Author Accepted Version || Final published version: J Epidemiol Community Health 2023; doi:10.1136/jech-2023-220801

Predictive value of metabolic profiling in cardiovascular risk scores: analysis of 75,000 adults in UK Biobank

Danyao Jin[1*]; Eirini Trichia[1,2*]; Nazrul Islam[3];

Sarah Lewington[1,2+]; Ben Lacey[1,4+]

1. Clinical Trial Service Unit & Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK

2. MRC Population Health Research Unit, University of Oxford, Oxford, UK

3. Faculty of Medicine, University of Southampton, Southampton, UK

4. UK Biobank, Stockport, Greater Manchester, UK

*Equal contribution; [+] Joint senior authors

Correspondence:

Danyao Jin

Oxford Population Health, Nuffield Department of Population Health (NDPH),

Big Data Institute Building, Old Road Campus, Oxford, OX3 7LF, UK

Email: danyao.jin@ndph.ox.ac.uk

Word count: abstract 229; text 2,508

2 May 2023

Abstract

Background: Metabolic profiling (the extensive measurement of circulating metabolites across multiple biological pathways) is increasing employed in clinical care. However, there is little evidence on the benefit of metabolic profiling as compared to established atherosclerotic cardiovascular disease (CVD) risk scores.

Methods: UK Biobank is a prospective study of 0.5 million participants, aged 40-69 at recruitment. Analyses were restricted to 74,780 participants with metabolic profiling (measured using nuclear magnetic resonance) and without CVD at baseline. Cox regression was used to compare model performance before and after addition of metabolites to QRISK3 (an established CVD risk score used in primary care in England); analyses derived three models, with metabolites selected by association significance or by employing two different machine-learning approaches.

Results: We identified 5,097 incident CVD events within the 10-year follow-up. Harrell's C-index of QRISK3 was 0.750 (95% CI, 0.739-0.763) for women and 0.706 (95% CI, 0.696-0.716) for men. Adding selected metabolites did not significantly improve measures of discrimination in women (Harrell's C-index of three models are 0.759 [0.747-0.772], 0.759 [0.746-0.770], and 0.759 [0.748-0.771], respectively) or men (0.710 [0.701-0.720], 0.710 [0.700-0.719], and 0.710 [0.701-0.719], respectively), and neither did it improve reclassification or calibration.

Conclusion: This large-scale study applied both conventional and machine-learning approaches to assess the potential benefit of metabolic profiling to well-established CVD risk scores. However, there was no evidence that metabolic profiling improved CVD risk prediction in this population.

Thumbnail Sketch

What is already known on this topic

Although previous studies have examined the associations of metabolic biomarkers with incidence and mortality of numerous common diseases, including CVD, there is little evidence on the benefit of metabolic profiling in clinical practice to identify those at high risk of CVD.

What this study adds

This study found no evidence of substantive improvement in prediction accuracy when adding metabolic profiling to a well-established CVD risk score (with information of cholesterol, blood pressure, BMI and medical history). This was despite the use of machine-learning methods to account for complex interactions of highly correlated metabolites.

How this study might affect research, practice and/or policy

As this prospective study of middle-aged adults from the UK general population found no evidence that metabolic profiling improved CVD risk prediction, it is unlikely that such measures would be value for CVD prediction in clinical practice (or as part of national screening programmes) in this population, although replication in other populations (or subgroups, such as young adults or the elderly) is warranted.

Introduction

Early identification of individuals at risk is important for primary prevention of major atherosclerotic cardiovascular disease (CVD). Several risk assessment algorithms have been developed, including the Framingham Risk Score, Systematic COronary Risk Evaluation (SCORE), and Pooled Cohort Equations (PCE) [1-3]. Among these established risk scores, QRISK3 is the most widely used across England's primary health service [4], and NICE are currently recommending that atorvastatin 20mg is considered for the primary prevention of CVD for people with a QRISK3 less than 10% who have non-modifiable CVD risk factors [5]. However, the discrimination of QRISK3 varies from 0.70 to 0.86 in different UK cohorts, and several studies suggested that QRISK3 may not perform very well in older and multi-morbid population [6-8]. Polygenic risk score and lipoprotein(a) have been added to QRISK3 but showed modest improvement in the risk discrimination [9,10]. Therefore, there is still considerable interest in finding new biomarkers to improve prediction accuracy.

Given the metabolic nature of atherosclerosis, circulating metabolic biomarkers are thought to have great potential to improve risk stratification [11]. However, current evidence on the predictive value of metabolites has only focused on a limited number of biomarkers with significant linear associations with CVD, which may not reflect the complex pathophysiology of atherosclerosis [12,13]. Nuclear magnetic resonance (NMR) spectroscopy is a high-throughput technology used for metabolic profiling of numerous metabolites across multiple biological pathways, and is being used in large-scale prospective studies [14]. Therefore, when assessing the predictive value, the large number of metabolites measured through NMR and their complex interrelations need to be accounted for. Machine learning has been increasingly used for development of prediction models, with the strengths of incorporating highly

correlated features and complex interactions that cannot be captured by traditional statistical models.

In this study, we aimed to evaluate whether adding circulating metabolic profiling to a well-established risk score using machine-learning methods improved the prediction of 10-year CVD risk.

Methods

Study design and population

UK biobank is a prospective cohort study of approximately 500,000 adults in the United Kingdom recruited from 2006 to 2010 [15,16]. All participants, aged 40-69 at study entry, completed questionnaires and physical measurements, and had biological samples collected at recruitment. Ethics approval was given by the North West Multicentre Research Ethics Committee, and the study was conformed to the principles embodied in the Declaration of Helsinki.

Measurement of metabolic profiling

NMR spectroscopy (Nightingale Health, Finland) was used for metabolic profiling of the baseline plasma samples of 117,980 participants (a random subset of the initial cohort) [17]. To decrease the interference from some unstable biomarkers and to avoid the overfitting due to large number of lipids-related biomarkers, of the metabolites available, the main analyses only included 39 metabolites all measured with comparable validity to clinical chemistry, as the candidate biomarkers (***Table S1***) [18]. In the sensitivity analyses, we expanded the candidate metabolites to a larger scope of NMR-derived metabolites that available in the UK Biobank (***Table S2***) [18].

Definition of risk scores

In the main analyses, the metabolites were added to QRISK3, an established risk score widely used across England's primary health service [4]. QRISK includes information on age, ethnicity, deprivation, systolic blood pressure (SBP), body mass index (BMI), total cholesterol to HDL cholesterol ratio (measured by traditional chemistry method), smoking status, family history of coronary heart disease, and medical history of a series of diseases, which were selected based on Bayes information criterion. In the sensitivity analyse, QRISK3 was replaced by SCORE2, which was another algorithm for risk prediction of CVD that widely used in European population, scoring by age, smoking status, SBP, and total and HDL cholesterol. Detailed definitions of QRISK3 and SCORE2 variables and mapping in the UK Biobank are provided in the ***Supplementary Methods*** and ***TableS3.***

Ascertainment of incident CVD

Incident CVD was defined as the first-ever coronary heart disease, ischemic stroke, or transient ischemic attack, identified from Hospital Episode Statistics (including diagnostic codes and relevant procedures) and the Office for National Statistics cause of death data, using codes of the 10th edition of the International Classification of Disease (ICD-10), and coronary-related procedures (coronary artery bypass surgery or percutaneous transluminal angioplasty stent placement) by the OPCS Classification of Interventions and Procedures (***Table S4***).

Statistical analysis

The analyses were restricted to participants without prior CVD and those not taking statins at baseline, and further excluded the participants with missing or outlying in QRISK3 variables (***Figure S1***). Since the participants in the UK Biobank are overall healthier (with lower incidence of CVD) than the general UK population, QRISK3 score was recalibrated by refitting the baseline survival function to the study population (***Supplementary Methods***).

The candidate metabolites were selected in three ways: (1) adding the metabolites that were significantly associated with CVD (independently from QRISK3 score) to QRISK3; (2) adding all metabolites to QRISK3 and penalized by elastic-net; (3) adding the novel metabolites selected by Boruta SHapley Additive exPlanations (BorutaSHAP) based on Extreme gradient boosting algorithm (XGBoost) to QRISK3. Elastic-net is a regression method that performs regularization and variable selection simultaneously, with the strength of handling highly correlated variables [19]. XGBoost is a tree-based machine-learning method where new models are created that predict the residuals or errors of prior models and then added together to make the final prediction [20,21]. It allows for including higher-order interactions and accounting for complex nonlinear relationships, and was chosen as our third model because of its modest computational cost and outstanding performance of risk prediction in recent studies involving a large number of proteins or metabolites [22,23]. BorutaSHAP is a wrapper feature selection method to explain how much each factor in a model has contributed to the prediction, and the combination with Boruta feature selection algorithm ensures a faster and more stable feature selection [24]. Detailed explanations of the machine-learning and feature selection methods are provide in ***Supplementary Methods***. The hyperparameters were fine-tuned using five-fold cross-validation (***Table S5***). In all three cases, prediction performance was assessed using Cox proportional hazards regression w/o the metabolites. Bootstrapping (500 times) was applied to evaluate the optimism of the models.

Harrell's C-index was used to assess the discriminatory ability (how the model separate cases from controls) of each model. The improvement in reclassification after adding metabolites was evaluated by the integrated discrimination improvement (IDI) and net reclassification improvement (NRI). IDI summarises the extent that a new model increases risk in events and decreases risk in non-events compared with the old model, while NRI quantifies the appropriateness of the change in predicted probabilities or categorised risk group when

changing from old to new model. 10-year probability of event > 10% was categorized as high risk and set as the cut-off for categorical NRI. The calibration, measuring how close the predicted probability is to the observed risk, was assessed with calibration plots at 10 years. All analyses followed the suggestions from TRIPOD [25], and all models were developed and evaluated separately for men and women in Python 3.9.12.

Results

After exclusions, 74,780 participants remained, with mean age of 55 years at study entry. The overall baseline characteristics of the study population was similar to the whole UK Biobank population (***Table S6***). Among the study population, 44% were men, 10% were current smokers, and 41% reported to have family history of heart disease. After a 10-year follow-up, 5,097 (6.8%) incident CVD events occurred, with about twice the rate in men than women (9.4% vs 4.8%). Compared to participants that did not have an incident CVD event, those with incident CVD were on average older, with higher BMI, systolic blood pressure and higher ratio of total cholesterol to HDL cholesterol, and more likely to be men and current moderate/heavy smokers. Participants who experienced CVD during follow-up were also more likely to have family heart disease history and baseline chronic disease history (***Table 1***).

The hazard ratio of the recalibrated QRISK3 score was 1.17 (95%CI, 1.15-1.18) per one point higher in women and 1.08 (1.07-1.09) in men. Independently from QRISK3 score, twelve metabolites (HDL cholesterol, two apolipoproteins biomarkers, six fatty acids ratio biomarkers, histidine, albumin and glycoprotein acetyls) in women and five (very-low-density lipoprotein [VLDL] cholesterol, ApolipoproteinB [ApoB] to ApolipoproteinA-1 [ApoA-1] ratio, omega-3 fatty acids concentration and its ratio to total fatty acids, albumin and glycoprotein acetyls) in men remained significantly associated with CVD (***Table 2***). In the two machine-learning models of both sexes, fewer fatty acids were selected, but some amino acids

and glycolysis-related metabolites were included as predictors. Compared with the selection criteria by association significance (first model), albumin and glycoprotein acetyls were also selected by the two machine-learning models for both sexes, while total triglycerides in women and glycine and leucine in men were newly selected as novel metabolites by the two machine-learning models (***Table S7***).

Harrell's C-index of the recalibrated QRISK3 was 0.750 (95% CI, 0.739-0.763) for women and 0.706 (95% CI, 0.696-0.716) for men (***Table 3***). Adding metabolites to QRISK, in all three models, did not improved the discrimination in women (C-index of three models are 0.759 [0.747-0.772], 0.759 [0.746-0.770], and 0.759 [0.748-0.771], respectively), or men (0.710 [0.701-0.720], 0.710 [0.700-0.719], and 0.710 [0.701-0.719], respectively). The reclassification showed no improvement after adding the metabolites, with statistically significant relative IDI, but less than 0.5% in all three models of both sexes. Although the continuous NRI showed statistically significant increase in most models, the categorical NRI (setting 10-year event probability≥10% as high risk), which is a better measure of reclassification, showed no improvement in either men or women. Calibration plots did not show any significant change either (***Figure***).

