THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Combining serum metabolomic profiles with traditional risk factors improves 10-year cardiovascular risk prediction in people with type 2 diabetes

Citation for published version:
Huang, Z, Klaric, L, Krasauskaite, J, Khalid, W, Strachan, MWJ, Wilson, JF & Price, JF 2023, 'Combining serum metabolomic profiles with traditional risk factors improves 10-year cardiovascular risk prediction in people with type 2 diabetes', *European Journal of Preventive Cardiology*.
https://doi.org/10.1093/eurjpc/zwad160

Digital Object Identifier (DOI):
10.1093/eurjpc/zwad160

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
European Journal of Preventive Cardiology

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 05. Jul. 2023

ESC European Society of Cardiology

European Journal of Preventive Cardiology (2023) **00**, 1–8
https://doi.org/10.1093/eurjpc/zwad160

FULL RESEARCH PAPER
Risk prediction/assessment & stratification

Combining serum metabolomic profiles with traditional risk factors improves 10-year cardiovascular risk prediction in people with type 2 diabetes

Zhe Huang [1]*, Lucija Klaric [2], Justina Krasauskaite[1], Wardah Khalid[1], Mark W. J. Strachan [3], James F. Wilson [1,2], and Jackie F. Price [1]

[1]Centre for Global Health, Usher Institute, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, UK; [2]MRC Human Genetics Unit, MRC Institute of Genetics and Cancer, University of Edinburgh, Western General Hospital, Crewe Road South, Edinburgh EH4 2XU, UK; and [3]Metabolic Unit, Western General Hospital, Crewe Road South, Edinburgh EH4 2XU, UK

Received 29 November 2022; revised 8 May 2023; accepted 10 May 2023; online publish-ahead-of-print 12 May 2023

Aims To identify a group of metabolites associated with incident cardiovascular disease (CVD) in people with type 2 diabetes and assess its predictive performance over-and-above a current CVD risk score (QRISK3).

Methods and results A panel of 228 serum metabolites was measured at baseline in 1066 individuals with type 2 diabetes (Edinburgh Type 2 Diabetes Study) who were then followed up for CVD over the subsequent 10 years. We applied 100 repeats of Cox least absolute shrinkage and selection operator to select metabolites with frequency >90% as components for a metabolites-based risk score (MRS). The predictive performance of the MRS was assessed in relation to a reference model that was based on QRISK3 plus prevalent CVD and statin use at baseline. Of 1021 available individuals, 255 (25.0%) developed CVD (median follow-up: 10.6 years). Twelve metabolites relating to fluid balance, ketone bodies, amino acids, fatty acids, glycolysis, and lipoproteins were selected to construct the MRS that showed positive association with 10-year cardiovascular risk following adjustment for traditional risk factors [hazard ratio (HR) 2.67; 95% confidence interval (CI) 1.96, 3.64]. The *c*-statistic was 0.709 (95%CI 0.679, 0.739) for the reference model alone, increasing slightly to 0.728 (95%CI 0.700, 0.757) following addition of the MRS. Compared with the reference model, the net reclassification index and integrated discrimination index for the reference model plus the MRS were 0.362 (95%CI 0.179, 0.506) and 0.041 (95%CI 0.020, 0.071), respectively.

Conclusion Metabolomics data might improve predictive performance of current CVD risk scores based on traditional risk factors in people with type 2 diabetes. External validation is warranted to assess the generalizability of improved CVD risk prediction using the MRS.

Lay summary This study looked at whether combining a group of new markers found in the blood (called metabolites) with traditional risk factors (such as high blood pressure and obesity) could more accurately predict how likely people with type 2 diabetes are to develop cardiovascular diseases in the next 10 years.

Key findings

- Twelve metabolites (including amino acids and lipids) showed strong association with 10-year cardiovascular risk in people with type 2 diabetes, and a metabolites-based risk score (MRS) was created by integrating these metabolites.
- Combining the MRS with traditional risk factors was better at predicting the risk of a person with T2D for developing cardiovascular diseases within the next 10 years than using traditional risk factors alone.

Keywords Cardiovascular diseases • Lipidomics • Metabolomics • Risk prediction model • Type 2 diabetes

Downloaded from https://academic.oup.com/eurjpc/advance-article/doi/10.1093/eurjpc/zwad160/7161066 by University of Edinburgh user on 30 June 2023

* Corresponding author. Tel: +44 7421817029, Email: zhe.huang@ed.ac.uk

Previous presentation: An abstract based on this work has been presented at the American Heart Association Scientific Session 2022.

© The Author(s) 2023. Published by Oxford University Press on behalf of the European Society of Cardiology.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

Cardiovascular disease (CVD) is common in people with type 2 diabetes (T2D) and is the leading contributor to premature mortality and economic burden in this high-risk population.[1,2] Substantial heterogeneity in levels of cardiovascular (CV) risk is found in people with T2D. It is critical to identify particularly high-risk sub-groups in order to better facilitate allocation of limited healthcare resources and target early intervention to prevent or delay the development of CVD.[3]

A variety of risk prediction algorithms for CVD have been developed in general populations and specifically in populations with T2D using traditional risk factors such as hypertension and hypercholesterolaemia. The UK National Institute for Health and Care Excellence clinical guideline recommends the QRISK2 risk score for assessing CV risk in people with T2D, and the American College of Cardiology/American Heart Association 2018 guideline advises a Pooled Cohort Equation to stratify populations with T2D by levels of CV risk.[4,5] However, studies on large independent cohorts have found the predictive performance of many existing CV risk scores in populations of T2D to be suboptimal.[6–8]

Instead of focusing on a small number of well-known biomarkers, metabolomics is the study of identifying and quantifying numerous low-molecular-weight endogenous compounds and exogenous chemicals in a single bio-sample.[9] As CVD is characterized by disturbances in cardiac and cerebrovascular metabolism, metabolomics might shed light into the molecular underpinnings of CVD.[10] Additionally, metabolomics is more proximal to disease phenotypes than other 'omics' (genomics, transcriptomics, and proteomics), and it could integrate upstream 'omics' variations and exposures to environmental factors. Metabolomics might therefore have potential to improve prediction of CVD that originates from interactions between genes and environments.[10]

