Aus dem Deutschen Krebsforschungszentrum Heidelberg

Wissenschaftlicher Stiftungsvorstand: Prof. Dr. med. Michael Baumann

Abteilung: Klinische Epidemiologie der Krebsfrüherkennung

Leiterin: Dr. Sigrid Carlsson

# Improvement of prediction models for cardiovascular disease and type 2 diabetes mellitus with multi-omics

**Inauguraldissertation**

zur Erlangung des *Doctor medicinae (Dr. med.)*

an der

Medizinischen Fakultät Heidelberg

der

Ruprecht-Karls-Universität

vorgelegt von

Ruijie Xie

aus

Hunan, Volksrepublik China

2025

Dekan: Prof. Dr. Hans-Georg Kräusslich

Doktorvater: Prof. Dr. Ben Schöttker

# Contents

List of figures ....9

List of tables ....11

List of abbreviations ....12

1 Introduction ....15

1.1 Public health relevance of cardiovascular disease and type 2 diabetes ....15

1.2 Risk prediction models in epidemiology ....17

1.3 Omics biomarkers as tools to improve prediction ....18

1.4 Integrating multi-omics data for improved prediction ....20

1.5 Aims of the dissertation ....22

2 Material and Methods ....23

2.1 Metabolomics for cardiovascular risk prediction in individuals with type 2 diabetes ....23

2.1.1 Study population ....23

2.1.2 In- and exclusion criteria ....23

2.1.3 Metabolites quantification ....24

2.1.4 Variables of the SCORE2-Diabetes model ....24

2.1.5 Outcome assessment ....25

2.1.6 Statistical analyses ....25

2.2 Metabolomics for cardiovascular risk prediction in the general population ....28

2.2.1 Study population ....28

2.2.2 In- and exclusion criteria ....29

2.2.3 Metabolites quantification ....29

2.2.4 SCORE2 model variables ....29

2.2.5 Outcome assessment ....29

2.2.6 Statistical analyses ....29

2.3 Metabolomics for type 2 diabetes risk prediction ....30

2.3.1 Study population ....30

2.3.2 In- and exclusion criteria ....31

2.3.3 Metabolites quantification ....31

2.3.4 Variables of the clinical CDRS ..... 31
2.3.5 Type 2 incidence ascertainment ..... 31
2.3.6 Statistical analyses ..... 32
2.4 Proteomics for type 2 diabetes risk prediction ..... 33
2.4.1 Study population ..... 33
2.4.2 In- and exclusion criteria ..... 33
2.4.3 Proteomics assessment ..... 33
2.4.4 Reference model: clinical Cambridge Diabetes Risk Score (CDRS) ..... 34
2.4.5 Ascertainment of incident type 2 diabetes ..... 34
2.4.6 Statistical analyses ..... 34
2.5 Proteomics for cardiovascular risk prediction in the general population ..... 36
2.5.1 Study population ..... 36
2.5.2 In- and exclusion criteria ..... 36
2.5.3 Proteomic profiling ..... 36
2.5.4 Assessment of covariates ..... 37
2.5.5 Outcome ascertainment ..... 37
2.5.6 Statistical analyses ..... 37
2.6 Inflammation-related proteins for cardiovascular risk prediction ..... 38
2.6.1 Study population ..... 38
2.6.2 In- and exclusion criteria ..... 38
2.6.3 Measurement of CRP and inflammation-related proteins ..... 38
2.6.4 Assessment of covariates ..... 38
2.6.5 Outcome ascertainment ..... 38
2.6.6 Statistical analyses ..... 39
2.7 Proteomics and Multi-omics for cardiovascular risk prediction in individuals with type 2 diabetes ..... 40
2.7.1 Study population ..... 40
2.7.2 In- and exclusion criteria ..... 40
2.7.3 Proteomics data ..... 40
2.7.4 Metabolomics data ..... 41

2.7.5 Polygenic risk score for CVD....41
2.7.6 Variables of the SCORE2-Diabetes model....41
2.7.7 Outcome ascertainment....41
2.7.8 Statistical analyses....41
2.8 Multi-omics for cardiovascular risk prediction in the general population....43
2.8.1 Study population....43
2.8.2 In- and exclusion criteria....43
2.8.3 Proteomics data....43
2.8.4 Metabolomics data....44
2.8.5 Polygenic risk score for CVD....44
2.8.6 Variables of the SCORE2 model....44
2.8.7 Outcome ascertainment....44
2.8.8 Statistical analyses....44
2.9 Multi-omics for type 2 diabetes risk prediction....46
2.9.1 Study population....46
2.9.2 In- and exclusion criteria....46
2.9.3 Proteomics data....47
2.9.4 Metabolomics data....47
2.9.5 Polygenic risk score for CVD....47
2.9.6 Reference model: The clinical CDRS....47
2.9.7 Outcome ascertainment....47
2.9.8 Statistical analyses....47
3 Results....50
3.1 Metabolomics for cardiovascular risk prediction in individuals with type 2 diabetes....50
3.1.1 Description of the population included....50
3.1.2 Associations of metabolomic biomarkers with MACE....51
3.1.3 MACE risk prediction by metabolomic biomarkers....52
3.1.4 Replacement of the clinical chemistry-based biomarkers of the original SCORE2-Diabetes model by NMR metabolomics....55
3.2 Metabolomics for cardiovascular risk prediction in the general population....57

3.2.1 Description of the population included .....57
3.2.2 Associations of metabolomic biomarkers with MACE .....57
3.2.3 MACE risk prediction by metabolomic biomarkers.....60
3.3 Metabolomics for type 2 diabetes risk prediction .....64
3.3.1 Description of the population included .....64
3.3.2 Associations of selected metabolites with incident T2D.....65
3.3.3 Improvements in T2D risk prediction by metabolomic biomarkers.....67
3.4 Proteomics for type 2 diabetes risk prediction .....71
3.4.1 Description of the population included .....71
3.4.2 Associations of selected proteins with incident type 2 diabetes.....72
3.4.3 Improvements in T2D risk prediction by metabolomic biomarkers.....75
3.5 Proteomics for cardiovascular risk prediction in the general population .....78
3.5.1 Description of the population included .....78
3.5.2 Associations of selected proteins with MACE .....78
3.5.3 MACE risk prediction by proteomic biomarkers .....80
3.6 Inflammation-related proteins for cardiovascular risk prediction .....84
3.6.1 Description of the population included .....84
3.6.2 Biomarker Selection and Model Derivation.....84
3.6.3 Associations of Selected Proteins with MACE .....85
3.6.4 Model Performance Validation .....86
3.6.5 Sensitivity analyses .....91
3.7 Proteomics and Multi-omics for cardiovascular risk prediction in individuals with type 2 diabetes .....92
3.7.1 Description of the population included .....92
3.7.2 Correlations among selected biomarkers and associations with MACE**Fehler! Textmarke nicht definiert.**
3.7.3 Predictive performance of proteomics and multi-omics extended SCORE2-Diabetes models .....93
3.8 Multi-omics for cardiovascular risk prediction in the general population .....101
3.8.1 Description of the population included .....101

3.8.2 Correlations among biomarkers and associations with MACE ..... 102
3.8.3 MACE risk prediction by multi-omics ..... 107
3.9 Multi-omics for type 2 diabetes risk prediction ..... 112
3.9.1 Description of the population included ..... 112
3.9.2 Correlations among selected omics biomarkers ..... 113
3.9.3 Associations of selected omics biomarkers with type 2 diabetes ..... 114
3.9.4 Improvements in Risk Prediction by Multi-Omics Integration ..... 116
4 Discussion ..... 120
4.1 Metabolomics for cardiovascular risk prediction in individuals with type 2 diabetes ..... 120
4.1.1 Comparison with previous studies ..... 120
4.1.2 Biological mechanisms of selected metabolites ..... 121
4.1.3 Strengths and limitations ..... 122
4.1.4 Potential clinical utility of the derived model ..... 123
4.2 Metabolomics for cardiovascular risk prediction in the general population ..... 124
4.2.1 Comparison with previous studies & novelty ..... 124
4.2.2 Biological mechanisms of selected metabolites ..... 125
4.2.3 Strengths and limitations ..... 127
4.2.4 Potential clinical implications of the derived model ..... 128
4.3 Metabolomics for type 2 diabetes risk prediction ..... 129
4.3.1 Comparison with previous studies & novelty ..... 129
4.3.2 Biological mechanisms of selected metabolites ..... 130
4.3.3 Strengths and limitations ..... 132
4.3.4 Potential clinical implications of the derived model ..... 133
4.4 Proteomics for type 2 diabetes risk prediction ..... 133
4.4.1 Comparison with previous studies & novelty ..... 134
4.4.2 Biological mechanisms of selected proteins ..... 135
4.4.3 Strengths and limitations ..... 137
4.5 Proteomics for cardiovascular risk prediction in the general population ..... 138
4.5.1 Comparison with previous studies & novelty ..... 138
4.5.2 Biological mechanisms of selected proteins ..... 139

4.5.3 Strengths and limitations ..... 141
4.6 Inflammation-related proteins for cardiovascular risk prediction ..... 141
4.6.1 Comparison with previous studies & novelty ..... 141
4.6.2 Biological mechanisms of selected proteins ..... 143
4.6.3 Strengths and limitations ..... 144
4.6.4 Potential clinical implications of the derived model ..... 145
4.7 Proteomics and Multi-omics for cardiovascular risk prediction in individuals with type 2 diabetes ..... 146
4.7.1 Comparison with previous studies & novelty ..... 146
4.7.2 Clinical Translation and Feasibility ..... 147
4.7.3 Biological mechanisms of selected proteins ..... 148
4.7.4 Strengths and limitations ..... 150
4.8 Multi-omics for cardiovascular risk prediction in the general population ..... 151
4.8.1 Comparison with previous studies & novelty ..... 151
4.8.2 Strengths and limitations ..... 153
4.9 Multi-omics for type 2 diabetes risk prediction ..... 153
4.9.1 Comparison with previous studies & novelty ..... 154
4.9.2 Strengths and limitations ..... 155
4.10 Conclusion ..... 156
5 Summary ..... 158
5.1 English summary ..... 158
5.2 Deutsche Zusammenfassung ..... 160
6 References ..... 162
7 Own publications and contributions ..... 187
Appendix ..... 192
Curriculum Vitae ..... 281
Acknowledgements ..... 282
Eidesstattliche Versicherung ..... 283

# List of figures

Figure 1. Study design for T2D model development and validation ..... 36
Figure 2. Flowchart of the two-stage biomarker selection process..... 43
Figure 3. Schematic of the multi-omics-extended SCORE2 model ..... 45
Figure 4. Schematic of the integrated clinical and multi-omics T2D prediction model ..... 49
Figure 5. Associations between selected metabolites and MACE across sexes in the derivation set (70% of UK biobank, N=7,180) ..... 54
Figure 6. The incremental discrimination of each metabolite for the model after the selected metabolites were added to SCORE2-Diabetes separately in different sexes in the internal validation set (30% of UK Biobank, N=3,077) ..... 56
Figure 7. Discriminative performance of four models for major adverse cardiovascular events (MACE) by sex and overall population in internal validation set (30% of UK Biobank, N=3,077) and the external validation set (ESTHER study, N=1,039) ..... 58
Figure 8. Associations between selected metabolites and MACE across sexes in the internal validation cohort (30% of UK Biobank, N=56,112) ..... 59
Figure 9. The incremental discrimination of each metabolite for the model after the selected metabolites were added to SCORE2 separately in different sexes in the internal validation cohort (30% of UK Biobank, N=56,112) ..... 65
Figure 10. Associations between selected metabolites and incident type 2 diabetes in the test set (30% of UK Biobank, N=25,870) and external validation set (ESTHER, N=5,904) ..... 68
Figure 11. The incremental discrimination of each metabolite for the model after the selected metabolites were added to the clinical Cambridge Diabetes Risk Score in the test set (30% of UK Biobank, N=25,870) and external validation set (ESTHER, N=4,383) ..... 72
Figure 12. Associations of 15 selected proteins from OLINK Explore with incident type 2 diabetes in the internal validation set (Panel 2A, 30% of UK Biobank, N=6,339) and associations of 6 selected proteins from OLINK Target Inflammation panel with incident type 2 diabetes in the internal validation set (Panel 2B, 30% of UK Biobank, N=6,339) and external validation set (Panel 2C, 70% of ESTHER, N=3,118)..... 76
Figure 13. Increase in C-index by each protein when added to the clinical CDRS for T2D risk prediction..... 79
Figure 14. Associations between selected proteins and major cardiovascular events across sexes in the validation cohort (30% of UK Biobank, N=14,216) ..... 79
Figure 15. Contribution of individual proteomic biomarkers to the C-index in MACE risk prediction stratified by sex in the validation cohort (30% of UK Biobank, N=14,216)..... 85
Figure 16. Spearman correlation coefficients of C-reactive protein and the selected inflammation-related proteins stratified by sex in the UK Biobank and ESTHER study ..... 87

Figure 17. Associations between selected proteins and MACE across sexes in the internal validation (30% of UK Biobank, N=14,216) and external validation (70% of ESTHER, N=4,397) .. 88
Figure 18. The incremental discrimination of each protein for the model after the selected proteins were added to SCORE2 separately in different sexes in the internal validation (30% of UK Biobank, N=14,216) and external validation (70% of ESTHER, N=4,397) ..... 92
Figure 19. Predictive performance of the SCORE2-Diabetes model extended with multi-omics biomarkers for 10-year MACE risk (N=990) ..... 96
Figure 20. Incremental predictive value of selected multi-omic biomarkers added to the SCORE2-Diabetes model for 10-year MACE risk prediction (N=990) ..... 97
Figure 21. Discrimination performance of single-omics and multi-omics extensions to the SCORE2-Diabetes in 10-year MACE risk prediction (N=990) ..... 96
Figure 22. Spearman correlation matrices for selected multi-omics biomarkers (N=990) ..... 972
Figure 23. Associations of selected multi-omics biomarkers and 10-year risk of MACE (N=990) ... 983
Figure 24. Sex-specific correlation matrices of selected biomarkers in the validation set (30% of UK Biobank, N=7,214) ..... 1046
Figure 25. Associations of selected biomarkers with 10-year MACE risk in males and females in the validation set (N=7,214) ..... 1068
Figure 26. Performance of SCORE2 with and without omics data extensions for 10-year MACE prediction ..... 109
Figure 27. Incremental changes in C-index attributed to individual biomarkers ..... 1102
Figure 28. C-index comparison of single-omics and multi-omics extensions to SCORE2 in MACE prediction ..... 1113
Figure 29. Correlation matrices of selected biomarkers in the validation set (30% of UK Biobank, N=6,998) ..... 1135
Figure 30. Associations of selected biomarkers with incident T2D in the test set (30% of UK Biobank, N = 6,998) ..... 1157
Figure 31. Improvements in 10-year type 2 diabetes risk prediction by integrating multi-omics data in the test set (30% of UK Biobank, N = 6,998) ..... 118
Figure 32. Incremental contributions of individual biomarkers to risk prediction ..... 119

# List of tables

Table 1. Overview on the variables included in the four evaluated models ..... 28
Table 2. Baseline characteristics of selected participants with type 2 diabetes from the UK Biobank and ESTHER study ..... 52
Table 3. Comparison of the predictive performance of model 1 “Original SCORE2-Diabetes model” and model 2 “Original SCORE2-Diabetes model extended by NMR metabolites” for 10-year MACE risk prediction ..... 56
Table 4. Baseline characteristics of selected participants from the UK Biobank and ESTHER study ..... 59
Table 5. Metrics of the predictive performance of the SCORE2 model for 10-year MACE risk without and with extension by metabolites ..... 63
Table 6. Baseline characteristics of selected participants from the UK Biobank and ESTHER study ..... 66
Table 7. Comparison of the predictive performance of the clinical Cambridge Diabetes Risk Score and the novel developed UK Biobank Diabetes Risk Score for 10-year type 2 diabetes risk prediction ..... 69
Table 8. Baseline characteristics of selected participants from the UK Biobank and ESTHER study ..... 73
Table 9. Comparison of the predictive performance of the clinical CDRS with and without proteomic biomarkers for 10-year T2D risk prediction ..... 78
Table 10. Baseline characteristics of selected participants from the UK Biobank ..... 78
Table 11. Metrics of the predictive performance of the SCORE2 model for 10-year MACE risk without and with extension by proteins ..... 84
Table 12. Baseline characteristics of the analyzed individuals from the UK Biobank and the ESTHER study ..... 86
Table 13. Metrics of the predictive performance of the SCORE2 model for 10-year MACE risk without and with extension by proteins ..... 88
Table 14. Baseline characteristics of participants included in the proteomics and multi-omics analyses from the UK Biobank ..... 94
Table 15. Metrics of the predictive performance of the SCORE2-Diabetes model for 10-year MACE risk without and with extension by proteins (N=1,751) ..... 95
Table 16. Baseline characteristics of the cardiovascular risk cohort ..... 1024
Table 17. Baseline characteristics of T2D cases and non-cases ..... 112

# List of abbreviations

| | |
|---|---|
| ACTA2 | Actin Alpha 2 |
| ADM | Adrenomedullin |
| ADAMTS13 | A Disintegrin and Metalloproteinase with Thrombospondin Motifs 13 |
| ADGRG | Adhesion G-protein coupled receptor G |
| ApoB/ApoA1 | Apolipoprotein B to Apolipoprotein A1 Ratio |
| AUC | Area Under the Curve |
| BCAN | Brevican Core Protein |
| BMI | Body Mass Index |
| CCL23 | Chemokine (C-C motif) Ligand 23 |
| CDCP1 | CUB Domain-Containing Protein 1) |
| CDRS | Cambridge Diabetes Risk Score |
| CDH2 | Cadherin 2 |
| CHGA | Chromogranin-A |
| CI | Confidence Interval |
| CKD-EPI | Chronic Kidney Disease Epidemiology Collaboration |
| CRNN | Cornulin |
| CRP | C-Reactive Protein |
| CRYBB2 | Beta-Crystallin B2 |
| CTRC | Chymotrypsin-C |
| CTSL1 | Cathepsin L1 |
| cTnI | Cardiac Troponin I |
| CVD | Cardiovascular Disease |
| CXC | C-C Motif Chemokine |
| DHA | Docosahexaenoic Acid |
| EDA2R | The Tumor Necrosis Factor Receptor Superfamily Member 27 |
| EGFR | Epidermal growth factor receptor |
| eGFR | Estimated Glomerular Filtration Rate |
| ENPP5 | Ectonucleotide Pyrophosphatase/Phosphodiesterase Family Member 5 |
| ESTHER | Epidemiologische Studie zu Chancen der Verhütung, Früherkennung und optimierten Therapie chronischer Erkrankungen in der älteren Bevölkerung |
| FAw3/FA | Omega-3 Fatty Acids to Total Fatty Acids Ratio |
| FDR | False Discovery Rate |
| FGF21 | Fibroblast Growth Factor 21 |
| FGF23 | Fibroblast Growth Factor 23 |
| FINDRISC | Finnish Diabetes Risk Score |
| Gal-9 | Galectin-9 |
| GAST | Gastrin |
| GDF15 | Growth Differentiation Factor 15 |
| GDRS | German Diabetes Risk Score |
| GlycA | Glycoprotein Acetyls |
| GP | General Practitioners |
| GWAS | Genome-Wide Association Study |
| HAVCR1 | Hepatitis A Virus Cellular Receptor 1 |
| $HbA_{1c}$ | Glycated Hemoglobin |
| HDL-C | High-Density Lipoprotein Cholesterol |
| HGF | Hepatocyte Growth Factor |
| HR | Hazard Ratio |