The hazard ratios (per one point higher) of the recalibrated SCORE2 were 1.12 (1.10-1.13) in women and 1.07 (1.06-1.07) in men (***Table S8***). Replacing QRISK3 by SCORE2 had limited impact on the selection of novel metabolites in all three models, of which XGBoost selected the exactly same metabolites as using QRISK3 as the basic score (***Table S9***). Meanwhile, adding metabolites to SCORE2 did not significantly improve the overall prediction accuracy, although some slight improvements were observed in continuous NRI, which may largely due to the poorer performance of SCORE2 in the study population (Harrell's C-index of SCORE2 were 0.731 [0.718-0.744] in women and 0.689 [0.679-0.699] in men) (***Table S10, Figure S2***). Similarly, there was no evidence of prediction improvement when expanding the scope of the

candidate metabolites (***Table S11, Figure S3***). Among individuals who currently identified as low-risk (10-year predicted risk less than 10%), risk categorisation (measured by categorical NRI) after adding metabolites to QRISK3 showed no improvement in women and limited improvement (less than 6%) in men.

Discussion

This large-scale prospective study examined the predictive value of adding high-throughput metabolic profiling to an established risk score among 75,000 participants in UK Biobank. To our knowledge, this is the first study to assess the additional predictive value of high-throughput circulating metabolites to a well-established CVD risk score. The application of machine-learning approaches allows for highly correlated variables and accounts for the complex interactions between metabolites in atherosclerosis. However, compared with the standard QRISK3 score, there was no evidence of substantive improvement in prediction of 10-year risk of CVD after adding the metabolic biomarkers.

Several previous studies have examined the value of metabolic profiling measured by NMR for the prediction of cardiovascular event or subclinical atherosclerosis [12,13,26]. Two of these studies, both of which used traditional statistical algorithms, found moderate improvement in discrimination or reclassification, but neither included BMI as an established risk factor in the basic models. One other recent study used risk factors including BMI in the basic model, and observed very slight C-index improvement of coronary heart disease prediction (0.003 [0.001, 0.004]) and no improvement of cerebral stroke prediction (0.001 [-0.003, 0.005]) when adding metabolomics [26]. However, the basic model of this study still lacked detailed information on several major risk factors, such as family history of heart disease. By contrast, QRISK3 is a score developed from more comprehensive risk factors including BMI, cholesterol level, family history and aspects of medical history and mediations.

Similarly, when using the SCORE2 (a risk score not including BMI and medical history as risk factors) as the basic score in our sensitivity, adding metabolites showed a slight improvement in continuous NRI due to the poorer performance of the original SCORE2, however, the overall prediction accuracy that measured by C-index were not significantly improved.

Two other cohorts have examined the predictive value of metabolites measured by mass spectrometry [27,28], which is another type of high-throughput technique for metabolic profiling with the capability of detecting thousands of metabolites [29]. One study used traditional statistical algorithms and the other applied elastic-net and principal components analysis, and they both observed modest improvement in the prediction of coronary heart disease or subclinical CVD. However, similar as the previous evidence on NMR-derived metabolites, neither of the studies compared the prediction performance with any established risk score. Moreover, because mass spectrometry is more expensive and time-consuming than NMR, the sample size of both studies was relatively small (less than 3,000 individuals).

As a result of selecting metabolites that were associated with CVD independently from the QRISK score, our study identified novel potential predictors for cardiovascular risk by using two different machine-learning algorithms. Elastic-net allows for handling highly correlated variables and enhances the prediction accuracy by regularization, while XGBoost is a novel tree-based model with the strength of incorporating complex variables interactions that cannot be captured by traditional statistics model. Additionally, BorutaSHAP is a relative stable feature selection algorithm using shapely value, which provides another way of measuring feature importance other than association. Although prediction performance was not improved in our results, applying machine-learning algorithms gave insight into the predictive value of some amino acids and glycolysis related metabolites that have previously been were overlooked in association analyses under linear assumption, and such selection were proved to

be robust because most of the metabolites remained to be select as novel biomarkers when changing to use SCORE2 as the basic score in the sensitivity analyses.

This study has a number of key strengths. It uses large-scale metabolite profiling and applies machine-learning algorithms. The linkage to NHS electronic health records and national death registries limited loss to follow-up and allowed reliable ascertainment of CVD events. In addition, the use of different analytical methods with different assumptions showed that our results were robust against different assumptions. However, as about 95% of participants are white in the UK Biobank, it's difficult to generalise our results to other ethnicities; more studies are needed in diverse populations and with longer follow-up to compare with other 10-year or life-time risk scores. Further, the UK Biobank are generally healthier than the wider UK population and only included participants aged 40-69. Future analyses should assess the benefit of metabolic profiling to cardiovascular risk in wider age range, in non-white and high-risk individuals, and explore the predictive value of other types of metabolites (e.g. gut microbiome).

Conclusion

This large-scale prospective study provides evidence that compared with an established risk score with information on BMI and medical history, adding circulating metabolic profiling measured by NMR spectroscopy is unlikely to lead to a substantive improvement in CVD risk prediction in primary care.

Statement and Declarations

Funding This research used UK Biobank data assets made available by National Safe Haven as part of the Data and Connectivity National Core Study, led by Health Data Research UK in partnership with the Office for National Statistics and funded by UK Research and Innovation (grant ref MC_PC_20058).

Completing Interests SL reports grants from the Medical Research Council (MRC) and research funding from the US Centers for Disease Control and Prevention Foundation (with support from Amgen) and from the World Health Organization during the conduct of the study. The Clinical Trial Service Unit and Epidemiological Studies Unit (CTSU) receives research grants from industry that are governed by University of Oxford contracts that protect its independence and has a staff policy of not taking personal payments from industry; further details can be found at https://www.ndph.ox.ac.uk/files/about/ndph-independence-of-research-policy-jun-20.pdf. All other authors declared no conflict of interest.

Author Contribution Jin and Trichia contributed equally to this work. Jin and Trichia had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Islam and Lacey critically revised the manuscript and contributed important intellectual content. Lacey and Lewington is the guarantor of this work.

Ethics Approval Ethical approval was obtained by the North West Research Ethics Committee (IRAS project ID: 299116).

Consent to Participate Informed consent was obtained from all individual participants included in the study**.**

Consent to Publish Written informed consent was provided by each participant in the study for publication**.**

Data statement This research used the UK Biobank resource (application number 31461). This work uses data provided by patients and collected by the NHS as part of their care and support. Data from the UK Biobank are available to researchers after registration at the UK Biobank server. The data cleaning and coding used to generate the findings of this study are available from the corresponding author on reasonable request.

References

1. Goff DC, Jr., Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, Greenland P, Lackland DT, Levy D, O'Donnell CJ, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation.* 2014;129:S49-73.
2. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation.* 1998;97:1837-1847.
3. Conroy RM, Pyörälä K, Fitzgerald AP, Sans S, Menotti A, De Backer G, De Bacquer D, Ducimetière P, Jousilahti P, Keil U, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *European heart journal.* 2003;24:987-1003.
4. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ (Clinical research ed).* 2017;357:j2099.
5. National Institute for Health and Care Excellence. NICE guideline CG181. Cardiovascular disease: risk assessment and reduction, including lipid modification-Evidence review for statins: efficacy and adverse effects. 2023; https://www.nice.org.uk/guidance/cg181/evidence/c-risk-assessment-and-reduction-including-lipid-modification-pdf-13065827440. Accessed August 17, 2023.
6. Livingstone SJ, Guthrie B, Donnan PT, Thompson A, Morales DR. Predictive performance of a competing risk cardiovascular prediction tool CRISK compared to QRISK3 in older people and those with comorbidity: population cohort study. *BMC medicine.* 2022;20:152.
7. Parsons RE, Liu X, Collister JA, Clifton DA, Cairns BJ, Clifton L. Independent external validation of the QRISK3 cardiovascular disease risk prediction model using UK Biobank. *Heart (British Cardiac Society).* 2023.
8. Livingstone S, Morales DR, Donnan PT, Payne K, Thompson AJ, Youn JH, Guthrie B. Effect of competing mortality risks on predictive performance of the QRISK3 cardiovascular risk prediction tool in older people and those with comorbidity: external validation population cohort study. *The lancet Healthy longevity.* 2021;2:e352-e361.
9. Elliott J, Bodinier B, Bond TA, Chadeau-Hyam M, Evangelou E, Moons KGM, Dehghan A, Muller DC, Elliott P, Tzoulaki I. Predictive Accuracy of a Polygenic Risk Score-Enhanced Prediction Model vs a Clinical Risk Score for Coronary Artery Disease. *Jama.* 2020;323:636-645.
10. Trinder M, Uddin MM, Finneran P, Aragam KG, Natarajan P. Clinical Utility of Lipoprotein(a) and LPA Genetic Risk Score in Risk Prediction of Incident Atherosclerotic Cardiovascular Disease. *JAMA cardiology.* 2020;6:1-9.
11. Shah SH, Kraus WE, Newgard CB. Metabolomic profiling for the identification of novel biomarkers and mechanisms related to common cardiovascular diseases: form and function. *Circulation.* 2012;126:1110-1120.
12. Würtz P, Havulinna AS, Soininen P, Tynkkynen T, Prieto-Merino D, Tillin T, Ghorbani A, Artati A, Wang Q, Tiainen M, et al. Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation.* 2015;131:774-785.
13. Würtz P, Raiko JR, Magnussen CG, Soininen P, Kangas AJ, Tynkkynen T, Thomson R, Laatikainen R, Savolainen MJ, Laurikka J, et al. High-throughput quantification of circulating metabolites improves prediction of subclinical atherosclerosis. *European heart journal.* 2012;33:2307-2316.
14. Soininen P, Kangas AJ, Würtz P, Suna T, Ala-Korpela M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circulation Cardiovascular genetics.* 2015;8:192-206.

15. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 2018;562:203-209.
16. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine.* 2015;12:e1001779.
17. The UK Biobank. Nightingale Health Metabolic Biomarkers: Phase 1 Release. 2020; https://biobank.ndph.ox.ac.uk/showcase/ukb/docs/nmrm_companion_doc.pdf. . Accessed May 31, 2022.
18. Nightingale Health. Biomarkers list analyzed by Nightingale Health. 2022; https://research.nightingalehealth.com/biomarkers. Accessed 6 Sept, 2022.
19. Zou H, Hastie T. Regularization and Variable Selection Via the Elastic Net. *Journal of the Royal Statistical Society Series B: Statistical Methodology.* 2005;67:301-320.
20. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016; San Francisco, California, USA.
21. Brownlee J. A Gentle Introduction to XGBoost for Applied Machine Learning. 2016; https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/. Accessed April 4, 2023.
22. Unterhuber M, Kresoja KP, Rommel KP, Besler C, Baragetti A, Klöting N, Ceglarek U, Blüher M, Scholz M, Catapano AL, et al. Proteomics-Enabled Deep Learning Machine Algorithms Can Enhance Prediction of Mortality. *Journal of the American College of Cardiology.* 2021;78:1621-1631.
23. Khera R, Haimovich J, Hurley NC, McNamara R, Spertus JA, Desai N, Rumsfeld JS, Masoudi FA, Huang C, Normand SL, et al. Use of Machine Learning Models to Predict Death After Acute Myocardial Infarction. *JAMA cardiology.* 2021;6:633-641.
24. Keany E. BorutaShap : A wrapper feature selection method which combines the Boruta feature selection algorithm with Shapley values. *Zenodo.* 2020.
25. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Annals of internal medicine.* 2015;162:W1-73.
26. Buergel T, Steinfeldt J, Ruyoga G, Pietzner M, Bizzarri D, Vojinovic D, Upmeier Zu Belzen J, Loock L, Kittner P, Christmann L, et al. Metabolomic profiles predict individual multidisease outcomes. *Nature medicine.* 2022;28:2309-2320.
27. Cruz DE, Tahir UA, Hu J, Ngo D, Chen ZZ, Robbins JM, Katz D, Balasubramanian R, Peterson B, Deng S, et al. Metabolomic Analysis of Coronary Heart Disease in an African American Cohort From the Jackson Heart Study. *JAMA cardiology.* 2022;7:184-194.
28. Murthy VL, Reis JP, Pico AR, Kitchen R, Lima JAC, Lloyd-Jones D, Allen NB, Carnethon M, Lewis GD, Nayor M, et al. Comprehensive Metabolic Phenotyping Refines Cardiovascular Risk in Young Adults. *Circulation.* 2020;142:2110-2127.
29. Emwas AH. The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. *Methods in molecular biology (Clifton, NJ).* 2015;1277:161-193.

