

1        **1. TITLE PAGE**

2

3        **A machine learning model to aid detection of familial hypercholesterolaemia**

4

5        Jasmine Gratton, MRes<sup>1</sup>; Marta Futema, PhD<sup>1,2</sup>; Steve E. Humphries, MD, PhD<sup>1</sup>; Aroon D.

6        Hingorani, MD, PhD<sup>1,3,4</sup>; Chris Finan, PhD\*<sup>1,3,5</sup>; A. Floriaan Schmidt, PhD\*<sup>1,3,4,5</sup>

7

8        \* Contributed equally

9        <sup>1</sup>Institute of Cardiovascular Science, University College London, London, United Kingdom

10       <sup>2</sup>Cardiology Research Centre, Molecular and Clinical Sciences Research Institute, St

11       George's University of London, London, United Kingdom

12       <sup>3</sup>UCL British Heart Foundation Research Accelerator

13       <sup>4</sup>Health Data Research, UK

14       <sup>5</sup>Department of Cardiology, Division Heart and Lungs, University Medical Centre Utrecht,

15       Utrecht University, the Netherlands

16

17       **Corresponding author:** Jasmine Gratton, MRes, Institute of Cardiovascular Science, UCL,

18       222 Euston Road, NW1 2DA, London, United Kingdom, [j.gratton@ucl.ac.uk](mailto:j.gratton@ucl.ac.uk), +44 20 3549

19       5969

20

21       **Article word count:** 3087 (max: 5000 words)

22

23

1           **2. TEXT ABSTRACT AND KEYWORDS (max word count: 250 words)**

2

3   **Background and Aims:** People with monogenic familial hypercholesterolaemia (FH) are at  
4 an increased risk of premature coronary heart disease and death. Currently there is no  
5 population screening strategy for FH, and most carriers are identified late in life, delaying  
6 timely and cost-effective interventions. The aim was to derive an algorithm to improve  
7 detection of people with monogenic FH.

8   **Methods:** A penalised (LASSO) logistic regression model was used to identify predictors  
9 that most accurately identified people with a higher probability of FH in 139,779 unrelated  
10 participants of the UK Biobank, including 488 FH carriers. Candidate predictors included  
11 information on medical and family history, anthropometric measures, blood biomarkers, and  
12 an LDL-C polygenic score (PGS). Model derivation and evaluation was performed using a  
13 random split of 80% training and 20% testing data.

14   **Results:** A 14-variable algorithm for FH was derived, where the top five variables included  
15 triglyceride, LDL-C, and apolipoprotein A1 concentrations, self-reported statin use, and an  
16 LDL-C PGS. Model evaluation in the test data resulted in an area under the curve (AUC) of  
17 0.77 (95% CI: 0.71; 0.83), and appropriate calibration (calibration-in-the-large: -0.07 (95%  
18 CI: -0.28; 0.13); calibration slope: 1.02 (95% CI: 0.85; 1.19)). Employing this model to  
19 prioritise people with suspected monogenic FH is anticipated to reduce the number of people  
20 requiring sequencing by 88% compared to a population-wide sequencing screen, and by 18%  
21 compared to prioritisation based on LDL-C and statin use.

22   **Conclusions:** The detection of individuals with monogenic FH can be improved with the  
23 inclusion of additional non-genetic variables and a PGS for LDL-C.

24   **Keywords:** FH, prediction, screening, PGS, UK Biobank

25

### 1           **3. INTRODUCTION**

2           Familial hypercholesterolaemia (FH) is an autosomal dominant disorder caused by  
3 variants in the *LDLR*, *APOB*, *PCSK9*, or *APOE* genes. It is characterised by elevated low-  
4 density lipoprotein (LDL-C) concentration and premature coronary heart disease (CHD).(1)  
5 FH-causing variants are found in about 1 in 250 individuals (95% CI: 1:345; 1:192),(2)  
6 however the condition remains highly underdiagnosed worldwide with only an estimated 1%  
7 to 10% of cases diagnosed.(3,4) Affected individuals are at increased risk of premature CHD,  
8 where early initiation of lipid-lowering treatment is paramount for risk management.(3)  
9 There is currently no systematic way of identifying new index FH cases in the general  
10 population, although cascade testing in families of affected individuals has been shown to be  
11 highly cost-effective in many countries.(5–8) Currently, patient diagnosis often happens after  
12 the development of CHD symptoms or by opportunistic measurement of lipid profile and at  
13 the discretion of clinicians. Diagnosis is made using tools such as the Dutch Lipid Clinical  
14 Network (DLCN) and the Simon Broome criteria, which have not been designed to be used  
15 as population screening tools.(1)

16           In 2016, Wald *et al.* suggested screening children aged 15 months of age by  
17 measurement of total or LDL-C to systematically identify index monogenic FH cases in the  
18 general population as a prelude to testing parents and other family members.(9) Futema *et al.*  
19 showed that measurement of LDL-C alone at age 9 may be insufficiently accurate in reliably  
20 distinguishing FH-variant carriers from those with an elevated cholesterol as a consequence  
21 diet and lifestyle factors, or carriage of a high burden of common cholesterol-raising alleles,  
22 and suggested adding a confirmatory targeted-sequencing step to reduce the number of false  
23 positive cases detected.(10)

24           The increased availability of routine health checks in adults either through work-place  
25 schemes or local healthcare providers offers an opportunity to systematically identify adult

1 carriers of FH-causing variants.(11) Positioning adult FH screening within routine health  
2 checks, which typically record a substantial number of other clinical measurements, offers  
3 the opportunity to consider additional predictors for FH. This may be important because,  
4 while the effect of FH on CHD risk is mediated through elevated circulating LDL-C  
5 concentration, it is well-known that LDL-C concentration associates with other variables  
6 such as blood and liver biomarkers, diet, and also with common, genetic variants.(12)  
7 Combining multiple environmental factors and a polygenic score for LDL-C raising genetic  
8 variants may improve the detection of people with monogenic FH for prioritisation for  
9 confirmatory genetic testing.(13,14) This is because individuals with monogenic FH are  
10 likely to have a measured LDL-C concentration that is higher than can be accounted for by  
11 these other variables.

12 In the current manuscript we utilise the UK Biobank data to evaluate the detection  
13 rate and testing burden of three prioritisation strategies to identify people with suspected FH-  
14 causing variants for confirmatory genetic testing: 1) no prioritisation (i.e., referring all  
15 participants for sequencing), 2) a plasma LDL-C-based prioritisation model adjusting for  
16 statin treatment, 3) a multivariable machine learning prioritisation model.