| | |
|---|---|
| HS3ST3B1 | Heparan Sulfate Glucosamine 3-O-Sulfotransferase 3B1 |
| IDI | Integrated Discrimination Improvement |
| IDL-CE-pct | The Cholesteryl Esters Percentage of Total Lipids in Intermediate-Density Lipoprotein |
| IGFBP | Insulin-like Growth Factor Binding Protein |
| IGSF | Immunoglobulin Superfamily Member |
| IL | Interleukin |
| IL18R1 | Interleukin-18 Receptor 1 |
| IL1RN | Interleukin-1 Receptor Antagonist Protein |
| IR | Incidence Rate |
| LA-pct | The Linoleic Acid Percentage of Total Fatty Acids |
| LASSO | Least Absolute Shrinkage and Selection Operator |
| LC-MS/MS | Liquid Chromatography-Mass Spectrometry/Mass Spectrometry |
| LDL-C | Low-Density Lipoprotein Cholesterol |
| LDL-TG | Triglycerides in LDL |
| LGALS4 | Galectin-4 |
| L-LDL-TG-pct | The Triglycerides Percentage of Total Lipids in Large LDL |
| LRRN1 | Leucine-Rich Repeat Neuronal Protein 1 |
| L-VLDL-TG-pct | The Triglycerides Percentage of Total Lipids in Large VLDL |
| MACE | Major Adverse Cardiovascular Events |
| MI | Myocardial Infarction |
| M-LDL-TG-pct | Triglycerides to Total Lipids in Medium LDL Percentage |
| MMP12 | Macrophage Metalloelastase |
| MUFA | Monounsaturated Fatty Acids |
| mGPS | Modified Glasgow Prognostic Score |
| NEFL | Neurofilament light polypeptide |
| NMR | Nuclear Magnetic Resonance |
| NPX | Normalized Protein eXpression |
| NRI | Net Reclassification Improvement |
| NT-proBNP | N-Terminal Prohormone of Brain Natriuretic Peptide |
| OLINK | Olink Proteomics Assay |
| Omega-3-pct | The Omega-3 Fatty Acids to Total Fatty Acids Percentage |
| PD-L1 | Programmed Death-Ligand 1 |
| PON3 | Serum Paraoxonase/Lactonase 3 |
| POSTN | Periostin |
| PRS | Polygenic Risk Score |
| PRSS8 | Prostasin |
| PUFA | Polyunsaturated Fatty Acids |
| PY | Person-Years |
| ROC | Receiver Operating Characteristic |
| RTN4R | Reticulon-4 Receptor |
| SBP | Systolic Blood Pressure |
| SCORE | Systematic Coronary Risk Evaluation |
| SD | Standard Deviation |
| S-HDL-CE | Cholesteryl Esters in Small High-Density Lipoprotein |
| SM | Sphingomyelins |
| SMOC1 | SPARC-Related Modular Calcium-Binding Protein 1 |
| SNP | Single-Nucleotide Polymorphism |
| SOST | Sclerostin |

| | |
|---|---|
| T2D | Type 2 Diabetes |
| TC | Total Cholesterol |
| TG | Triglycerides |
| TNC | Tenascin-C |
| TNFRSF11B | Tumor Necrosis Factor Receptor Superfamily Member 11B |
| TRIM21 | E3 Ubiquitin-Protein Ligase TRIM21 |
| UKB | UK Biobank |
| VLDL | Very Low-Density Lipoprotein |
| VLDL-size | Average Diameter of VLDL Particles |
| XL-HDL-P | Very Large HDL Particles |
| SD | Standard Deviation |
| SEPTIN8 | Septin-8 |
| SOST | Sclerostin |
| SM | Sphingomyelins |
| SNP | Single-Nucleotide Polymorphism |
| T2D | Type 2 Diabetes |
| TC | Total Cholesterol |
| TG | Triglycerides |
| TNFRSF11B | Tumor Necrosis Factor Receptor Superfamily Member 11B |
| TWEAK | Tumor Necrosis Factor-Related Weak Inducer of Apoptosis |
| UKB | UK Biobank |
| UKB-DRS | UK Biobank Diabetes Risk Score |
| VLDL | Very Low-Density Lipoprotein |
| VLDL-size | The Average Diameter for Very Low-Density Lipoprotein Particles |
| WFDC2 | WAP Four-Disulfide Core Domain Protein 2 |
| XL-HDL-FC | Free Cholesterol in Very Large High-Density Lipoprotein |
| XL-HDL-P | Very Large HDL Particles |

# 1 Introduction

Cardiovascular disease (CVD) and type 2 diabetes mellitus (T2D) represent two of the most prevalent and burdensome chronic diseases worldwide (Balakumar et al. 2016; Einarson et al. 2018). CVD remains the leading cause of morbidity and mortality globally, while the incidence of T2D continues to rise sharply, reflecting changes in lifestyle, aging populations, and increasing rates of obesity (Ahmad et al. 2022; Chen et al. 2012; Roth et al. 2020b). Although these conditions have distinct clinical manifestations, CVD and T2D share multiple pathophysiological mechanisms such as insulin resistance, chronic low-grade inflammation, endothelial dysfunction, and dyslipidemia (De Rosa et al. 2018; Laakso und Kuusisto 2014; Lu et al. 2024). As both diseases are largely preventable through early intervention, identifying individuals at elevated risk before the onset of clinical symptoms remains a critical public health priority (DeFronzo und Abdul-Ghani 2011; Kahn et al. 2008).

Despite the widespread implementation of traditional risk assessment tools based on demographic and clinical variables, the accuracy of these models is limited by their inability to fully account for individual-level biological heterogeneity (Abbasi et al. 2012; Buijsse et al. 2011; Tokgozoglu und Torp-Pedersen 2021). In recent years, high-throughput molecular technologies such as metabolomics, proteomics, and genomics have advanced substantially, enabling comprehensive profiling of biological traits that capture multiple aspects of disease pathogenesis (Karczewski und Snyder 2018; Wörheide et al. 2021). These developments have created new opportunities for improving risk stratification and enabling more personalized prevention strategies (Lin et al. 2025). However, the clinical translation of such omics-based approaches remains limited, and their additional value over established prediction models needs to be evaluated in large, population-based studies (Chen et al. 2023a).

## 1.1 Public health relevance of cardiovascular disease and type 2 diabetes

CVD is the leading global cause of death, encompassing coronary heart disease, stroke, and other vascular conditions (Vaduganathan et al. 2022). It causes about 20.5 million deaths annually, nearly one third of all deaths worldwide (Di Cesare et al. 2024). The burden is not limited to high-income countries, as more than 75% of cardiovascular deaths occur in low- and middle-income settings, largely due to

common risk factors such as tobacco use, unhealthy diet, physical inactivity, and elevated blood pressure (Di Cesare et al. 2024; Roth et al. 2020a). Alongside its impact on mortality, CVD also contributes substantially to disability and health care costs.

In recent decades, T2D has emerged as a major global health concern. According to the International Diabetes Federation, around 537 million adults (approximately 10.5% of the adult population) were living with diabetes in 2021, and this number is projected to increase to 783 million by 2045 (Sun et al. 2022). T2D is a leading cause of complications such as visual impairment, kidney failure, and lower-limb amputations (Low Wang et al. 2016). In addition to its direct disease burden, T2D is strongly associated with cardiovascular complications. Adults with T2D have a two- to four-fold higher risk of developing CVD than those without diabetes (Low Wang et al. 2016). Meta-analyses and large-scale cohort studies have shown that T2D is an independent risk factor for coronary heart disease, stroke, and peripheral arterial disease (Sarwar et al. 2010). Approximately 50% of all diabetes-related deaths are due to cardiovascular causes (Roglic 2016). As a result, T2D is frequently regarded as a cardiovascular risk equivalent, indicating that individuals with diabetes have a similar myocardial infarction (MI) risk to non-diabetic individuals with a history of heart attack (Haffner et al. 1998).

CVD and T2D share numerous epidemiological features (Siam et al. 2024). Both are driven by overlapping behavioral and metabolic risk profiles, including obesity, hypertension, and dyslipidemia (Grundy et al. 2004). Their prevalence continues to rise globally, influenced by rapid urbanization, aging populations, and changing lifestyle patterns (De Rosa et al. 2018). The frequent co-occurrence of CVD and T2D adds complexity to disease prevention and management and highlights the need for integrated public health strategies.

Efforts to reduce the burden of these diseases rely on both population-wide and high-risk approaches (Cooney et al. 2009a). Public health campaigns that promote healthy behaviors have shown modest success, but targeted prevention remains essential (Shah Ebrahim und Smith 2001). For this purpose, accurate identification of individuals at increased risk is critical. Epidemiological risk prediction models are widely used to estimate an individual's probability of developing disease and to guide preventive interventions (Binuya et al. 2022). For example, individuals with a high 10-year CVD risk may benefit

from statin therapy or lifestyle counseling, while those at elevated risk for T2D may be offered preventive strategies such as weight loss programs or pharmacologic intervention (Golovaty et al. 2023; Visseren et al. 2022). Therefore, improving risk stratification tools for CVD and T2D is a priority in both clinical and public health contexts.

## 1.2 Risk prediction models in epidemiology

Risk prediction models play a central role in epidemiology by estimating the likelihood that an individual will develop a specific disease over a defined period of time (Spitz et al. 2005). These models are widely used to inform clinical decisions and guide public health interventions, particularly in the prevention of chronic diseases such as CVD and T2D (Rezaee et al. 2020). By integrating multiple demographic, clinical, and laboratory variables into a single predictive algorithm, such models enable stratification of individuals into different risk categories, facilitating targeted prevention strategies.

In the field of cardiovascular prevention, several well-established risk models have been developed over the past decades (Farzadfar 2019). One of the earliest and most widely known is the Framingham Risk Score, originally derived from the Framingham Heart Study in the United States. This model estimates the 10-year risk of developing coronary heart disease based on factors such as age, sex, systolic blood pressure (SBP), total cholesterol (TC), smoking status, and diabetes (Wilson et al. 1998). Although influential, the Framingham Risk Score was later found to overestimate risk in European populations, prompting the development of region-specific tools (Brindle et al. 2005).

In Europe, the Systematic Coronary Risk Evaluation (SCORE) project led to the creation of country-adapted models for estimating 10-year risk of fatal CVD based on age, sex, smoking, TC, and SBP (Conroy et al. 2003). More recently, the SCORE2 model has been introduced to improve the prediction of both fatal and non-fatal CVD events across diverse European populations. SCORE2 incorporates recalibrated risk equations for different European risk regions and uses contemporary cohort data (Visseren et al. 2022). The SCORE2-Diabetes model was subsequently developed to address the distinct risk profiles of individuals with T2D (Pennells et al. 2023). This model adjusts for the duration of diabetes and includes diabetes-specific factors in addition to classical cardiovascular risk variables.

In parallel, several tools have been proposed for predicting the onset of T2D. The Finnish Diabetes Risk Score (FINDRISC) is a widely used non-invasive tool that estimates 10-year risk of developing T2D based on age, BMI, waist circumference, physical activity, dietary habits, family history of diabetes, and antihypertensive medication use (Lindstrom und Tuomilehto 2003). Other models such as the German Diabetes Risk Score (GDRS) and the Cambridge Diabetes Risk Score (CDRS) integrate similar clinical and behavioral variables (Griffin et al. 2000; Schulze et al. 2007). While easy to implement, these scores often have limited discriminatory power, especially in populations with intermediate risk, and their performance varies across ethnic and geographic groups.

Despite their clinical utility, traditional risk prediction models have several limitations. First, they rely on a restricted set of variables that may not fully capture the underlying biological complexity of disease development (Folsom 2013). Second, the generalizability of many models is limited, as their performance may be reduced when applied to populations other than those in which they were developed (Singh et al. 2022). Third, the discriminatory ability of classical models remains modest, with C-indices often below 0.80, indicating room for improvement (Cooney et al. 2009b). Finally, many models treat risk factors as static variables, although risk can change over time with age, lifestyle modifications, or emerging comorbidities (Kasartzian und Tsiampalis 2025).

To address these limitations, efforts have been made to update prediction models with additional clinical biomarkers, imaging data, or novel algorithmic approaches such as machine learning (Gore et al. 2020). However, a growing body of research suggests that molecular biomarkers derived from high-throughput omics technologies may offer a promising avenue for further improving risk stratification (Sun et al. 2024).

## 1.3 Omics biomarkers as tools to improve prediction

Recent progress in high-throughput omics technologies, including genomics, proteomics, and metabolomics, has made it possible to measure thousands of biological markers in a single assay, offering novel insights into disease pathogenesis and risk stratification (Chen et al. 2023a). In

epidemiological research, these biomarkers have shown potential to improve individual-level prediction for chronic diseases such as CVD and T2D (Chen und Gerszten 2020; Jacob et al. 2019).

Genomic biomarkers, particularly in the form of polygenic risk scores (PRS), summarize the cumulative effects of multiple genetic variants associated with disease susceptibility (Torkamani et al. 2018). Numerous genome-wide association studies (GWAS) have identified genetic loci linked to CVD and T2D (Visscher et al. 2017). By aggregating hundreds of thousands of single nucleotide polymorphisms (SNPs), PRSs can quantify inherited risk at the individual level. For example, studies have shown that individuals with high PRS for coronary artery disease or T2D have substantially elevated lifetime risk, independent of conventional risk factors (Khera et al. 2018; Läll et al. 2017). While PRSs are fixed from birth and not modifiable, they can aid in early risk stratification, especially in younger individuals with minimal clinical signs (Knowles und Ashley 2018).

Proteomic biomarkers reflect the downstream expression of genes and provide a dynamic measure of pathophysiological processes, including inflammation, vascular remodeling, and metabolic dysregulation (Anderson und Anderson 2002). Advances in affinity-based platforms such as proximity extension assays (e.g., Olink) and aptamer-based technologies (e.g., SomaScan) have enabled the quantification of hundreds to thousands of proteins in small volumes of blood (Assarsson et al. 2014). Several studies have demonstrated the predictive utility of circulating proteins in population-based cohorts (Ganz et al. 2016). For instance, adding selected proteomic markers to clinical risk scores has been associated with significant improvements in CVD and T2D prediction, as measured by increases in C-index and net reclassification (Hoogeveen et al. 2020; Yao et al. 2024). Moreover, some protein biomarkers may serve as both predictors and potential causal mediators, offering mechanistic insights.

Metabolomic biomarkers capture a wide range of small-molecule intermediates and end products of metabolism (Patti et al. 2012). Nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry have been used to profile circulating lipids, amino acids, and other metabolites associated with cardiometabolic risk. Studies have shown that specific metabolites, such as branched-chain amino acids, ketone bodies, and lipoprotein subclasses, are strongly associated with incident CVD and T2D (Ahola-

Olli et al. 2019; Wang et al. 2011). Incorporating metabolite profiles into clinical models has yielded modest to substantial gains in prediction performance across different populations.

While each omics layer provides powerful, distinct insights, they each capture only one dimension of a complex biological system. Genomics offers a static view of inherited predisposition, which does not account for the dynamic influence of lifestyle and environment. Conversely, proteomics and metabolomics provide a real-time snapshot of the body's physiological state but lack the underlying genetic context. A truly comprehensive understanding, therefore, requires integrating these complementary data types. By combining the genetic blueprint (genomics) with its functional consequences (proteomics and metabolomics), a multi-omics approach has the potential to create more robust and accurate predictive models that capture a far greater degree of individual biological variation. This integrative strategy represents a logical next step and a promising new frontier for enhancing risk stratification.

## 1.4 Integrating multi-omics data for improved prediction

The rationale for multi-omics integration is grounded in the biological complementarity of different molecular layers. Genetic variation contributes to disease susceptibility but does not reflect environmental exposures or dynamic physiological changes. In contrast, proteomic and metabolomic profiles capture downstream effects of both genetic and environmental influences, providing functional and time-sensitive markers of disease activity (Karczewski und Snyder 2018). By integrating these layers, researchers aim to identify more informative biomarker signatures that can enhance risk stratification beyond what is possible with any single omics modality (Hasin et al. 2017).

Early evidence from epidemiological studies indicates that multi-omics integration can yield modest yet significant gains in cardiometabolic risk prediction. In the EPIC-Norfolk cohort, combining top features from genomics, proteomics, and metabolomics with clinical variables improved 10-year T2D prediction compared with the clinical model alone, with a higher concordance index and better reclassification (Carrasco-Zanini et al. 2024c). In large population studies of CVD, integrating polygenic risk with circulating molecular profiles similarly enhanced discrimination beyond established risk scores, and

lipidomic or proteomic panels captured risk dimensions that were weakly correlated with genetic risk, indicating complementary information (Lauber et al. 2022).

Beyond overall accuracy, multi-omics can refine risk in subgroups that conventional screening may classify as low risk. For example, adding polygenic and proteo-metabolomic markers improves prediction among individuals without biochemical evidence of dysglycemia at baseline, indicating potential for earlier identification of T2D risk before clinical abnormalities emerge (Carrasco-Zanini et al. 2024c). Polygenic risk also contributes most in settings where clinical signals are weak, such as younger or normoglycemic adults, because inherited risk is present long before metabolic perturbations are detectable. Together, these findings suggest that integrating genetic and circulating molecular profiles with standard factors can uncover latent risk that traditional tools may miss, thereby enhancing early prevention.

Despite encouraging results, studies applying multi-omics risk prediction for CVD and T2D remain scarce and are often limited to relatively small, homogenous populations. To ensure robustness and reproducibility, there is an urgent need to establish standardized laboratory procedures and harmonized analytical workflows across omics platforms. Furthermore, prospective, real-world evaluations are essential to determine whether omics-informed risk stratification can enhance clinical decision-making and lead to improved patient outcomes in diverse healthcare settings.

Looking ahead, declining assay costs, expanding population biobanks with longitudinal follow-up, and advances in data-integration methods are likely to accelerate robust multi-omics model development and external validation for CVD and T2D. Population resources such as UK Biobank (UKB) provide the scale to compare single-omics versus integrated models, while modern machine-learning frameworks can incorporate non-linear and interaction effects across layers to improve calibration and transportability (Anagnostakis et al. 2025; Chen et al. 2024; Garg et al. 2024; Hong et al. 2024). As evidence accumulates in large, diverse cohorts, multi-omics–augmented risk scores may complement existing tools to support more precise prevention strategies for cardiometabolic disease.

## 1.5 Aims of the dissertation

The overarching aim of this dissertation is to enhance the prediction of CVD and T2D by integrating multi-omics biomarkers into established clinical risk models. Leveraging data from the large-scale UKB and ESTHER (German name: Epidemiologische Studie zu Chancen der Verhütung, Früherkennung und optimierten Therapie chronischer Erkrankungen in der älteren Bevölkerung) cohorts, this work systematically evaluates the incremental predictive value of metabolomics, proteomics, and integrated multi-omics approaches. The specific aims are as follows:

**A) Metabolomics**

i. To evaluate the added predictive value of metabolomic biomarkers for CVD risk prediction in individuals with T2D.

ii. To evaluate the added predictive value of metabolomic biomarkers for CVD risk prediction in the general population without prior CVD or diabetes

iii. To investigate the utility of metabolomic biomarkers in improving 10-year T2D risk prediction in the general population.

**B) Proteomics**

iv. To evaluate the predictive utility of proteomic biomarkers for incident T2D in the general population.

v. To assess the contribution of proteomic biomarkers to CVD risk prediction in the general population.

vi. To determine whether proteomic biomarkers enhance 10-year CVD risk prediction in individuals with T2D.

**C) Multi-omics integration**

vii. To investigate whether a combined model incorporating proteomics, metabolomics, and a PRS improves CVD risk prediction in individuals with T2D.

viii. To evaluate the incremental benefit of integrated multi-omics data for CVD risk prediction in the general population.

ix. To assess whether the integration of proteomics, metabolomics, and a T2D-PRS improves 10-year T2D risk prediction beyond clinical and single-omics models.

# 2 Material and Methods

## 2.1 Metabolomics for cardiovascular risk prediction in individuals with type 2 diabetes

This section corresponds to the first aim of this dissertation: to evaluate the added predictive value of metabolomic biomarkers for cardiovascular risk prediction in individuals with T2D.

### 2.1.1 Study population

This study utilized data from two large European cohorts: the UKB and the ESTHER study. Both studies were conducted according to the Declaration of Helsinki, and all participants provided written informed consent.

The UKB, a large prospective cohort in the United Kingdom, aims to bolster knowledge, preventive measures, diagnostic modalities, and therapeutic approaches across diverse disease spectra (Sudlow et al. 2015). In brief, between 13 March 2006 and 1 October 2010, over half million participants aged 37-73 were enrolled across England, Scotland, and Wales. The UKB was approved by the North-West Multi-center Research Ethics Committee.

The ESTHER study is a large population-based cohort study conducted in Saarland, Germany. Participants aged between 50 and 75 years were recruited during a regular health check by local general practitioners (GP) between 1 July 2000 and 30 June 2002. After that, participants were followed up every two to three years. The study effectively collected data on 9,940 participants (Schöttker et al. 2013; Xuan et al. 2019). The ESTHER study was approved by the ethics committees of the Medical Faculty of Heidelberg and the state medical board of Saarland, Germany.

### 2.1.2 In- and exclusion criteria

We included N=10,257 and N=1,039 individuals with T2D and measured metabolomic biomarkers from the UKB and ESTHER study, respectively (**Appendix 1**). Subjects with potential type 1 diabetes, ascertained by diabetes diagnosis before the 40$^{th}$ birthday, were excluded. UKB participants outside the

age range of 50-74 years were excluded to make the age distribution of two cohorts comparable. Furthermore, those with a history of MI or stroke before baseline or with a missing date on these events of interest were excluded.