Figure Legends

Figure. Calibration of risk prediction models for 10-year CVD risk

Calibration of risk prediction models for 10-year CVD risk. For each model, the observed and predicted CVD event rates are shown for each of 10 equally sized groups of absolute predicted risk. Vertical lines represent 95% CIs (bootstrap percentile confidence interval, bootstrap for 500 times).

Table 1. Characteristics of baseline QRISK factors by 10-year incident CVD

	Incident CVD		
	No	Yes	All
No. of participants	69,683	5,097	74,780
Age, sex and socioeconomic factors			
Men, %	42.0	59.8	43.3
Baseline age, years	55.0 (8.0)	59.6 (7.0)	55.3 (8.0)
White, %	94.8	95.7	94.9
Townsend Deprivation Index*	-1.5 (2.9)	-1.3 (3.0)	-1.4 (3.0)
Anthropometry, blood pressure, and lipids by clinical chemistry			
Body Mass Index, kg/m^2	26.9 (4.5)	27.9 (4.6)	27.0 (4.6)
Systolic blood pressure, mmHg	136.2 (17.9)	143.7 (18.1)	136.7 (18.4)
Standard deviation between two readings†, mmHg	5.1 (4.0)	5.5 (4.2)	5.2 (4.0)
Total cholesterol to HDL-C ratio	4.2 (1.1)	4.5 (1.2)	4.2 (1.1)
Smoking intensity, %			
Ex-smoker	32.4	35.7	32.7
Light smoker (< 10 per day)	4.8	4.9	4.8
Moderate smoker (10-19 per day)	2.9	4.7	3.0
Heavy smoker (≥20 per day)	2.3	4.6	2.4
Family history of heart disease‡, %	39.5	50.5	40.2
Disease and medication history, %			
Type 1 diabetes	0.3	0.6	0.3
Type 2 diabetes	1.6	3.6	1.7
Chronic kidney disease (stage 3,4,5)	1.5	2.8	1.6
Atrial fibrillation	0.8	2.6	0.9
Migraines	4.5	5.2	4.6
Rheumatoid arthritis	1.1	2.4	1.2
Systemic lupus erythematosus	0.1	0.3	0.1
Severe mental illness§	5.0	5.5	5.0
Erectile dysfunction	0.2	0.5	0.2
Hypertension treatment	11.6	23.0	12.3
Atypical antipsychotic medication	0.2	0.2	0.2
Regular steroid tablets	0.7	1.8	0.8

Sex adjusted characteristics of QRISK factors at baseline by 10-year incident ASCVD. Continuous variables are presented as mean (standard deviation) and categorical variables are presented as column percentages. *Higher values indicate higher levels of material deprivation; †QRISK asks for standard deviation of systolic blood pressure values recorded in the five years before study entry, but UK biobank only provided two automated or manual readings at study entry; ‡QRISK asks for the family history in first degree relatives aged less than 60 years, but UK biobank only identified family history in first degree relatives in all ages; §Includes schizophrenia, bipolar disorder and moderate/severe depression. HDL-C=high-density lipoproteins cholesterol.

Table 2. Associations of clinical metabolites independent from QRISK3 score

	Hazard ratio (95% CI)	
	Women	**Men**
Recalibrated QRISK3 score	**1.17 (1.15, 1.18)***	**1.08 (1.07, 1.09)***
Cholesterols & Triglycerides		
Total cholesterol	0.96 (0.91, 1.00)	1.01 (0.97, 1.05)
VLDL cholesterol	1.04 (0.99, 1.08)	1.01 (0.97, 1.05)
LDL cholesterol	0.98 (0.94, 1.03)	1.01 (0.98, 1.05)
HDL cholesterol	**0.89 (0.85, 0.93)***	0.98 (0.94, 1.02)
Total triglycerides	1.02 (0.98, 1.07)	0.97 (0.94, 1.02)
Fatty acids		
Total fatty acids	1.01 (0.96, 1.05)	0.98 (0.95, 1.02)
Omega-3 fatty acids	0.96 (0.92, 1.00)	**0.94 (0.91, 0.97)***
Omega-6 fatty acids	0.97 (0.93, 1.01)	1.00 (0.96, 1.04)
Polyunsaturated fatty acids	0.96 (0.92, 1.01)	0.98 (0.95, 1.02)
Monounsaturated fatty acids	1.05 (1.00, 1.09)	0.98 (0.95, 1.02)
Saturated fatty acids	1.01 (0.97, 1.06)	0.98 (0.95, 1.02)
Docosahexenoic acid	0.95 (0.91, 0.99)	0.95 (0.92, 0.98)
Linoleic acid	0.96 (0.92, 1.00)	1.00 (0.96, 1.03)
Omega-3 to total fatty acids	0.95 (0.91, 0.99)	**0.94 (0.90, 0.97)***
Omega-6 to total fatty acids	**0.93 (0.89, 0.97)***	1.03 (0.99, 1.07)
Polyunsaturated to total fatty acids	**0.92 (0.88, 0.95)***	1.01 (0.97, 1.05)
Monounsaturated to total fatty acids	**1.13 (1.08, 1.18)***	1.00 (0.96, 1.04)
Saturated to total fatty acids	1.02 (0.98, 1.06)	0.99 (0.96, 1.03)
Docosahexaenoic acid to total fatty acids	**0.94 (0.89, 0.98)***	0.96 (0.92, 0.99)
Linoleic acid to total fatty acids	**0.92 (0.88, 0.96)***	1.02 (0.99, 1.06)
Polyunsaturated to monounsaturated fatty acids	**0.88 (0.84, 0.92)***	1.00 (0.96, 1.04)
Omega-6 to omega-3 fatty acids	1.02 (0.98, 1.07)	1.04 (1.01, 1.08)
Apolipoproteins		
Apolipoprotein B	1.02 (0.97, 1.06)	1.02 (0.99, 1.06)
Apolipoprotein A-1	**0.91 (0.87, 0.95)***	0.96 (0.93, 1.00)
Apolipoprotein B to apolipoproteinA-1	**1.07 (1.02, 1.12)***	**1.06 (1.02, 1.10)***
Amino acids		
Alanine	1.02 (0.98, 1.07)	0.98 (0.94, 1.01)
Glycine	0.95 (0.91, 0.99)	0.96 (0.92, 0.99)
Histidine	**0.91 (0.87, 0.95)***	0.97 (0.93, 1.00)
Isoleucine	1.02 (0.98, 1.06)	1.01 (0.98, 1.05)
Leucine	1.00 (0.96, 1.05)	1.00 (0.97, 1.04)
Valine	0.99 (0.95, 1.03)	0.98 (0.95, 1.02)
Total branched-chain amino acids	1.00 (0.96, 1.04)	0.99 (0.96, 1.03)
Phenylalanine	1.05 (1.01, 1.09)	1.04 (1.01, 1.08)
Tyrosine	1.00 (0.96, 1.05)	1.01 (0.97, 1.04)
Glycolysis related metabolites		
Glucose	1.02 (0.98, 1.06)	1.01 (0.98, 1.04)
Lactate	1.03 (0.99, 1.08)	0.99 (0.95, 1.02)
Fluid balance		
Creatinine	1.02 (0.98, 1.06)	1.01 (0.98, 1.04)
Albumin	**0.88 (0.84, 0.92)***	**0.91 (0.88, 0.94)***
Inflammation		
Glycoprotein acetyls	**1.14 (1.09, 1.19)***	**1.06 (1.02, 1.10)***

Hazard ratios (HR) per one score higher of concentration. HR of each metabolite was calculated by Cox proportional-hazards regression with adjustment of QRISK3 score. *Associations remained significant (p-value<0.01) by correction of false discovery rate using Benjamini-Hochberg method.

Table 3. Comparing prediction performance of 10-year CVD risk w/o metabolites

Prediction Performance	Women (95% CI*)	Men (95% CI)
Recalibrated QRISK3		
Harrell's C-index †	0.750 (0.739, 0.763)	0.706 (0.696, 0.716)
Adding metabolites associated with CVD independently from QRISK3 score		
C-statistics	0.759 (0.747, 0.772)	0.710 (0.701, 0.720)
IDI‡ (%)	0.30 (0.17, 0.41)	0.20 (0.12, 0.28)
Continuous NRI§ (%)	12.4 (6.7, 16.6)	6.8 (2.7,11.6)
events	6.5 (1.0, 10.8)	4.0 (0.0, 8.3)
non-events	5.9 (5.0, 6.8)	2.8 (1.8, 3.9)
Categorical NRI (%)	0.3 (-1.8, 0.9)	0.9 (-0.2, 2.0)
events	0.4 (-1.2, 1.5)	0.4 (-0.7, 1.4)
non-events	-0.7 (-0.8, -0.5)	0.5 (0.3, 0.8)
Adding metabolites with regularization (using Elastic-net)		
Harrell's C-index	0.759 (0.746, 0.770)	0.710 (0.700, 0.719)
IDI (%)	0.16 (0.03, 0.26)	0.16 (0.04, 0.25)
Continuous NRI (%)	4.4 (-0.7, 9.6)	7.4 (3.3,11.0)
events	4.7 (-0.3, 9.9)	5.2 (1.4, 8.8)
non-events	-0.3 (-1.3, 0.7)	2.2 (1.2, 3.3)
Categorical NRI (%)	-0.3 (-1.6, 1.1)	0.7 (-0.8, 1.8)
events	0.2 (-1.2, 1.5)	0.3 (-1.1, 1.5)
non-events	-0.4 (-0.5, -0.3)	0.4 (0.1, 0.7)
Adding metabolites selected by BorutaSHAP from XGBoost		
Harrell's C-index	0.759 (0.748, 0.771)	0.710 (0.701, 0.719)
IDI (%)	0.26 (0.11, 0.38)	0.13 (0.03, 0.20)
Continuous NRI (%)	14.7 (9.2, 19.7)	5.5 (1.7, 9.5)
events	2.7 (-2.9, 7.7)	-0.1 (-4.0, 3.4)
non-events	12.0 (11.0, 12.9)	5.9 (4.9, 6.9)
Categorical NRI (%)	0.0 (-1.6, 1.3)	0.7 (-0.5, 1.8)
events	0.6 (-0.9, 1.9)	0.3 (-0.9, 1.2)
non-events	-0.6 (-0.7, -0.5)	0.5 (0.2, 0.7)

Comparing prediction performance of 10-year CVD risk w/o metabolites. In all models, metabolites are added to recalibrated QRISK3 using Cox proportional-hazards regression. Hyper-parameters of each model are in appendix. *Bootstrap percentile confidence interval, bootstrap for 500 times; †Harrell's C-index, measuring the probability that a randomly selected subject with shorter time-to-event will have a higher predicted probability of event than a randomly selected subject with longer time-to-event; ‡Integrated discrimination improvement, summarising the extent a new model increases risk in events and decreases risk in non-event compared with the old model; §Net reclassification improvement, quantifying the appropriateness of the change in predicted probabilities or categorised risk group when changing from old to new model; Categorical NRI is based on a 10% risk threshold.