17

## 18 **4. METHODS**

### 19 **Available genomics data and FH ascertainment**

20 We identified 472,147 UK Biobank participants of White British ancestry (data-field  
21 21000) as part of the approved project ID 40721. After performing genomic quality control  
22 steps (see Supplementary Material page 1), 341,515 individuals remained, including 140,439  
23 with whole-exome sequencing (WES) data necessary to identify those who carry an FH-  
24 causing variant. Causal FH variants were searched for in the WES data encompassing the  
25 *LDLR*, *APOB*, *PCSK9* and *APOE* genes (Online Methods section of the Supplementary

1 Material and Supplementary Table 1). A total of 488 pathogenic and likely pathogenic FH  
2 variants were identified (Supplementary Table 2). Additionally, 660 participants were found  
3 to carry FH variants of uncertain significance (VUS) (Supplementary Table 3). These were  
4 excluded from the analysis because more evidence is required to interpret the effect of those  
5 VUS.

6

### 7 **LDL-C PGS generation**

8 We next generated a PGS for LDL-C concentration using an independent data subset  
9 of 173,672 White British participants without lipid-lowering medication or WES data  
10 (Supplementary Figure 1). An initial list of 10,137 genetic variants with a p-value threshold  
11 of  $<5 \times 10^{-4}$  was obtained from the Global Lipids Genetics Consortium (GLGC) genome-wide  
12 association study (GWAS) summary statistics for LDL-C.(15) To reduce the number of  
13 potentially redundant variants and optimise LDL-C prediction, we next applied a least  
14 absolute shrinkage and selection operator (LASSO) regression algorithm using the biglasso  
15 package in R.(16) The degree of penalisation was determined through 15-fold cross-  
16 validation, maximising the explained variance (R-squared), which resulted in a 1,466 genetic  
17 variant LDL-C PGS.

18

### 19 **Deriving a machine learning algorithm to prioritise participants with FH**

20 We extracted data on a total of 24 candidate FH predictors, specifically: LDL-C, high-  
21 density lipoprotein cholesterol (HDL-C), total cholesterol, triglycerides, lipoprotein A  
22 (Lp(a)), apolipoprotein A1 (Apo-A1), apolipoprotein B (Apo-B), C-reactive protein (CRP),  
23 aspartate aminotransferase (AST), alanine aminotransferase (ALT), alkaline phosphatase  
24 (ALP), sex, body mass index (BMI), age, self-reported statin use, alcohol use, systolic blood  
25 pressure (SBP), diastolic blood pressure (DBP), Townsend deprivation index, smoking status,

1 family history of CHD, type 2 diabetes diagnosis, hypertension, and LDL-C PGS. This was  
2 expanded by including 10 product terms between: age and LDL-C, age and LDL-C PGS,  
3 LDL-C PGS and LDL-C, age<sup>2</sup>, LDL-C<sup>2</sup>, statin use and LDL-C, family history of CHD and  
4 sex, family history of CHD and statin use, family history of CHD and alcohol use, family  
5 history of CHD and hypertension. The limited missing data (Supplementary Table 4) were  
6 singly imputed using the R package MICE.(17)

7 Model derivation was performed using the WES data, applying a 80% training data  
8 split of 111,824 subjects, retaining 20% testing data (containing 93 carriers out of 27,955  
9 subjects) to unbiasedly evaluate model performance (Supplementary Figure 1). To prevent  
10 potential model instability, highly correlated variables (i.e. multicollinear) were removed.  
11 These included Apo-B and total cholesterol (Supplementary Figure 2). Variables were  
12 standardised to mean zero and standard deviation (SD) one. Finally, we applied a binomial  
13 regression model with LASSO penalisation to derive a discrimination-optimised FH  
14 prediction model. Specifically, optimal penalisation was determined through 15-fold cross-  
15 validation maximising the c-statistic (i.e., the area under the receiver operating characteristic  
16 (AUC-ROC) curve).(16)

17 Model performance was evaluated using the 20% testing data based on its  
18 discriminative ability (c-statistic), appropriate calibration of predicted and observed  
19 probability of having an FH variant (using calibration plots, calibration-in-the-large, and  
20 calibration slope), and classification metrics (sensitivity, specificity (or its complement the  
21 false positive rate), positive predictive value, and the negative predicted value).

22

### 23 **Evaluating the burden of genomic sequencing for FH**

24 While genetic sequencing is the gold standard for FH diagnosis, it may often be  
25 prohibitively expensive to offer it to an entire population as a screening strategy. We

1 therefore explored whether prioritising people with suspected FH can reduce the screening  
2 burden with an acceptable number of false-negative results. We evaluated the following  
3 prioritisation strategies: 1) no prioritisation (i.e. referring all participants for sequencing), 2)  
4 prioritisation based on LDL-C concentration (adjusting for statin use), 3) a multivariable  
5 model built from genetic, clinical biomarkers and environmental predictors.

6 These prioritisation strategies were evaluated on the number of subjects that would  
7 need to be sequenced, the proportion of FH carriers who would be missed, and the ratio of  
8 FH carriers correctly prioritised by the number of non-carriers unnecessarily offered  
9 sequencing. Additionally, prioritisation based on LDL-C concentrations (adjusted for statin  
10 use) was compared to prioritisation using the multivariable model with the help of a net  
11 reclassification index (NRI) analysis.

12

## 13 **5. RESULTS**

### 14 **Participant characteristics of our study cohort**

15 Using the UK Biobank WES data, we identified 488 pathogenic or likely pathogenic  
16 FH variant carriers (list of variants shown in Supplementary Table 2) and 139,291 non-  
17 carriers; 0.35% (95% confidence interval (CI): 0.32; 0.38). FH variant carriers had a  
18 significantly higher frequency of a family history of coronary heart disease (CHD) (62.7%  
19 versus 48.1% in controls), higher prevalence (8.2% versus 2.8% in controls) and incidence  
20 (6.6% versus 3.9% in controls) of CHD (Supplementary Material and Table 1).

21

### 22 **Multivariable machine learning model to prioritise FH variant carriers**

23 14 out of the 32 variables were retained by the LASSO regression model for the  
24 prediction of FH (Figure 1.a, Supplementary Figure 3, Supplementary Table 5), including  
25 triglyceride, Apo-A1, ALT and CRP concentrations, statin use, LDL-C PGS, family history

1 of CHD, DBP, BMI, and prevalent T2D. Additionally, the following product terms were  
2 selected: LDL-C<sup>2</sup>, statin use and LDL-C, age and LDL-C PGS. Retention of these product  
3 terms indicated the presence of non-linear associations with FH, for example the LDL-C  
4 association with the presence of a monogenic FH variant was found to be quadratic  
5 (Supplementary Figure 4).