### 2.1.3 Metabolites quantification

Nightingale Health's high-throughput nuclear magnetic resonance (NMR) metabolomics platform performed the metabolomic profiling on 274,353 baseline plasma samples from randomly selected UKB participants, alongside all available 8,308 baseline serum samples from the ESTHER cohort with sufficient blood sample quality (Soininen et al. 2015). This process enabled the quantification of 250 metabolomic biomarkers, including lipids, fatty acids, amino acids, ketone bodies, and other essential low-molecular-weight markers (Davyson et al. 2023). However, glycerol was excluded because it could not be measured in most of the participants of both cohorts, leaving n=249 metabolites for the analyses.

### 2.1.4 Variables of the SCORE2-Diabetes model

The SCORE2-Diabetes is a sex-specific competing risk-adjusted algorithm recently developed and validated by the European Society of Cardiology, designed for adults with T2D aged 40 to 69 years. It includes age, sex, SBP, smoking status, age at diabetes diagnosis, high-density lipoprotein cholesterol (HDL-C), TC, glycated hemoglobin ($HbA_{1c}$), and the estimated glomerular filtration rate (eGFR) (Pennells et al. 2023).

Information on demographic characteristics, lifestyle factors and medical history, including age, sex, age at diabetes diagnosis and smoking status (current or non-current) was collected by standardized questionnaires in both cohorts. The SBP measurements were conducted by automated reading of the Omron digital blood pressure monitor at the left upper arm in UKB and from the physician's medical conditions report of the health check-up in the ESTHER study. The laboratory methods used to measure HDL-C, TC, $HbA_{1c}$ and creatinine in both cohorts are shown in **Appendix 2**. The eGFR was calculated from creatinine values using the CKD-Epi 2009 equation in both cohorts (Levey et al. 2009).

### 2.1.5 Outcome assessment

The primary outcome was MACE, which included cardiovascular mortality, non-fatal MI, and non-fatal stroke (see detailed definition in **Appendix 3**). It was constructed in line with the endpoint definition used to derive the SCORE2-Diabetes model. The only exception was that the non-fatal stroke definition was broader in the ESTHER study because stroke subtypes cannot be distinguished in this cohort.

In the UKB cohort, non-fatal MI and or non-fatal stroke were identified from primary care records or hospital episode statistics. The date and cause of death were determined by referring to death registries of the National Health Service Information Centre for participants in the UKB study from England and Wales, and the National Health Service Central Register of Scotland for those from Scotland.

As previously described (Niersmann et al. 2020), study participants from the ESTHER study reported on the occurrence of incident MI and stroke in standardized questionnaires at 2-, 5, 8-, and 11-year follow-ups. Self-reported cases underwent validation by the dissemination of standardized questionnaires to the GPs of the study participants. The proportions of physician-validated incident MI and stroke cases were 89% and 91%, respectively. A vital status inquiry was made at registration offices of residents, with death certificates of deceased individuals being procured from local health authorities.

The follow-up period extended until the first recorded non-fatal MI, non-fatal stroke, death, or the end of the event registration period, which was capped at ten years (Pennells et al. 2023).

### 2.1.6 Statistical analyses

#### 2.1.6.1 General remarks

All analyses were conducted using R software, version 4.3.0 (R Foundation for Statistical Computing, Vienna, Austria), with statistical significance set at *P*-values<0.05 for two-sided tests. Missing values were imputed using the random forest imputation method, as implemented in the r-package *missForest* (version 1.5) (Stekhoven und Bühlmann 2012). Most variables of the SCORE2-Diabetes model and the NMR metabolites were complete. The highest missing value rate was observed for $HbA_{1c}$ in the UKB at 5.1%. In the ESTHER study, aside from the age at diabetes diagnosis (39.6%), no other variable's missing rate exceeded 10%.

### 2.1.6.2 Metabolites selection and model derivation

Initially, all metabolite concentrations underwent log-transformation to approximate normal distributions for analysis. These measures were then scaled to standard deviation units independently within each cohort. The UKB dataset was partitioned into a derivation set (70%) and an internal validation set (30%); the ESTHER study served as the external validation cohort. The derivation set was used for metabolite selection and model derivation.

The Least Absolute Shrinkage and Selection Operator (LASSO), a regularization technique adept at identifying strong predictors among high-dimensional and correlated predictors, was implemented via the r-package *glmnet* (version 4.1-7) (Tibshirani 1996). Ten-fold cross-validation was employed to determine the optimal tuning parameter λ for LASSO, based on the lowest model validation error. A bootstrap LASSO approach was then undertaken, involving the creation of 1000 derived sets with the LASSO procedure applied to each resampled dataset (Bach 2008). Metabolites selected by LASSO in at least 95% of instances were designated as our metabolites of interest to enhance model generalization and mitigate overfitting. These selected metabolites were subsequently incorporated into model 1 "Original SCORE2-Diabetes" to construct a new sex-specific competing risk model 2 "Original SCORE2-Diabetes extended by NMR metabolites" using the r-package *cmprsk* (version 2.2-11).

### 2.1.6.3 Model's predictive performance validation

The predictive performance of model 2 "Original SCORE2-Diabetes extended by NMR metabolites" was validated by using 30% of the UKB as the internal validation cohort and the ESTHER cohort as the external validation cohort. The incremental discrimination contributed by each metabolite was evaluated based on the internal validation results.

The discriminative ability was assessed using Harrell's C-index, adjusted for competing risks (via the r-package *riskRegression*, version 2023.03.2) (van Geloven et al. 2022). Furthermore, the net reclassification index (NRI) and the integrated discrimination improvement (IDI) were evaluated for the final model (using the r-package *nricens*, version 1.6) (Kerr et al. 2014). Pre-specified CVD risk categories (0–10%, >10–20%, and >20%) were employed for the NRI to indicate the proportion of

individuals correctly reclassified compared with the SCORE2-Diabetes model. Model calibration was examined by plotting observed MACE event rates against predicted event rates across deciles of absolute predicted risk.

To further assess whether the clinical chemistry parameters of the SCORE2-Diabetes model could be replaced by NMR-derived metabolites in cardiovascular risk prediction, the predictive performance of two further models were tested. **Table 1** gives an overview of the variables included in the in total 4 evaluated models. Model 3 "SCORE2-Diabetes with TC, HDL-C and creatinine measured by NMR" replaces those clinical chemistry parameters of the original SCORE2-Diabetes model which have NMR equivalents which excludes $HbA_{1c}$, which is not available from NMR. Model 4 "Age, smoking, SBP, diabetes age at diagnosis, and NMR metabolites" only retains the non-laboratory features of the SCORE2-Diabetes model and does not include any parameters measured by clinical chemistry. Instead it includes TC, HDL-C and creatinine measured by NMR, it replaces $HbA_{1c}$ by glucose, which can also be measured by NMR, and it adds the NMR metabolites selected by LASSO regression.

**Table 1. Overview on the variables included in the four evaluated models**

| Variables | Model 1 - "Original SCORE2-Diabetes" | Model 2 - "Original SCORE2-Diabetes extended by NMR metabolites" | Model 3 - "SCORE2-Diabetes with TC, HDL-C and creatinine measured by NMR" | Model 4 - "Age, smoking, SBP, diabetes age at diagnosis, and NMR metabolites" |
|---|---|---|---|---|
| Age | X | X | X | X |
| Current smoking | X | X | X | X |
| Systolic blood pressure | X | X | X | X |
| SBP interaction with age | X | X | X | X |
| Smoking interaction with age | X | X | X | X |
| Diabetes | X | X | X | X |
| Diabetes age at diagnosis | X | X | X | X |
| Total cholesterol | X CC | X CC | X NMR | X NMR |
| Total cholesterol interaction with age | X CC | X CC | X NMR | X NMR |
| HDL cholesterol | X CC | X CC | X NMR | X NMR |
| HDL interaction with age | X CC | X CC | X NMR | X NMR |
| Ln eGFR | X CC | X CC | X NMR | X NMR |
| Ln $eGFR^2$ | X CC | X CC | X NMR | X NMR |
| Ln eGFR interaction with age | X CC | X CC | X NMR | X NMR |
| $HbA_{1c}$ | X CC | X CC | X CC | - |
| $HbA_{1c}$ interaction with age | X CC | X CC | X CC | - |
| **Added NMR metabolites** | | | | |

| | | | | |
|---|---|---|---|---|
| Creatinine (for men and women) | - | X NMR | - | X NMR |
| Albumin (for men and women) | - | X NMR | - | X NMR |
| GlycA (for men and women) | - | X NMR | - | X NMR |
| Acetate (for men only) | - | X NMR | - | X NMR |
| Omega-3-pct (for men only) | - | X NMR | - | X NMR |
| VLDL-size (for men only) | - | X NMR | - | X NMR |
| Lactate (for women only) | - | X NMR | - | X NMR |
| Glucose (for men and women) | - | - | - | X NMR |
| Glucose interaction with age (for men and women) | - | - | - | X NMR |

**Abbreviations:** CC, clinical chemistry measured; GlycA, glycoprotein acetyls; NMR, nuclear magnetic resonance spectrometry measured; Omega-3-pct; the Omega-3 fatty acids to total fatty acids percentage; VLDL-size, the average diameter for very low-density lipoprotein particles

#### 2.1.6.4 Associations of selected metabolites with MACE

To elucidate the hazard ratios (HRs) and 95% confidence intervals (CIs) of selected metabolites (per one standard deviation increment) for enhanced 10-year MACE prediction in male and female patients with T2D, metabolites were individually added to Cox proportional hazards regression models in the derivation cohort, the internal validation cohort, and the external validation cohort, respectively. These models were adjusted for SCORE2-Diabetes model variables, using the r-package *survival* (version 3.5-5). The Benjamini-Hochberg procedure was applied to control the false discovery rate (FDR) in this analysis.

## 2.2 Metabolomics for cardiovascular risk prediction in the general population

This section corresponds to the second aim of the dissertation: to evaluate the added predictive value of metabolomic biomarkers for cardiovascular risk prediction in the general population without prior CVD or diabetes.

### 2.2.1 Study population

This study also utilized data from the UKB and the ESTHER cohort. Descriptions of the cohort design, recruitment procedures, and ethical approvals are provided in *Chapter 2.1.1*. A cohort study design was applied, and participants without a history of CVD or diabetes at baseline were included.

### 2.2.2 In- and exclusion criteria

In the UKB and ESTHER cohort N=247,353 and N=8,308 study participants were randomly selected for metabolomics measurements, respectively. To ensure comparability in age distribution between two cohorts, participants not in the age range of 50-69 years were excluded (see **Appendix 4** for study flow-charts). We further excluded individuals with missing data on diabetes or those diagnosed with diabetes at the baseline. Additionally, we excluded participants with a prior history of MI or stroke before baseline or those lacking precise dates for these cardiovascular events. Finally, we included N=187,039 and N=5,578 individuals with measured metabolomics data from the UKB and ESTHER study, respectively.

### 2.2.3 Metabolites quantification

A total of 249 metabolites were analyzed after excluding glycerol due to missingness, as described in *Chapter 2.1.3*.

### 2.2.4 SCORE2 model variables

The variables of the clinical SCORE2 model, including age, sex, systolic blood pressure, total cholesterol, HDL cholesterol, and smoking status, were measured and harmonized using identical methods described in *Chapter 2.1.3*.

### 2.2.5 Outcome assessment

The outcome of interest was the occurrence of MACE. The approaches for outcome ascertainment in both UKB and ESTHER were described in *Chapter 2.1.5*, and remain unchanged.

### 2.2.6 Statistical analyses

#### 2.2.6.1 General remarks

General remarks, encompassing analytical software used, significance levels, and methods for handling missing data persist unchanged from the description provided in *Chapter 2.1.6.1*.

#### 2.2.6.2 Metabolite selection and model derivation

A 70/30 split was applied to the UKB sample to create a derivation and internal validation set, with ESTHER used as the external validation cohort. Metabolites were selected using bootstrap LASSO regression with 1,000 iterations. Those selected in ≥95% of iterations were included in the final model. These selected metabolites were then added to the sex-specific Fine–Gray models built upon the SCORE2 algorithm, as previously described in *Chapter 2.1.6.2*.

#### 2.2.6.3 Evaluation of prediction performance

The predictive performance of the extended SCORE2 model was assessed using Harrell's C-index (adjusted for competing risks), category-based NRI (risk categories: 0–5%, 5–10%, >10%), and IDI. These analyses were conducted in both validation sets. Observed and predicted event rates were compared by deciles to assess calibration.

#### 2.2.6.4 Associations of selected metabolites with MACE

As described in *Chapter 2.1.6.4*, associations were evaluated using cause-specific Cox models adjusted for SCORE2 variables.

## 2.3 Metabolomics for type 2 diabetes risk prediction

For the third aim of this dissertation, we aimed to evaluate whether circulating metabolomic biomarkers can improve the 10-year prediction of T2D in the general population.

### 2.3.1 Study population

This study also utilized data from the UKB and the ESTHER cohort. Descriptions of the cohort design, recruitment procedures, and ethical approvals are provided in *Chapter 2.1.1*.

### 2.3.2 In- and exclusion criteria

The inclusion and exclusion criteria for the study population are shown in **Appendix 5.** Finally, we included 86,232 individuals from the UKB and 4,383 individuals from the ESTHER study.

### 2.3.3 Metabolites quantification

Quantification of 249 metabolites were analyzed as described in *Chapter 2.1.3*.

### 2.3.4 Variables of the clinical CDRS

The CDRS is a predictive tool used to assess the risk of individuals for future T2D development. This scoring system includes age, sex, BMI, family history of diabetes, smoking status, prescription of anti-hypertensive medication and steroids.(Rahman et al. 2008) If blood samples can be taken, the clinical CDRS is recommended to use, which additionally includes the $HbA_{1c}$.(Carrasco-Zanini et al. 2024c) The assessment methods of the variables of the clinical CDRS and of type 2 incidence in both cohorts are shown in **Appendix 6**.

### 2.3.5 Type 2 incidence ascertainment

In both the UKB and ESTHER study, information on demographic characteristics, lifestyle factors and medical history, including age, sex, family history of diabetes and smoking status were collected by standardized questionnaires. In the UKB, prescribed medication was obtained from self-report, primary care records and verbal interviews. In the ESTHER study, GPs completed a standardized health check-up form and documented current drug prescriptions. Weight and height were measured in both studies – in the UKB by study staff in in the assessment centers and in the ESTHER study by GPs during the health check-up. The $HbA_{1c}$ was measured from whole-blood samples with high-performance liquid chromatography on the Variant II (Bio-Rad) in both cohorts.

In the UKB, incident T2D was ascertained by three sources. First, participants self-reported diagnoses of type 2 diabetes or use of glucose-lowering medications during follow-up surveys. Second, medical codes for type 2 diabetes in primary care records, hospital admissions, and death registry data were

utilized for case identification. Prescription data for glucose-lowering medications from primary care were also used to identify cases.

In the ESTHER study, the ascertainment of incident T2D was identified via four sources. Participants provided self-reports of diabetes diagnoses and glucose-lowering medication use via standardized questionnaires during the 2-, 5-, 8-, and 11-year follow-ups. Self-reported diagnoses at the 2- and 5-year marks were validated through questionnaires sent to the GPs of the participants, with unconfirmed cases being excluded. Similarly, at the 8- and 11-year follow-ups, GPs received questionnaires inquiring about new diabetes diagnoses within the last three years and unconfirmed self-reports were disregarded.

Study participants were censored at the first occurrence of a diabetes diagnosis, date of death, data of loss to follow-up, or when the maximum follow-up time of 10 years was reached.

### 2.3.6 Statistical analyses

#### 2.3.6.1 General remarks

Statistical software, significance thresholds, and missing data imputation methods were the same as those described in *Chapter 2.1.6.1*.

#### 2.3.6.2 Feature selection and model derivation

The UKB data were randomly split into a 70% training set and a 30% internal validation set. Metabolites were selected via bootstrapped LASSO Cox regression models with 1,000 resamples in the training set, using incident T2D as the outcome and adjusting for CDRS variables. Metabolites selected in at least 95% of resamples were retained for model construction. The selected metabolites were added to the clinical CDRS to derive the new metabolomics-augmented prediction model.

#### 2.3.6.3 Validation of the model's predictive performance

Model performance was evaluated using Harrell's C-index, time-dependent receiver operating characteristic (ROC) curves, and model calibration in the 30% UKB validation set and the independent ESTHER cohort. Improvements in the C-index were tested using the r-package *compareC*. Risk reclassification was assessed using the continuous NRI and IDI. A concise model was also developed

using only metabolites with an incremental C-index > 0.005 in the external validation set. Subgroup analyses were conducted by age, sex, obesity status, and baseline CDRS risk.

#### 2.3.6.4 Associations of selected metabolites

To interpret the contribution of individual metabolites, Cox regression models were applied as previously described in *Chapters 2.1.6.4*, adjusting for the variables of the clinical CDRS.

## 2.4 Proteomics for type 2 diabetes risk prediction

For the fourth aim of this dissertation, we evaluated whether circulating proteomic biomarkers improve 10-year prediction of incident T2D beyond the clinical CDRS.

### 2.4.1 Study population

The description of the UKB and ESTHER cohorts and the statement of ethical approval were described in *Chapter 2.1.1*.

### 2.4.2 In- and exclusion criteria

The inclusion and exclusion criteria for the study population are shown in **Appendix 7.** Finally, we included 21,898 individuals from the UKB and 4,454 individuals from the ESTHER study.

### 2.4.3 Proteomics assessment

In the UKB, proteomic profiling was performed on EDTA plasma samples collected at baseline. The Olink Explore 3072 platform (Olink Proteomics, Uppsala, Sweden) was used, which allowed the detection of 2,923 unique proteins. Measurements are reported as Normalised Protein eXpression (NPX) units. Detailed methods on the assay, sample selection, and handling have been described previously (Sun et al. 2023). Proteins with more than 20% missing values or more than 25% of values below the limit of detection were excluded (n=838). Ultimately, 2,085 proteins from OLINK Explore were included in our analysis. In the ESTHER study, levels of 92 inflammation-related proteins were measured in baseline serum samples with the Olink Target 96 inflammation panel (Trares et al. 2022;

Trares et al. 2024). The 92 proteins of the Olink Target 96 inflammation panel are also part of the Olink Explore 3072 panel. Overall, 19 proteins from the inflammation panel were excluded after applying the same cutoffs as for the OLINK Explore, leaving n=73 for analysis.

### 2.4.4 Reference model: clinical Cambridge Diabetes Risk Score (CDRS)

The CDRS definition, included variables, and harmonization procedures were identical to those described in *Chapter 2.3.3*.

### 2.4.5 Ascertainment of incident type 2 diabetes

Follow-up procedures, censoring rules, and T2D case ascertainment were as described in *Chapter 2.3.5*.

### 2.4.6 Statistical analyses

#### 2.4.6.1 General remarks

Analytical software, significance thresholds, and handling of missing data were as described in Chapter 2.1.6.1.

#### 2.4.6.2 Feature selection and model derivation

Feature selection and model derivation followed the procedures described in *Chapter 2.3.6.2*. Briefly, LASSO Cox regression with nested cross-validation in 200 bootstrap samples was used in the UKB training set to identify proteins most predictive of incident T2D, and selected proteins were added to the refitted clinical CDRS.

#### 2.4.6.3 Validation of the model's predictive performance

Internal validation of both novel models was performed in 30% of the UKB set (**Figure 1**). The model derived from the OLINK Target Inflammation panel was additionally externally validated in the ESTHER study. As the relative NPX values of the OLINK panels are cohort specific, the UKB training set coefficients could not be directly applied to the ESTHER cohort. Therefore, the ESTHER cohort was

randomly split into a calibration set (30%) and a validation set (70%) and the calibration set was used to obtain cohort specific ß coefficients for model's factors. Internal validation procedures in the UKB and external validation in ESTHER followed the approach outlined in *Chapter 2.3.6.3*, including model recalibration, assessment of Harrell's C-index, NRI, IDI, and calibration plots.