Author Accepted Version || Final published version: J Epidemiol Community Health 2023; doi:10.1136/jech-2023-220801

Predictive value of metabolic profiling in cardiovascular risk scores: analysis of 75,000 adults in UK Biobank

Supplementary Appendix

Danyao Jin, MSc[1*]; Eirini Trichia, PhD[1,2*]; Nazrul Islam, PhD[1,3]; Sarah Lewington, DPhil[1,2+]; Ben Lacey, DPhil[1,4+]

Table of Contents

Supplementary Methods 2
QRISK3 variables and mapping in UK Biobank 2
SCORE2 variables[3] 3
Recalibration of QRISK3 and SCORE2 4
Elastic-net 5
XGBoost 6
BorutaSHAP 7
Assessment of prediction performance 8
Reference 9
Table S1: List of clinically validated metabolites for main analyses 10
Table S2: List of metabolites for sensitivity analyses 11
Table S3: Disease and medication codes of QRISK3 variables in UK biobank 12
Table S4: ICD-10 and operation code of cardiovascular disease 13
Figure S1: Flowchart of exclusion criteria for study population in UK Biobank 15
Table S5: Fine-tuning of hyper-parameters (QRISK3) 16
Table S6: Baseline characteristics in UK Biobank versus in study population 17
Table S7: List of selected novel metabolites by different methods (QRISK3) 18
Table S8. Associations of clinical metabolites independent from SCORE2 19
Table S9: List of selected metabolites using different methods (SCORE2) 20
Table S10: Comparing prediction performance of 10-year CVD risk w/o metabolites (SCORE2) 21
Figure S2: Calibration of risk prediction models for 10-year CVD risk (SCORE2) 22
Table S11: Prediction performance of 10-year ASCVD risk w/o metabolites (QRISK3 and wider scope of candidate metabolites) 23
Figure S3: Calibration of risk prediction models for 10-year ASCVD risk (QRISK3 and wider scope of candidate metabolites) 24

Supplementary Methods

QRISK3 variables and mapping in UK Biobank[1,2]

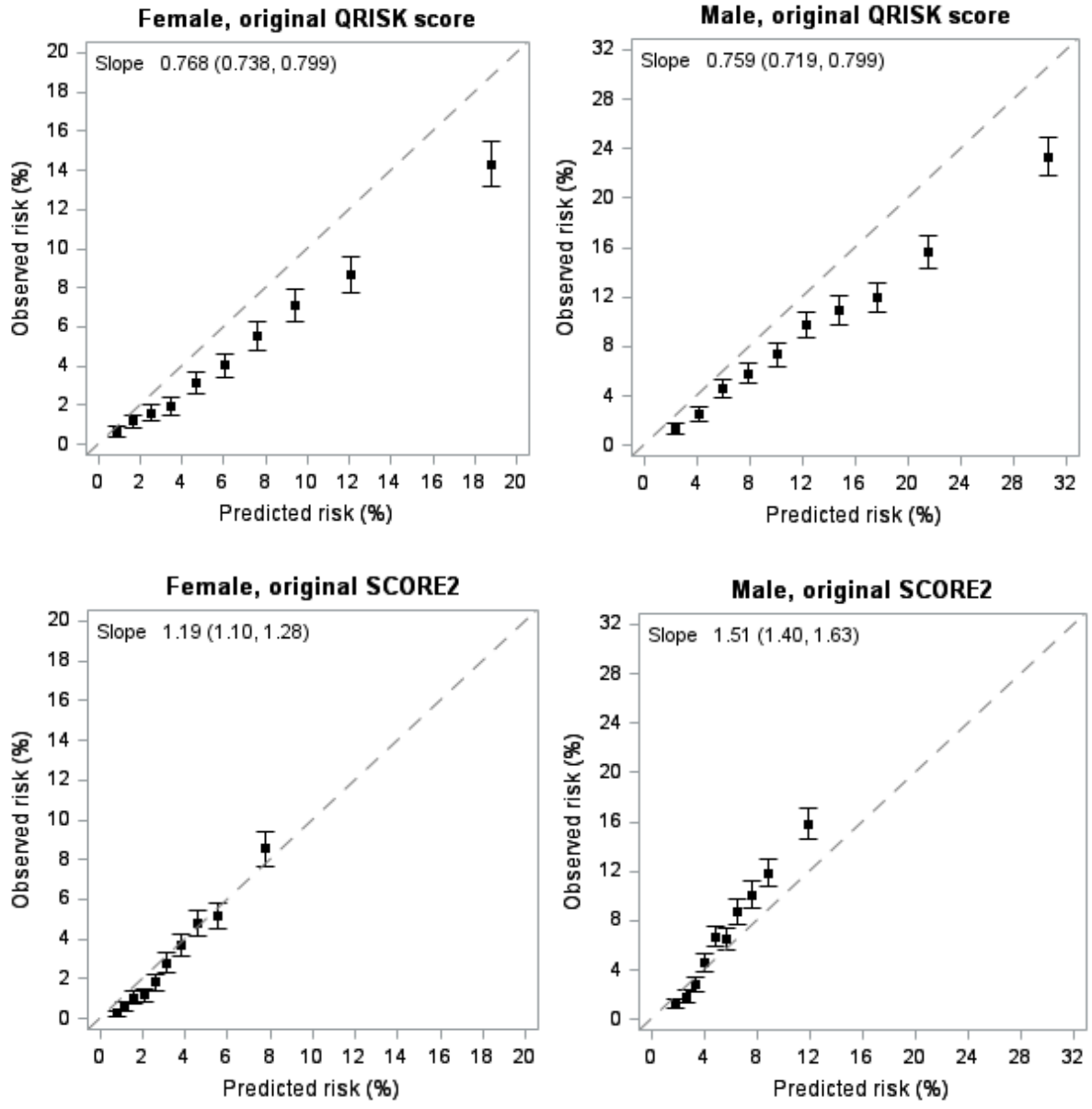
- Age at study entry (years)
- Ethnic origin (White or not state; Indian; Pakistani; Bangladeshi; Other Asian; Black Caribbean; Black African; Chinese; Other ethnic group): *our study only included White participants for analyses*
- Deprivation (as measured by the Townsend score, where higher values indicate higher levels of material deprivation)
- Systolic blood pressure (SBP) (mmHg)
- Measure of systolic blood pressure variability (standard deviation of repeated measures): *UK Biobank does not include information on variability in SBP. Our study derived this variable by the standard deviation between two automated or manual SBP readings at baseline (Variable ID 4080 and 93).*
- Body mass index (kg/m^2)
- Total cholesterol-to-high density lipoprotein cholesterol ratio
- Smoking status (non-smoker, former smoker, light smoker (1-9/day), moderate smoker (10-19/day), or heavy smoker (≥20/day)):
- Family history of coronary heart disease in a first-degree relative aged less than 60 years: *UK Biobank includes illnesses in father (Variable ID 20107), illnesses in mother (Variable ID 20110), and illnesses of siblings (Variable ID 20111), but does not have information on age at diagnosis. Our study assumed age less than 60 years at diagnosis.*
- Diabetes (type 1, type 2, or no diabetes)
- Treated hypertension (diagnosis of hypertension and treatment with at least one antihypertensive drug)
- Rheumatoid arthritis (diagnosis of rheumatoid arthritis, Felty's syndrome, Caplan's syndrome, adult onset Still's disease, or inflammatory polyarthropathy not otherwise specified)
- Atrial fibrillation (including atrial fibrillation, atrial flutter, and paroxysmal atrial fibrillation)
- Chronic kidney disease (stage 3, 4 or 5) and major chronic renal disease (including nephrotic syndrome, chronic glomerulonephritis, chronic pyelonephritis, renal dialysis, and renal transplant)
- Diagnosis of migraine (including classic migraine, atypical migraine, abdominal migraine, cluster headaches, basilar migraine, hemiplegic migraine, and migraine with or without aura)
- Corticosteroid use (including oral or parenteral prednisolone, betamethasone, cortisone, depo-medrone, dexamethasone, deflazacort, efcortesol, hydrocortisone, methylprednisolone, or triamcinolone)
- Systemic lupus erythematosus (including diagnosis of SLE, disseminated lupus erythematosus, or Libman-Sacks disease)
- Second generation "atypical" antipsychotic use (including amisulpride, aripiprazole, clozapine, lurasidone, olanzapine, paliperidone, quetiapine, risperidone, sertindole, or zotepine)
- Diagnosis of severe mental illness (including psychosis, schizophrenia, or bipolar affective disease)
- Diagnosis of erectile dysfunction or treatment for erectile dysfunction (including alprostadil, phosphodiesterase type 5 inhibitors, papaverine, or phentolamine)

SCORE2 variables[3]

- Age at study entry (years)
- Smoking (current vs. other)
- Systolic blood pressure (SBP) (mmHg)
- Diabetes (yes or no)
- Total cholesterol (mmol/L)
- HDL cholesterol (mmol/L)
- Smoking x age interaction
- SBP x age interaction
- Total cholesterol x age interaction
- HDL cholesterol x age interaction
- Diabetes x age interaction

Recalibration of QRISK3 and SCORE2

The participants in UK Biobank are in overall healthier than the general UK population, with lower incidence of CVD in both men and women, and the calibration plot also showed that the original QRISK3 score was overestimated and original SCORE2 was underestimated when applying to the study population (Figure below). Therefore, following TRIPOD guidelines,[4] our study only used the predicted hazard ratios calculated by the original algorithm[5], and refitted the baseline survival function from the study population to obtain recalibrated predicted probabilities. After refitting the baseline risk, the recalibrated predicted risk from QRISK3 and SCORE2 was well calibrated to the observed risk of each individual (in main Figure and Figure S2, respectively).



Elastic-net[6,7]

Elastic-net is a regularization and variable selection method that linearly combines the L1 and L2 penalties in the regression model. The method overcomes the limitations of the LASSO when dealing with highly correlated variables.

In our study, elastic-net was applied in Cox proportional hazards model, using the Python package of *sksurv.linear_model.CoxnetSurvivalAnalysis*.[8] The key parameters include:

- n_alphas (int, default: 100) – Number of alphas along the regularization path.
- alphas (array-like or None) – List of alphas where to compute the models.
- alpha_min_ratio (float or "auto", default: "auto") –Determines minimum alpha of the regularization path if alphas is None. The smallest value for alpha is computed as the fraction of the data derived maximum alpha (i.e. the smallest value for which all coefficients are zero).If set to “auto”, the value will depend on the sample size relative to the number of features. If n_samples > n_features, the default value is 0.0001 If n_samples <= n_features, 0.01 is the default value.
- l1_ratio (float, default: 0.5) – The ElasticNet mixing parameter, with 0 < l1_ratio <= 1. For l1_ratio = 0 the penalty is an L2 penalty. For l1_ratio = 1 it is an L1 penalty. For 0 < l1_ratio < 1, the penalty is a combination of L1 and L2.

XGBoost[9,10]

XGBoost (eXtreme Gradient Boosting) is a gradient boosting decision tree algorithm that can include higher-order interactions and account for complex nonlinear relationships of variables. Boosting is an ensemble technique where new models are added to correct the errors made by existing models. Models are added sequentially until no further improvements can be made. Gradient boosting is an approach where new models are created that predict the residuals or errors of prior models and then added together to make the final prediction, using a gradient descent algorithm to minimize the loss when adding new models. This approach supports both regression and classification predictive modeling problems, including hazard risk prediction. XGBoost handles sparse data and enables quicker model exploration, and often achieves higher accuracy than a single decision tree.

In our study, XGBoost was applied in Cox proportional hazards model, using the Python package of *xgboost*.[11] The key parameters include:

- objective: Learning objective.
    - survival:cox: Cox regression for right censored survival time data
- eval_metric: Evaluation metrics for validation data
    - cox-nloglik: negative partial log-likelihood for Cox proportional hazards regression
- n_estimators (range: (0,∞], default: 100): The number of trees (or rounds)
- learning_rate (range: [0,1], default: 0.3): Step size shrinkage used in update to prevents overfitting.
- max_depth (range: [0,∞], default: 6): Maximum depth of a tree. Increasing this value will make the model more complex and more likely to overfit. 0 indicates no limit on depth.
- subsample (range: (0,1], default: 1): Subsample ratio of the training instances. Setting it to 0.5 means that XGBoost would randomly sample half of the training data prior to growing trees. and this will prevent overfitting. Subsampling will occur once in every boosting iteration.
- colsample_bytree (range: (0,1], default: 1): Subsample ratio of columns when constructing each tree. Subsampling occurs once for every tree constructed.
- min_child_weight (range: [0,∞], default: 1): Minimum sum of instance weight (hessian) needed in a child. If the tree partition step results in a leaf node with the sum of instance weight less than min_child_weight, then the building process will give up further partitioning. The larger min_child_weight is, the more conservative the algorithm will be.
- reg_lambda (default: 1): L2 regularization term on weights. Increasing this value will make model more conservative.
- reg_alpha (default: 0): L1 regularization term on weights. Increasing this value will make model more conservative.

BorutaSHAP[12]

SHAP (SHapley Additive exPlanations) is a unified approach to explain how much each factor in a model has contributed to the prediction, in other words, it measures the impact in model predictions with and without a particular feature. BorutaSHAP is a wrapper feature selection method, which combines both the Boruta feature selection algorithm with shapley values. This combination has proven to outperform the original Permutation Importance method in both speed, and the quality of the feature subset produced. Not only does this algorithm provide a better subset of features, but it can also simultaneously provide the most accurate and consistent global feature rankings, which can be used for model inference too. BorutaSHAP allows the user to choose any Tree Based learner as the base model in the feature selection process.

In our study, BorutaSHAP was applied in XGBoost survival model, using the Python package of *BorutaShap*,[13] The key parameters include:

- importance_measure ("shap", "gain"or "permutation", default: "shap"): BorutaShap object
- n_trials (range: (0,∞], default: 100): Number of iterations for Boruta algorithm

Assessment of prediction performance

Discrimination: The ability of a model to separate cases from controls

Harrell's C-index: Goodness of fit measure to evaluate risk models in survival analysis. It measures the probability that a randomly selected subject with shorter time-to-event will have a higher predicted probability of event than a randomly selected subject with longer time-to-event.