6 The median predicted probability of having monogenic FH by the multivariable  
7 model was ~3 fold higher in FH carriers (0.64%, interquartile range (IQR): 0.31; 1.62)  
8 compared to non-carriers (0.23%, IQR: 0.14; 0.38), with partial overlap between FH carriers  
9 and non-carriers (Figure 1.b). The test data AUC for this model was 0.77 (95% CI: 0.71;  
10 0.83), with a training data AUC of 0.78 (95% CI: 0.75; 0.81). Calibration statistics  
11 (calibration-in-the-large: -0.073 (95% CI: -0.28; 0.13) and calibration slope: 1.02 (95% CI:  
12 0.85; 1.19)) indicated the predicted probability agreed well with the observed probability  
13 (Figure 2.a).

14 The multivariable machine learning model outperformed a model which only consider  
15 LDL-C (AUC: 0.62, 95% CI: 0.56; 0.68), as well as a model which additionally included a  
16 statin indicator (AUC: 0.71, 95% CI: 0.65; 0.77), both evaluated in the test data. (Figure 2.b).

17

### 18 **Model FH classification**

19 Next, we evaluated the FH classification performance of the multivariable model  
20 using six cut-off values of having an FH variant (from 0.001 to 0.10) in the test dataset. The  
21 sensitivity increased from 1.1% (95% CI: 0.2; 5.8) for a predicted probability of 0.10, to  
22 94.6% (95% CI: 88.0; 97.7) for a predicted probability of 0.001; with the false positive rate  
23 similarly increasing from 0.1% (95% CI: 0.0; 0.1) to 87.0% (95% CI: 86.6; 87.4) (Table 2).  
24 We further compared the performance of these thresholds to a simpler model of LDL-C  
25 concentration adjusted for statin, which underperformed (Table 2).



1           The net reclassification index (NRI) comparing the LDL-C and statin use model to the  
2 multivariable model, indicated that the improved performance of the latter was due to it  
3 assigning a higher predicted probability to FH variant carriers. At a predicted probability  
4 threshold of 0.006, the probability for FH carriers being reclassified as having an FH variant  
5 was equal to 0.097 (95% CI: 0.038; 0.159), as opposed to the probability of 0.075 (95% CI:  
6 0.026; 0.130) of being down-classified as not having an FH variant (Table 3).

7

### 8 **Prioritising individuals for FH genomic testing in a two-stage population screening** 9 **strategy**

10           Finally, we evaluated the performance of a two-stage population screen for  
11 identifying new index FH cases, where the second stage consisted of targeted sequencing of  
12 FH variants (Supplementary Figure 5). The multivariable and LDL-C with statin use models  
13 were compared using a common threshold of 0.006, where on average, seven additional FH  
14 carriers would be detected for 100,000 individuals screened when using the multivariable  
15 model compared to the LDL-C and statin use model. Per 100,000 individuals screened, the  
16 multivariable model would refer 12,033 individuals (12%) for genetic sequencing, compared  
17 to 14,730 (15%) with the LDL-C and statin use model, resulting in a 18% reduction in  
18 genetic testing.

19           Furthermore, if we assume that FH has a population prevalence of 1 in 286 (equal to  
20 our cohort's prevalence) and that one FH case has on average 1.5 first-degree relatives ((2  
21 children + 1 sibling) / 2) who are also affected by FH (discovered through cascade  
22 testing),(18) then overall one FH case would be identified for every ~219 people screened  
23 when using the multivariable model, compared to one FH case for every ~228 individuals  
24 screened with the LDL-C and statin use model.

25

## 1        6. DISCUSSION

2            In the current manuscript we derived a multivariable machine learning model to  
3 identify people with suspected FH for confirmatory DNA sequencing in the context of  
4 population screening. Using LASSO regression, we derived a 14-feature model consisting of  
5 LDL-C, Apo-A1, triglyceride, ALT, and CRP concentrations, self-reported statin use, family  
6 history of CHD, DBP, BMI, type 2 diabetes diagnosis, three product terms, and an LDL-C  
7 PGS. The multivariable algorithm was able to discriminate between FH and non-FH carriers  
8 with an AUC of 0.77 (95% CI: 0.71; 0.83), with good calibration, outperforming a simpler  
9 model consisting of LDL-C and an indicator for statin prescription.

10            Independent of the classification threshold applied, the multivariable algorithm was  
11 able to substantially decrease the number of subjects referred to genetic sequencing (e.g. from  
12 100,000 individuals without any prioritisation, to 14,730 with prioritisation using the LDL-C  
13 and statin use model, and to 12,033 with prioritisation using the multivariable model for a  
14 predicted probability threshold of carrying a variant for monogenic FH of 0.006; equivalent  
15 to approximately a 18% decrease in individuals needed to be sequenced between the last two  
16 models). These differences become especially significant if extrapolating the values to a  
17 population-wide scale comprising of millions of participants screened. Our results provide  
18 support for opportunistic screening and seeding of cascade testing for FH, which could be  
19 integrated within existing health checks offered to employers or local healthcare  
20 providers.(11)

21            Previously, Banda *et al.* used a machine learning method to detect monogenic FH  
22 cases from electronic health records (EHR).(19) While their model showed an impressive  
23 AUC of 0.94, one of their most important features was referral to a cardiology clinic, which  
24 is in very close proximity to confirmatory FH testing, limiting the model's utility as a  
25 prospective tool for FH diagnosis. Besseling *et al.* developed a multivariable model to

1 identify FH carriers validated in study participants consisting of FH cases and their relatives,  
2 again limiting applicability to the general population.(20) Our model instead considers FH  
3 prioritisation in a non-GP-referred population and is more generalisable as a systematic  
4 population screening tool.

5 Our multivariable model included three terms for LDL-C (LDL-C itself, LDL-C  
6 squared, and an interaction with statin prescription), which combined makes it the most  
7 important predictor. Additionally, our model also identified novel predictors for FH such as  
8 triglyceride and Apo-A1 concentrations, with triglycerides having the largest absolute OR per  
9 SD (0.60). High triglyceride levels are often linked to poor diet and a sedentary lifestyle.(21–  
10 24) Here we find that FH carriers had lower triglyceride concentrations than non-carriers  
11 (Table 1), which resulted in a protective association, indicating that triglyceride  
12 concentrations can be useful in discriminating between individuals who have  
13 hypercholesterolaemia due to lifestyle factors as opposed to an FH-causing variant. We also  
14 found that higher Apo-A1 concentrations, a protein found on HDL particles, was associated  
15 with a decreased probability of FH. Finally, we note that our multivariable FH model retained  
16 a squared term for LDL-C, suggesting that LDL-C is not linearly related with carrying an FH  
17 variant, but rather has a quadratic relationship (Supplementary Figure 4).