The incremental value of adding metabolites to traditional CV risk factors for risk prediction has been assessed in general populations. The metabolites investigated, mainly lipid species, provided some additional value for CVD prediction.[11] In people with T2D, although several studies have explored association between metabolomic profiles and CVD, few have further evaluated the improved predictive performance of metabolites over traditional risk factors and/or existing CVD risk scores.[9] The limited studies that have addressed this important issue have been restricted by relatively short follow-up period, inclusion of only particular subclasses of metabolites (e.g. fatty acids), or inclusion of participants from unrepresentative populations of T2D.[12–15]

We aimed to investigate a wide range of potentially important metabolites, to identify from these a group of metabolites associated with 10-year CVD risk, and then assess the predictive performance of this panel of metabolites over-and-above a current CVD risk score (QRISK3) in a representative cohort of men and women with T2D.

Methods

Study population

The Edinburgh Type 2 Diabetes Study (ET2DS) is a prospective cohort study of 1066 men and women aged 60–75 years with T2D at baseline (2006/2007). Participants were randomly recruited from a representative diabetes register, the Lothian Diabetes Register, which captures almost everyone with diagnosed diabetes in Lothian, Scotland. Details on recruitment have been described previously.[16] Questionnaires and physical examinations were used to collect data at baseline, and again at Year 4 ($n = 831$) and Year 10 ($n = 581$). Hospital discharge data and death records for each participant were obtained at baseline, Year 4 and Year 8 from the Information and Services Division of NHS Scotland, and Year 10 from the National Records of Scotland.[17] As shown in *Figure 1*, after excluding individuals missing metabolomic profiling and/or information on CVD, 1021 individuals were finally included in this study. Ethical permission was granted by the Lothian Medical Research Ethics Committee, and written informed consent was obtained from all participants.

Metabolomic profiling

Fasting venous blood samples were collected at baseline, and serum was extracted for global metabolomic profiling using a high throughput targeted nuclear magnetic resonance platform (Nightingale, Helsinki, Finland) that has been described previously and applied in several large epidemiological studies.[18] The platform simultaneously quantifies 228 metabolites and derived ratios including lipid particles and subclasses, fatty acids, glycolysis related metabolites, amino acids, ketone bodies, fluid balance molecules, and inflammation marker. Concentrations of some metabolites were recorded as zero because they were below detection levels, and this resulted in some

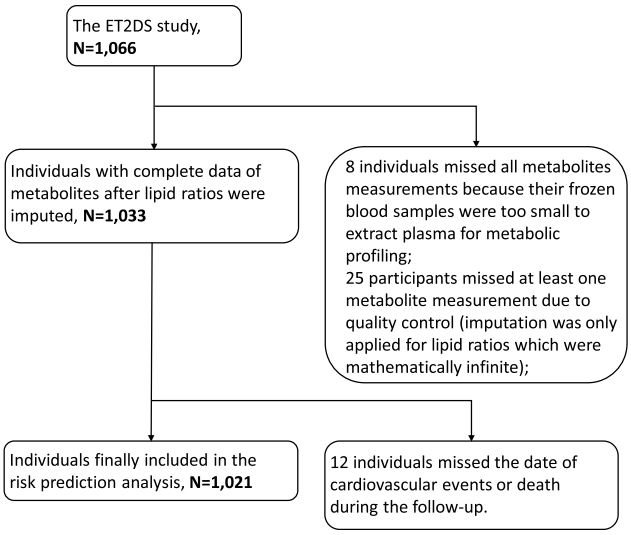

Figure 1 Flowchart of the study population.

Downloaded from https://academic.oup.com/eurjpc/advance-article/doi/10.1093/eurjpc/zwad160/7161066 by University of Edinburgh user on 30 June 2023

lipid ratios being mathematically infinite, and potential issues when log-transformation was applied. Thus, a value equal to half of the minimum concentration recorded in the dataset for the affected metabolite was used to replace these zero values, and then all affected metabolite ratios were re-calculated using the updated values.

Assessment of the outcome

Details of the assessments undertaken to identify CV events, and the criteria used to confirm such events, have been described previously.[19] Briefly, in this study, CVD is a composite phenotype consisting of myocardial infarction (MI), angina, transient ischaemic attack, stroke, and coronary intervention. Prevalent CVD at baseline was assessed using a combination of questionnaires, electrocardiograph, and hospital discharge data. Similarly, incident CV events were identified during follow-up using data from the same data sources plus scrutiny of death records and review of clinical case notes from general practices and hospitals as required to confirm events. Incident CVD in this study refers to the first new CV event for individuals free of CVD at baseline or the first recurrent CV event for those with existing CVD at baseline.

The reference cardiovascular disease risk score (QRISK3)

Developed and validated in 2017, the QRISK3 algorithm is an update of QRISK2 with the improvement for CVD prediction over its predecessor.[20] The components of QRISK3 include age, sex, ethnicity, deprivation, smoking, body mass index (BMI), systolic blood pressure (SBP), SBP variability [standard deviation (SD) of repeated measures], ratio of total cholesterol to cholesterol in high density lipoproteins (HDL), diabetes, chronic kidney disease (CKD, stage 3, 4, or 5), treated hypertension, family history of CVD, atrial fibrillation, rheumatoid arthritis, migraine, severe mental illness, systemic lupus erythematosus (SLE), erectile dysfunction (ED), atypical antipsychotic use, and corticosteroid use.

As predictive performance of original QRISK3 algorithm was not satisfactory in ET2DS [*c*-statistics: 0.647 (0.615, 0.679)], and people with existing CVD and/or with statin prescriptions were excluded in the derivation population of QRISK3, we decided to derive a reference model based on QRISK3 algorithm. In this reference risk model, diabetes and SLE were excluded for their non-variability within the population (all are T2D and non-SLE in ET2DS), and ED was removed to avoid decrease in statistical power due to sex-specific modelling. Prevalent CVD and lipid-lowering drug use was reported in a considerable proportion of population in ET2DS, which also represents a common situation in populations of T2D in the real world, so these two variables were added to the reference score. We tried to make definitions of individual components of the reference score in this study be as close as possible to that in the original study of QRISK3, and they were ascertained by a combination of questionnaires, physical examinations, laboratory tests, and data linkage with hospitalization and death records (see Supplementary material online, *Table S1*).