Reclassification: The ability of a new model to improve on an old model

Integrated discrimination improvement (IDI): It summarises the extent a new model increases risk in events and decreases risk in non-event compared with the old model. ($\bar{\hat{P}}$ represents the average predicted probability for that group)

- IDI = $(\bar{\hat{P}}_{\text{new,events}} - \bar{\hat{P}}_{\text{old,events}}) - (\bar{\hat{P}}_{\text{new,non-events}} - \bar{\hat{P}}_{\text{old,non-events}})$

Net reclassification improvement (NRI): It quantifies the appropriateness of the change in predicted probabilities or categorised risk group when changing from old to new model ($\hat{P}$ represents the proportion and D the occurrence of death)

- Continuous NRI = Continuous $\text{NRI}_{\text{event}}$ + Continuous $\text{NRI}_{\text{non-event}}$
  Continuous $\text{NRI}_{\text{event}} = \hat{P}_{\text{higher predicted prob, D=1}} - \hat{P}_{\text{lower predicted prob, D=1}}$
  Continuous $\text{NRI}_{\text{non-event}} = \hat{P}_{\text{lower predicted prob, D=0}} - \hat{P}_{\text{higher predicted prob, D=0}}$
- Categorical NRI = Categorical $\text{NRI}_{\text{event}}$ + Categorical $\text{NRI}_{\text{non-event}}$
  Categorical $\text{NRI}_{\text{event}} = \hat{P}_{\text{higher risk group, D=1}} - \hat{P}_{\text{lower risk group, D=1}}$
  Categorical $\text{NRI}_{\text{non-event}} = \hat{P}_{\text{lower risk group, D=0}} - \hat{P}_{\text{higher risk group, D=0}}$

Calibration: How close the predicted probability is to the actual (observed) risk

Calibration plot: It reflects how close the predicted probability is to the actual risk in each decile group of predicted probability.

Reference

1. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ (Clinical research ed).* 2017;357:j2099.
2. Parsons RE, Colopy GW, Clifton DA, Clifton L. Clinical Prediction Models in Epidemiological Studies: Lessons from the Application of QRISK3 to UK Biobank Data. *Journal of Data Science.* 2022;20(1):1-13.
3. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. *European heart journal.* 2021;42(25):2439-2454.
4. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Annals of internal medicine.* 2015;162(1):W1-73.
5. Julkunen H, Cichońska A, Tiainen M, et al. Atlas of plasma NMR biomarkers for health and disease in 118,461 individuals from the UK Biobank. *Nature communications.* 2023;14(1):604.
6. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *Journal of statistical software.* 2011;39(5):1-13.
7. Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology).* 2005;67(2):301-320.
8. Clarke R, Emberson JR, Breeze E, et al. Biomarkers of inflammation predict both vascular and non-vascular mortality in older men. *European heart journal.* 2008;29(6):800-809.
9. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016; San Francisco, California, USA.
10. Brownlee J. A Gentle Introduction to XGBoost for Applied Machine Learning. 2016; https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/.
11. Plummer M. Improved estimates of floating absolute risk. *Statistics in medicine.* 2004;23(1):93-104.
12. Keany E. BorutaShap : A wrapper feature selection method which combines the Boruta feature selection algorithm with Shapley values. *Zenodo.* 2020.
13. Lewington S, Clarke R, Qizilbash N, Peto R, Collins R. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet (London, England).* 2002;360(9349):1903-1913.

Table S1: List of clinically validated metabolites for main analyses

	Clinically validated metabolites*	Abbreviation
	Cholesterols, mmol/L	
1	Total cholesterol	Total_C
2	VLDL cholesterol	VLDL_C
3	LDL cholesterol	LDL_C
4	HDL cholesterol	HDL_C
	Triglycerides, mmol/L	
5	Total triglycerides	Total_TG
	Fatty acids, mmol/L	
6	Total fatty acids	TotFA
7	Omega-3 fatty acids	FAw3
8	Omega-6 fatty acids	FAw6
9	Polyunsaturated fatty acids	PUFA
10	Monounsaturated fatty acids	MUFA
11	Saturated fatty acids	SFA
12	Docosahexenoic acid	DHA
13	Linoleic acid	LA
	Fatty acids ratios	
14	Omega-3 fatty acids to total fatty acids	FAw3_FA
15	Omega-6 fatty acids to total fatty acids	FAw6_FA
16	Polyunsaturated fatty acids to total fatty acids	PUFA_FA
17	Monounsaturated fatty acids to total fatty acids	MUFA_FA
18	Saturated fatty acids to total fatty acids	SFA_FA
19	Docosahexaenoic acid to total fatty acids	DHA_FA
20	Linoleic acid to total fatty acids	LA_FA
21	Polyunsaturated to monounsaturated fatty acids	PUFA_MUFA
22	Omega-6 fatty acids to omega-3 fatty acids	FAw6_FAw3
	Apolipoproteins	
23	Apolipoprotein B, g/l	ApoB
24	Apolipoprotein A1, g/l	ApoA1
25	Apolipoprotein B ratio to apolipoprotein A1	ApoB_ApoA1
	Amino acids, mmol/L	
26	Alanine	Ala
27	Glycine	Gly
28	Histidine	His
	Branched-chain amino acids, mmol/L	
29	Isoleucine	Ile
30	Leucine	Leu
31	Valine	Val
32	Total concentration of branched-chain amino acids	BCAA
	Aromatic amino acids, mmol/L	
33	Phenylalanine	Phe
34	Tyrosine	Tyr
	Glycolysis related metabolites, mmol/L	
35	Glucose	Glc
36	Lactate	Lac
	Fluid balance	
37	Creatinine, mmol/L	Crea
38	Albumin, g/L	Alb
	Inflammation, mmol/L	
39	Glycoprotein acetyls	GlycA

*Clinically and analytically validated biomarkers, which are comparable with other clinically and analytically validated laboratory method, such as photometric or enzymatic methods.

Table S2: List of metabolites for sensitivity analyses

Biomarker name	Abbreviation	Biomarker name	Abbreviation
Total lipids, mmol/L		**Fatty acids (concentration), mmol/L**	
Total cholesterol*	Total_C	Polyunsaturated fatty acids*	PUFA
VLDL cholesterol*	VLDL_C	Monounsaturated fatty acids*	MUFA
IDL cholesterol	IDL_C	Saturated fatty acids*	SFA
LDL cholesterol*	LDL_C	Docosahexaenoic acid*	DHA
HDL cholesterol*	HDL_C	Linoleic acid*	LA
Total cholesterol minus HDL-C	non_HDL_C	Omega-3 fatty acids*	FAw3
Remnant cholesterol	Remnant_C	Omega-6 fatty acids*	FAw6
Total esterified cholesterol	Total_CE	Total fatty acids*	TotFA
Total free cholesterol	Total_FC	**Fatty acids ratio, %**	
Total phospholipids	Total_PL	Polyunsaturated fatty acids to total*	PUFA_FA
Total triglycerides*	Total_TG	Monounsaturated fatty acids to total*	MUFA_FA
Lipoprotein particle concentration, mmol/L		Saturated fatty acids to total*	SFA_FA
Chylomicrons&extremely large VLDL	XXL_VLDL_P	Docosahexaenoic acid to total*	DHA_FA
Very large VLDL	XL_VLDL_P	Linoleic acid to total*	LA_FA
Large VLDL	L_VLDL_P	Omega-3 fatty acids to total*	FAw3_FA
Medium VLDL	M_VLDL_P	Omega-6 fatty acids to total*	FAw6_FA
Small VLDL	S_VLDL_P	Polyunsaturated to monounsaturated fatty acids*	PUFA_MUFA
Very small VLDL	XS_VLDL_P	Omega-6 to omega-3 fatty acids*	FAw6_FAw3
Total VLDL	VLDL_P	**Cholines, mmol/L**	
IDL	IDL_P	Total cholines	TotCho
Large LDL	L_LDL_P	Phosphatidylcholine	PC
Medium LDL	M_LDL_P	Sphingomyelins	SM
Small LDL	S_LDL_P	Phosphoglycerides	Phosphoglyc
Total LDL	LDL_P	**Amino acids, mmol/L**	
Very large HDL	XL_HDL_P	Alanine*	Ala
Large HDL	L_HDL_P	Glutamine	Gln
Medium HDL	M_HDL_P	Glycine*	Gly
Small HDL	S_HDL_P	Histidine*	His
Total HDL	HDL_P	Isoleucine*	Ile
Mean lipoportein particle size, nm		Leucine*	Leu
VLDL	VLDL_D	Valine*	Val
LDL	LDL_D	Branched-chain amino acids*	BCAA
HDL	HDL_D	Phenylalanine*	Phe
Lipoprotein particle composition		Tyrosine*	Tyr
Esterified cholesterol in VLDL	VLDL_CE	**Glycolysis related metabolites, mmol/L**	
Free cholesterol in VLDL	VLDL_FC	Lactate*	Lac
Phospholipids in VLDL	VLDL_PL	Citrate	Cit
Triglycerides in VLDL	VLDL_TG	Glucose*	Glc
Esterified cholesterol in IDL	IDL_CE	Pyruvate	Pyruvate
Free cholesterol in IDL	IDL_FC	**Ketone bodies, mmol/L**	
Phospholipids in IDL	IDL_PL	Acetate	Ace
Triglycerides in IDL	IDL_TG	Aceto acetate	AcAce
Esterified cholesterol in LDL	LDL_CE	Acetone	Acetone
Free cholesterol in LDL	LDL_FC	Beta-hydroxybutyrate	bOHBut
Phospholipids in LDL	LDL_PL	**Fluid balance**	
Triglycerides in LDL	LDL_TG	Albumin*, g/L	Alb
Esterified cholesterol in HDL	HDL_CE	Creatinine*, mmol/L	Crea_nmr
Free cholesterol in HDL	HDL_FC	**Inflammation, mmol/L**	
Triglycerides in HDL	HDL_TG	Glycoprotein acetyls*	Gp
Phospholipids in HDL	HDL_PL		
Apolipoproteins, g/L			
Apolipoprotein A-I*	ApoA-1		
Apolipoprotein B*	ApoB		
Apolipoprotein B to A-1 ratio*	ApoB_ApoA-1		

*The clinical-validated metabolites used in the main analyses

Table S3: Disease and medication codes of QRISK3 variables in UK biobank

	ICD-10 code	Verbal interview or questionnaire code	Medication code or other measurement
Diabetes			
Type 1	E10	1222; Variable ID 2443=1 & age≤20	If recode both type1 & type2, then categorize as type1
Type 2	E11;E13;E14	1220;1223; Variable ID 2443=1 & age>20	HbA1c≥48 & ≤184 mmol/mol
Chronic kidney disease (stage 3, 4, 5)	N183; N184; N185; N180	1193	eGFR <60 ml/min
Atrial fibrillation	I48	1471;1483	
Hypertension treatment		Variable ID 6177, 6153 =2	1140860192, 1140860292, 1140860696, 1140860728, 1140860750, 1140860806, 1140860882, 1140860904, 1140861088, 1140861190, 1140861276, 1140866072, 1140866078, 1140866090, 1140866102, 1140866108, 1140866122, 1140866138, 1140866156, 1140866162, 1140866724, 1140866738, 1140868618, 1140872568, 1140874706, 1140874744, 1140875808, 1140879758, 1140879760, 1140879762, 1140879802, 1140879806, 1140879810, 1140879818, 1140879822, 1140879826, 1140879830, 1140879834, 1140879842, 1140879866, 1140884298, 1140888552, 1140888556, 1140888560, 1140888646, 1140909706, 1140910442, 1140910614, 1140916356, 1140923272, 1140923336, 1140923404, 1140923712, 1140926778, 1140928226, 1141145660, 1141146126, 1141152998, 1141153026, 1141164276, 1141165470, 1141166006, 1141169516, 1141171336, 1141180592, 1141180772, 1141180778, 1141184722, 1141193282, 1141194794, 1141194810
Migraines	G43	1265	
Rheumatoid arthritis	M05; M06	1464	
Systemic lupus erythematosus	M32	1381	
Severe mental illness*	F20; F31; F331; F332; F333	1289;1291; Variable ID 20126=1,2,3,4	
Atypical antipsychotic medication			1140867420, 1140867444, 1140927956, 1140928916, 1141152848, 1141153490, 1141169714, 1141195974
Regular steroid tablets			1140874790, 1140874816, 1140874896, 1140874930, 1140874976, 1141145782, 1141173346
Erectile disfunction	N484	1518	1141168936, 1141168948, 1141168944, 1141168946, 1140869100, 1140883010