18 The variables included in our multivariable algorithm should not be interpreted as  
19 causal risk factors for monogenic FH; they simply help to distinguish non-monogenic  
20 sources of variation in LDL-C concentrations from monogenic causes (as was discussed in  
21 more detail previously with triglyceride concentrations). This also provides the rationale for  
22 including an LDL-C PGS in the model: a large discrepancy between predicted LDL-C  
23 concentrations (by the LDL-C PGS) and observed LDL-C concentrations might be indicative  
24 of FH carriership,(13,14) demonstrated here by a negative coefficient for LDL-C PGS in the  
25 model (Supplementary Table 5). We note that a previous LDL-C PGS by Wu *et al.* had a

1 substantially larger R-squared (0.21 (95% CI: 0.20-0.22)) than reported here (0.14 (95% CI:  
2 0.13-0.15)).(25) Unlike Wu *et al.* who identified genetic variants from an internal UK  
3 Biobank LDL-C GWAS overlapping with the PGS training data; we identified variants based  
4 on an independent dataset from GLGC,(15) guarding against overfitting through ‘data-  
5 leakage’ between the training and testing datasets and providing a more robust estimate of  
6 explained variance.

7 A study limitation to consider is the exclusion of individuals with VUS from our  
8 study cohort. There is conflicting evidence as to the causal effects of these VUS in FH. We  
9 anticipate that some are likely to be FH-causing while others are not, but more research is  
10 needed. As more VUS are classified as either FH-causing or not, the model can be readily  
11 updated to reflect our growing understanding of FH. Additionally, it is impossible to know  
12 whether some study participants have been genetically tested for carrying an FH variant, and  
13 whether they might have modified their behaviour (e.g. diet) following their diagnosis. This  
14 could potentially impact the accuracy of the multivariable model developed here; however,  
15 considering that only approximately 7% of FH cases have been diagnosed in the UK,(26) this  
16 low number of diagnoses is unlikely to have a significant effect on the model and results  
17 presented here. Currently, PGS information is not routinely used or collected in clinical  
18 practice, which is why we also derived a penalised multivariable model without an LDL-C  
19 PGS, which did not meaningfully decrease performance (Supplementary Table 6). Previous  
20 studies have suggested that PGS could be used to identify individuals with a rare variant for  
21 certain diseases, such as FH.(13,14) Our study confirms the utility of the PGS for FH  
22 prioritisation; however given its correlation with environmental variables (e.g. lipid levels),  
23 this genetic information can be readily replaced with information from non-genetic data.

24 We have tested our multivariable model in a dataset which was independent from the  
25 training data, with no significant difference between training and testing AUC (difference of

1 0.01), suggesting limited model overfitting to the current sample. Nevertheless, considering  
2 the health discrepancies observed between the UK Biobank and the general UK  
3 population,(27) we suggest that this model is locally validated and updated before applying it  
4 to distinct settings. Model validation should especially be conducted when considering  
5 populations of non-European ancestry. Irrespective of the important considerations regarding  
6 model transferability, prior to integrating the model in clinical care, an informed decision  
7 should be made on the optimal predicted probability threshold for monogenic FH  
8 classification. We wish to highlight that our choice of 0.006 as a threshold is purely  
9 pragmatic, and a more optimised threshold could further increase benefit. Given that  
10 monogenic FH is relatively rare in the general population, we would expect the optimum  
11 probability threshold to be low, similar to the one employed here. While Youden's J statistic  
12 can be used to identify the optimal threshold balancing sensitivity and specificity, this implies  
13 equal costs between false-positive and false-negative predictions which is unlikely to be true.  
14 The choice of threshold should be supported by (local) health-technology assessments  
15 incorporating direct and indirect costs.

16 In conclusion, we derived a multivariable classification model for detecting monogenic  
17 FH variant carriers that outperformed a model based on LDL-C concentration (adjusted for  
18 statin use) for FH screening, and that offers an opportunity to prioritise suspected FH carriers  
19 for genetic sequencing.

20

## 21 **7. FUNDING**

22 JG is supported by the BHF studentship FS/17/70/33482. AFS is supported by the BHF grant  
23 PG/18/5033837 and the UCL BHF Research Accelerator AA/18/6/34223. CF and AFS  
24 received additional support from the National Institute for Health Research University  
25 College London Hospitals Biomedical Research Centre. ADH is an NIHR Senior

1 Investigator. SH and MF were supported by a grant from the British Heart Foundation (BHF  
2 grant PG 08/008) and by funding from the Department of Health's NIHR Biomedical  
3 Research Centers funding scheme.

4

## 5 **8. ACKNOWLEDGEMENTS**

6 This research has been conducted using data from UK Biobank (application number 40721),  
7 a major biomedical database.(28) The authors are grateful to UK Biobank participants. UK  
8 Biobank was established by the Wellcome Trust medical charity, Medical Research Council,  
9 Department of Health, Scottish Government, and the Northwest Regional Development  
10 Agency. It has also had funding from the Welsh Assembly Government and the British Heart  
11 Foundation.

12

## 13 **9. CONFLICT OF INTEREST**

14 None to declare.

15

## 16 **10. REFERENCES**


- 17 1. McGowan MP, Hosseini Dehkordi SH, Moriarty PM, Duell PB. Diagnosis and  
18 treatment of heterozygous familial hypercholesterolemia. J Am Heart Assoc [Internet].  
19 2019 Dec 17 [cited 2022 Jan 12];8(24). Available from:  
20 <https://www.ahajournals.org/doi/abs/10.1161/JAHA.119.013225>
- 21 2. Akioyamen LE, Genest J, Shan SD, Reel RL, Albaum JM, Chu A, et al. Estimating the  
22 prevalence of heterozygous familial hypercholesterolaemia: a systematic review and  
23 meta-analysis. BMJ Open [Internet]. 2017 Sep 1 [cited 2022 May 9];7(9):e016461.  
24 Available from: <https://bmjopen.bmj.com/content/7/9/e016461>
- 25 3. Vallejo-Vaz AJ, Stevens CAT, Lyons ARM, Dharmayat KI, Freiburger T, Hovingh