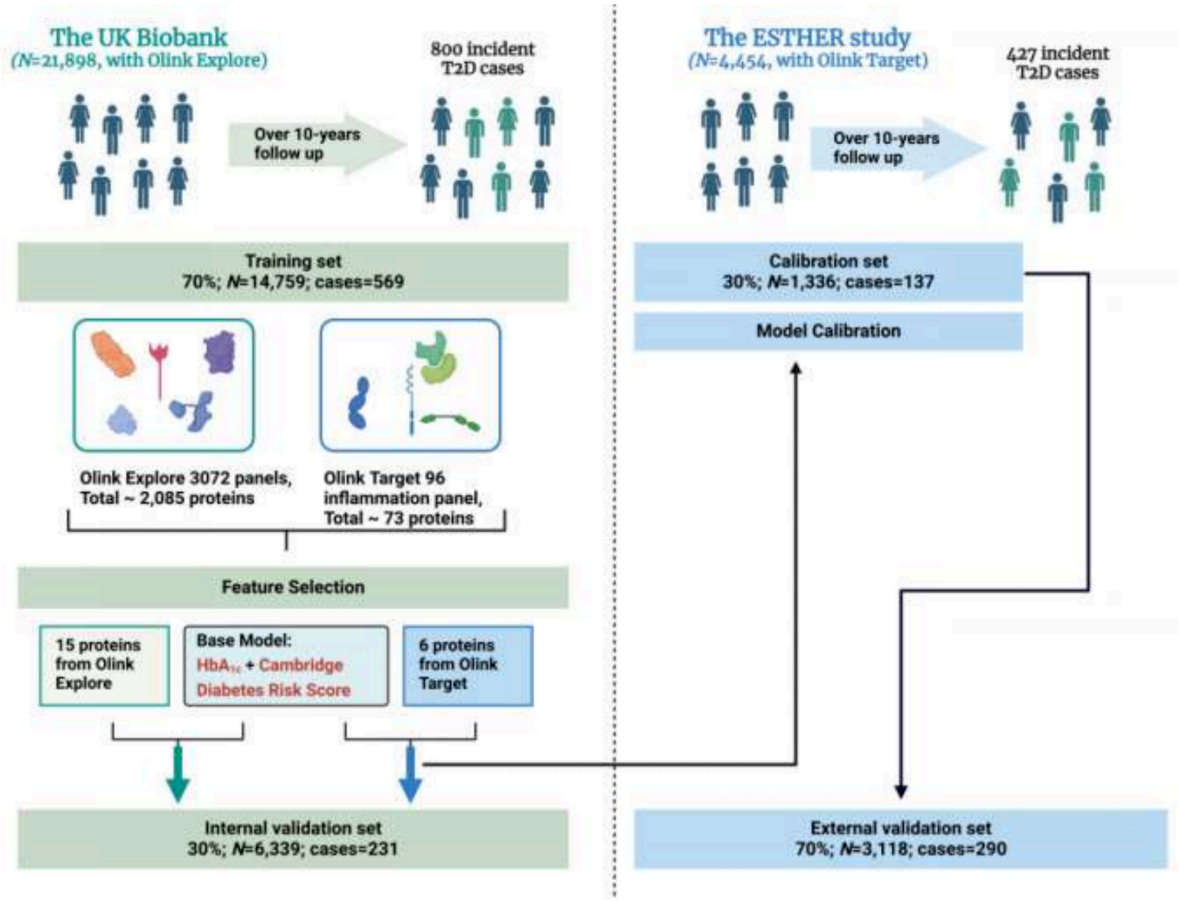



**Figure 1. Study design for T2D model development and validation**

**Abbreviations:** $HbA_{1c}$, glycated hemoglobin; T2D, type 2 diabetes.

### 2.4.6.4 Sensitivity analysis

To test the robustness and generalizability of our findings, we conducted several additional analyses. First, we repeated all analyses after excluding 2,655 participants who were not randomly selected for proteomics measurements. Second, we stratified participants by baseline prediabetes status ($HbA_{1c}$ <5.7% vs. 5.7–6.4%) and re-evaluated model performance in each subgroup to assess whether adding proteomic biomarkers conferred consistent predictive gains. Third, we investigated whether incorporating CRP (C-reactive protein) into the clinical CDRS could further improve predictive accuracy. Finally, we compared our results to those from Carrasco-Zanini et al. (Carrasco-Zanini et al.

2022; Carrasco-Zanini et al. 2024c), incorporating their suggested proteins and screening thresholds to determine whether these alternative approaches yielded similar or superior model performance.

#### 2.4.6.5 Associations of selected proteins with incident type 2 diabetes

As described in Chapter 2.3.6.5, associations of individual selected proteins with T2D risk were estimated using Cox regression models adjusted for the clinical CDRS variables. In addition, we calculated Spearman correlation coefficients to evaluate correlations of these proteins with BMI and $HbA_{1c}$.

## 2.5 Proteomics for cardiovascular risk prediction in the general population

For the fifth aim of this dissertation, we evaluated whether circulating proteomic biomarkers improve 10-year prediction of MACE beyond the SCORE2 model in the general population without prior CVD or diabetes.

### 2.5.1 Study population

The description of the UK Biobank and the statement of ethical approval were described in *Chapter 2.1.1*.

### 2.5.2 In- and exclusion criteria

The inclusion and exclusion criteria for the analyzed study population are shown in **Appendix 8**. After applying these criteria, 47,382 participants were included in the analysis.

### 2.5.3 Proteomic profiling

Proteomic profiling of plasma samples was performed using the Olink Explore 3072 platform. Detailed procedures, including sample handling, quality control, and protein selection thresholds, were as described in *Chapter 2.4.3*.

### 2.5.4 Assessment of covariates

Variables included in the SCORE2 model (age, sex, HDL-C, TC, SBP, and smoking status) were assessed as described in *Chapter 2.2.4*.

### 2.5.5 Outcome ascertainment

The definition of MACE, data sources, and follow-up procedures were identical to those described in *Chapter 2.2.5*.

### 2.5.6 Statistical analyses

#### 2.5.6.1 General remarks

General analytical procedures, including statistical software used, significance thresholds, and handling of missing data, were as described in *Chapter 2.1.6.1*.

#### 2.5.6.2 Biomarker selection and model derivation

Proteomic biomarker selection and model derivation were performed separately for male and female participants, following the procedures described in *Chapter 2.4.5.2*.

#### 2.5.6.3 Model performance validation

Model discrimination, reclassification, and calibration were evaluated as described in *Chapter 2.2.6.3*. The incremental contribution of each selected protein to model discrimination was assessed based on the change in C-index within the same framework.

#### 2.5.6.4 Associations of selected proteins with MACE

Associations between selected proteins and MACE were examined using Cox proportional hazards regression models adjusted for SCORE2 variables, as described in *Chapter 2.2.6.4*.

## 2.6 Inflammation-related proteins for cardiovascular risk prediction

This section corresponds to the fifth aim of this dissertation, as in *Chapter 2.5,* but focuses specifically on inflammation-related proteomic biomarkers and includes external validation in an independent cohort.

### 2.6.1 Study population

Analyses were conducted in the UKB and ESTHER cohorts. Cohort descriptions and ethical approvals are provided in *Chapter 2.1.1*.

### 2.6.2 In- and exclusion criteria

The inclusion and exclusion criteria for the analyzed study population are shown in **Appendix 9**, with final study populations of 47,382 UKB participants and 6,280 ESTHER participants.

### 2.6.3 Measurement of CRP and inflammation-related proteins

CRP was measured using immunoturbidimetric analysis with a Beckman Coulter AU5800 in UKB and an ADVIA 2400 in ESTHER. In UKB, proteins were quantified using the Olink Explore 3072 platform; in ESTHER, the Olink Target 96 Inflammation panel was used. Only inflammation-related proteins available in both datasets (N = 73) were retained.

### 2.6.4 Assessment of covariates

Covariates comprised SCORE2 variables, as in *Chapter 2.2.4*.

### 2.6.5 Outcome ascertainment

The primary endpoint was MACE, defined in accordance with SCORE2; ascertainment procedures are described in *Chapter 2.2.5*.

### 2.6.6 Statistical analyses

#### 2.6.6.1 General remarks

General statistical considerations were identical to those in *Chapter 2.1.6.1*.

#### 2.6.6.2 Biomarker selection and model derivation

In UKB, sex-specific derivation and validation sets (70%/30%) were used. Candidate proteins were selected via LASSO regression with 200 bootstrap iterations, retaining those with selection frequency ≥95%. Selected proteins were added to recalibrated SCORE2 models to derive extended sex-specific models. **Figure 1** gives an overview of the study design.

#### 2.6.6.3 Model performance validation

Model discrimination, reclassification, and calibration were evaluated as in *Chapter 2.2.6.3*, with internal validation in UKB and external validation in ESTHER following recalibration.

#### 2.6.6.4 Sensitivity analyses

Sensitivity analyses included: (1) repeating the main analysis after excluding non-randomly selected participants; (2) comparing the extended model with models including cardiac biomarkers (NT-proBNP and cTnI); and (3) assessing model performance stratified by statin use and examining associations between statin use and protein concentrations.

#### 2.6.6.5 Associations of selected proteins with MACE

Associations of selected proteins witth MACE were examined using cause-specific Cox models adjusted for SCORE2 variables, as described in *Chapter 2.2.6.4*.

## 2.7 Proteomics and Multi-omics for cardiovascular risk prediction in individuals with type 2 diabetes

This section addresses aims vi and vii, evaluating whether circulating proteomic biomarkers improve 10-year CVD risk prediction in individuals with T2D and whether an integrated model combining proteomics, metabolomics, and a PRS provides additional predictive value in this population.

### 2.7.1 Study population

The study population was drawn from the UKB, as described in *Chapter 2.1.1*. For the proteomics analysis, a case-cohort study design was implemented. This involved including all participants who experienced a MACE during follow-up, supplemented by a randomly selected subcohort from the full UKB cohort with available proteomic data.

### 2.7.2 In- and exclusion criteria

From this case-cohort sample, we excluded participants with >50% missing proteomics data. Subsequently, we applied the study-specific criteria: exclusion of individuals without T2D, those with type 1 diabetes, a prior history of MACE, or those outside the 40–69 year age range. This resulted in a proteomics-based analysis cohort of 1,751 participants. For the multi-omics analysis, an additional 761 participants with missing metabolomics or PRS data were excluded, resulting in a multi-omics analysis subset of 990 participants (**Appendix 10**).

### 2.7.3 Proteomics data

Proteomic profiling used the Olink Explore 3072 platform; preprocessing, NPX units, quality control, and filtering thresholds (proteins with >20% missing or >25% below LOD excluded) followed *Chapter 2.4.3*. After filtering, 2,085 proteins were available.

### 2.7.4 Metabolomics data

Metabolomics quantification followed the procedures described in Chapter 2.1.3. For the present multi-omics analysis, we utilized a pre-selected parsimonious signature of seven sex-specific metabolites identified in our previous work (*Chapter 2.1*), which demonstrated improved prediction when added to the SCORE2-Diabetes model.

### 2.7.5 Polygenic risk score for CVD

The PRS for CVD were provided by the UK Biobank PRS Release (Thompson et al. 2024). These scores were derived using a Bayesian modelling approach, trained on meta-analysed summary statistics from 9 external GWAS datasets, and subsequently optimized using internal UK Biobank data. The final PRS for CVD, which included 1.42 million variants with non-zero weights, demonstrated strong predictive performance in internal benchmarking and outperformed a range of existing algorithms (Thompson et al. 2024).

### 2.7.6 Variables of the SCORE2-Diabetes model

Identical to *Chapter 2.1.4*.

### 2.7.7 Outcome ascertainment

MACE definition, sources, and 10-year follow-up were identical to *Chapter 2.1.5*.

### 2.7.8 Statistical analyses

#### 2.7.8.1 General remarks

Software, significance thresholds, and missing-data handling mirrored *Chapter 2.1.6.1*.

#### 2.7.8.2 Biomarker selection and model evaluation strategy

Model development and evaluation followed a two-stage approach (Figure 2). In the first stage, proteomics-based models were derived and evaluated in models adjusted for the variables of the

SCORE2-Diabetes model using the subset of T2D patients with available proteomics data (N = 1,751) for best statistical power.

In the second stage, the additional predictive value of signatures derived from genomic, metabolomic, and proteomic data, when added to the SCORE2-Diabetes model, was assessed in the subset of T2D patients with complete multi-omics data (N = 990). These three multi-omics layers were added sequentially to the SCORE2-Diabetes model, and model performance was evaluated. Finally, the best-performing omics layers were added one by one to the SCORE2-Diabetes model until the addition of another layer did not result in a statistically significant improvement in model discrimination.

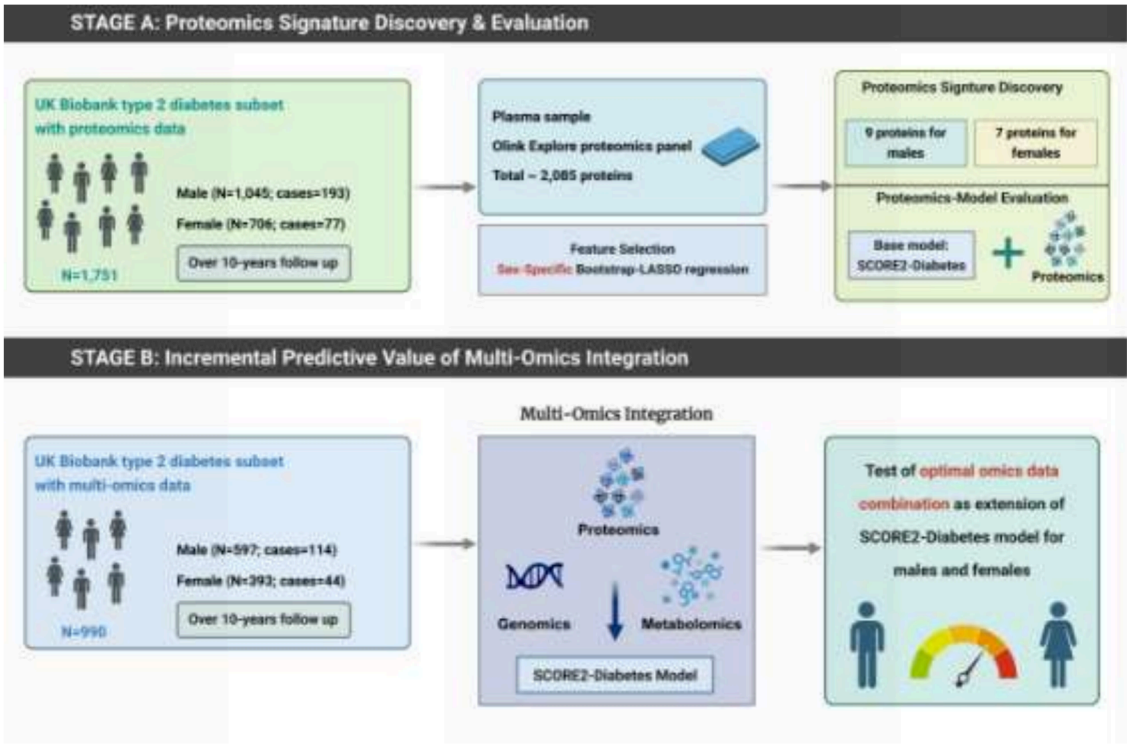



**Figure 2. Flowchart of the two-stage biomarker selection process**

### 2.7.8.3 Model performance assessment

Model performance was evaluated for discrimination (Harrell's C-index, with significance testing via the r-package *compareC*), reclassification (NRI with risk categories of 0–15%, >15–30%, and >30%; and IDI), and calibration (plots of observed vs. predicted risk).

#### 2.7.8.4 Association and correlation analyses

Associations of individual biomarkers with MACE were quantified using sex-stratified Cox models adjusted for SCORE2-Diabetes variables to derive HRs. Inter-biomarker redundancy was assessed using Spearman correlation coefficients.

## 2.8 Multi-omics for cardiovascular risk prediction in the general population

For the eighth aim of this dissertation, this study evaluated the incremental benefit of integrating proteomic, metabolomic, and polygenic risk score data for 10-year MACE risk prediction in the general population without prior CVD or diabetes.

### 2.8.1 Study population

This study was conducted within the UKB cohort, utilizing the case-cohort sample established for proteomics analyses, as detailed in *Chapter 2.7.1*.

### 2.8.2 In- and exclusion criteria

The detailed inclusion and exclusion flowchart for the present analysis is provided in **Appendix 11**. In brief, from the case-cohort sample, we excluded participants with missing multi-omics data, those with prevalent diabetes, a prior history of MACE, or those outside the 40–69 year age range of the SCORE2 model. The final study population comprised 24,042 participants.

### 2.8.3 Proteomics data

The proteomic analysis utilized a pre-selected, sex-specific signature of 18 proteins that were previously identified to enhance CVD risk prediction in the SCORE2 model, as detailed in *Chapter 2.5*. The proteomic profiling methods were described in *Chapter 2.4.3*.

### 2.8.4 Metabolomics data

The metabolomic analysis utilized a pre-selected, sex-specific signature of 13 metabolites previously identified to improve risk stratification, as described in *Chapter 2.2*. The metabolite quantification methods were detailed in *Chapter 2.1.3*.

### 2.8.5 Polygenic risk score for CVD

The analysis incorporated the pre-computed CVD-PRS provided by the UKB. The derivation and validation of this score were described in *Chapter 2.7.5*.

### 2.8.6 Variables of the SCORE2 model

The variables of the SCORE2 model were assessed as described in *Chapter 2.2.4.*

### 2.8.7 Outcome ascertainment

MACE definition, sources, and 10-year follow-up were identical to *Chapter 2.1.5*.

### 2.8.8 Statistical analyses

#### 2.8.8.1 General remarks

Statistical procedures, including software, significance thresholds, and missing value imputation, were performed as described in *Chapter 2.1.6.1*.

#### 2.8.8.2 Model derivation and validation

The analytical strategy is illustrated in **Figure 3**. For this multi-omics analysis, we utilized pre-selected, sex-specific biomarker panels (18 proteins, 13 metabolites, CVD-PRS). The dataset was randomly stratified by sex into a training set (70%) for model derivation and a testing set (30%) for validation. The model building process was systematic: we first assessed the performance of each omics layer (proteomics, metabolomics, PRS) when added individually to the SCORE2 model. Following this, a full

model incorporating all three omics layers was constructed. To arrive at a final, optimal, and parsimonious model, the best-performing omics layers were added sequentially to the base model until no further statistically significant improvement in the C-index was observed.

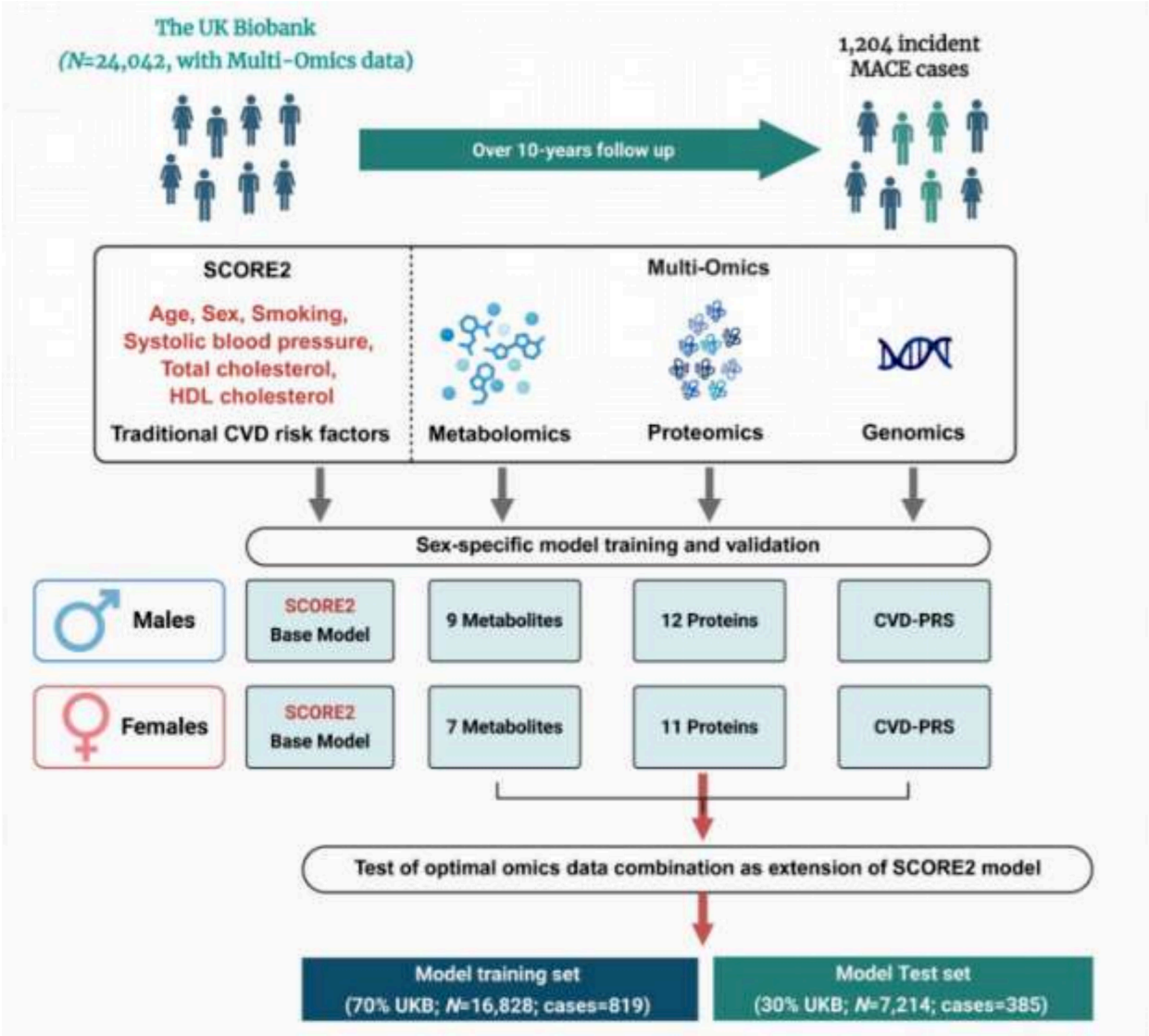



**Figure 3. Schematic of the multi-omics-extended SCORE2 model**

**Abbreviations:** HDL, high-density lipoprotein; MACE, major adverse cardiovascular events; PRS, polygenic risk score.

#### 2.8.8.3 Model performance evaluation

Model performance was evaluated for discrimination (Harrell's C-index, with significance testing via the r-package *compareC*), reclassification (NRI with risk categories of 0–7.5%, >7.5–15%, and >15%; and IDI), and calibration (plots of observed vs. predicted risk).