Statistical analyses

Characteristics of the population are shown as mean ± SD, median [lower interquartile range (IQR) and upper IQR], or *n* (%) for normally distributed variables, variables with skewed distribution, and categorical variables, respectively. Levels of metabolites were transformed into the natural logarithm scale and then standardized before regression modelling. Single imputation was used to address missing values of predictors in the reference score using the *mice* package (version 3.14.0) in R.

To construct a single metabolites-based risk score (MRS) that would integrate most informative metabolites associated with incident CVD, the least absolute shrinkage and selection operator (LASSO) with the Cox regression model was used for metabolites selection. The LASSO minimizes the residual sum of squares subject to the sum of the absolute value of the coefficients being less than a constant, and thus, predictors selection is achieved by forcing coefficients of some variables to be exactly zero.[21] Given this property, LASSO is a useful tool to handle highly correlated and high-dimensional data.

The tuning parameter (λ) in LASSO was chosen by five-fold cross-validation with one standard error rule using *cv.glmnet* function in *glmnet* package (version 4.0–2) in R. To obtain a stable group of components for the MRS, 100 repeats of Cox LASSO were applied using Bootstrap with replacement by changing the fold of cross-validation and random seeds (see Supplementary material online, *Figure S1*). Only metabolites with top frequencies (being selected more than 90% times) in the 100 repeats of selection were regarded as components of the MRS. An unadjusted and unpenalized Cox model with all selected metabolites and 10-year incident CVD was further performed, and the linear predictor (i.e. sum of regression coefficients multiplied by corresponding metabolites' concentration) in this Cox model was calculated as a weighted MRS for each participant.

To explore how adding the MRS will affect the associations between traditional CV risk factors with incident CVD, Cox regression was performed to estimate hazard ratios (HRs) and corresponding 95% confidence interval (CI) for associations between individual components of the reference score and 10-year CVD before and after the MRS was added into the reference model. Association of the MRS with individual components of the reference was visualized in a heatmap.

Predictive performance of the reference model and its combination with the MRS was assessed separately by discrimination (Harrell *c*-statistics) and calibration. Internal validation with 500 bootstrap repeats was performed to take account of overfitting. Added predictive value of the MRS over traditional risk factors for CVD prediction was assessed by continuous net reclassification improvement (NRI) and integrated discrimination improvement (IDI).[22,23] According to pre-specified CV risk categories (0–10%, 10–20%, and >20%), categorical NRIs (i.e. NRI_{event} and $NRI_{non\text{-}event}$) were calculated separately to indicate the fraction of individuals correctly reclassified by the updated model within the event and non-event groups in comparison with the reference model.[19]

The proportional hazard assumptions of Cox regression models were checked based on the scaled Schoenfeld residuals, and if assumptions were violated for some predictors, interactions between age and these predictors would be added into the Cox model. A *P*-value of <0.05 was considered significant. All analyses were performed using R version 4.0.3 (R Foundation for Statistical Computing, Vienna, Austria).

Results

Characteristics of the study population

Major demographics and CV risk factors of the study population at baseline are described in *Table 1*. During the median follow-up of 10.6 years, 255 (25.0%) individuals had a CV event, and 116 of these represented a first incidence of CVD. Distribution of serum metabolites is shown in Supplementary material online, *Table S2*, and substantial inter-correlations (correlation coefficients ranged between −0.998 and +0.999) between individual metabolites are displayed in Supplementary material online, *Figure S2*.

Construction of the metabolites-based risk score

In the construction of the MRS, the number of metabolites selected in the optimum subset in each single repeat of Cox LASSO ranged between 4 and 26. Overall, 33 metabolites (mainly subclasses of lipoproteins, amino acids, and ketone bodies) were selected after 100 repeats of Cox LASSO as being associated with incident CVD in at least one repeat of the analysis (see Supplementary material online, *Table S3*). Twelve metabolites with top frequencies (>90%) in the 100 repeats of selection were included as final components of the MRS.

The HRs of all 12 selected metabolites in the unpenalized and unadjusted Cox model are shown in *Figure 2*. Seven metabolites showed a positive direction of association with incident CVD and of these, three (3-hydroxybutyrate, creatinine, and phenylalanine) were statistically significantly associated at $P < 0.05$ [HRs (95%CI): 1.20 (1.06, 1.37), 1.34 (1.19, 1.51), and 1.17 (1.02, 1.35)]. Five metabolites showed an inverse direction of association with incident CVD, and of these, two ratios of free cholesterol [free cholesterol to total lipids ratio in IDL

Downloaded from https://academic.oup.com/eurjpc/advance-article/doi/10.1093/eurjpc/zwad160/7161066 by University of Edinburgh user on 30 June 2023

Table 1 Baseline characteristics of the study population ($n = 1021$)

Characteristics	
Demographics	
Age (years)	67.9 ± 4.2
Male	528 (51.7)
Ethnicity	
White	1007 (98.6)
Others	14 (1.4)
SIMD	
Quintile 1 (most deprived)	119 (11.7)
Quintile 2	202 (19.8)
Quintile 3	183 (17.9)
Quintile 4	185 (18.1)
Quintile 5 (least deprived)	332 (32.5)
Traditional CV risk factors	
Smoking	
Non-smoker	398 (39.0)
Ever-smoker	484 (47.4)
Light smoker (<10 cigarettes or equivalent/day)	29 (2.8)
Moderate smoker (10–19 cigarettes or equivalent/day)	46 (4.5)
Heavy smoker (20+ cigarettes or equivalent/day)	64 (6.3)
BMI (kg/m^2)	31.5 ± 5.6
SBP (mmHg)	133.2 ± 16.4
SBP variability[a]	13.3 ± 4.2
Ratio of total cholesterol to HDL cholesterol	3.6 ± 1.1
Chronic kidney disease	247 (24.2)
Treated hypertension	835 (81.8)
Lipid-lowering drug use	874 (85.6)
Additional CV risk factors in the reference model	
Probability of having family history of CVD[b]	0.36 ± 0.04
Atrial fibrillation	67 (6.6)
Rheumatoid arthritis	39 (3.8)
Migraine	5 (0.5)
Severe mental illness	26 (2.5)
Atypical antipsychotic use	6 (0.6)
Corticosteroid use	143 (14.0)
Prevalent CVD	365 (35.7)
Diabetes-related characteristics	
Plasma glucose (mmol/L)	7.6 ± 2.1
HbA1c (%)	7.4 ± 1.1
HbA1c (mmol/mol)	57.4 ± 12.2
Duration of diabetes (years)	6.0 (3.0, 11.0)