- 1 GK, et al. Global perspective of familial hypercholesterolaemia: a cross-sectional  
2 study from the EAS Familial Hypercholesterolaemia Studies Collaboration (FHSC).  
3 Lancet [Internet]. 2021 Nov 6 [cited 2022 Mar 15];398(10312):1713–25. Available  
4 from: <http://www.thelancet.com/article/S0140673621011223/fulltext>
- 5 4. Tromp TR, Hartgers ML, Hovingh GK, Vallejo-Vaz AJ, Ray KK, Soran H, et al.  
6 Worldwide experience of homozygous familial hypercholesterolaemia: retrospective  
7 cohort study. Lancet [Internet]. 2022 Feb 19 [cited 2022 Mar 15];399(10326):719–28.  
8 Available from: <http://www.thelancet.com/article/S0140673621020018/fulltext>
- 9 5. Kerr M, Pears R, Miedzybrodzka Z, Haralambos K, Cather M, Watson M, et al. Cost  
10 effectiveness of cascade testing for familial hypercholesterolaemia, based on data from  
11 familial hypercholesterolaemia services in the UK. Eur Heart J [Internet]. 2017 Jun 14  
12 [cited 2021 Nov 10];38(23):1832–9. Available from:  
13 <https://academic.oup.com/eurheartj/article/38/23/1832/3104460>
- 14 6. Marquina C, Lacaze P, Tiller J, Riaz M, Sturm AC, Nelson MR, et al. Population  
15 genomic screening of young adults for familial hypercholesterolaemia: a cost-  
16 effectiveness analysis. Eur Heart J [Internet]. 2021 Nov 11 [cited 2022 Mar 15];  
17 Available from: [https://academic.oup.com/eurheartj/advance-](https://academic.oup.com/eurheartj/advance-article/doi/10.1093/eurheartj/ehab770/6425622)  
18 [article/doi/10.1093/eurheartj/ehab770/6425622](https://academic.oup.com/eurheartj/advance-article/doi/10.1093/eurheartj/ehab770/6425622)
- 19 7. Jackson CL, Huschka T, Borah B, Agre K, Zordok M, Farwati M, et al. Cost-  
20 effectiveness of cascade genetic testing for familial hypercholesterolemia in the United  
21 States: A simulation analysis. Am J Prev Cardiol. 2021 Dec 1;8:100245.
- 22 8. Lázaro P, Pérez de Isla L, Watts GF, Alonso R, Norman R, Muñiz O, et al. Cost-  
23 effectiveness of a cascade screening program for the early detection of familial  
24 hypercholesterolemia. J Clin Lipidol. 2017 Jan 1;11(1):260–71.
- 25 9. Wald DS, Bestwick JP, Morris JK, Whyte K, Jenkins L, Wald NJ. Child–Parent

- 1           Familial Hypercholesterolemia Screening in Primary Care. *N Engl J Med* [Internet].  
2           2016 Oct 26 [cited 2021 Nov 6];375(17):1628–37. Available from:  
3           <https://www.nejm.org/doi/10.1056/NEJMoa1602777>
- 4   10.   Futema M, Cooper JA, Charakida M, Boustred C, Sattar N, Deanfield J, et al.  
5           Screening for familial hypercholesterolaemia in childhood: Avon Longitudinal Study  
6           of Parents and Children (ALSPAC). *Atherosclerosis* [Internet]. 2017 May 1 [cited  
7           2021 Nov 8];260:47. Available from: [/pmc/articles/PMC5414415/](https://pubmed.ncbi.nlm.nih.gov/35414415/)
- 8   11.   Duddy C, Wong G, Gadsby EW, Krska J, Hibberd V. NHS Health Check programme:  
9           a protocol for a realist review. *BMJ Open* [Internet]. 2021 Apr 1 [cited 2022 Feb  
10          28];11(4):e048937. Available from: <https://bmjopen.bmj.com/content/11/4/e048937>
- 11   12.   Talmud PJ, Drenos F, Shah S, Shah T, Palmen J, Verzilli C, et al. Gene-centric  
12          Association Signals for Lipids and Apolipoproteins Identified via the HumanCVD  
13          BeadChip. *Am J Hum Genet*. 2009;
- 14   13.   Zhou D, Yu D, Scharf JM, Mathews CA, McGrath L, Cook E, et al. Contextualizing  
15          genetic risk score for disease screening and rare variant discovery. *Nat Commun*  
16          [Internet]. 2021 Jul 20 [cited 2022 Jan 27];12(1):1–14. Available from:  
17          <https://www.nature.com/articles/s41467-021-24387-z>
- 18   14.   Lu T, Forgetta V, Richards JB, Greenwood CMT. Polygenic risk score as a possible  
19          tool for identifying familial monogenic causes of complex diseases. *Genet Med*  
20          [Internet]. 2022 Apr [cited 2022 Apr 26];0(0):1–11. Available from:  
21          <http://www.gimjournal.org/article/S1098360022007171/fulltext>
- 22   15.   Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al.  
23          Discovery and refinement of loci associated with lipid levels. *Nat Genet* [Internet].  
24          2013 Oct 6 [cited 2022 Jan 19];45(11):1274–83. Available from:  
25          <https://www.nature.com/articles/ng.2797>



- 1 16. Zeng Y, Breheny P. The biglasso Package: A Memory- and Computation-Efficient  
2 Solver for Lasso Model Fitting with Big Data in R. R J [Internet]. 2017 Jan 20 [cited  
3 2022 Jan 19];12:1–14. Available from: <https://arxiv.org/abs/1701.05936v2>
- 4 17. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained  
5 equations in R. J Stat Softw [Internet]. 2011 Dec 12 [cited 2020 Oct 17];45(3):1–67.  
6 Available from:  
7 <https://www.jstatsoft.org/index.php/jss/article/view/v045i03/v45i03.pdf>
- 8 18. Families and households in the UK - Office for National Statistics [Internet]. [cited  
9 2022 Mar 1]. Available from:  
10 [https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/fa  
11 milies/bulletins/familiesandhouseholds/2020](https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/families/bulletins/familiesandhouseholds/2020)
- 12 19. Banda JM, Sarraju A, Abbasi F, Parizo J, Pariani M, Ison H, et al. Finding missed  
13 cases of familial hypercholesterolemia in health systems using machine learning. npj  
14 Digit Med [Internet]. 2019 [cited 2022 Jan 12]; Available from:  
15 <https://doi.org/10.1038/s41746-019-0101-5>
- 16 20. Besseling J, Reitsma JB, Gaudet D, Brisson D, Kastelein JJP, Hovingh GK, et al.  
17 Selection of individuals for genetic testing for familial hypercholesterolaemia:  
18 development and external validation of a prediction model for the presence of a  
19 mutation causing familial hypercholesterolaemia. Eur Heart J [Internet]. 2017 Feb 21  
20 [cited 2022 Jan 12];38(8):565–73. Available from:  
21 <https://academic.oup.com/eurheartj/article/38/8/565/3056905>
- 22 21. Laufs U, Parhofer KG, Ginsberg HN, Hegele RA. Clinical review on triglycerides. Eur  
23 Heart J [Internet]. 2020 Jan 1 [cited 2022 Feb 22];41(1):99-109c. Available from:  
24 <https://academic.oup.com/eurheartj/article/41/1/99/5640489>
- 25 22. Futema M, Ramaswami U, Tichy L, Bogsrud MP, Holven KB, Roeters van Lennep J,