#### 2.8.8.4 Association and correlation analyses

Associations of individual biomarkers with MACE were quantified using sex-stratified Cox models adjusted for SCORE2 variables to derive HRs. Inter-biomarker redundancy was assessed using Spearman correlation coefficients.

## 2.9 Multi-omics for type 2 diabetes risk prediction

This section corresponds to the ninth aim of this dissertation, evaluating the incremental predictive value of an integrated multi-omics approach for the 10-year risk of T2D beyond established clinical risk factors.

### 2.9.1 Study population

The study population was drawn from the UKB cohort, which was described in *Chapter 2.1.1*.

### 2.9.2 In- and exclusion criteria

From the initial subsample of participants with available proteomics data (N=54,219), we applied several exclusion criteria as detailed in **Appendix 12**. This included removing participants with incomplete multi-omics data, those with prevalent T2D (either diagnosed or defined by an $HbA_{1c} \geq 6.5\%$), or those non-randomly selected for proteomics measurements. The final analytical cohort comprised 23,325 participants.

### 2.9.3 Proteomics data

The proteomic analysis utilized a pre-selected panel of 15 proteins. This protein signature was previously identified for its predictive value for T2D, as detailed in *Chapter 2.4*. The proteomic profiling methods were described in *Chapter 2.4.3*.

### 2.9.4 Metabolomics data

The metabolomic analysis utilized a pre-selected panel of 11 metabolites. This signature was also originally identified for its ability to predict T2D, as described in *Chapter 2.3*. The metabolite quantification methods were detailed in *Chapter 2.1.3*.

### 2.9.5 Polygenic risk score for CVD

The analysis incorporated the pre-computed T2D-PRS provided by the UK Biobank, which includes 2.12 million variants. The derivation of this score via a Bayesian modeling approach is methodologically similar to that described for the CVD-PRS in *Chapter 2.7.5*.

### 2.9.6 Reference model: The clinical CDRS

The clinical CDRS was used as the reference model. The variables and their assessment were previously described in *Chapter 2.3.4*.

### 2.9.7 Outcome ascertainment

The primary outcome was incident T2D, ascertained over a 10-year follow-up period using methods identical to those described in *Chapter 2.3.5*.

### 2.9.8 Statistical analyses

#### 2.9.8.1 General remarks

Standard statistical procedures were followed, as described in *Chapter 2.1.6.1*.

### 2.9.8.2 Model derivation and validation

The analytical strategy is illustrated in **Figure 4**. The cohort was randomly split into a training set (70%) and a testing set (30%). In the training set, we first evaluated the performance of each pre-selected omics panel (15 proteins, 11 metabolites, T2D-PRS) when added individually to the clinical CDRS model. Subsequently, we identified the single best-performing omics layer and used this extended model as a new reference, testing whether the addition of further omics layers provided significant incremental value. Model performance was validated in the testing set. Furthermore, subgroup analyses were conducted for participants with and without prediabetes at baseline ($HbA_{1c}$ < 5.7% vs. 5.7–6.4%).

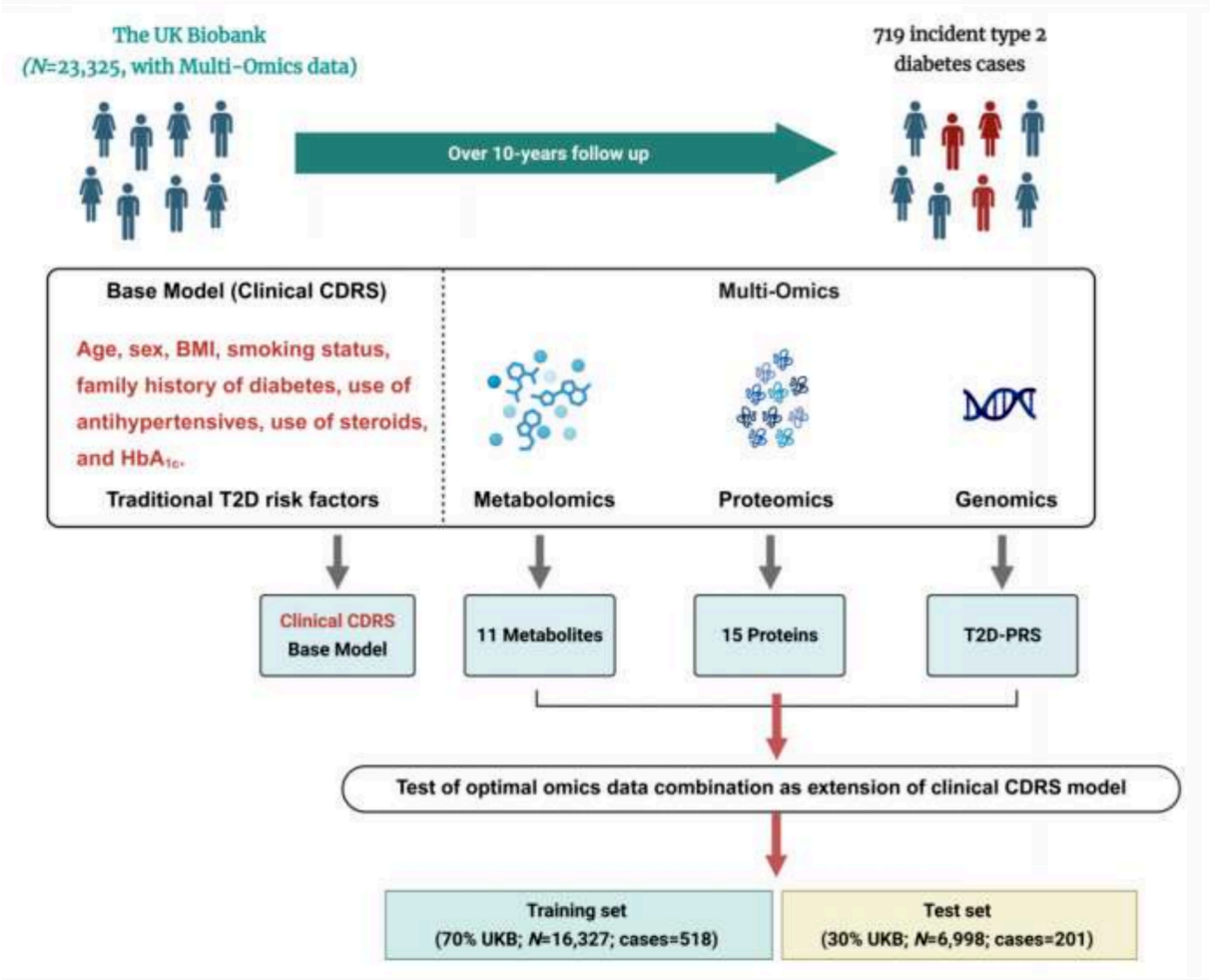



**Figure 4. Schematic of the integrated clinical and multi-omics T2D prediction model**

**Abbreviations:** BMI, body mass index; CDRS, Cambridge Diabetes Risk Score; PRS, polygenic risk score; T2D, type 2 diabetes.

#### 2.9.8.3 Model performance evaluation

Model performance was assessed using standard metrics for discrimination, reclassification, and calibration, as detailed in *Chapter 2.3.6.3*.

#### 2.9.8.4 Association and correlation analyses

The association of each individual biomarker with incident T2D was quantified using Cox proportional hazards models, adjusted for the clinical CDRS variables, to derive HRs per SD increment. To assess redundancy, Spearman correlation coefficients between all pairs of selected biomarkers were calculated in the test set.

# 3 Results

## 3.1 Metabolomics for cardiovascular risk prediction in individuals with type 2 diabetes

### 3.1.1 Description of the population included

To provide an overview of our study populations, **Table 2** summarizes the characteristics of the participants for all variables of the SCORE2-Diabetes model from the 10,257 and 1,039 included participants with T2D from the UKB and ESTHER cohorts, respectively. The included UKB participants had an average age of 61.5 years (SD 5.2), with males constituting 59.9% of the cohort. Over a follow-up time limited to 10 years, a total of 1,366 MACE events were recorded in the UKB. The ESTHER sample had a mean age of 64.0 years (SD 6.3), and 50.8% of participants were males. During the first 10 years of ESTHER follow-up, 175 MACE events were recorded.

**Table 2. Baseline characteristics of selected participants with type 2 diabetes from the UK Biobank and ESTHER study**

| **Baseline characteristics** | UK Biobank (n=10,257) | ESTHER (n=1,039) |
|---|---|---|
| Male sex, N (%) | 6,140 (59.9) | 528 (50.8) |
| Age (years), mean (SD) | 61.5 (5.2) | 64.0 (6.3) |
| Current smoker, N (%) | 1,024 (10.0) | 157 (15.7) |
| Systolic blood pressure (mmHg), mean (SD) | 144.8 (18.2) | 144.2 (19.5) |
| Total cholesterol (mmol/L), mean (SD) | 3.7 (0.9) | 5.6 (1.1) |
| HDL cholesterol (mmol/L), mean (SD) | 1.1 (0.3) | 1.3 (0.3) |
| Age at diabetes diagnosis (years), mean (SD) | 55.4 (6.4) | 55.6 (5.7) |
| $HbA_{1c}$ (mmol/mol), mean (SD) | 51.8 (12.8) | 53.1 (14.4) |
| eGFR (ml/min/1.73m$^2$)[a], mean (SD) | 87.7 (15.4) | 78.6 (17.7) |

[a] The eGFR was calculated with the CKD-EPI 2009 equation.

**Abbreviations:** eGFR, estimated Glomerular Filtration Rate; $HbA_{1c}$, glycated hemoglobin; HDL, high-density lipoprotein; IQR, interquartile range; SD, standard deviation.

### 3.1.2 Associations of metabolomic biomarkers with MACE

A total of 7 metabolites were selected to enhance MACE risk prediction in the SCORE2-diabetes model in the UKB cohort by LASSO analyses and bootstrapping with a frequency of more than 950 times in the 1000 replicate selection (**Appendix 13**). Of these metabolites, three metabolites improved MACE risk prediction in both men and women (creatinine, albumin, and GlycA (glycoprotein acetyls)). Furthermore, lactate emerged as an additional predictive metabolite solely in women, while the three metabolites - acetate, VLDL-size (the average diameter for very low-density lipoprotein particles), and Omega-3-pct (the Omega-3 fatty acids to total fatty acids percentage) - exhibited enhanced predictive capabilities exclusively in males.

As illustrated in **Figure 5**, the associations of the 7 selected metabolites with MACE in multivariate Cox regression models adjusted for SCORE2-Diabetes model variables separately in males and females in the derivation cohort. GlycA, and creatinine were positively, and omega-3-pct and albumin were inversely, statistically significantly associated with MACE at an FDR-adjusted $P<0.05$ level in both sexes. The results were generally consistent in the internal validation cohort but some results differed like for lactate for which results showed a sex difference in the internal validation study (**Appendix 14**) which was not seen in the derivation cohort (**Figure 5**) or the external validation cohort (**Appendix 15**). With the lower sample size of the ESTHER study, of the 7 metabolites, only albumin was statistically significantly associated with MACE in males.

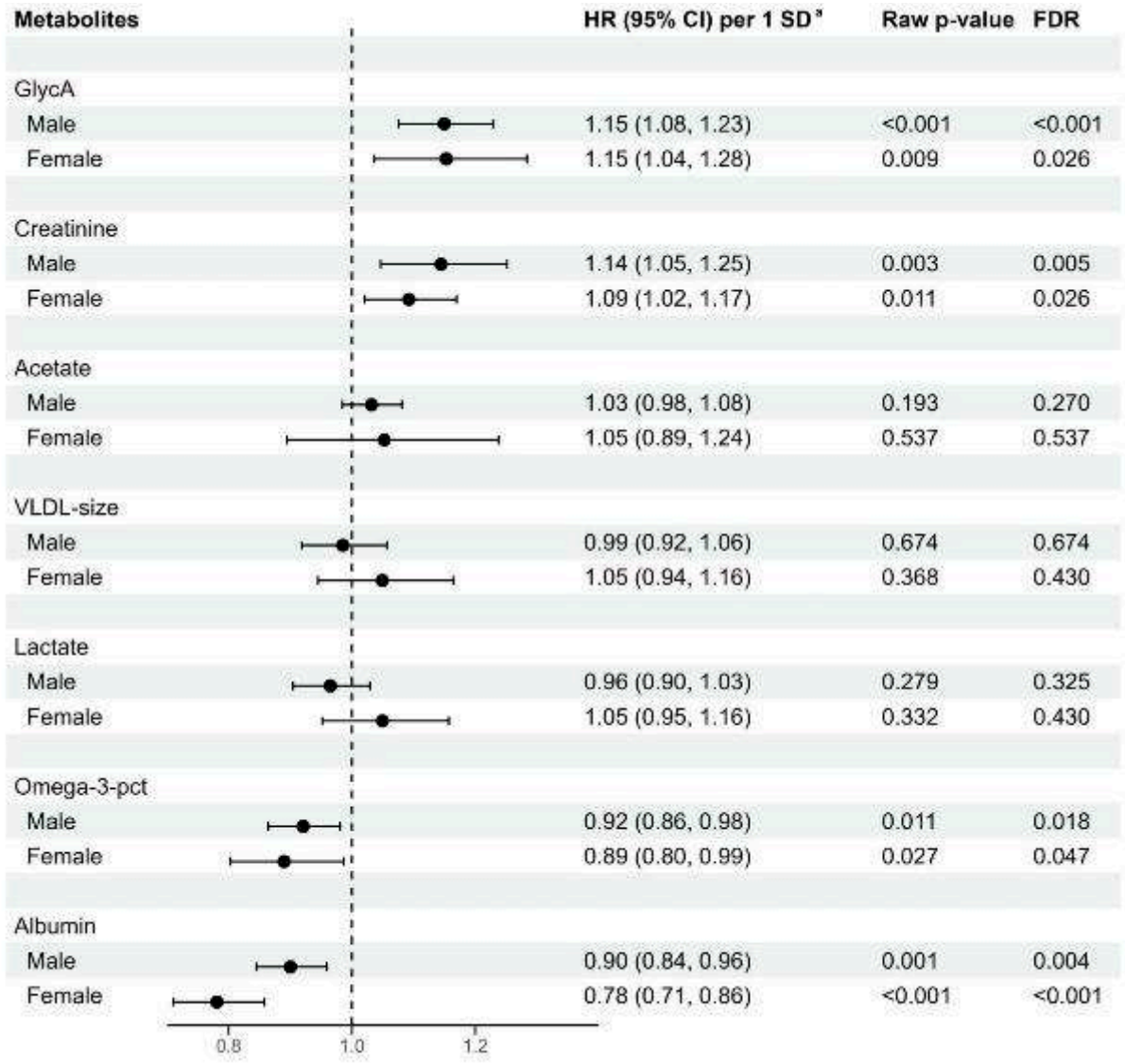



**Figure 5. Associations between selected metabolites and MACE across sexes in the derivation set (70% of UK biobank, N=7,180)**

**Abbreviations:** CI, confidence interval; FDR, false discovery rate; GlycA, glycoprotein acetyls; HR, hazard ratio; Omega-3-pct, Omega-3 fatty acids to total fatty acids percentage; SD, standard deviation; VLDL-size, average diameter for very-low-density lipoprotein particles.

[a] Hazard ratios are expressed per 1 standard deviation of the respective metabolite concentration and are adjusted for age, systolic blood pressure, smoking status, diabetes age at diagnosis, glycated hemoglobin, and the estimated glomerular filtration rate. The standard deviations of creatinine, albumin, GlycA, acetate, omega-3-pct, VLDL-size, and lactate were 0.02 mmol/L, 3.60 mmol/L, 0.13 mmol/L, 0.04 mmol/L, 1.51%, 1.41 mmol/L, and 1.23 mmol/L, respectively.

### 3.1.3 MACE risk prediction by metabolomic biomarkers

To evaluate the impact of adding metabolomic biomarkers on risk prediction, we examined the changes in the C-statistic with the sequential inclusion of each metabolite. As shown in **Figure 6**, the incremental

improvement in the C-statistic upon the sequential inclusion of the selected 7 metabolites into the models for males and females one by one in the internal validation set. Especially, the addition of albumin and Omega-3-pct significantly improved discrimination power in males. GlycA, creatinine, acetate, and VLDL-size showed notable but non-significant enhancements in males. In females, only lactate exhibited a notable, yet non-significant, C-index increase.

**Table 3** presents the predictive performance metrics of model 1 "Original SCORE2-Diabetes model" and model 2 "Original SCORE2-Diabetes extended by NMR metabolites" for 10-year MACE risk prediction. The ß-coefficients of all variables required for calculation of the extended risk model are shown in **Appendix 16**. In the derivation set, by adding the selected metabolites, the C-index (95% CI) of the SCORE2-Diabetes model statistically significantly improved (*P*=0.007) from 0.662 (0.644-0.681) to 0.678 (0.660-0.697) in the total sample (**Table 3**). The C-index increase was more pronounced in males (+0.017) than females (+0.007), with statistical significance solely noted for males (*P* =0.020). Internal validation in the UKB yielded very similar results for the C-indices. Notably, an enhancement in risk reclassification statistics was primarily observed in female participants with a statistically significant IDI. As shown in **Appendix 17**, model calibration was comparably good for the original SCORE2-Diabetes model and the NMR metabolites extended model in the internal validation set.

In the external validation with the total ESTHER study, the model demonstrated an even greater improvement in discriminatory power compared to the UKB, with a likewise statistically significant result (the C-index increased from 0.612 to 0.655, *P*=0.011; **Table 3**). C-index increases were comparable for men (+0.031) and women (+0.030). Beyond these improvements, the reclassification metrics for males indicated that 9.6% of events were correctly reclassified by the extended model compared with the original SCORE2-Diabetes model [NRI= 9.6% (0.9%, 16.1%)]. The model calibration was comparably good for the SCORE2-Diabetes model and the extended model in the external validation set (**Appendix 18**).

