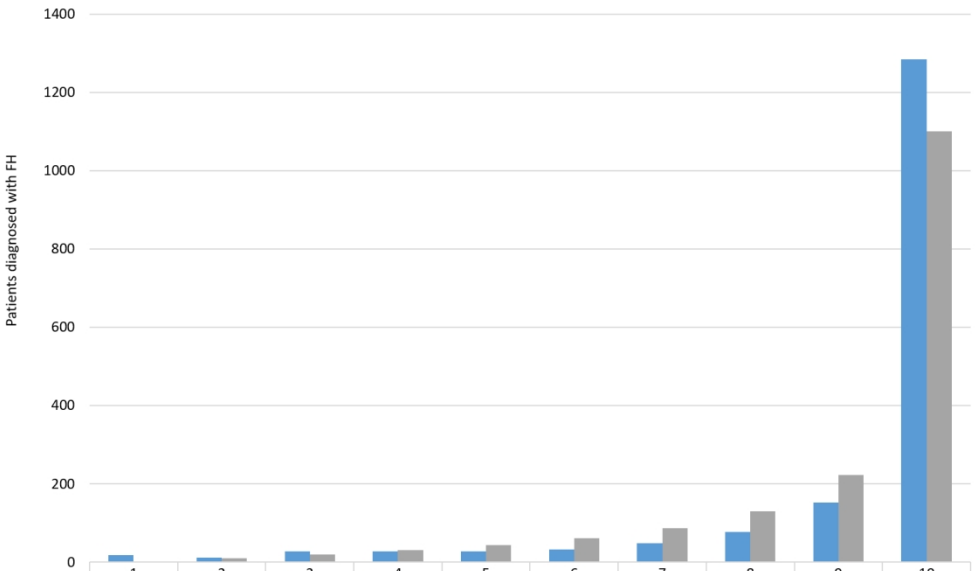


Figure 2. FAMCAT model calibration of observed versus predicted cases of familial hypercholesterolaemia in the external validation cohort by deciles of predicted probability

132x85mm (300 x 300 DPI)



	1	2	3	4	5	6	7	8	9	10
Observed	18	11	28	27	27	33	48	78	152	1285
Predicted	2.6	9.6	19.5	30.7	43.9	61.3	86.7	130	222.7	1100

Figure 3. Calibration of observed versus predicted cases of Familial Hypercholesterolaemia (FH) in the external validation cohort by deciles of predicted probability using the optimised FAMCAT model

135x86mm (300 x 300 DPI)