- 1 et al. Comparison of the mutation spectrum and association with pre and post  
2 treatment lipid measures of children with heterozygous familial hypercholesterolaemia  
3 (FH) from eight European countries. *Atherosclerosis*. 2021 Feb 1;319:108–17.
- 4 23. gavinband / bgen / wiki / bgenix — Bitbucket [Internet]. [cited 2020 Mar 20].  
5 Available from: <https://bitbucket.org/gavinband/bgen/wiki/bgenix>
- 6 24. Haralambos K, Whatley SD, Edwards R, Gingell R, Townsend D, Ashfield-Watt P, et  
7 al. Clinical experience of scoring criteria for Familial Hypercholesterolaemia (FH)  
8 genetic testing in Wales. *Atherosclerosis* [Internet]. 2015 May 1 [cited 2022 Feb  
9 22];240(1):190–6. Available from:  
10 <http://dx.doi.org/10.1016/j.atherosclerosis.2015.03.003>
- 11 25. Wu H, Forgetta V, Zhou S, Bhatnagar SR, Paré G, Richards BJ. Polygenic Risk Score  
12 for Low-Density Lipoprotein Cholesterol Is Associated with Risk of Ischemic Heart  
13 Disease and Enriches for Individuals with Familial Hypercholesterolemia. *Circ  
14 Genomic Precis Med* [Internet]. 2021 [cited 2021 Nov 24];14:3106. Available from:  
15 <https://www.ahajournals.org/doi/abs/10.1161/CIRCGEN.120.003106>
- 16 26. NHS England and NHS Improvement London  » Familial Hypercholesterolemia (FH)  
17 [Internet]. [cited 2022 May 9]. Available from:  
18 [https://www.england.nhs.uk/london/london-clinical-networks/our-  
19 networks/cardiac/familial-hypercholesterolaemia/](https://www.england.nhs.uk/london/london-clinical-networks/our-networks/cardiac/familial-hypercholesterolaemia/)
- 20 27. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al.  
21 Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank  
22 Participants With Those of the General Population. *Am J Epidemiol* [Internet]. 2017  
23 [cited 2020 Mar 20];186(9). Available from: <http://creativecommons.org/licenses/by-nc-nd/4.0/>.
- 24 28. UK Biobank - UK Biobank [Internet]. [cited 2022 Mar 4]. Available from:  
25 <https://www.ukbiobank.ac.uk/>

1

## 2 **11. FIGURE LEGENDS**

3

4 **Figure 1. Feature importance of the variables retained by LASSO regression predicting**  
5 **monogenic FH, and the density predicted probability distributions from this model for**  
6 **unaffected and affected FH individuals in White British participants of the UK**

7 **Biobank. a)** The 14 predictors retained by LASSO regression ordered by absolute log odds  
8 ratio (OR) per standard deviation (SD). The “x” sign is used to indicate an interaction term.  
9 Abs = absolute; LDL-C = low-density lipoprotein cholesterol; Apo-A1 = apolipoprotein A1;  
10 PGS = polygenic score; CHD = coronary heart disease; DBP = diastolic blood pressure; ALT  
11 = alanine aminotransferase; BMI = body mass index; CRP = C-reactive protein; T2D = type  
12 2 diabetes. **b)** The density predicted probability distributions for affected (in orange) and  
13 unaffected (in blue) FH participants in our test cohort as predicted by the multivariable  
14 model. 14 unaffected individuals had a monogenic FH predicted probability above 0.12 and  
15 are not shown on the plot for legibility purposes. The vertical dotted lines represent the  
16 various FH predicted probability thresholds evaluated in Table 2.

17

18 **Figure 2. Discrimination and calibration of a multivariable algorithm predicting FH**  
19 **carriership using independent testing data. a)** The calibration plot for the multivariable  
20 model where the mean predicted and mean observed probability for each decile of the test  
21 data are depicted by the datapoints with their 95% confidence intervals (CI). Perfect  
22 calibration is indicated by the vertical black line. The calibration-in-the-large (CIL) and the  
23 calibration slope (CS) values are indicated on the plot with their 95% CI in brackets. The  
24 loess line was fitted with FH-causing variant status as the outcome and mean predicted  
25 probability as the predictor. **b)** The receiver operating characteristic (ROC) curves for the

1 multivariable model (in red), LDL-C concentration and statin model (in green), and LDL-C  
 2 concentration only model (in blue). The area under the curve (AUC) for each of these models  
 3 are equal to 0.77 (95% CI: 0.71; 0.83), 0.71 (95% CI: 0.65; 0.77) and 0.62 (95% CI: 0.56;  
 4 0.68) respectively.

5

## 6 12. TABLES

7

8 **Table 1. UK Biobank participant characteristics stratified by carrying a familial**  
 9 **hypercholesterolaemia (FH)-causing variant.** The p-values shown in the table are from the  
 10 Kruskal-Wallis Rank Sum test for continuous variables, and from the Man-Whitney U test  
 11 for binary variables. IQR = interquartile range; BMI = body mass index; CHD = coronary  
 12 heart disease; LDL-C = low-density lipoprotein cholesterol; HDL-C = high-density  
 13 lipoprotein cholesterol; CVD = cardiovascular disease; PGS = polygenic score.

14

|   | FH-variant negative | FH-variant positive | p-value | Missing (%) |
|---|---------------------|---------------------|---------|-------------|
| n   | 139291              | 488                 |         |             |
| Sex (male) (%)                            | 63382 (45.5)        | 207 (42.4)          | 0.187   | 0.0         |
| Age (median [IQR])                        | 58.0 [51.0, 63.0]   | 58.0 [51.0, 63.0]   | 0.803   | 0.0         |
| Townsend deprivation index (median [IQR]) | -2.4 [-3.8, 0.0]    | -2.2 [-3.7, 0.2]    | 0.346   | 0.1         |
| BMI, kg/m <sup>2</sup> (median [IQR])     | 26.7 [24.1, 29.8]   | 27.1 [23.9, 29.8]   | 0.689   | 0.3         |
| Smoking status (%)                        |                     |                     | 0.685   | 3.7         |
| Non-smoker                                | 76862 (57.3)        | 262 (56.2)          |         |             |
| Former smoker                             | 49302 (36.7)        | 171 (36.7)          |         |             |
| Light smoker (<10 cigarettes/day)         | 1952 (1.5)          | 6 (1.3)             |         |             |
| Moderate smoker (10-19 cigarettes/day)    | 3296 (2.5)          | 13 (2.8)            |         |             |
| Heavy Smoker (>20 cigarettes/day)         | 2796 (2.1)          | 14 (3.0)            |         |             |
| Alcohol use (%)                           |                     |                     | 0.492   | 0.0         |
| Prefer not to answer                      | 88 (0.1)            | 1 (0.2)             |         |             |
| 1/day                                     | 29719 (21.3)        | 93 (19.1)           |         |             |
| 3-4 times/week                            | 34015 (24.4)        | 135 (27.7)          |         |             |
| 1-2 times/week                            | 36823 (26.4)        | 130 (26.6)          |         |             |
| 1-3 times/month                           | 15498 (11.1)        | 54 (11.1)           |         |             |