Data are presented as mean ± SD, *n* (%), or median (lower IQR and upper IQR).
SD, standard deviation; IQR, interquartile range; SIMD, Scottish Index of Multiple Deprivation; BMI, body mass index; SBP, systolic blood pressure; CVD, cardiovascular diseases; HbA1c, haemoglobin A1c.
[a]Based on standard deviation of ≥2 historic systolic blood pressure values within 5 years before baseline.
[b]A proxy was used here to approximate the conditional probability of having a family history of CVD.

(intermediate density lipoproteins) and free cholesterol to total lipids ratio in small HDL] showed significant association with incident CVD [HRs (95%CI): 0.87 (0.77, 0.98) and 0.74 (0.64, 0.85)].

The MRS varies from a minimum of −1.62 to a maximum of 2.26, and mean ± SD MRS was 0.0 ± 0.6. Its distribution is approximately normal and is displayed in Supplementary material online, *Figure S3*.

Association of individual cardiovascular risk factors and the metabolites-based risk score with 10-year incident cardiovascular disease

Associations between individual components of the reference model and 10-year incident CVD, both before and after addition of the MRS, are described in Supplementary material online, *Table S4*. In the presence of the MRS, most factors that showed significant association with 10-year CVD remained as significant predictors, except for the ratio of total cholesterol to HDL cholesterol and CKD. The MRS showed strong association with CVD with a HR of 2.67 (95%CI 1.96, 3.64) after adjustment for individual components of the reference model.

Predictive values of the metabolites-based risk over-and-above the reference model

Stratified CVD-free survival curves by predicted risk quartiles in the reference model and the model combining the reference model with the MRS are shown separately in *Figure 3*, illustrating that the two models had similar discriminative ability for 10-year CVD but the model with the MRS showed a slightly better performance for low-risk groups (the 1st and 2nd quartiles).

Table 2 lists metrics of predictive performance for the reference model and its combination with the MRS. The *c*-statistic for the reference model was 0.709 (95%CI 0.679, 0.739), and it was slightly increased to 0.728 (95%CI 0.700, 0.757) after the MRS was added. Similarly, the improvement in discriminative ability by adding the MRS were internally validated by bootstrapping 500 times, with *c*-statistics of 0.704 (95%CI 0.675, 0.732) for the combination model and 0.684 (95%CI 0.654, 0.714) for the reference model (*Table 2*).

In terms of reclassification, the addition of the MRS to the reference model resulted in a continuous NRI (category-free) of 0.362 (95%CI 0.179, 0.506) with 8.5% of non-events correctly reclassified for the combination model compared with the reference model alone [$NRI_{non\text{-}event} = 0.085$ (95%CI 0.017, 0.157)]. The IDI of 0.041 (95%CI 0.020, 0.071) indicates that the combination model has an absolute increase of 4.1% in mean predicted risk for participants with incident CV events compared with participants without events over the reference model. Calibration performance of the two models is shown in Supplementary material online, *Figure S4* that illustrates good agreement between observed risk and predicted risk in both two models.

Association of the metabolites-based risk with individual components of the reference model

Association of the MRS and its constitutive metabolites with individual components of the reference model was illustrated in Supplementary material online, *Figure S5*. The correlation coefficients of the association between the MRS and individual CV risk factors ranged between −0.338 and +0.590, and the strongest correlation was for the MRS and ratio of total cholesterol to HDL cholesterol.

Discussion

In older people with T2D, a metabolomic risk score (which integrated information on lipoproteins, amino acids, ketone bodies, and markers of fluid balance) provided some improvement in the prediction of

Downloaded from https://academic.oup.com/eurjpc/advance-article/doi/10.1093/eurjpc/zwad160/7161066 by University of Edinburgh user on 30 June 2023