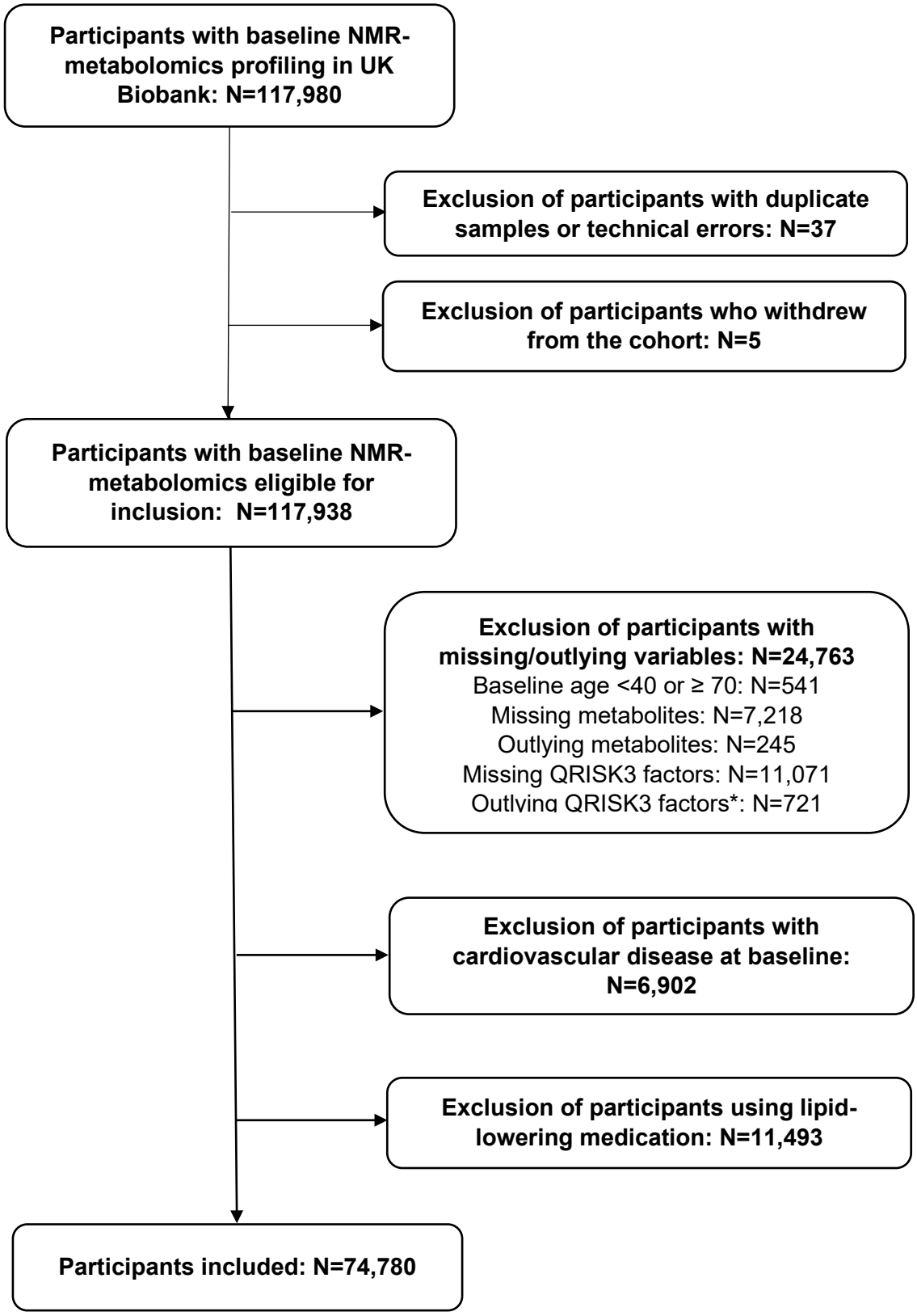
* Includes schizophrenia, bipolar disorder and moderate/severe depression.

Table S4: ICD-10 and operation code of cardiovascular disease

ICD/OPCS category	Disease category	Code definition
I20	Angina pectoris	I20.0 Unstable angina I20.1 Angina pectoris with documented spasm I20.8 Other forms of angina pectoris I20.9 Angina pectoris, unspecified angina
I21	Acute myocardial infarction	I21.0 Acute transmural myocardial infarction of anterior wall I21.1 Acute transmural myocardial infarction of inferior wall I21.2 Acute transmural myocardial infarction of other sites I21.3 Acute transmural myocardial infarction of unspecified site I21.4 Acute subendocardial myocardial infarction I21.9 Acute myocardial infarction, unspecified
I22	Subsequent myocardial infarction	I22.0 Subsequent myocardial infarction of anterior wall I22.1 Subsequent myocardial infarction of inferior wall I22.8 Subsequent myocardial infarction of other sites I22.9 Subsequent myocardial infarction of unspecified site
I23	Certain current complications following acute myocardial infarction	I23.0 Haemopericardium as current complication following acute myocardial infarction; I23.1 Atrial septal defect as current complication following acute myocardial infarction; I23.2 Ventricular septal defect as current complication following acute myocardial infarction; I23.3 Rupture of cardiac wall without haemopericardium as current complication following acute myocardial infarction; I23.4 Rupture of chordae tendineae as current complication following acute myocardial infarction I23.5 Rupture of papillary muscle as current complication following acute myocardial infarction I23.6 Thrombosis of atrium, auricular appendage, and ventricle as current complications following acute myocardial infarction; I23.8 Other current complications following acute myocardial infarction
I24	Other acute ischaemic heart diseases	I24.0 Coronary thrombosis not resulting in myocardial infarction I24.1 Dressler's syndrome I24.8 Other forms of acute ischaemic heart disease I24.9 Acute ischaemic heart disease, unspecified (excl. ischaemic heart disease (chronic) NOS)
I25	Chronic ischaemic heart disease	I25.0 Atherosclerotic cardiovascular disease, so described I25.1 Atherosclerotic heart disease I25.2 Old myocardial infarction I25.3 Aneurysm of heart I25.4 Coronary artery aneurysm I25.5 Ischaemic cardiomyopathy I25.6 Silent myocardial ischaemia I25.8 Other forms of chronic ischaemic heart disease - Any condition in I21-I22 and I24.- specified as chronic I25.9 Chronic ischaemic heart disease, unspecified - Ischaemic heart disease (chronic) NOS
I63	Cerebral infarction	I63.0 Cerebral infarction due to thrombosis of precerebral arteries I63.1 Cerebral infarction due to embolism of precerebral arteries I63.2 Cerebral infarction due to unspecified occlusion or stenosis of precerebral arteries I63.3 Cerebral infarction due to thrombosis of cerebral arteries I63.4 Cerebral infarction due to embolism of cerebral arteries I63.5 Cerebral infarction due to unspecified occlusion or stenosis of cerebral arteries

		I63.6 Cerebral infarction due to cerebral venous thrombosis, nonpyogenic I63.8 Other cerebral infarction I63.9 Cerebral infarction, unspecified
I64	Stroke	Stroke, not specified as haemorrhage or infaction
G45	Transient cerebral ischaemic attacks	G45.0 Vertebro-basilar artery syndrome G45.1 Carotid artery syndrome (hemispheric) G45.2 Multiple and bilateral precerebral artery syndromes G45.3 Amaurosis fugax G45.4 Transient global amnesia G45.8 Other transient cerebral ischaemic attacks and related syndromes G45.9 Transient cerebral ischaemic attack, unspecified
K40		Saphenous vein graft replacement of coronary artery
K41		Other autograft replacement of coronary artery
K42		Allograft replacement of coronary artery
K43		Prosthetic replacement of coronary artery
K44		Other replacement of coronary artery
K45		Connection of thoracic artery to coronary artery
K46		Other bypass of coronary artery
K47		Repair of coronary artery
K49		Transluminal balloon angioplasty of coronary artery
K50		Other therapeutic transluminal operations on coronary artery
K75		Percutaneous transluminal balloon angioplasty and insertion of stent into coronary artery

Figure S1: Flowchart of exclusion criteria for study population in UK Biobank



*Exclude outlying of baseline standard deviation of systolic blood pressure>20, most missing in QRISK3 variables come from clinical chemistry measurement of total cholesterol to HDL cholesterol ratio (n=10,794)

Table S5: Fine-tuning of hyper-parameters (QRISK3)

Model	Package (Python)	Hyperparameter*	Tuning Range	Tuning Step†	Selected Value	
					Women	Men
Model 1	CoxPHSurvivalAnalysis	-	-	-		
Model 2	CoxnetSurvivalAnalysis (Penalized Cox Model)	L1_ratio	(0.7, 1.0)	0.05	0.9	0.9
		alphas	alpha_min_ratio=0.01, max_iter=1000 to search for 1000 α values up to 1% of estimated maximum.		0.00041	0.0012
Model 3	XGBoost	objective	-	-	survival:cox	
		eval_metric	-	-	cox-nloglik	
		learning_rate	[0.01, 0.02, 0.05, 0.1]		0.01	0.01
		max_depth	(2,5)	1	4	4
		n_estimators	(50,2000)	50	550	450
		subsample	(0.5, 1.0)	0.1	0.8	0.6
		colsample_bytree	(0.5, 1.0)	0.1	0.5	0.9
		min_child_weight	(6,30)	1	21	18
		reg_lambda	Start from [1e-6, 1e-4, 0.01, 0.1, 1, 10, 100]		2.3	1.6
		reg_alpha	Start from [0, 1e-6, 1e-4, 0.01, 0.1, 1, 10, 100]		0.18	0.0036
	BorutaShap	importance_measure	-	-	SHAP	
		n_trials	-	-	100	
	CoxPHSurvivalAnalysis	-	-	-		

*The meaning of each hyperparameter is explained in eMethods; † Tuning the hyperparameter from the lowest value to the highest value in the tuning range, with increase of the tuning step each time, and selected the hyperparmeter with the best performance.

Table S6: Baseline characteristics in UK Biobank versus in study population

	UK Biobank population		Study Population	
	Women	**Men**	**Women**	**Men**
No. of participants	232,744	169,405	42,427	32,353
Age and socioeconomic factors				
Baseline Age, years	55.5 (8.0)	55.3 (8.2)	55.4 (7.9)	55.1 (8.2)
White, %	94.4	94.1	95.0	94.8
Townsend deprivation index*	-1.4 (3.0)	-1.3 (3.1)	-1.5 (3.0)	-1.4 (3.0)
Anthropometry, blood pressure, and lipids by clinical chemistry				
Body Mass Index, kg/m^2	26.7 (5.0)	27.4 (4.0)	26.7 (4.9)	27.4 (4.0)
Systolic blood pressure, mmHg	134.4 (19.2)	140.3 (17.3)	134.0 (18.9)	140.2 (17.2)
Standard deviation between two readings†, mmHg	5.4 (4.5)	5.2 (4.3)	5.2 (4.1)	5.0 (3.9)
Total cholesterol to HDL-C ratio	3.9 (1.0)	4.6 (1,1)	3.9 (1.0)	4.6 (1.1)
Smoking intensity, %				
Ex-smoker	30.7	34.9	30.9	35.0
Light smoker (< 10 per day)	3.9	6.1	3.9	5.9
Moderate smoker (10-19 per day)	3.0	3.1	3.0	3.0
Heavy smoker (≥20 per day)	1.9	3.4	1.8	3.3
Family history of heart disease‡, %	42.8	36.1	43.2	36.3
Parents	40.1	33.7	40.5	33.9
Siblings	7.2	5.4	7.2	5.3
Disease and medication history, %				
Type 1 diabetes	0.5	0.2	0.5	0.2
Type 2 diabetes	1.3	2.4	1.3	2.3
Chronic kidney disease	1.7	1.1	1.8	1.2
Atrial fibrillation	0.6	1.4	0.6	1.3
Migraines	6.0	2.2	6.3	2.3
Rheumatoid arthritis	1.5	0.8	1.5	0.8
Systemic lupus erythematosus	0.2	<0.1	0.2	<0.1
Severe mental illness§	0.5	0.5	5.9	3.8
Erectile dysfunction	-	0.5	-	0.6
Hypertension treatment	12.5	12.6	12.2	12.5
Atypical antipsychotic medication	0.2	0.3	0.2	0.3
Regular steroid tablets	0.8	0.7	0.8	0.8

Characteristics of QRISK factors at baseline by sex. Continuous variables are presented as mean (standard deviation) and categorical variables are presented as column percentages. *Higher values indicate higher levels of material deprivation; †QRISK asks for standard deviation of systolic blood pressure values recorded in the five years before study entry, but UK biobank only provided two automated or manual readings at study entry; ‡QRISK asks for the family history in first degree relatives aged less than 60 years, but UK biobank only identified family history in first degree relatives in all ages; §Includes schizophrenia, bipolar disorder and moderate/severe depression. HDL-C=high-density lipoproteins cholesterol.