|  |                      |                      |        |      |
|--|----------------------|----------------------|--------|------|
| Special occasions  | 14383 (10.3)         | 45 (9.2)             |        |      |
| Never  | 8765 (6.3)           | 30 (6.1)             |        |      |
| Family history of CHD (%)                                | 67013 (48.1)         | 306 (62.7)           | <0.001 | 0.0  |
| Systolic blood pressure, mmHg (median [IQR])             | 136.5 [125.0, 149.5] | 135.0 [124.5, 148.5] | 0.119  | 0.2  |
| Diastolic blood pressure, mmHg (median [IQR])            | 82.0 [75.0, 89.0]    | 81.0 [74.0, 87.0]    | 0.024  | 0.2  |
| Statin use (%)   | 18139 (13.0)         | 165 (33.8)           | <0.001 | 0.0  |
| Hypertension (median [IQR])                              | 7946 (5.7)           | 35 (7.2)             | 0.195  | 0.0  |
| LDL-C PGS (median [IQR])                                 | 3.7 [3.5, 3.9]       | 3.7 [3.5, 3.9]       | 0.652  | 0.0  |
| <b>Biomarkers</b>  |                      |                      |        |      |
| LDL-C (unadjusted for statin use), mmol/L (median [IQR]) | 3.5 [3.0, 4.1]       | 3.9 [3.2, 4.9]       | <0.001 | 5.0  |
| HDL-C, mmol/L (median [IQR])                             | 1.4 [1.2, 1.7]       | 1.4 [1.2, 1.6]       | 0.086  | 12.5 |
| Total cholesterol, mmol/L (median [IQR])                 | 5.7 [4.9, 6.4]       | 6.1 [5.2, 7.3]       | <0.001 | 4.8  |
| Lipoprotein(a), nmol/L (median [IQR])                    | 20.0 [9.3, 59.8]     | 27.6 [10.3, 59.2]    | 0.083  | 24.3 |
| Apolipoprotein A1, g/L (median [IQR])                    | 1.5 [1.4, 1.7]       | 1.5 [1.3, 1.6]       | <0.001 | 13.0 |
| Apolipoprotein B, g/L (median [IQR])                     | 1.0 [0.9, 1.2]       | 1.2 [1.0, 1.4]       | <0.001 | 5.3  |
| Triglycerides, mmol/L (median [IQR])                     | 1.5 [1.1, 2.2]       | 1.3 [0.9, 1.9]       | <0.001 | 4.9  |
| C-reactive protein, mg/L (median [IQR])                  | 1.3 [0.7, 2.7]       | 1.2 [0.6, 2.3]       | 0.065  | 5.1  |
| Aspartate aminotransferase, um (median [IQR])            | 24.4 [21.0, 28.8]    | 25.1 [21.0, 29.6]    | 0.111  | 5.2  |
| Alanine aminotransferase, um (median [IQR])              | 20.1 [15.4, 27.3]    | 20.2 [15.6, 27.2]    | 0.848  | 4.9  |
| Alkaline phosphatase, um (median [IQR])                  | 80.1 [67.1, 95.4]    | 80.6 [66.8, 96.1]    | 0.506  | 4.8  |
| <b>Disease prevalence &amp; incidence</b>                |                      |                      |        |      |
| CHD prevalence (%)                                       | 3890 (2.8)           | 40 (8.2)             | <0.001 | 0.0  |
| CHD incidence (%)  | 5370 (3.9)           | 32 (6.6)             | 0.003  | 0.0  |
| CVD prevalence (%)                                       | 5686 (4.1)           | 45 (9.2)             | <0.001 | 0.0  |
| CVD incidence (%)  | 9038 (6.5)           | 46 (9.4)             | 0.011  | 0.0  |
| Type 2 diabetes prevalence (%)                           | 3593 (2.6)           | 11 (2.3)             | 0.757  | 0.0  |
| Type 2 diabetes incidence (%)                            | 4948 (3.6)           | 19 (3.9)             | 0.776  | 0.0  |

1

1 **Table 2. The classification accuracy of an algorithm for predicting monogenic familial**  
 2 **hypercholesterolaemia (FH) using the multivariable model and LDL-C concentration**  
 3 **accounting for statin use.** There are 93 FH-causing variant positive participants in the test  
 4 data comprising of a total of 27,955 participants.

5

| <b>Predicted probability cut-off</b>           | <b>% sensitivity (95%CI)</b> | <b>% false positive rate (95%CI)</b> | <b>% positive predictive value (95%CI)</b> | <b>% negative predictive value (95%CI)</b> | <b>FH-causing variants below threshold</b> | <b>FH-causing variants above threshold</b> | <b>Controls above threshold</b> |
|--|------------------------------|--------------------------------------|--|--|--|--|---------------------------------|
| <b>Multivariable model</b>                     |                              |                                      |  |  |  |  |                                 |
| <b>0.1</b>                                     | 1.1 (0.2;5.8)                | 0.1 (0.0;0.1)                        | 5.6 (1.0;25.8)                             | 99.7 (99.6;99.7)                           | 92   | 1  | 17                              |
| <b>0.05</b>                                    | 7.5 (3.7;14.7)               | 0.2 (0.1;0.2)                        | 13.0 (6.4;24.4)                            | 99.7 (99.6;99.8)                           | 86   | 7  | 47                              |
| <b>0.02</b>                                    | 20.4 (13.5;29.7)             | 1.1 (0.9;1.2)                        | 6.0 (3.9;9.2)                              | 99.7 (99.7;99.8)                           | 74   | 19   | 296                             |
| <b>0.01</b>                                    | 41.9 (32.4;52.1)             | 4.5 (4.2;4.7)                        | 3.0 (2.2;4.1)                              | 99.8 (99.7;99.8)                           | 54   | 39   | 1244                            |
| <b>0.006</b>                                   | 54.8 (44.7;64.6)             | 11.9 (11.5;12.3)                     | 1.5 (1.2;2.0)                              | 99.8 (99.8;99.9)                           | 42   | 51   | 3311                            |
| <b>0.001</b>                                   | 94.6 (88.0;97.7)             | 87.0 (86.6;87.4)                     | 0.4 (0.3;0.4)                              | 99.9 (99.7;99.9)                           | 5  | 88   | 24240                           |
| <b>Model: LDL-C concentration + statin use</b> |                              |                                      |  |  |  |  |                                 |
| <b>0.1</b>                                     | 0.0 (0.0;4.0)                | 0.0 (0.0;0.1)                        | 0.0 (0.0;35.4)                             | 99.7 (99.6;99.7)                           | 93   | 0  | 7                               |
| <b>0.05</b>                                    | 1.1 (0.2;5.8)                | 0.1 (0.1;0.2)                        | 3.2 (0.6;16.2)                             | 99.7 (99.6;99.7)                           | 92   | 1  | 30                              |
| <b>0.02</b>                                    | 12.9 (7.5;21.2)              | 1.1 (1.0;1.2)                        | 3.8 (2.2;6.5)                              | 99.7 (99.6;99.8)                           | 81   | 12   | 304                             |
| <b>0.01</b>                                    | 38.7 (29.4;48.9)             | 5.6 (5.4;5.9)                        | 2.2 (1.6;3.1)                              | 99.8 (99.7;99.8)                           | 57   | 36   | 1574                            |
| <b>0.006</b>                                   | 52.7 (42.6;62.5)             | 14.6 (14.2;15.0)                     | 1.2 (0.9;1.6)                              | 99.8 (99.8;99.9)                           | 44   | 49   | 4067                            |
| <b>0.001</b>                                   | 90.3 (82.6;94.8)             | 84.0 (83.5;84.4)                     | 0.4 (0.3;0.4)                              | 99.8 (99.6;99.9)                           | 9  | 84   | 23393                           |