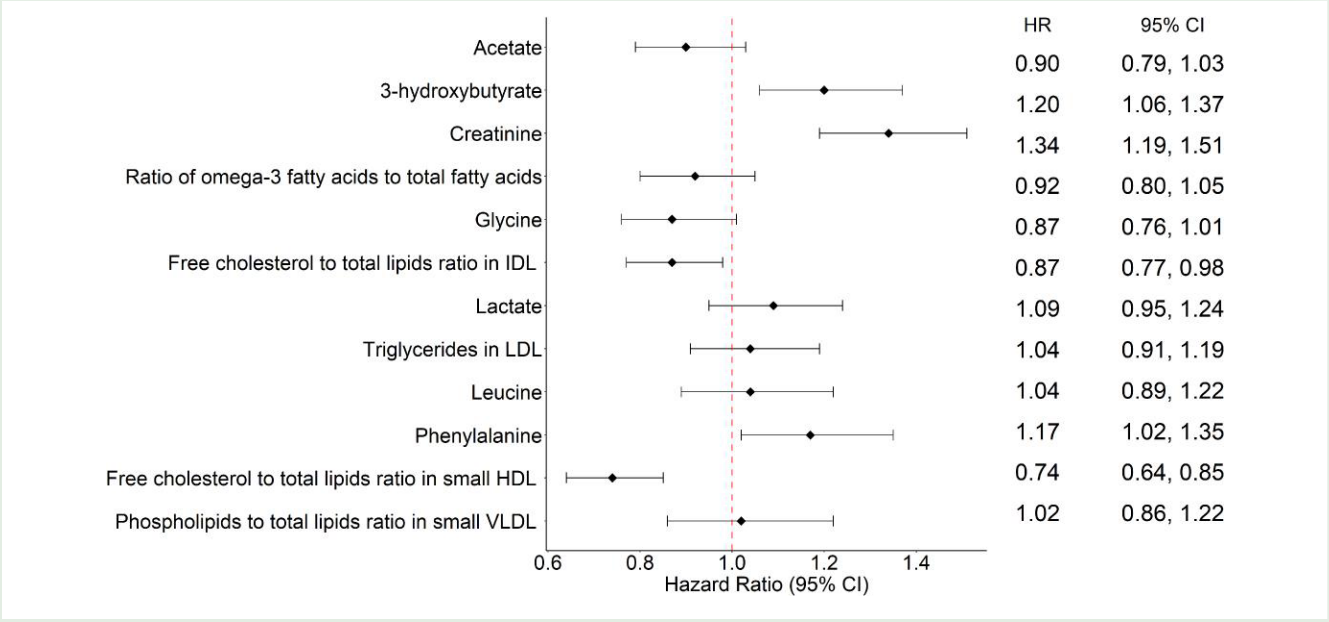

Figure 2 Association between 12 selected individual metabolites and 10-year cardiovascular disease in the unpenalized and unadjusted Cox model. The 12 metabolites entered the Cox model together, and no traditional cardiovascular risk factors were adjusted. HDL, high-density lipoprotein; IDL, intermediate-density; LDL, low-density lipoprotein; VLDL, very low-density lipoprotein; HR, hazard ratio; CI confidence interval.

10-year CV risk over and above traditional CV risk factors. Slight improvement was seen in predictive performance of the updated model with the MRS in terms of discrimination and calibration, and there was better reclassification, in comparison with the reference model based on QRISK3.

Individual metabolites and improved cardiovascular risk prediction

Prior studies are often done in the general population and tend to assess predictive performance by adding individual metabolites into reference models, rather than integrating them into one index,[11] with candidate metabolites primarily reflecting lipid species and/or amino acids.[24–27] The *c*-statistics in these studies ranged from 0.71 to 0.76, and after metabolites were added into the model, they were slightly improved to between 0.72 and 0.79.[24–27] Of note, Mundra *et al.*[28] investigated predictive performance of metabolites in both general and populations of T2D and suggested that predictive models should be optimized for the diabetes population because of evident interactions between diabetes status and identified lipids.

Metabolomics studies focusing on improving CV risk prediction in people with T2D are even more scarce, with the majority based on a single population from the ADVANCE clinical trial.[12,13,28,29] Rather than investigating global metabolomic profiles as in our study, analyses from ADVANCE separately assessed predictive performance of specific subclasses of metabolites. In an exploration of the predictive performance of fatty acids for CV events over 5 years, the percentage of omega-3 fatty acids to total fatty acids resulted in a very limited increase in *c*-statistic (from 0.692 to 0.695) together with increased accuracy in reclassification (continuous NRI = 0.144).[13] Notably, this fatty acid ratio was also identified as a component of the MRS in our study. In ADVANCE, there was also a modest incremental value of adding seven amino acids into the reference model.[12] Although three of these amino acids (glycine, leucine, and phenylalanine) were also identified in our study, other metabolites found to increase prediction of CVD in people with T2D in previous studies were not measured in our study.[14,29]

Predictive values of the metabolites-based risk score

A meta-analysis suggests that metabolites scores tend to have stronger association with CVD than individual metabolites.[11] An MRS-like index has been assessed in several studies in the general population,[30–32] but we found only one small study in a population of T2D that applied the 'score' technique to explore incremental values of six amino acids for predicting CVD.[33]

The key contributors to the MRS in our study were creatinine, free cholesterol to total lipids ratio in small HDL, 3-hydroxybutyrate, and phenylalanine based on their weights. Of note, four components of the MRS were lipid-related metabolites or ratios, indicating that high-resolution lipidomic profile might capture the complexity of the altered lipid metabolism underlying the increased CV risk in T2D and thus contribute to improved risk prediction. Interestingly, although associations of two strong predictors (CKD and the ratio of total to HDL cholesterol) with incident CVD were largely attenuated after the MRS was added to the model, the updated model still had some improvement in predictive performance. On the one hand, this might suggest that some components of the MRS could lie on the pathophysiological pathways of these well-known risk factors, which could be further confirmed by correlations between components of the MRS and these risk factors. On the other hand, residual risk, which could not be reflected by traditional risk factors, might be explained by the MRS, as metabolomic profiles might serve as mediators of association between CVD and some unmeasured risk factors such as physical activities and diet.[34,35]

The limited improvement in discrimination might be explained by good performance of the reference model, since *c*-statistics are often insensitive in reflecting improved model performance after adding novel biomarkers into a model that already contains strong

Downloaded from https://academic.oup.com/eurjpc/advance-article/doi/10.1093/eurjpc/zwad160/7161066 by University of Edinburgh user on 30 June 2023