Table S7: List of selected novel metabolites by different methods (QRISK3)

Clinically validated metabolites	Women				Men			
	Significant associated*	Independent associated†	Elestic -net‡	Boruta SHAP¶	Significant associated	Independent associated	Elestic -net	Boruta SHAP
Cholesterols&Triglycerides								
Total_C					✓			
VLDL_C	✓				✓			✓
LDL_C	✓			✓	✓			
HDL_C	✓	✓			✓		✓	
Total_TG	✓		✓	✓	✓		✓	
Fatty acids								
Total FA	✓				✓			
Omega-3 FA			✓			✓	✓	
Omega-6 FA					✓			
PUFA					✓			
MUFA	✓				✓			
SFA	✓				✓			
DHA								
LA					✓			
Omega-3 FA to total FA					✓	✓	✓	
Omega-6 FA to total FA	✓	✓		✓				
PUFA to total FA	✓	✓			✓			
MUFA to total FA	✓	✓		✓	✓		✓	
SFA to total FA				✓				✓
DHA to total FA	✓	✓			✓			
LA to total FA	✓	✓	✓					
PUFA to MUFA	✓	✓	✓		✓			
Omega-6 to omega-3 FA			✓					✓
Apolipoproteins								
ApoB	✓		✓		✓			
ApoA-1	✓	✓			✓			
ApoB to ApoA-1	✓	✓		✓	✓	✓	✓	✓
Amino acids								
Alanine							✓	
Glycine	✓		✓		✓		✓	✓
Histidine	✓	✓	✓					
Isoleucine								
Leucine							✓	✓
Valine			✓				✓	
BACC								
Phenylalanine	✓		✓		✓		✓	
Tyrosine				✓				✓
Glycolysis related								
Glucose				✓				✓
Lactate							✓	
Fluid balance								
Creatinine					✓			
Albumin	✓	✓	✓	✓	✓	✓	✓	✓
Inflammation								
Glycoprotein acetyls	✓	✓	✓	✓	✓	✓	✓	✓

*Association was calculated using Cox proportional-hazards regression with adjustment of established risk factors, including age, education, region, townsend deprivation index, smoking, alcohol intake, body mass index, systolic blood pressure, and baseline diabetes; Significant association was defined as p-value<0.01 after correction of false discovery rate using Benjamini-Hochberg method; †Association was calculated using Cox proportional-hazards regression with adjustment of QRISK3 score; ‡Novel metabolites selected by elastic-net based on Cox proportional-hazards regression, when adding all metabolites into the model; ¶Novel metabolites selected by BorutaSHAP from XGBoost survival model, when adding all metabolites into the model

Table S8. Associations of clinical metabolites independent from SCORE2

	Hazard ratio (95% CI)	
	Women	**Men**
Recalibrated SCORE2	**1.12 (1.10, 1.13)**	**1.07 (1.06, 1.07)**
Cholesterols & Triglycerides		
Total cholesterol	0.96 (0.92, 1.01)	0.98 (0.95, 1.02)
VLDL cholesterol	1.06 (1.01, 1.11)	1.01 (0.97, 1.05)
LDL cholesterol	0.99 (0.95, 1.04)	0.99 (0.96, 1.03)
HDL cholesterol	**0.87 (0.83, 0.92)***	0.95 (0.91, 0.99)
Total triglycerides	1.05 (1.01, 1.10)	0.99 (0.95, 1.03)
Fatty acids		
Total fatty acids	1.03 (0.99, 1.08)	0.98 (0.95, 1.02)
Omega-3 fatty acids	0.96 (0.92, 1.01)	**0.93 (0.90, 0.96)***
Omega-6 fatty acids	0.98 (0.94, 1.03)	0.99 (0.95, 1.02)
Polyunsaturated fatty acids	0.97 (0.93, 1.02)	0.97 (0.93, 1.00)
Monounsaturated fatty acids	**1.08 (1.03, 1.12)***	1.00 (0.96, 1.04)
Saturated fatty acids	1.04 (0.99, 1.08)	0.98 (0.95, 1.02)
Docosahexenoic acid	0.94 (0.90, 0.98)	**0.93 (0.90, 0.96)***
Linoleic acid	0.97 (0.93, 1.02)	0.99 (0.95, 1.02)
Omega-3 to total fatty acids	0.94 (0.90, 0.98)	**0.92 (0.89, 0.96)***
Omega-6 to total fatty acids	**0.91 (0.87, 0.95)***	1.01 (0.98, 1.05)
Polyunsaturated to total fatty acids	**0.89 (0.85, 0.93)***	0.98 (0.95, 1.02)
Monounsaturated to total fatty acids	**1.16 (1.11, 1.22)***	1.03 (0.99, 1.07)
Saturated to total fatty acids	1.03 (0.99, 1.08)	1.00 (0.96, 1.04)
Docosahexaenoic acid to total fatty acids	**0.92 (0.88, 0.96)***	**0.93 (0.90, 0.97)***
Linoleic acid to total fatty acids	**0.90 (0.86, 0.94)***	1.00 (0.97, 1.04)
Polyunsaturated to monounsaturated fatty acids	**0.85 (0.81, 0.89)***	0.97 (0.94, 1.01)
Omega-6 to omega-3 fatty acids	1.03 (0.99, 1.08)	**1.06 (1.02, 1.09)***
Apolipoproteins		
Apolipoprotein B	1.03 (0.98, 1.08)	1.01 (0.98, 1.05)
Apolipoprotein A-1	**0.90 (0.86, 0.94)***	**0.94 (0.90, 0.97)***
Apolipoprotein B to apolipoproteinA-1	**1.09 (1.04, 1.14)***	**1.07 (1.03, 1.11)***
Amino acids		
Alanine	1.03 (0.99, 1.08)	0.98 (0.95, 1.02)
Glycine	0.94 (0.89, 0.98)	0.95 (0.92, 0.99)
Histidine	**0.91 (0.87, 0.95)***	0.97 (0.93, 1.00)
Isoleucine	1.04 (0.99, 1.08)	1.03 (0.99, 1.06)
Leucine	1.02 (0.98, 1.06)	1.01 (0.98, 1.05)
Valine	1.01 (0.97, 1.05)	1.00 (0.96, 1.03)
Total branched-chain amino acids	1.02 (0.98, 1.06)	1.01 (0.97, 1.04)
Phenylalanine	1.06 (1.02, 1.11)	**1.06 (1.03, 1.10)***
Tyrosine	1.02 (0.97, 1.06)	1.02 (0.99, 1.06)
Glycolysis related metabolites		
Glucose	1.03 (0.99, 1.07)	1.03 (0.99, 1.06)
Lactate	1.03 (0.99, 1.08)	0.99 (0.95, 1.02)
Fluid balance		
Creatinine	1.05 (1.01, 1.09)	1.03 (1.00, 1.06)
Albumin	**0.86 (0.82, 0.90)***	**0.89 (0.86, 0.93)***
Inflammation		
Glycoprotein acetyls	**1.18 (1.13, 1.23)***	**1.08 (1.04, 1.12)***

Hazard ratios (HR) per one score higher of concentration. HR of each metabolite was calculated by Cox proportional-hazards regression with adjustment of SCORE2. *Associations remained significant (p-value<0.01) by correction of false discovery rate using Benjamini-Hochberg method.

Table S9: List of selected metabolites using different methods (SCORE2)

Clinically validated metabolites	Women				Men			
	Significant associated*	Independent associated†	Elestic -net‡	Boruta SHAP¶	Significant associated	Independent associated	Elestic -net	Boruta SHAP
Cholesterols&Triglycerides								
Total_C					✓			
VLDL_C	✓				✓			✓
LDL_C	✓			✓	✓			
HDL_C	✓	✓			✓			
Total_TG	✓		✓	✓	✓		✓	
Fatty acids								
Total FA	✓				✓			
Omega-3 FA			✓			✓	✓	
Omega-6 FA					✓			
PUFA					✓			
MUFA	✓	✓			✓			
SFA	✓				✓			
DHA						✓		
LA					✓			
Omega-3 FA to total FA					✓	✓		
Omega-6 FA to total FA	✓	✓		✓				
PUFA to total FA	✓	✓			✓			
MUFA to total FA	✓	✓		✓	✓		✓	
SFA to total FA				✓				✓
DHA to total FA	✓	✓			✓	✓		
LA to total FA	✓	✓	✓					
PUFA to MUFA	✓	✓	✓		✓			
Omega-6 to omega-3 FA			✓			✓		✓
Apolipoproteins								
ApoB	✓		✓		✓			
ApoA-1	✓	✓			✓	✓		
ApoB to ApoA-1	✓	✓	✓	✓	✓	✓	✓	✓
Amino acids								
Alanine							✓	
Glycine	✓		✓		✓		✓	✓
Histidine	✓	✓	✓					
Isoleucine								
Leucine								✓
Valine			✓				✓	
BACC								
Phenylalanine	✓		✓		✓	✓	✓	
Tyrosine			✓	✓			✓	✓
Glycolysis related								
Glucose			✓	✓				✓
Lactate			✓				✓	
Fluid balance								
Creatinine			✓		✓		✓	
Albumin	✓	✓	✓	✓	✓	✓	✓	✓
Inflammation								
Glycoprotein acetyls	✓	✓	✓	✓	✓	✓	✓	✓

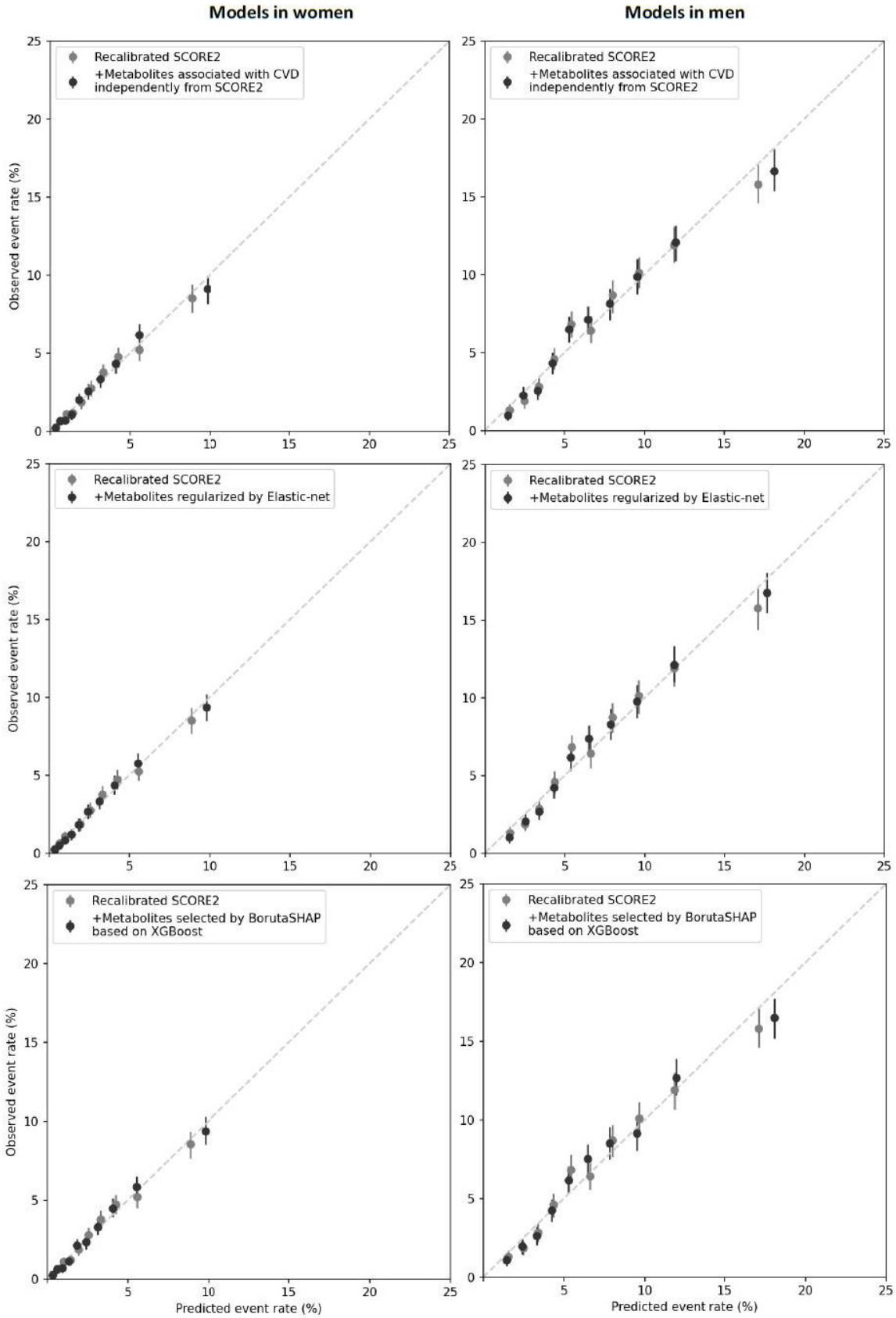
*Association was calculated using Cox proportional-hazards regression with adjustment of established risk factors, including age, education, region, townsend deprivation index, smoking, alcohol intake, body mass index, systolic blood pressure, and baseline diabetes; Significant association was defined as p-value<0.01 after correction of false discovery rate using Benjamini-Hochberg method; †Association was calculated using Cox proportional-hazards regression with adjustment of SCORE2; ‡ Novel metabolites selected by elastic-net based on Cox proportional-hazards regression, when adding all metabolites into the model; ¶ Novel metabolites selected by BorutaSHAP from XGBoost survival model, when adding all metabolites into the model.