6

7

8

1 **Table 3. Net reclassification index (NRI) table and estimates for a predicted probability**  
 2 **threshold of 0.006 comparing the multivariable model to a simpler model of LDL-C**  
 3 **concentration and statin use.** The predicted probability threshold of 0.006 was chosen to  
 4 illustrate the NRI analysis between the multivariable model and the LDL-C with statin use  
 5 model. This threshold choice was chosen as it had a false positive rate close to 10% (11.9%).  
 6 The test dataset of 27,955 participants was used, which included 93 FH variant carriers. NRI  
 7 estimates were obtained via percentile bootstrap method. Pr = probability; Up = reclassified  
 8 to the higher category; Down = reclassified to the lower category; Case = FH-causing variant  
 9 positive; Ctrl = control (negative for an FH-causing variant).

10

| LDL-C + statin use model                 | Multivariable model                     |  | Total  |
|--|---|--|--------|
|  | < 0.006 predicted probability threshold | >= 0.006 predicted probability threshold |        |
| < 0.006 predicted probability threshold  | 22,846                                  | 993                                      | 23,839 |
| >= 0.006 predicted probability threshold | 1,747                                   | 2,369                                    | 4,116  |
| <b>Total</b>                             | 24,593                                  | 3,362                                    | 27,955 |

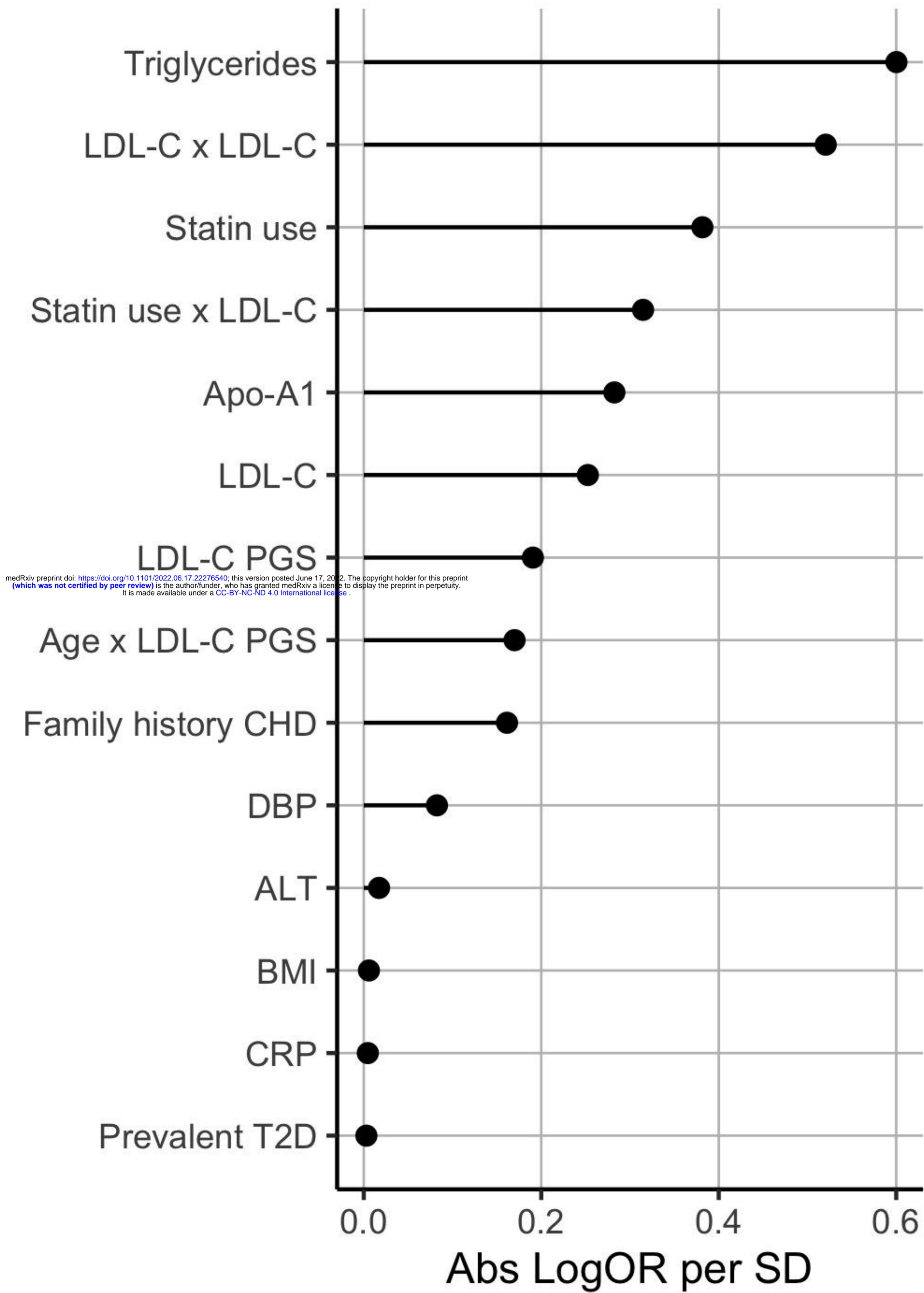
**NRI estimates**

|                       |                       |
|-----------------------|-----------------------|
| <b>NRI:</b>           | 0.049 (-0.037; 0.131) |
| <b>Event NRI:</b>     | 0.022 (-0.063; 0.104) |
| <b>Non-event NRI:</b> | 0.027 (0.023; 0.031)  |
| <b>Pr(Up Case)</b>    | 0.097 (0.038; 0.159)  |
| <b>Pr(Down Case)</b>  | 0.075 (0.026; 0.130)  |
| <b>Pr(Down Ctrl)</b>  | 0.062 (0.060; 0.065)  |
| <b>Pr(Up Ctrl)</b>    | 0.035 (0.033; 0.037)  |

11

12

13

**a****b**