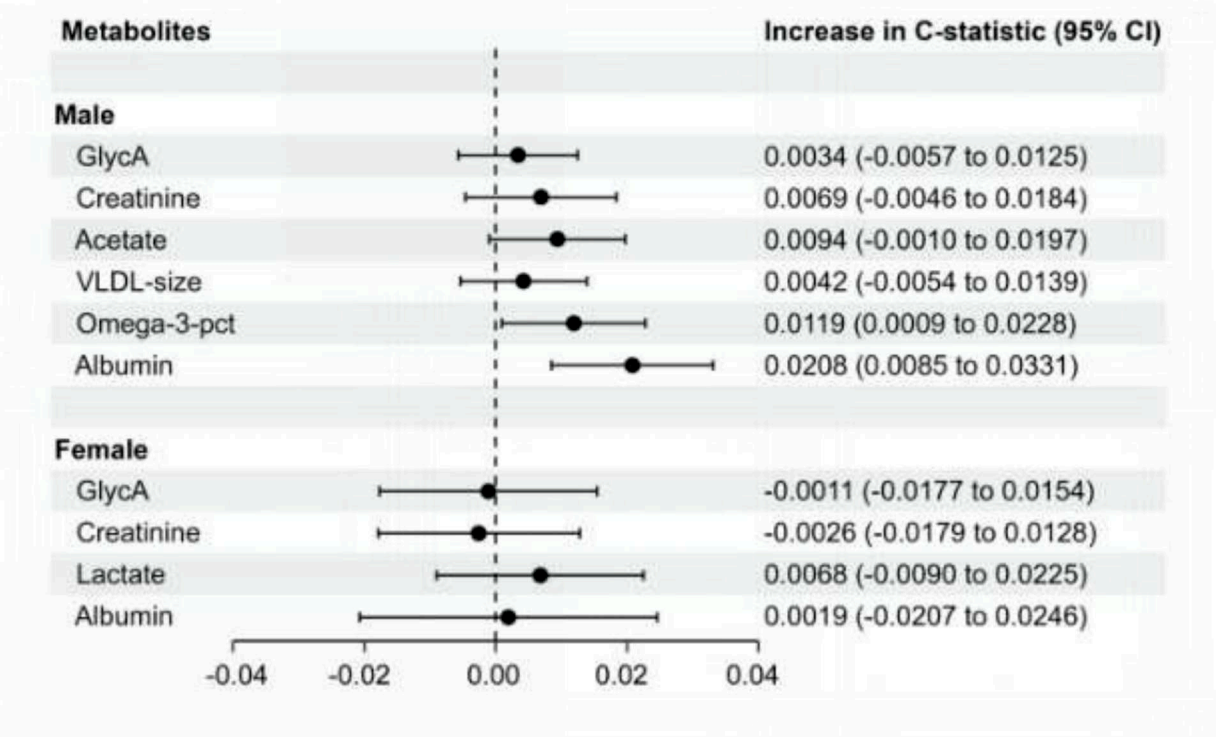



**Figure 6. The incremental discrimination of each metabolite for the model after the selected metabolites were added to SCORE2-Diabetes separately in different sexes in the internal validation set (30% of UK Biobank, N=3,077)**

**Abbreviations:** CI, confidence interval; GlycA, glycoprotein acetyls; HR, hazard ratio; Omega-3-pct, Omega-3 fatty acids to total fatty acids percentage; VLDL-size, average diameter for very-low-density lipoprotein particles.

**Table 3. Comparison of the predictive performance of model 1 "Original SCORE2-Diabetes model" and model 2 "Original SCORE2-Diabetes model extended by NMR metabolites" for 10-year MACE risk prediction**

| Metrics | Men | Women | Overall |
|---|---|---|---|
| **Derivation set (70% of UK Biobank)** | | | |
| Total sample size / MACE cases | 4,298/691 | 2,882/249 | 7,180/940 |
| C-index (Model 1) | 0.630 (0.606, 0.653) | 0.673 (0.638, 0.707) | 0.662 (0.644, 0.681) |
| C-index (Model 2[a]) | 0.647 (0.624, 0.670) | 0.680 (0.645, 0.714) | 0.678 (0.660, 0.697) |
| *P*-values for C-index comparisons[b] | **0.020** | 0.053 | **0.007** |
| **Internal validation set (30% of UK Biobank)** | | | |
| Total sample size / MACE cases | 1,842/313 | 1,235/113 | 3,077/426 |
| C-index (Model 1) | 0.632 (0.597, 0.668) | 0.670 (0.617, 0.722) | 0.660 (0.632, 0.688) |
| C-index (Model 2[a]) | 0.657 (0.623, 0.692) | 0.679 (0.625, 0.733) | 0.678 (0.650, 0.706) |
| *P*-values for C-index comparisons[b] | **0.005** | 0.716 | **0.037** |
| NRI[c] total (%) | 1.1 (-7.0, 9.3) | 7.0 (-8.8, 24.5) | 5.7 (-2.4, 13.9) |

| | | | |
|---|---|---|---|
| NRI[c] events (%) | 3.6 (-4.0, 7.5) | 2.5 (-13.7, 26.9) | 0.7 (-3.5, 8.1) |
| NRI[c] non-events (%) | -2.5 (-7.4, 4.0) | 4.5 (-4.7, 12.8) | 5.0 (-1.6, 9.4) |
| IDI | 0.001 (-0.014, 0.015) | **0.024 (0.002, 0.052)** | **0.012 (0.001, 0.025)** |
| **External validation set (ESTHER)** | | | |
| Total sample size / MACE cases | 528/107 | 511/68 | 1,039/175 |
| C-index (Model 1) | 0.632 (0.577, 0.687) | 0.618 (0.545, 0.691) | 0.612 (0.566, 0.657) |
| C-index (Model 2[a]) | 0.663 (0.608, 0.719) | 0.648 (0.575, 0.721) | 0.655 (0.611, 0.699) |
| *P*-values for C-index comparisons[b] | 0.178 | 0.241 | **0.011** |
| NRI[c] total (%) | 7.5 (-6.6, 16.0) | -6.3 (-23.1, 17.5) | 4.5 (-5.5, 13.1) |
| NRI[c] events (%) | **9.6 (0.9, 16.1)** | -2.8 (-15.8, 15.9) | 4.4 (-3.8, 11.6) |
| NRI[c] non-events (%) | -2.1 (-16.6, 10.3) | -3.5 (-17.0, 10.8) | 0.1 (-7.8, 8.6) |
| IDI | 0.004 (-0.025, 0.035) | 0.001 (-0.030, 0.041) | 0.013 (-0.005, 0.029) |

**Abbreviations:** IDI, integrated discrimination improvement; MACE, major cardiovascular events; NRI, net reclassification improvement.

[a] Metabolites that were included in model 2 were creatinine, albumin, GlycA, acetate, Omega-3-pct, and VLDL-size for men, and lactate, creatinine, albumin, and GlycA for women (see Table 1).

[b] The statistical method for comparing differences in C-index is based on U-statistics estimation, using asymptotic normal approximation for hypothesis testing.

[c] NRI calculated with the pre-specified 10-year MACE risk categories of 0–10%, >10–20%, and >20%.

### 3.1.4 Replacement of the clinical chemistry-based biomarkers of the original SCORE2-Diabetes model by NMR metabolomics

In the first analysis, we examined whether NMR-based measurements of total cholesterol, HDL-cholesterol, and creatinine could replace their counterparts measured with clinical chemistry in the SCORE2-Diabetes model. Model 3 "SCORE2-Diabetes with TC, HDL-C and creatinine measured by NMR" achieved C-indices of 0.660 in the internal validation and 0.611 in the external validation, both of which were nearly identical to those observed for model 1 "Original SCORE2-Diabetes" (**Figure 7**). In the second analysis, we investigated whether $HbA_{1c}$ could be replaced by glucose measured by NMR as well, retaining only the non-laboratory parameters of the SCORE2-Diabetes model and added the 7 NMR metabolites selected by LASSO regression. The ß-coefficients of all variables required for calculation of the resulting model 4 "Age, smoking, SBP, diabetes age at diagnosis, and NMR

metabolites" are shown in **Appendix 19**. Model 4 achieved a C-index of 0.677 in the internal validation and a C-index of 0.656 in the external validation, which were almost identical with the C-indices observed for model 2 "Original SCORE2-Diabetes extended by NMR metabolites" (**Figure 3**). The NRI and IDI of model 4 (**Appendix 20**) were also almost identical with those obtained from model 2 (**Table 3**). In summary, the replacement of the clinical chemistry-based biomarkers of the original SCORE2-Diabetes model by NMR metabolomics was possible and did not lead to a worse MACE prediction.

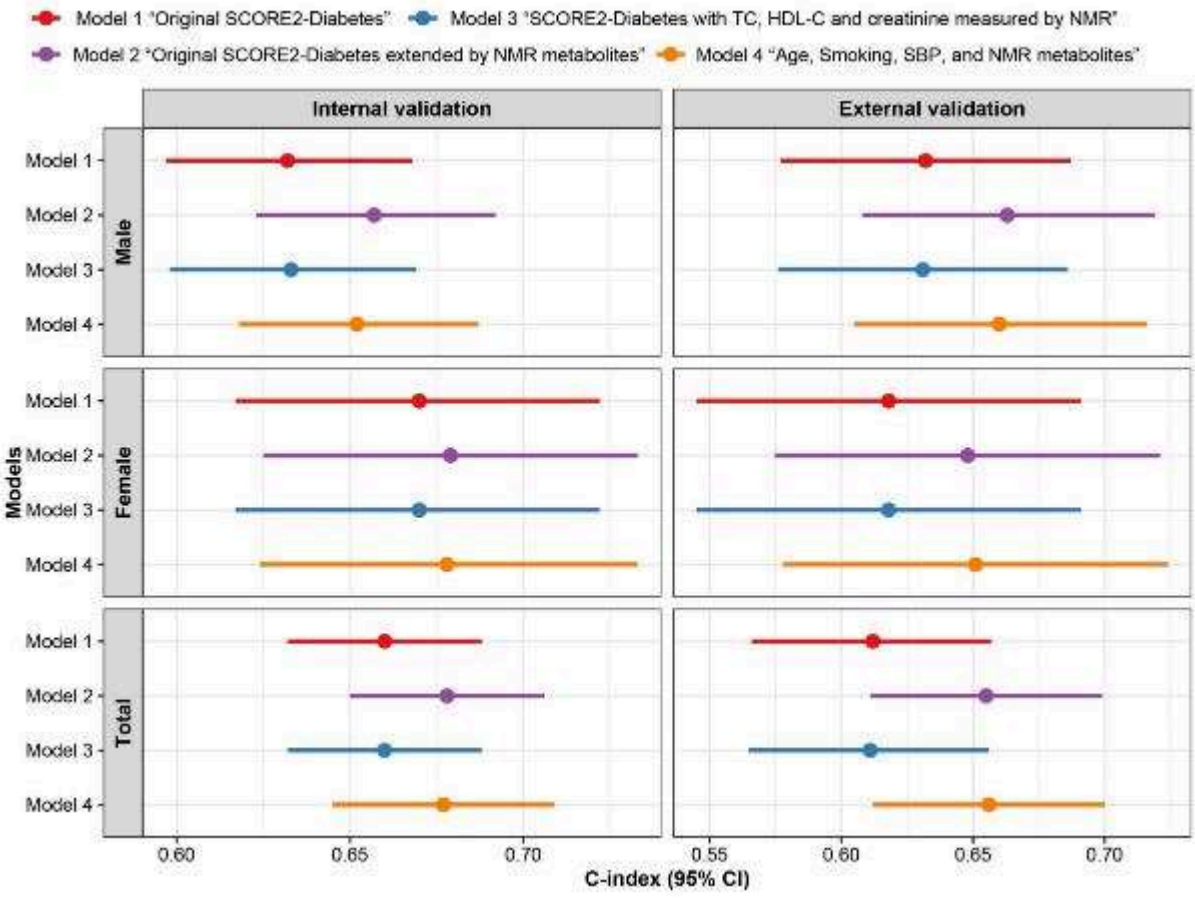



**Figure 7. Discriminative performance of four models for major adverse cardiovascular events (MACE) by sex and overall population in internal validation set (30% of UK Biobank, N=3,077) and the external validation set (ESTHER study, N=1,039)**

**Note:** Table 1 lists the variables included in each of the four models.

**Abbreviations:** NMR, nuclear magnetic resonance; CI, confidence interval; HDL, high-density lipoprotein.

## 3.2 Metabolomics for cardiovascular risk prediction in the general population

### 3.2.1 Description of the population included

**Table 4** summarizes the characteristics of the 187,039 included participants from the UKB and the 5,578 included participants from the ESTHER study for all variables included in the SCORE2 model. The UKB participants had an average age of 59.8 years (SD 5.4) and 56.1% were male. Over a follow-up time limited to 10 years, 9,396 MACE events were recorded in the UKB. The ESTHER participants had a mean age of 60.0 years (SD 5.5), and 58.3% of the participants were male. In total, 474 MACE events were recorded in the first 10 years of follow-up in ESTHER.

**Table 4. Baseline characteristics of selected participants from the UK Biobank and ESTHER study**

| **Baseline characteristics** | UK Biobank (n=187,539) | ESTHER (n= 5,578) |
|---|---|---|
| Male sex, N (%) | 105,301 (56.1) | 3,253 (58.3) |
| Age (years), mean (SD) | 59.8 (5.4) | 60.0 (5.5) |
| Current smoker, N (%) | 17,307 (9.2) | 897 (16.1) |
| Systolic blood pressure (mmHg), mean (SD) | 142.4 (19.7) | 138.1 (19.0) |
| Total cholesterol (mmol/L), mean (SD) | 4.8 (0.9) | 5.8 (1.1) |
| HDL cholesterol (mmol/L), mean (SD) | 1.4 (0.3) | 1.4 (0.3) |

**Abbreviations:** HDL, high-density lipoprotein; SD, standard deviation.

### 3.2.2 Associations of metabolomic biomarkers with MACE

A total of 13 metabolites were selected to enhance MACE risk prediction in the SCORE2 model in the derivation cohort by LASSO analyses and bootstrapping with a frequency of more than 950 times in the 1000 replicate selection (**Appendix 21**). These metabolites include three that improved MACE risk prediction in both sexes — albumin, GlycA, and IDL-CE-pct (the cholesteryl esters percentage of total lipids in intermediate-density lipoprotein). Additionally, four metabolites were specifically predictive in females - acetoacetate, LA-pct (the linoleic acid percentage of total fatty acids), L-LDL-TG-pct (the triglycerides percentage of total lipids in large low-density lipoprotein), and valine - and six were

exclusive to males: acetate, L-VLDL-TG-pct (the triglycerides percentage of total lipids in large very-low-density lipoprotein), Omega-3-pct, S-HDL-CE (cholesteryl esters in small high-density lipoprotein), VLDL-size, and XL-HDL-FC (free cholesterol in very large high-density lipoprotein).

**Figure 8** shows the associations of the 13 selected metabolites with MACE in multivariate Cox regression models adjusted for SCORE2 model variables assessed separately in males and females in the internal validation set of the UKB. GlycA, L-LDL-TG-pct and VLDL-size were positively, and LA-pct, Omega-3-pct, albumin, IDL-CE-pct, L-VLDL-TG-pct and S-HDL-CE were inversely, statistically significantly associated with MACE at an FDR-adjusted $P<0.05$ level in both sexes. Acetate was only statistically significant in males and XL-HDL-FC was only statistically significant in females with an FDR-adjusted $p<0.05$. Due to the lower statistical power of the ESTHER study, of the 13 selected metabolites, only GlycA, acetate, IDL-CE-pct and L-LDL-TG-pct were statistically significantly associated with MACE in women (**Appendix 22**).

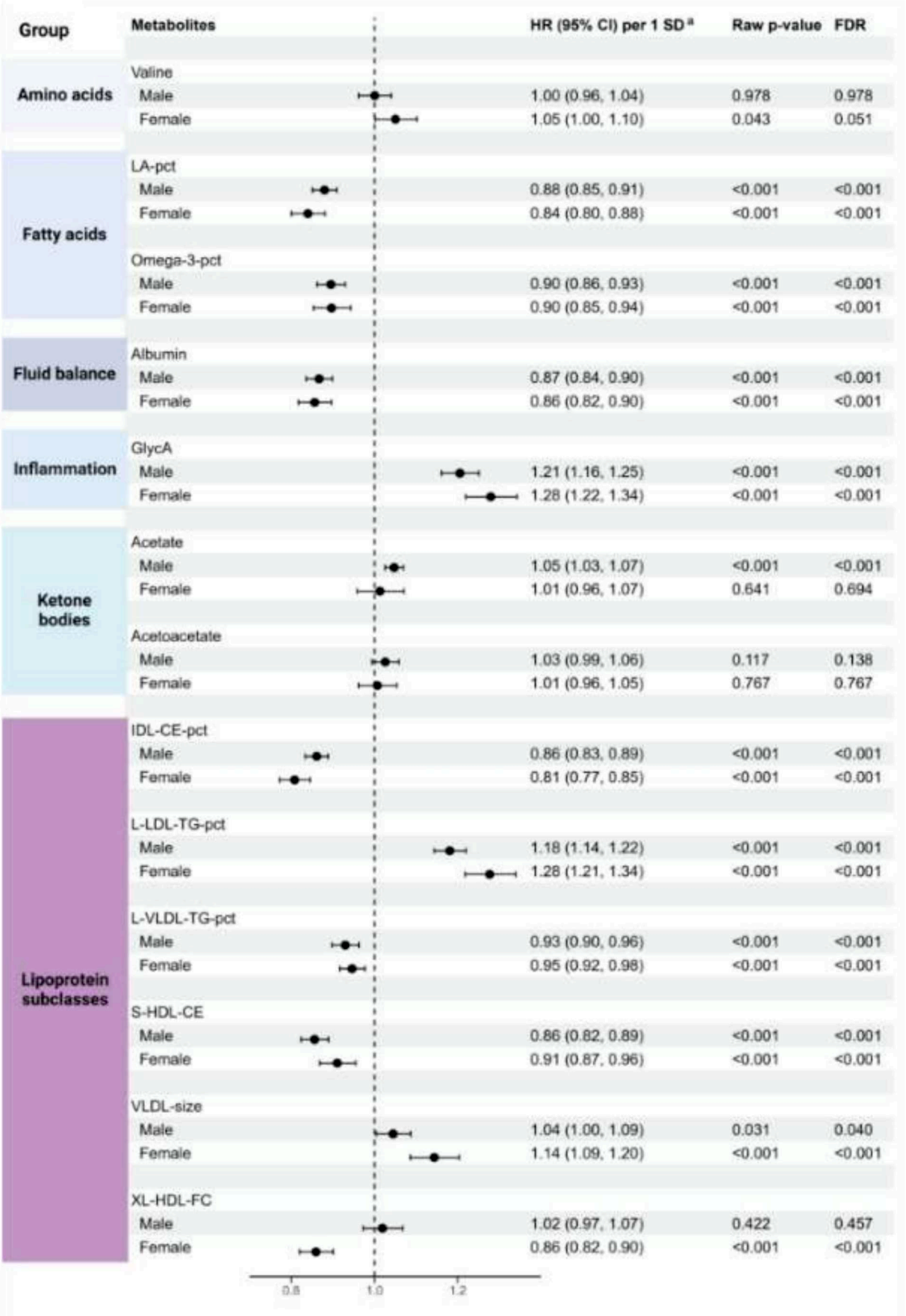



**Figure 8. Associations between selected metabolites and MACE across sexes in the internal validation cohort (30% of UK Biobank, N=56,112)**

**Abbreviations:** CI, confidence interval; FDR, false discovery rate; GlycA, glycoprotein acetyls; IDL-CE-pct, the cholesteryl esters percentage of total lipids in intermediate-density lipoprotein; LA-pct, the linoleic acid percentage of total fatty acids; L-LDL-TG-pct, triglycerides to total lipids in large low-density lipoprotein percentage; L-VLDL-TG-pct, the triglycerides percentage of total lipids in large very-low-density lipoprotein; Omega-3-pct, the Omega-3 fatty acids percentage of total fatty acids; SD, standard deviation; S-HDL-CE, cholesteryl esters in small high-density lipoprotein; VLDL-size, average diameter for very-low-density lipoprotein particles; XL-HDL-FC, free cholesterol in very large high-density lipoprotein.

[a] The standard deviations of the metabolites acetate, albumin, GlycA, IDL-CE-pct, L-VLDL-TG-pct, Omega-3-pct, S-HDL-CE, VLDL-size, XL-HDL-FC, acetoacetate, LA-pct, L-LDL-TG-pct, and valine were 0.03 mmol/L, 3.35 mmol/L, 0.12 mmol/L, 2.31%, 5.84%, 1.58nm, 0.05 mmol/L, 1.21 mmol/L, 0.01 mmol/L, 0.01 mmol/L, 3.33mmol/L, 1.68%, and 0.04 mmol/L, respectively.

### 3.2.3 MACE risk prediction by metabolomic biomarkers

First, we assessed the predictive capacity of the 13 selected metabolites without further covariates. The metabolites model had a C-index (95%CI) of 0.662 (0.662, 0.663) in the derivation set, of 0.661 (0.658, 0.664) in the internal validation set, and of 0.646 (0.620, 0.671) in the external validation set.

**Table 5** presents the main results with the predictive performance metrics of the SCORE2 model for 10-year MACE risk prediction, augmented by 13 metabolomic biomarkers. The coefficients for all variables in the extended risk model are detailed in **Appendix 23**. In the derivation set, the integration of these metabolites statistically significantly bolstered the C-index of the SCORE2 model ($P$<0.001) from 0.691 (95% CI: 0.691, 0.692) to 0.712 (95% CI: 0.712, 0.712) for the total sample of women and men. This improvement was more pronounced in males (+0.034) than in females (+0.021), with statistical significance observed in both sexes ($P$<0.001). Internal validation within the UKB cohort confirmed these findings, showcasing very similar C-index values. Moreover, the enhanced model demonstrated significant gains in reclassification statistics for both sexes, evidenced by statistically significant improvements in NRIs and IDIs ($P$<0.01). Calibration of both the SCORE2 and extended models exhibited satisfactory performance in the internal validation set, as depicted in **Appendix 24**.