Table S10: Comparing prediction performance of 10-year CVD risk w/o metabolites (SCORE2)

Prediction Performance	Women (95% CI*)	Men (95% CI)
Recalibrated SCORE2		
Harrell's C-index †	0.731 (0.718, 0.744)	0.689 (0.679, 0.699)
Adding metabolites associated with CVD independently from SCORE2		
Harrell's C-index	0.745 (0.732, 0.758)	0.695 (0.686, 0.705)
IDI‡ (%)	0.39 (0.24, 0.52)	0.34 (0.20, 0.44)
Continuous NRI§ (%)	21.1 (15.7, 26.3)	15.3 (11.3, 19.3)
events	9.2 (3.9, 14.3)	6.2 (2.3, 10.0)
non-events	12.0 (11.1, 12.9)	9.1 (8.0, 10.2)
Categorical NRI (%)	1.5 (-0.1, 2.8)	0.4 (-1.0, 1.8)
events	2.3 (0.8, 3.7)	-0.2 (-1.5, 1.2)
non-events	-0.9 (-1.0, -0.7)	0.6 (0.3, 0.8)
Adding metabolites with regularization (using Elastic-net)		
Harrell's C-index	0.746 (0.734, 0.758)	0.695 (0.685, 0.705)
IDI (%)	0.36 (0.20, 0.49)	0.21 (0.10, 0.30)
Continuous NRI (%)	20.1 (14.5, 25.1)	7.6 (3.3,11.6)
events	7.5 (2.0,12.3)	5.1 (0.6, 9.0)
non-events	12.7 (11.7, 13.6)	2.5 (1.4, 3.6)
Categorical NRI (%)	1.4 (-0.1, 3.0)	0.2 (-1.2, 1.6)
events	2.3 (0.8, 3.8)	-0.4 (-1.7, 1.0)
non-events	-0.9 (-1.0, -0.7)	0.6 (0.3, 0.9)
Adding metabolites selected by BorutaSHAP from XGBoost		
Harrell's C-index	0.747 (0.734, 0.758)	0.694 (0.685, 0.704)
IDI (%)	0.36 (0.21, 0.48)	0.27 (0.14, 0.36)
Continuous NRI (%)	21.9 (16.1, 27.3)	13.3 (9.1 17.7)
events	5.2 (-0.4, 10.4)	2.9 (-1.2, 6.8)
non-events	16.7 (15.7, 17.6)	10.4 (9.2, 11.5)
Categorical NRI (%)	1.4 (0, 2.8)	0.5 (-0.8, 1.7)
events	2.2 (0.9, 3.7)	-0.1 (-1.4, 1.0)
non-events	-0.8 (-1.0, -0.7)	0.6 (0.3, 0.9)

Comparing prediction performance of 10-year CVD risk w/o metabolites. In all models, metabolites are added to recalibrated SCORE2 using Cox proportional-hazards regression. Hyper-parameters of each model are in appendix. *Bootstrap percentile confidence interval, bootstrap for 500 times; †Harrell's C-index, measuring the probability that a randomly selected subject with shorter time-to-event will have a higher predicted probability of event than a randomly selected subject with longer time-to-event; ‡Integrated discrimination improvement, summarising the extent a new model increases risk in events and decreases risk in non-event compared with the old model; §Net reclassification improvement, quantifying the appropriateness of the change in predicted probabilities or categorised risk group when changing from old to new model; Categorical NRI is based on a 10% risk threshold.

Figure S2: Calibration of risk prediction models for 10-year CVD risk (SCORE2)



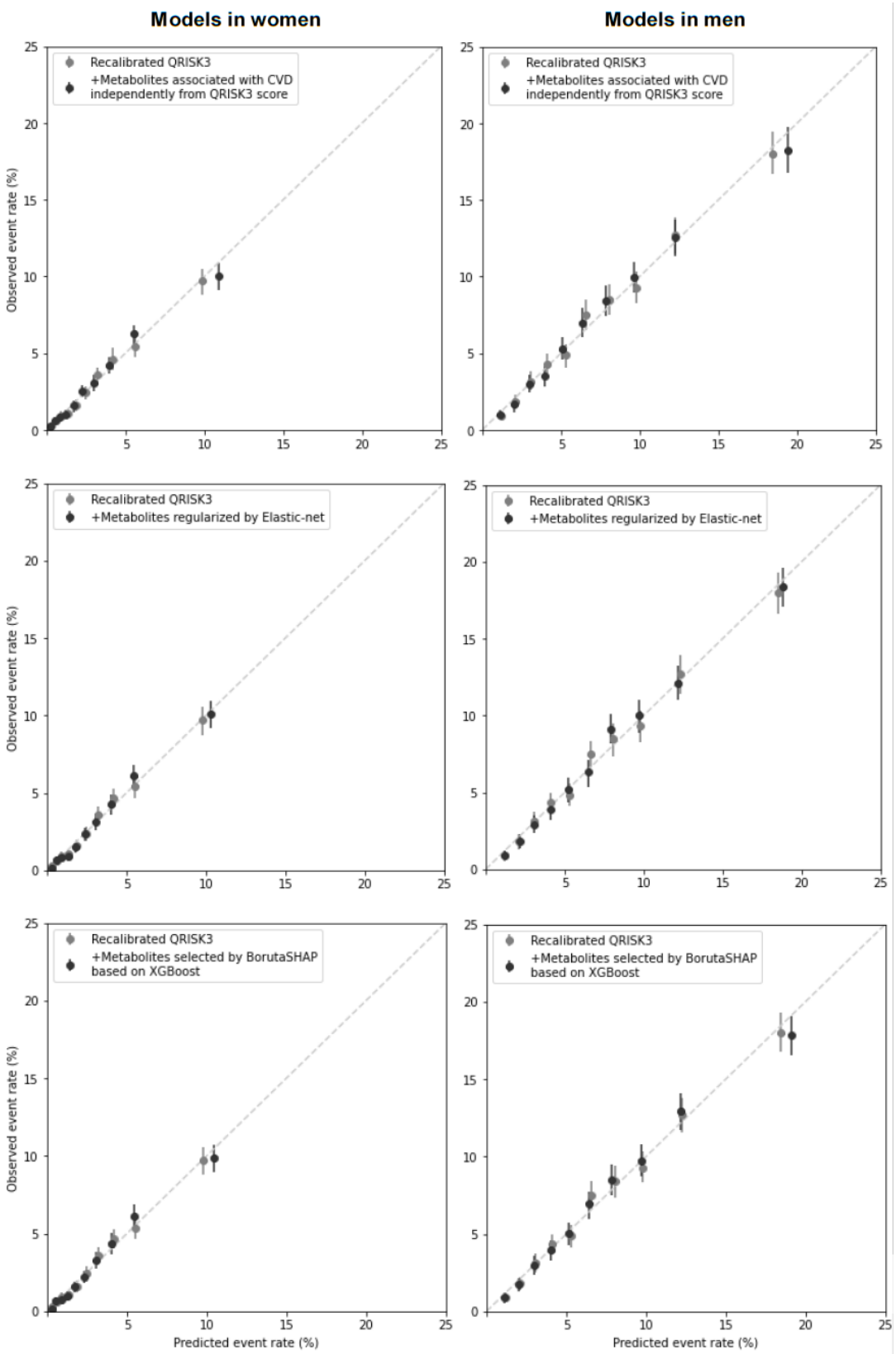
Calibration of risk prediction models for 10-year CVD risk. For each model, the observed and predicted CVD event rates are shown for each of 10 equally sized groups of absolute predicted risk. Vertical lines represent 95% CIs (bootstrap percentile confidence interval, bootstrap for 500 times).

Table S11: Prediction performance of 10-year ASCVD risk w/o metabolites (QRISK3 and wider scope of candidate metabolites)

Prediction Performance	Women (95% CI*)	Men (95% CI)
Recalibrated QRISK3		
Harrell's C-index †	0.750 (0.739, 0.763)	0.706 (0.696, 0.716)
Adding metabolites associated with CVD independently from QRISK3 score		
Harrell's C-index	0.759 (0.748, 0.770)	0.712 (0.702, 0.722)
IDI‡ (%)	0.49 (0.21, 0.65)	0.31 (0.18, 0.40)
Continuous NRI§ (%)	17.3 (11.6, 22.2)	10.0 (5.5,13.8)
events	7.1 (1.4, 12.2)	1.6 (-2.6, 5.2)
non-events	10.2 (9.3, 11.2)	8.4 (7.3, 9.6)
Categorical NRI (%)	1.5 (-0.2, 3.0)	0.8 (-0.7, 2.2)
events	2.3 (0.6, 3.8)	0.2 (-1.3, 1.6)
non-events	-0.8 (-1.0, -0.7)	0.6 (0.3, 0.9)
Adding metabolites with regularization (using Elastic-net)		
Harrell's C-index	0.760 (0.749, 0.772)	0.711 (0.701, 0.720)
IDI (%)	0.24 (0.08, 0.36)	0.12 (0.01, 0.21)
Continuous NRI (%)	6.7 (1.5, 11.9)	2.8 (-1.5, 7.1)
events	6.7 (1.6, 11.8)	7.2 (3.1, 11.4)
non-events	-0.5 (-1.0, 0.9)	-4.4 (-5.4, -3.3)
Categorical NRI (%)	0.8 (-0.9, 2.2)	0.6 (-0.7, 1.8)
events	1.2 (-0.4, 2.7)	0.1 (-1.2, 1.3)
non-events	-0.5 (-0.6, -0.3)	0.5 (0.2, 0.8)
Adding metabolites selected by BorutaSHAP from XGBoost		
Harrell's C-index	0.760 (0.748, 0.771)	0.710 (0.700, 0.720)
IDI (%)	0.35 (0.20, 0.47)	0.19 (0.09, 0.27)
Continuous NRI (%)	17.4 (12.0, 23.7)	9.2 (5.0, 13.6)
events	5.6 (0.4, 11.1)	1.4 (-2.6, 5.4)
non-events	11.8 (10.8, 12.7)	7.8 (6.7, 8.9)
Categorical NRI (%)	0.6 (-0.7, 2.0)	1.0 (-0.2, 2.2)
events	1.3 (-0.1, 2.7)	0.6 (-0.7, 1.7)
non-events	-0.7 (-0.8, -0.5)	0.5 (0.2, 0.7)

Comparing prediction performance of 10-year ASCVD risk w/o metabolites. In all models, metabolites are added to recalibrated QRISK3 using Cox proportional-hazards regression. Hyper-parameters of each model are in appendix. *Bootstrap percentile confidence interval, bootstrap for 500 times; †Harrell's C-index, measuring the probability that a randomly selected subject with shorter time-to-event will have a higher predicted probability of event than a randomly selected subject with longer time-to-event; ‡Integrated discrimination improvement, summarising the extent a new model increases risk in events and decreases risk in non-event compared with the old model; §Net reclassification improvement, quantifying the appropriateness of the change in predicted probabilities or categorised risk group when changing from old to new model; Categorical NRI is based on a 10% risk threshold.

Figure S3: Calibration of risk prediction models for 10-year ASCVD risk (QRISK3 and wider scope of candidate metabolites)



Calibration of risk prediction models for 10-year ASCVD risk. For each model, the observed and predicted CVD event rates are shown for each of 10 equally sized groups of absolute predicted risk. Vertical lines represent 95% CIs (bootstrap percentile confidence interval, bootstrap for 500 times).