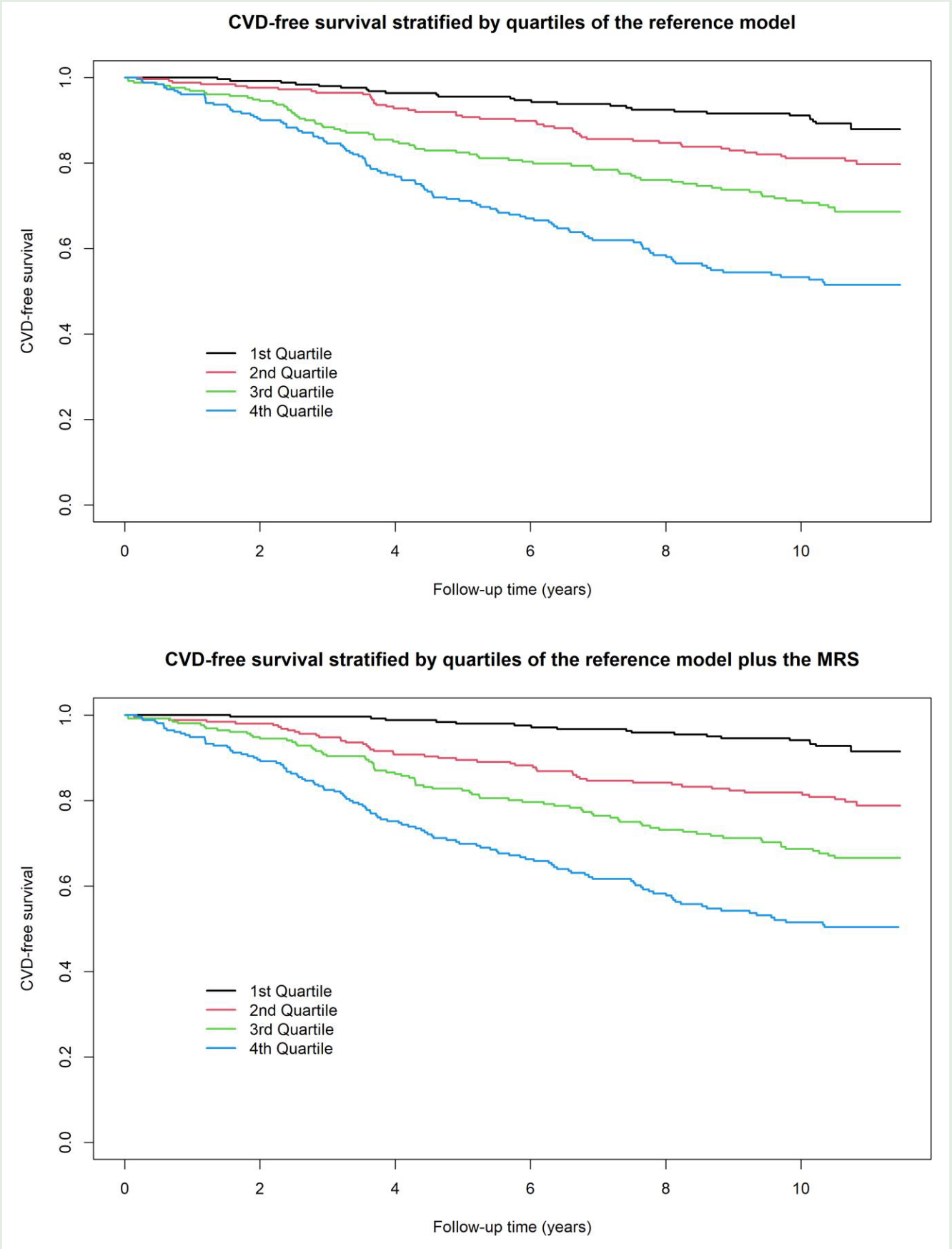

Figure 3 Kaplan–Meier curves of cardiovascular disease-free survival probability stratified by predicted risk in the reference model and its combination with the MRS. CVD, cardiovascular disease; MRS, metabolites-based risk score.

Downloaded from https://academic.oup.com/eurjpc/advance-article/doi/10.1093/eurjpc/zwad160/7161066 by University of Edinburgh user on 30 June 2023

Table 2 Metrics of predictive performance of the reference model and its combination with the MRS for 10-year CVD risk

Metrics	Reference model	Reference model plus the MRS
c-Statistics	0.709 (0.679, 0.739)	0.728 (0.700, 0.757)
Optimism-adjusted *c*-statistics (internal validation)	0.684 (0.654, 0.714)	0.704 (0.675, 0.732)
Continuous NRI	Ref.	0.362 (0.179, 0.506)
NRI_{event}	Ref.	−0.010 (−0.053, 0.068)
$NRI_{non\text{-}event}$	Ref.	0.085 (0.017, 0.157)
IDI	Ref.	0.041 (0.020, 0.071)

MRS, metabolites-based risk score; NRI, net reclassification improvement; IDI, integrated discrimination improvement.

predictors.[19,36] Thus, it has been suggested that reclassification metrics might be more suitable to reflect the incremental predictive value of novel biomarkers in this situation.[36] The model with the MRS in our study showed an improved net reclassification of non-events [$NRI_{non\text{-}event} = 0.085$ (95%CI 0.017, 0.157)], which means net reclassification of 65 of 766 people without CVD assigned a lower risk in the combination model. Given that the categorized NRI relies on selection of risk groups and events rates,[37] continuous NRI and IDI were also calculated in our study, which indicates improved reclassification of the model with the MRS.

Potential clinical utility

Although the improvement in discrimination by adding the MRS to traditional risk factors is modest in this study, a slight improvement may still be clinically important when the disease impacts a large population (e.g. CVD). Further, significantly improved reclassification in the non-event group might help more people, who will not develop CVD (the majority of T2D population), avoid unnecessary medication treatments. However, identifying potentially important metabolites to include in a risk score is just the first step to evaluating their clinical utility, and further investigation on its generalizability and cost-effectiveness should be thoroughly assessed before the MRS is implemented in clinic.

Strengths and limitations

In terms of study strengths, our study was based on a representative population of T2D and has the longest follow-up period to date among metabolomics studies for predicting CV risks in people with T2D. Also, this study included various well-defined phenotypes enabling us to thoroughly depict CV risk of participants. Most importantly, we used LASSO, a prediction-oriented modern technique, to select candidate metabolites and construct the MRS, which could take account of overfitting.