In the external validation using the ESTHER cohort, the model displayed also a pronounced improvement in discrimination (C-index increase from 0.673 to 0.688; **Table 5**), reaching statistical significance in the overall sample of men and women ($P$=0.042). This improvement was more pronounced in males (+0.027) than in females (+0.009) but was not statistically significant in either of the sexes ($P$>0.05). Regarding reclassification, although the categorical NRI did not show significant improvement (total categorical NRI: 8.0, 95% CI: -2.0, 14.0), both the continuous NRI and IDI exhibited

significant enhancements in the total cohort (continuous NRI: 16.2, 95% CI: 1.9, 25.8; IDI: 0.011, 95% CI: 0.002, 0.023). These findings still illustrate the superior performance of the extended model over the SCORE2 model in risk reclassification. However, in the external validation set, both the SCORE2 and extended models demonstrated suboptimal calibration by underestimating the observed risk in both male and female participants (**Appendix 25**).

**Figure 9** illustrates the incremental enhancement in the C-statistic achieved by sequentially incorporating the 13 selected metabolites into sex-specific models in the internal validation set. Notably, in males, the inclusion of acetate, GlycA, IDL-CE-pct, Omega-3-pct, S-HDL-CE, and XL-HDL-FC significantly improved the models' discriminative capabilities. While albumin and L-VLDL-TG-pct also showed improvement in males, these changes were not statistically significant. In females, the addition of GlycA, IDL-CE-pct, and L-LDL-TG-pct significantly augmented discrimination. Additionally, albumin and LA-pct contributed to an appreciable, albeit non-significant, increase in the C-index. **Appendix 26** shows the results of the same analysis for the external validation. All 9 metabolites led to a non-significantly improved C-index in men, and in women, 4 out of 7 showed a non-significant increase in the C-index. Taking the results for men and women together, both in internal validation and external validation, the inflammation biomarker GlycA contributed the largest share in C-index increase of all selected metabolites.

**Table 5. Metrics of the predictive performance of the SCORE2 model for 10-year MACE risk without and with extension by metabolites**

| Metrics | Male [a] | Female [b] | Overall |
|---|---|---|---|
| **Derivation set (70% of UK Biobank)** | | | |
| **Total sample size (N=130,927)/MACE case number (N=6,573)** | | | |
| C-Statistics (SCORE2) | 0.632 (0.630, 0.635) | 0.688 (0.686, 0.690) | 0.691 (0.691, 0.692) |
| C-Statistics (+Metabolites) | 0.666 (0.664, 0.668) | 0.709 (0.707, 0.711) | 0.712 (0.712, 0.712) |
| *P*-values for C-Statistics comparisons | **<0.001** | **<0.001** | **<0.001** |
| **Internal validation set (30% of UK Biobank)** | | | |
| **Total sample size (N=56,112)/MACE case number (N=2,823)** | | | |
| C-Statistics (SCORE2) | 0.638 (0.625, | 0.686 (0.670, 0.702) | 0.691 (0.688, |

| | | | |
|---|---|---|---|
| | 0.650) | | 0.694) |
| C-Statistics (+Metabolites) | 0.665 (0.652, 0.677) | 0.709 (0.693, 0.724) | 0.710 (0.707, 0.713) |
| *P*-values for C-Statistics comparisons | **<0.001** | **<0.001** | **<0.001** |
| NRI continuous (%) | **13.1 (6.1, 16.9)** | **12.9 (1.7, 21.9)** | **10.5 (1.5, 15.1)** |
| NRI categorical total (%) | **8.4 (4.8, 12.0)** | **3.9 (1.0, 8.8)** | **6.7 (4.8, 9.1)** |
| NRI categorical events (%) | **3.5 (1.2, 6.3)** | 3.4 (-1.1, 8.6) | **5.1 (3.1, 7.7)** |
| NRI categorical non-events (%) | **4.9 (1.8, 6.9)** | 0.5 (-1.1, 2.5) | **1.6 (1.1, 2.1)** |
| IDI | **0.009 (0.005, 0.012)** | **0.007 (0.003, 0.011)** | **0.007 (0.003, 0.009)** |
| **External validation set (ESTHER)** | | | |
| **Total sample size (N=5,578)/MACE case number (N=474)** | | | |
| C-Statistics (SCORE2) | 0.602 (0.562, 0.642) | 0.715 (0.680, 0.749) | 0.673 (0.647, 0.698) |
| C-Statistics (+Metabolites) | 0.629 (0.589, 0.669) | 0.724 (0.690, 0.757) | 0.688 (0.663, 0.713) |
| *P*-values for C-Statistics comparisons | 0.099 | 0.317 | **0.042** |
| NRI continuous (%) | **13.0 (0.6, 20.7)** | 7.8 (-13.7, 23.3) | **16.2 (1.9, 25.8)** |
| NRI categorical total (%) | 3.4 (-4.0, 11.9) | 7.7 (-1.0, 14.0) | 8.0 (-2.0, 14.0) |
| NRI categorical events (%) | -0.1 (-7.4, 6.9) | 4.8 (-3.1, 8.2) | 3.2 (-2.6, 6.7) |
| NRI categorical non-events (%) | 3.5 (-5.4, 12.7) | 2.9 (-2.7, 8.1) | 4.8 (-4.2, 11.6) |
| IDI | **0.010 (0.000, 0.018)** | 0.006 (-0.008, 0.023) | **0.011 (0.002, 0.023)** |

**Abbreviations:** IDI, integrated discrimination improvement; NRI, net reclassification improvement.

[a] Metabolites that were included in the combined model for male were acetate, albumin, GlycA, IDL-CE-pct, L-VLDL-TG-pct, Omega-3-pct, S-HDL-CE, VLDL-size, and XL-HDL-FC.

[b] Metabolites that were included in the combined model for female were acetoacetate, albumin, GlycA, IDL-CE-pct, LA-pct, L-LDL-TG-pct, and valine.

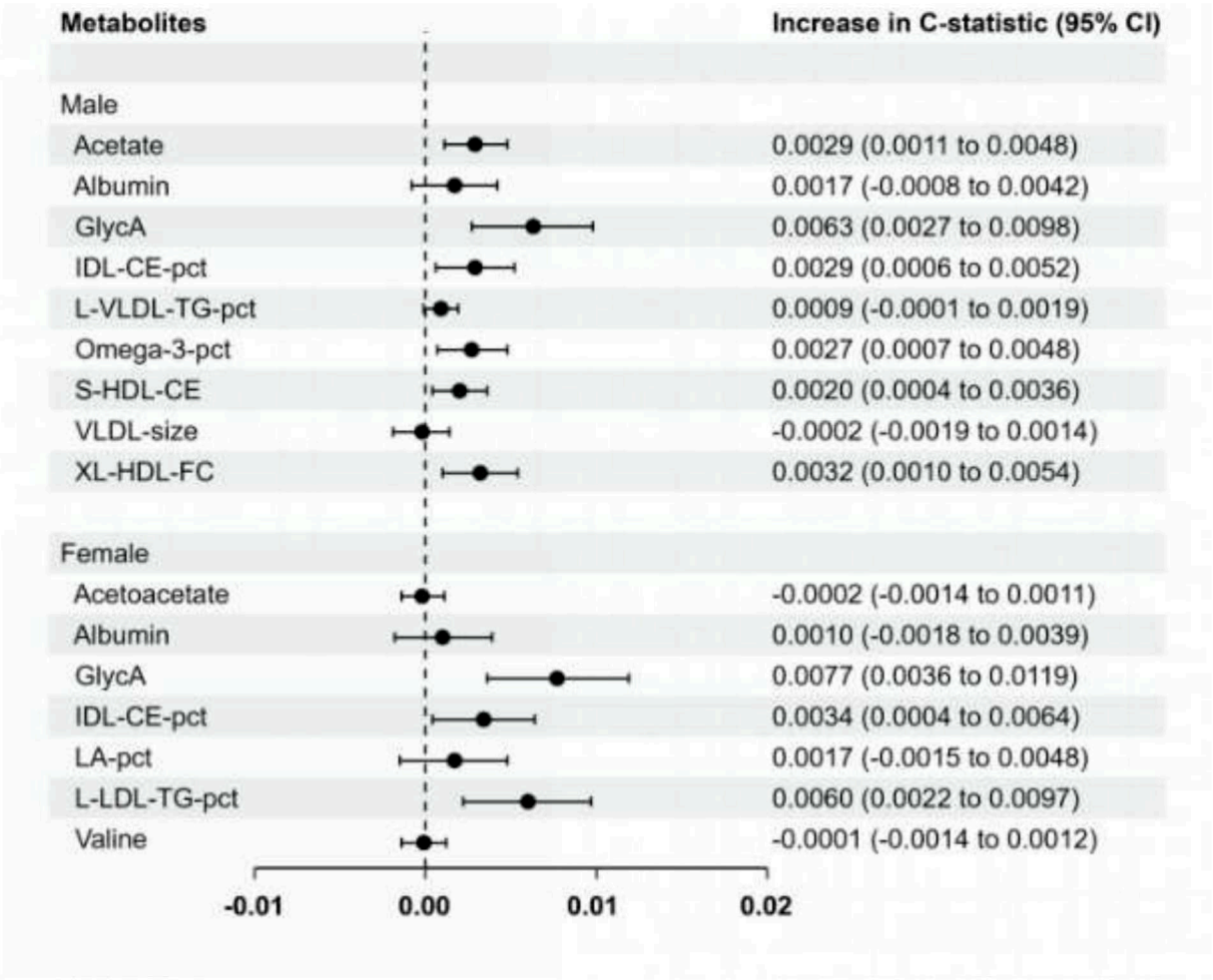



**Figure 9. The incremental discrimination of each metabolite for the model after the selected metabolites were added to SCORE2 separately in different sexes in the internal validation cohort (30% of UK Biobank, N=56,112)**

**Abbreviations:** CI, confidence interval; GlycA, glycoprotein acetyls; IDL-CE-pct, the cholesteryl esters percentage of total lipids in intermediate-density lipoprotein; LA-pct, the linoleic acid percentage of total fatty acids; L-LDL-TG-pct, triglycerides to total lipids in large low-density lipoprotein percentage; L-VLDL-TG-pct, the triglycerides percentage of total lipids in large very-low-density lipoprotein; Omega-3-pct, the Omega-3 fatty acids percentage of total fatty acids; S-HDL-CE, cholesteryl esters in small high-density lipoprotein; VLDL-size, average diameter for very-low-density lipoprotein particles; XL-HDL-FC, free cholesterol in very large high-density lipoprotein.

## 3.3 Metabolomics for type 2 diabetes risk prediction

### 3.3.1 Description of the population included

**Table 6** summarizes the characteristics of the clinical CDRS variables among 86,232 participants in the UKB cohort and 4,383 participants in the ESTHER study. The average age and sex distribution were similar across the two cohorts, with participants in the UKB having an average age of 59.9 years (SD 4.4) and comprising 44.3% males. In the ESTHER cohort, the mean age was 60.2 years (SD 5.5), with 42.7% of the participants being male. Furthermore, the BMI and $HbA_{1c}$ levels were comparable. Compared to the UKB, a higher proportion of participants in the ESTHER study were current smokers (17.5% in ESTHER vs. 9.3% in the UKB), more had a family history of diabetes (36.9% vs. 16.8%) and more were prescribed anti-hypertensive medication (37.8% in ESTHER vs. 13.3% in the UKB). Steroid prescriptions were equally rare in both cohorts (0.8% in ESTHER vs. 1.3% in UKB).

**Table 6. Baseline characteristics of selected participants from the UK Biobank and ESTHER study**

| Baseline characteristics | Derivation cohort | Validation cohort |
|---|---|---|
| | UK Biobank | ESTHER |
| Number of participants | 86,232 | 4,383 |
| Male sex, n (%) | 39,442 (44.3) | 1,871 (42.7) |
| Age (years), mean (SD) | 59.9 (5.4) | 60.2 (5.5) |
| $HbA_{1c}$ (%), mean (SD) | 3.1 (0.4) | 3.2 (0.5) |
| BMI ($kg/m^2$), mean (SD) | 27.4 (4.5) | 27.4 (4.3) |
| BMI category, n (%) | | |
| <25 $kg/m^2$ | 27,626 (32.0) | 1,302 (29.7) |
| 25-27.49 $kg/m^2$ | 21,529 (25.0) | 1,193 (27.2) |
| 27.5-29.99 $kg/m^2$ | 16,666 (19.3) | 958 (21.7) |
| ≥ 30 $kg/m^2$ | 20,411 (23.7) | 984 (22.4) |
| Cigarette smoking, n (%) | | |
| Never | 46,150 (53.5) | 2,233 (50.9) |
| Ex-smoker | 32,075 (37.2) | 1,385 (31.6) |
| Current smoker | 8,007 (9.3) | 765 (17.5) |
| Family history of diabetes, n (%) | | |
| None | 71,783 (83.2) | 2,768 (63.2) |
| Parent or sibling | 12,982 (15.1) | 1,384 (31.6) |

| | | |
|---|---|---|
| Parent and sibling | 1,467 (1.7) | 231 (5.3) |
| Prescribed medication, n (%) | | |
| Anti-hypertensive | 11,502 (13.3) | 1,658 (37.8) |
| Steroid | 1,104 (1.3) | 33 (0.8) |

**Abbreviations:** BMI, body mass index; $HbA_{1c}$, glycated hemoglobin; SD, standard deviation.

### 3.3.2 Associations of selected metabolites with incident T2D

Over a follow-up of up to 10 years, 3,537 of 86,232 study participants of the UKB developed diabetes (incidence rate (IR) per 10,000 person-years (PY), 55.0) and 495 of 4,383 participants of the ESTHER study (IR per 10,000 PY, 145.2).

Through LASSO analyses and bootstrapping, 11 metabolites were selected to enhance the predictive power of the clinical CDRS for the risk of T2D in the training set. These metabolites include four glycolysis related metabolites (citrate, glucose, lactate and pyruvate), two ketone bodies (3-hydroxybutyrate and acetate), two amino acids (glutamine and tyrosine), two lipoprotein related metabolites (IDL-CE-pct and M-LDL-TG-pct (triglycerides to total lipids in medium LDL percentage), and a fatty acid-related metabolite LA-pct. **Appendix 27** displays the Pearson correlation matrix of the selected 11 metabolites in their log-transformed and standardized format, which transformed their distributions close to normal distributions. Three biomarkers of glycolysis, pyruvate, glucose and lactate, were highly correlated with each other (all one-on-one Pearson's $r > 0.5$ in ESTHER and for pyruvate and lactate in UKB). Furthermore, the three lipoprotein related biomarkers M-LDL-TG-pct, IDL-CE-pct, and LA-pct were highly correlated in both cohorts (all Pearson's $r > 0.5$). Except for these two biomarker clusters, the correlations were low to moderate.

The 11 selected metabolites alone (without any covariates like age and sex) already predicted incident diabetes cases in the next 10 years well, with a C-index (95% CI) of 0.733 (0.715; 0.750) in the internal validation set and a C-index of 0.735 (0.713; 0.757) in the external validation set. **Figure 10** presents the associations between the 11 selected metabolites and incident T2D within the test set of the UKB and the ESTHER cohort, adjusted for the variables of the clinical CDRS graphically, while the HRs with 95%CIs per SD increments and p-values can be found in **Appendix 28**. All biomarkers except the ketone

bodies 3-hydroxybutyrate and acetate were statistically significantly associated with T2D in the UKB. Except for 3-hydroxyruvate, for which confidence intervals did not overlap, the results were remarkably similar in the ESTHER study. Results were also comparable for citrate and pyruvate although these biomarkers lacked statistical significance in the ESHTER study.

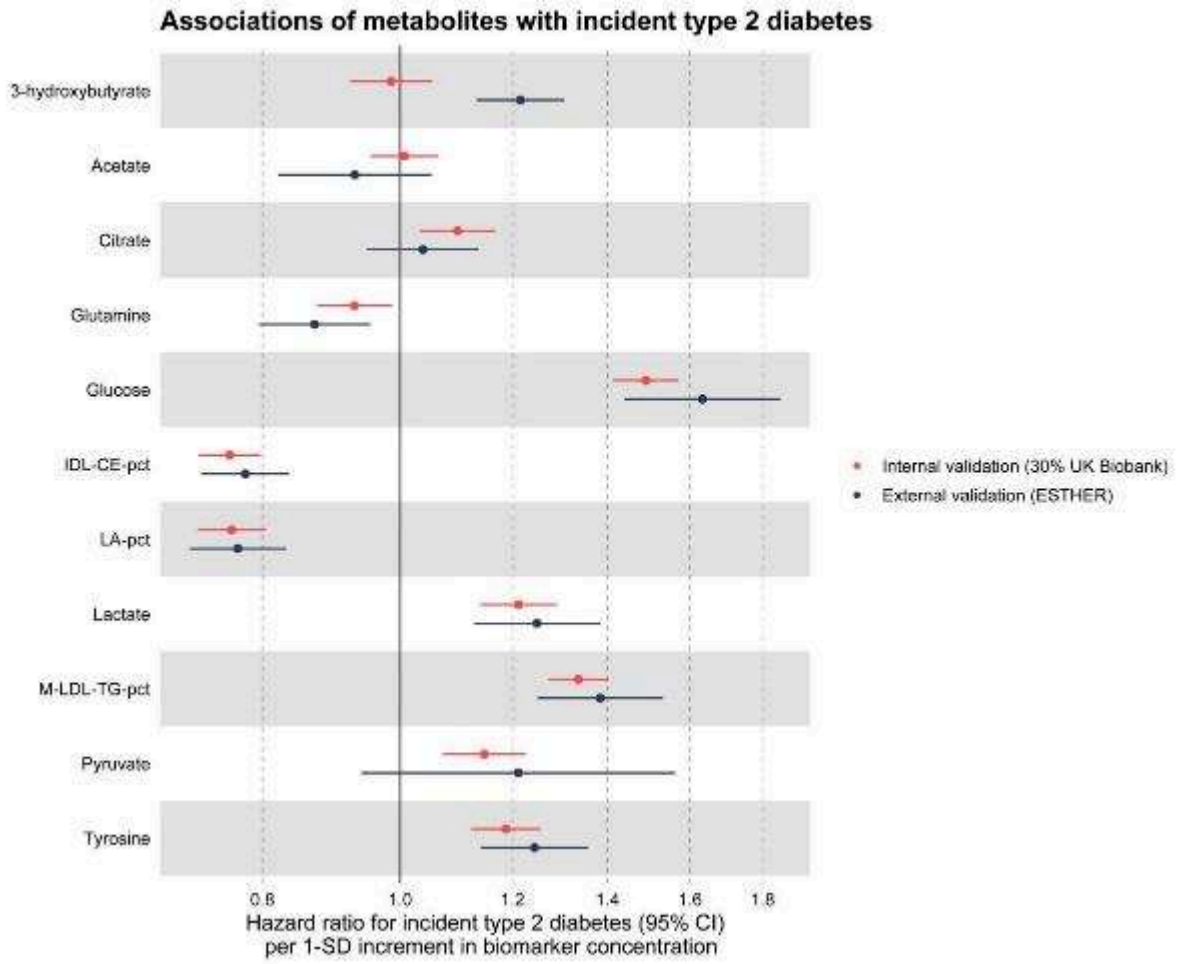



**Figure 10. Associations between selected metabolites and incident type 2 diabetes in the test set (30% of UK Biobank, N=25,870) and external validation set (ESTHER, N=5,904)**

**Notes:** Hazard ratios are expressed per 1 standard deviation of the respective metabolite concentration and are adjusted for the variables of the clinical CDRS. The exact HRs with 95%CIs and p-values per SD increments and the SDs of the metabolites are shown in **Appendix 28.**
**Abbreviations:** CI, confidence interval; IDL-CE-pct, cholesteryl esters to total lipids in IDL percentage; LA-pct, linoleic acid to total fatty acids percentage; M-LDL-TG-pct, triglycerides to total lipids in medium LDL percentage; SD, standard deviation.

### 3.3.3 Improvements in T2D risk prediction by metabolomic biomarkers

In the training set, the C-index for the clinical CDRS [0.830 (0.821, 0.839)] was much higher than for the CDRS [0.740 (0.731, 0.749)]. Thus, we compared the predictive performance of the novel model of the clinical CDRS extended by 11 metabolites to the clinical CDRS, only. We call the novel model "UK Biobank Diabetes Risk Score (UKB-DRS)" in the following. Coefficients for all variables in the UKB-DRS are listed in **Appendix 29**. In the training set, test set and external validation set, the UKB-DRS consistently had a statistically significantly ($P$<0.001) higher C-index (95%CI) than the clinical CDRS, with comparable C-index increases of 0.017, 0.019 and 0.028 units, respectively **(Table 7)**. The ROC are shown in **Appendix 30**. Moreover, an improved risk stratification ability was observed, with a continuous NRI (95% CI) of 39.8% (34.6%, 45.0%) and an IDI (95% CI) of 0.0216 (0.0170, 0.0262) in internal validation and a continuous NRI of 33.8% (26.4%, 41.2%), and an IDI of 0.0016 (0.0012, 0.0019) in external validation. The model calibration of both the clinical CDRS and the UKB-DRS was good in the internal and the external validation set and the confidence interval bands overlapped (**Appendix 31**).