However, findings in this study should be interpreted in the context of several limitations. This study is limited by little variation of some predictors, and despite being one of the largest metabolomics studies for predicting CVD in people with T2D, it still has limited statistical power due to sample size. Another potential limitation of our study is the relatively selective set of co-variables that we chose to include in our models, and which, were this to have been analysed with less focus on the additive value of the MRS over-and-above an established risk score, could be viewed as confounding variables. However, we were able to include most predictors of the QRISK3 plus prevalent CVD and lipid-lowering drug use, which met the overall aim of our approach. Despite this, it is entirely possible that residual confounders might remain that could lead to a less 'competitive' reference model compared with a model that did not rely on the improvement of an existing model alone. Moreover, definitions of some predictors and CVD in this study are not exactly the same as that in original QRISK3 study. To this end, we referred to definitions or proxies used in prior peer-reviewed literature, but this might still restrict the predictive performance of the reference model and introduce measurement errors. Whilst we applied internal validations on this representative population of T2D, it is just a minimum prerequisite of generalizability, so further external validation using independent populations of T2D is now required. Furthermore, clinical interpretation of some metrics (i.e. continuous NRI) is still unclear,[38] and we were unable to quantify the impact of the updated predictive model in practice by evaluating individuals' behaviours and cost-effectiveness of care, so this merits randomized clinical trials and economic modelling in the future.

In conclusion, in this study of 1021 participants from a representative cohort of T2D population, we identified 12 metabolites associated with 10-year CV risk and constructed a risk score based on them. The model combining the MRS and traditional CV risk factors performed better than the reference model with slight improvement in discrimination and reclassification, but generalizability of the model needs to be validated in external populations of T2D.

Supplementary material

Supplementary material is available at *European Journal of Preventive Cardiology*.

Acknowledgements

We would like to thank all the participants, staff, and researchers of the ET2DS.

Author contributions

J.F.P. and M.W.J.S. were involved in the conception and design of the ET2DS and oversaw acquisition and analysis of data. Z.H. performed data analysis and interpreted results in this paper. Z.H. and J.F.P. wrote the first draft of the paper and finalized the manuscript based on feedbacks from all co-authors. L.K., J.K., W.K., and J.F.W. provided substantial advice on study design, statistical methods, and interpretation of findings. L.K., J.K., W.K., M.W.J.S., and J.F.W. critically revised and commented on the manuscript. All authors approved the final version of the manuscript. Z.H. is the guarantor of this work and, as such, takes responsibility for the integrity of the data and the accuracy of the data analysis.

Funding

The ET2DS was funded by the grant from the Medical Research Council (UK) (Project Grant G0500877) and the Chief Scientist Office of Scotland (Program Support Grant CZQ/1/38). Z.H. was supported by a PhD studentship from the Darwin Trust of Edinburgh. The work of L.K. was supported by an RCUK Innovation Fellowship from the National Productivity Investment Fund (MR/R026408/1). All these funding bodies played no role in study design, data analysis, and manuscript writing.

Conflict of interest: None declared.

Data availability

The data set analysed during current study is not publicly available due to it containing information that could compromise research participant privacy/

Downloaded from https://academic.oup.com/eurjpc/advance-article/doi/10.1093/eurjpc/zwad160/7161066 by University of Edinburgh user on 30 June 2023

consent, but aggregate data and analytical plan might be available from the corresponding author on reasonable request.

References

1. Einarson TR, Acs A, Ludwig C, Panton UH. Prevalence of cardiovascular disease in type 2 diabetes: a systematic literature review of scientific evidence from across the world in 2007–2017. *Cardiovasc Diabetol* 2018;**17**:83.
2. Einarson TR, Acs A, Ludwig C, Panton UH. Economic burden of cardiovascular disease in type 2 diabetes: a systematic review. *Value Health* 2018;**21**:881–890.
3. Bachmann KN, Wang TJ. Biomarkers of cardiovascular disease: contributions to risk prediction in individuals with diabetes. *Diabetologia* 2018;**61**:987–995.
4. NICE CG181. Cardiovascular disease: risk assessment and reduction, including lipid modification. https://www.nice.org.uk/guidance/cg181/chapter/1-Recommendations#identifying-and-assessing-cardiovascular-disease-cvd-risk-2 (06/27 2022)
5. Grundy SM, Stone NJ, Bailey AL, Beam C, Birtcher KK, Blumenthal RS, *et al.* 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the management of blood cholesterol: a report of the American College of Cardiology/American Heart Association task force on clinical practice guidelines. *Circulation* 2019;**139**:e1082–e1143.
6. Read SH, van Diepen M, Colhoun HM, Lindsay RS, McKnight JA, McAllister DA, *et al.* Performance of cardiovascular disease risk scores in people diagnosed with type 2 diabetes: external validation using data from the National Scottish Diabetes Register. *Diabetes Care* 2018;**41**:2010–2018.
7. Dziopa K, Asselbergs FW, Gratton J, Chaturvedi N, Schmidt AF. Cardiovascular risk prediction in type 2 diabetes: a comparison of 22 risk scores in primary care settings. *Diabetologia* 2022;**65**:644–656.
8. van der Leeuw J, van Dieren S, Beulens JWJ, Boeing H, Spijkerman AM, van der Graaf Y, *et al.* The validation of cardiovascular risk scores for patients with type 2 diabetes mellitus. *Heart* 2015;**101**:222–229.
9. Jin Q, Ma RCW. Metabolomics in diabetes and diabetic complications: insights from epidemiological studies. *Cells* 2021;**10**:2832.
10. McGarrah RW, Crown SB, Zhang GF, Shah SH, Newgard CB. Cardiovascular metabolomics. *Circ Res* 2018;**122**:1238–1258.
11. McGranaghan P, Saxena A, Rubens M, Radenkovic J, Bach D, Schleußner L, *et al.* Predictive value of metabolomic biomarkers for cardiovascular disease risk: a systematic review and meta-analysis. *Biomarkers* 2020;**25**:101–111.
12. Welsh P, Rankin N, Li Q, Mark PB, Würtz P, Ala-Korpela M, *et al.* Circulating amino acids and the risk of macrovascular, microvascular and mortality outcomes in individuals with type 2 diabetes: results from the ADVANCE trial. *Diabetologia* 2018;**61**:1581–1591.
13. Harris K, Oshima M, Sattar N, Würtz P, Jun M, Welsh P, *et al.* Plasma fatty acids and the risk of vascular disease and mortality outcomes in individuals with type 2 diabetes: results from the ADVANCE study. *Diabetologia* 2020;**63**:1637–1647.
14. Shah HS, Moreno LO, Morieri ML, Tang Y, Mendonca C, Jobe JM, *et al.* Serum orotidine: a novel biomarker of increased CVD risk in T2D discovered through metabolomics studies. *Diabetes Care* 2022:**45**;1882–1892.
15. Chen Y, Jia H, Qian X, Wang J, Yu M, Gong Q, *et al.* Circulating palmitoyl sphingomyelin is associated with cardiovascular disease in individuals with type 2 diabetes: findings from the China Da Qing diabetes study. *Diabetes Care* 2022;**45**:666–673.
16. Price JF, Reynolds RM, Mitchell RJ, Williamson RM, Fowkes FGR, Deary IJ, *et al.* The Edinburgh Type 2 Diabetes Study: study protocol. *BMC Endocr Disord* 2008;**8**:18.
17. Sluiman AJ, McLachlan S, Forster RB, Strachan MWJ, Deary IJ, Price JF. Higher baseline inflammatory marker levels predict greater cognitive decline in older people with type 2 diabetes: year 10 follow-up of the Edinburgh Type 2 Diabetes Study. *Diabetologia* 2022;**65**:467–476.
18. Wurtz P, Kangas AJ, Soininen P, Lawlor DA, Davey Smith G, Ala-Korpela M. Quantitative serum nuclear magnetic resonance metabolomics in large-scale epidemiology: a primer on -omic technologies. *Am J Epidemiol* 2017;**186**:1084–1096.
19. Price AH, Weir CJ, Welsh P, McLachlan S, Strachan MWJ, Sattar N, *et al.* Comparison of non-traditional biomarkers, and combinations of biomarkers, for vascular risk prediction in people with type 2 diabetes: the Edinburgh Type 2 Diabetes Study. *Atherosclerosis* 2017;**264**:67–73.
20. 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* 2017;**357**:j2099.
21. Tibshirani R. Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society: Series B (Methodological)* 1996;**58**:267–288.
22. Kerr KF, McClelland RL, Brown ER, Lumley T. Evaluating the incremental value of new biomarkers with integrated discrimination improvement. *Am J Epidemiol* 2011;**174**:364–374.
23. Pencina MJ, D'Agostino RB, Pencina KM, Janssens AC, Greenland P. Interpreting incremental value of markers added to risk prediction models. *Am J Epidemiol* 2012;**176**:473–481.
24. Stegemann C, Pechlaner R, Willeit P, Langley SR, Mangino M, Mayr U, *et al.* Lipidomics profiling and risk of cardiovascular disease in the prospective population-based Bruneck study. *Circulation* 2014;**129**:1821–1831.
25. Paynter NP, Balasubramanian R, Giulianini F, Wang DD, Tinker LF, Gopal S, *et al.* Metabolic predictors of incident coronary heart disease in women. *Circulation* 2018;**137**:841–853.
26. Wurtz P, Havulinna AS, Soininen P, Tynkkynen T, Prieto-Merino D, Tillin T, *et al.* Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation* 2015;**131**:774–785.
27. Wang Z, Zhu C, Nambi V, Morrison AC, Folsom AR, Ballantyne CM, *et al.* Metabolomic pattern predicts incident coronary heart disease. *Arterioscler Thromb Vasc Biol* 2019;**39**:1475–1482.
28. Mundra PA, Barlow CK, Nestel PJ, Barnes EH, Kirby A, Thompson P, *et al.* Large-scale plasma lipidomic profiling identifies lipids that predict cardiovascular events in secondary prevention. *JCI Insight* 2018;**3**:e121326.
29. Alshehry ZH, Mundra PA, Barlow CK, Mellett NA, Wong G, McConville MJ, *et al.* Plasma lipidomic profiles improve on traditional risk factors for the prediction of cardiovascular events in type 2 diabetes mellitus. *Circulation* 2016;**134**:1637–1650.
30. Cheng ML, Wang CH, Shiao MS, Liu M-H, Huang Y-Y, Huang C-Y, *et al.* Metabolic disturbances identified in plasma are associated with outcomes in patients with heart failure: diagnostic and prognostic value of metabolomics. *J Am Coll Cardiol* 2015;**65**:1509–1520.
31. Vaarhorst AA, Verhoeven A, Weller CM, Böhringer S, Göraler S, Meissner A, *et al.* A metabolomic profile is associated with the risk of incident coronary heart disease. *Am Heart J* 2014;**168**:45–52 e47.
32. Wang DD, Toledo E, Hruby A, Rosner BA, Willett WC, Sun Q, *et al.* Plasma ceramides, Mediterranean diet, and incident cardiovascular disease in the PREDIMED trial (Prevencion con Dieta Mediterranea). *Circulation* 2017;**135**:2028–2040.
33. Kume S, Araki S, Ono N, Shinhara A, Muramatsu T, Araki H, *et al.* Predictive properties of plasma amino acid profile for cardiovascular disease in patients with type 2 diabetes. *PLoS One* 2014;**9**:e101219.
34. Pang Y, Kartsonaki C, Du H, Millwood IY, Guo Y, Chen Y, *et al.* Physical activity, sedentary leisure time, circulating metabolic markers, and risk of major vascular diseases. *Circ Genom Precis Med* 2019;**12**:386–396.
35. Lara-Guzman OJ, Alvarez R, Munoz-Durango K. Changes in the plasma lipidome of healthy subjects after coffee consumption reveal potential cardiovascular benefits: a randomized controlled trial. *Free Radic Biol Med* 2021;**176**:345–355.
36. Moons KG, Kengne AP, Woodward M, *et al.* Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart* 2012;**98**:683–690.
37. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med* 2011;**30**:11–21.
38. Kerr KF, Wang Z, Janes H, McClelland RL, Psaty BM, Pepe MS. Net reclassification indices for evaluating risk prediction instruments: a critical review. *Epidemiology* 2014;**25**:114–121.

Downloaded from https://academic.oup.com/eurjpc/advance-article/doi/10.1093/eurjpc/zwad160/7161066 by University of Edinburgh user on 30 June 2023