To check, whether glucose is the main source of the model improvement, we compared the predictive performance metrics of the UKB-DRS with the clinical CDRS augmented with glucose as the reference model (**Appendix 32**). The UKB-DRS had a higher C-index (95% CI) [0.834 (0.820, 0.847) vs 0.820 (0.805, 0.834) in internal and 0.798 (0.779, 0.817) vs 0.773 (0.755, 0.791) in external validation set] and a statistically significant NRI (95% CI) [26.3% (21.2%, 31.4%) in internal and 13.2% (7.1%, 19.2%) in external validation set] compared to the clinical CDRS extended by glucose.

To further explore whether the addition of the 11 metabolites was particularly useful in specific subgroups, we conducted subgroup analyses by age, sex, obesity, and clinical CDRS risk. As shown in **Appendix 33**, the UKB-DRS consistently outperformed the clinical CDRS across all subgroups in both internal and external validation. There was no subgroup with a particularly low or high C-index change.

**Figure 11** illustrate the incremental improvement in the C-index and **Appendix 34** the NRI and IDI for each of the 11 selected metabolites after their incorporation into the clinical CDRS in internal and external validation. Especially glucose, IDL-CE-pct, LA-pct and M-LDL-TG-pct enhanced the

discriminative ability of the model in both cohorts with C-index changes > 0.005 in the external validation cohort. The other seven metabolites showed low increases in C-statistic at best. Thus, we additionally developed a simplified model, termed the concise UKB-DRS, which incorporated only these four key metabolites (glucose, IDL-CE-pct, LA-pct, and M-LDL-TG-pct) alongside the traditional CDRS variables. With a C-index (95% CI) of 0.830 (0.815, 0.844) in internal validation and 0.786 (0.766, 0.805) in external validation, the concise UKB-DRS had a comparable discriminative performance compared to the full UKB-DRS, including 11 metabolites. **Appendix 30** shows that, compared to the ROC curves of the full UKB-DRS, the ROC curves of the concise UKB-DRS were almost identical (in internal validation) or just a little lower (in external validation). The NRI (95%CI) and IDI (95%CI) of the concise UKB-DRS in internal validation (36.7 (31.4, 41.9) and 0.0205 (0.0162, 0.0248), respectively) and external validation (24.4 (16.7, 32.2) and 0.0011 (0.0008, 0.0015), respectively) were also just a little weaker than for the full UKB-DRS. Moreover, the calibration curves of the full and concise UKB-DRS were also comparable (**Appendix 31**).

**Table 7. Comparison of the predictive performance of the clinical Cambridge Diabetes Risk Score and the novel developed UK Biobank Diabetes Risk Score for 10-year type 2 diabetes risk prediction**

| Metrics | |
|---|---|
| **Derivation set (70% of UK Biobank)** | |
| **Sample size / incident type 2 diabetes** | **N=60,362 / 2,477** |
| C-index (Clinical CDRS[a]) | 0.830 (0.821, 0.839) |
| C-index (Concise UKB-DRS[b]) | 0.846 (0.837, 0.856) |
| C-index (UKB-DRS[c]) | 0.847 (0.838, 0.855) |
| Δ C-index (UKB-DRS[c] vs. Clinical CDRS[a]); *P*-value | +0.017; *P*<0.0001 |
| **Internal validation set (30% of UK Biobank)** | |
| **Sample size / incident type 2 diabetes** | **N=25,870 / 1,059** |
| C-index (Clinical CDRS[a]) | 0.815 (0.800, 0.829) |
| C-index (Concise UKB-DRS[b]) | 0.830 (0.815, 0.844) |
| C-index (UKB-DRS[c]) | 0.834 (0.820, 0.847) |
| Δ C-index (UKB-DRS[c] vs. Clinical CDRS[a]); *P*-value | +0.019; *P*<0.0001 |
| Continuous NRI (%) (UKB-DRS[c] vs. Clinical CDRS[a]); *P*-value | 39.8 (34.6, 45.0); *P*<0.0001 |
| IDI (UKB-DRS[c] vs. Clinical CDRS[a]); *P*-value | 0.0216 (0.0170, 0.0262); *P*<0.0001 |
| **External validation set (ESTHER)** | |
| **Sample size / incident type 2 diabetes** | **N=4,383 / 495** |
| C-index (Clinical CDRS[a]) | 0.770 (0.750, 0.791) |
| C-index (Concise UKB-DRS[b]) | 0.786 (0.766, 0.805) |
| C-index (UKB-DRS[c]) | 0.798 (0.779, 0.817) |
| Δ C-index (UKB-DRS[c] vs. Clinical CDRS[a]); *P*-value | +0.028; *P*<0.0001 |
| Continuous NRI (%) (UKB-DRS[c] vs. Clinical CDRS[a]); *P*-value | 33.8 (26.4, 41.2); *P*<0.0001 |
| IDI (UKB-DRS[c] vs. Clinical CDRS[a]); *P*-value | 0.0016 (0.0012, 0.0019); *P*<0.0001 |

**Abbreviations:** CDRS, Cambridge Diabetes Risk Score; IDI, integrated discrimination improvement; NRI, net reclassification improvement, UKB-DRS (UK Biobank Diabetes Risk Score)

[a] The clinical CDRS consists of the following diabetes risk factors: (age, sex, BMI, smoking status, family history of diabetes, prescription of anti-hypertensive medication and steroids, and $HbA_{1c}$).

[b] The concise UKB-DRS is based on the traditional diabetes risk factors of the clinical CDRS (see above) and the following 4 metabolites: glucose, IDL-CE-pct (cholesteryl esters to total lipids in intermediate density lipoproteins percentage), LA-pct (linoleic acid to total fatty acids percentage), and M-LDL-TG-pct (triglycerides to total lipids in medium low-density lipoproteins percentage).

[c] The UKB-DRS is based on the traditional diabetes risk factors of the clinical CDRS (see above) and the following 11 metabolites: glucose, citrate, lactate, pyruvate, 3-hydroxybutyrate, acetate, glutamine, tyrosine, IDL-CE-pct, LA-pct, and M-LDL-TG-pct (see footnote above for abbreviations).

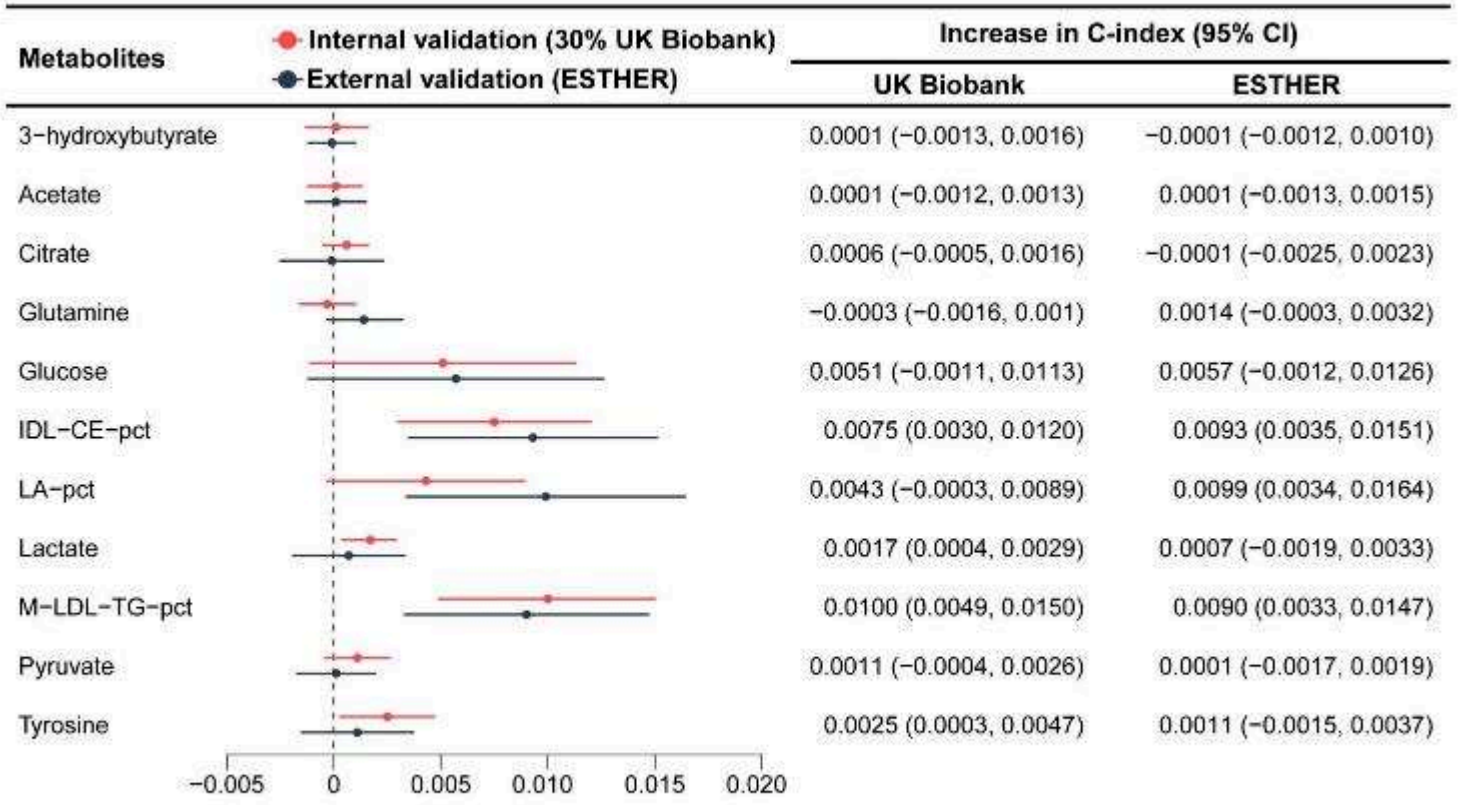



**Figure 11. The incremental discrimination of each metabolite for the model after the selected metabolites were added to the clinical Cambridge Diabetes Risk Score in the test set (30% of UK Biobank, N=25,870) and external validation set (ESTHER, N=4,383)**

**Abbreviations:** CI, confidence interval; IDL-CE-pct, cholesteryl esters to total lipids in IDL percentage; LA-pct, linoleic acid to total fatty acids percentage; M-LDL-TG-pct, triglycerides to total lipids in medium LDL percentage.

## 3.4 Proteomics for type 2 diabetes risk prediction

### 3.4.1 Description of the population included

**Table 8** summarizes the characteristics of the clinical CDRS variables among 21,898 participants in the UKB cohort and 4,454 participants in the ESTHER study. The sex distribution was comparable between the two cohorts with a little more women than men included. The UKB participants were on average 5 years younger than the ESTHER participants (means: 56.6 v s.61.7 years). This was associated with higher proportions of family history of diabetes and prescribed antihypertensive medication in the ESTHER study. Furthermore, there were more current smokers (16.8%) among the ESTHER participants than among the UKB participants (10.3%). Despite these differences, the BMI and $HbA_{1c}$ distribution were comparable. Steroid prescriptions were rare in both cohorts (0.8% in ESTHER vs. 1.6% in UKB).

**Table 8. Baseline characteristics of selected participants from the UK Biobank and ESTHER study**

| Baseline characteristics | Derivation cohort | Validation cohort |
|---|---|---|
| | UK Biobank | ESTHER |
| Number of participants | 21,898 | 4,454 |
| Male sex, n (%) | 9,794 (44.7) | 1,896 (42.6) |
| Age (years), mean (SD) | 56.6 (8.2) | 61.7 (6.6) |
| $HbA_{1c}$ (%), mean (SD) | 5.4 (0.3) | 5.5 (0.4) |
| BMI ($kg/m^2$), mean (SD) | 27.3 (4.6) | 27.3 (4.3) |
| BMI category, n (%) | | |
| <25 $kg/m^2$ | 7,273 (33.2) | 1,348 (30.4) |
| 25-27.49 $kg/m^2$ | 5,457 (24.9) | 1,151 (25.8) |
| 27.5-29.99 $kg/m^2$ | 4,076 (18.6) | 964 (21.6) |
| ≥ 30 $kg/m^2$ | 5,092 (23.3) | 991 (22.2) |
| Cigarette smoking, n (%) | | |
| Never | 11,759 (53.7) | 2,339 (52.5) |
| Ex-smoker | 7,888 (36.0) | 1,368 (30.7) |
| Current smoker | 2,251 (10.3) | 747 (16.8) |
| Family history of diabetes, n (%) | | |
| None | 18,275 (83.4) | 2,941 (66.0) |

| | | |
|---|---|---|
| Parent or sibling | 3,270 (14.9) | 1,280 (28.8) |
| Parent and sibling | 353 (1.6) | 233 (5.2) |
| Prescribed medication, n (%) | | |
| Anti-hypertensive drugs | 2,702 (12.3) | 1,775 (39.9) |
| Steroids | 340 (1.6) | 37 (0.8) |

**Abbreviations:** BMI, body mass index; $HbA_{1c}$, glycated hemoglobin; SD, standard deviation.

### 3.4.2 Associations of selected proteins with incident type 2 diabetes

Over a follow-up period of up to 10 years, 800 of 21,898 study participants of the UKB and 427 of 4,454 participants of the ESTHER study developed T2D (IR per 10,000 PY 41.6 and 124.6, respectively). The higher IR in the ESTHER study is related to the older average age of participants compared to the UKB.

Through LASSO analyses and bootstrapping, 15 proteins from the Olink Explore panel and 6 inflammation-related proteins from the Olink Target Inflammation panel were selected in the training set to improve the predictive ability of the clinical CDRS for T2D risk. As illustrated in **Appendix 35**, the selected proteins had low Spearman correlations with BMI and $HbA_{1c}$ in both cohorts.

**Figure 12** shows the HRs (95%CI) for the associations between the selected proteins and the incidence of T2D per 1 SD increase of the respective protein, adjusted for the variables in the clinical CDRS. All 15 proteomic biomarkers selected from the Olink Explore platform were significantly associated with the incidence of T2D in the internal validation set from the UKB (**Figure 12A**). Likewise, all 6 proteomic biomarkers selected from the Olink Target platform were statistically significantly associated with incident T2D in internal validation (**Figure 12B**). Interestingly, the hepatocyte growth factor (HGF), interleukin-18 receptor 1 (IL18R1), and TNF-related weak inducer of apoptosis (TWEAK) were selected from both platforms underscoring the important role of inflammation in T2D risk. In external validation in the ESTHER study, 3 out of 6 inflammation biomarkers (HGF, IL18R1, and fibroblast growth factor 21 (FGF21)) were confirmed as risk predictors of T2D with statistical significance (**Figure 12C**).

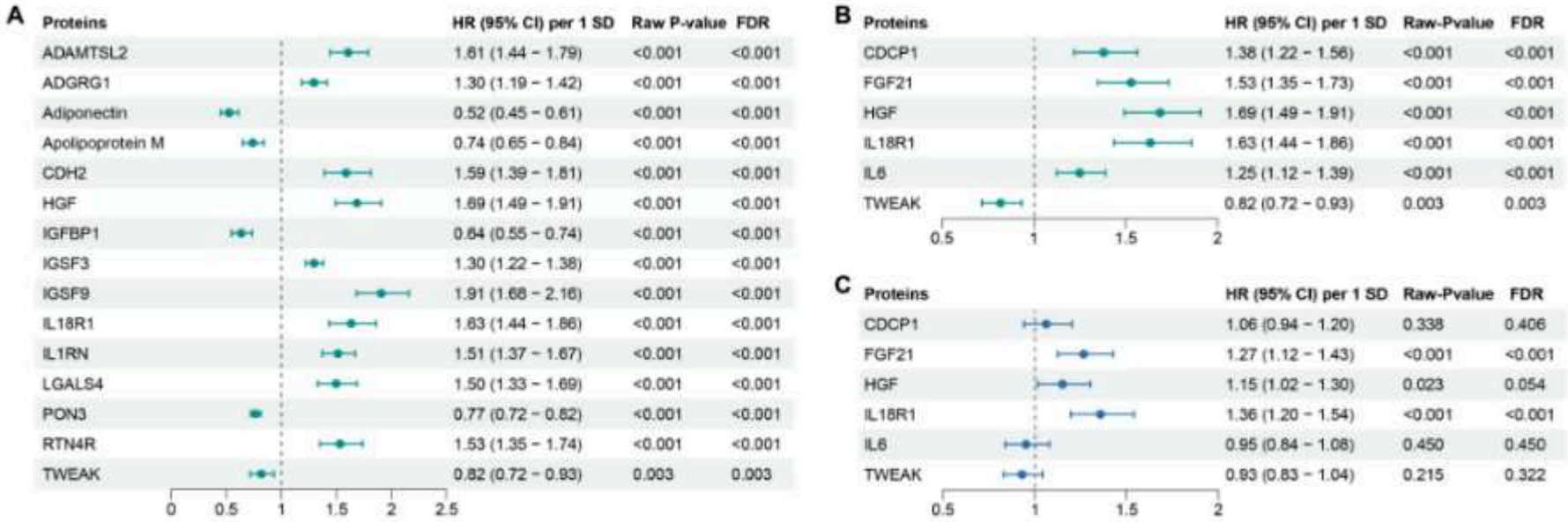



**Figure 12. Associations of 15 selected proteins from OLINK Explore with incident type 2 diabetes in the internal validation set (Panel 2A, 30% of UK Biobank, N=6,339) and associations of 6 selected proteins from OLINK Target Inflammation panel with incident type 2 diabetes in the internal validation set (Panel 2B, 30% of UK Biobank, N=6,339) and external validation set (Panel 2C, 70% of ESTHER, N=3,118)**

**Abbreviations:** ADAMTSL2, ADAMTS-like protein 2; ADGRG1, Adhesion G-protein coupled receptor G1; CDH2, Cadherin-2; CDRS, Cambridge Diabetes Risk Score; CI, confidence interval; FDR, false discovery rate; HGF, Hepatocyte growth factor; IGFBP1, Insulin-like growth factor-binding protein 1; IGSF, Immunoglobulin superfamily member; IL1RN, Interleukin-1 receptor antagonist protein; IL18R1, Interleukin-18 receptor 1; LGALS4, Galectin-4; NPX, Normalised Protein eXpression; PON3, Serum paraoxonase/lactonase 3; RTN4R, Reticulon-4 receptor; Serum paraoxonase/lactonase 3; SD, standard deviation; TWEAK, TNF-related weak inducer of apoptosis.

### 3.4.3 Improvements in T2D risk prediction by metabolomic biomarkers

The clinical CDRS demonstrated good discriminative ability, with C-indices of 0.831 in the internal validation set and 0.748 in the external validation set (**Table 9**). The addition of proteomic biomarkers significantly improved the area under the ROC curves of the clinical CDRS (**Appendix 36**). Adding 15 or 6 proteins to the clinical CDRS improved its C-index by 0.029 and 0.016 in internal validation (both $P<0.001$). The ß-coefficients of the variables included in the two models are provided in **Appendix 37** and **Appendix 38**. The 15-protein-extended model yielded a higher C-index than the 6-protein-extended model ($\Delta$C-index = 0.013, $P<0.001$). External validation in ESTHER showed a 0.014 improvement for the 6-protein-extended model ($P<0.01$).

Risk stratification improved substantially with the addition of proteomic biomarkers. In the UKB internal validation, the NRI was 23.0% for the 15-protein-extended model and 29.0% for the 6-protein-extended model. In the ESTHER external validation, the 6-protein-extended model achieved an NRI of 19.1%. Moreover, the IDI was statistically significant in the external validation (p=0.039).

Model calibration of the two protein-extended models was good in the internal and the external validation sets (**Appendix 39**).

**Figure 13** illustrates the incremental improvement in the C-index for each selected protein. Among the 15 proteins from the OLINK Explore panel, 12 significantly improved predictive ability. Among the 6 proteins from OLINK Inflammation panel, 4 improved discrimination in the UKB, and 2 were significant in both cohorts. Differences in mean protein concentrations of the 6 selected proteins were marginal between fasting and non-fasting study participants of the ESTHER study (**Appendix 40**), which is important to note because the non-fasting status was higher in the UKB (95.9%) than in the ESTHER study (12.9%).

**Appendix 40-43** describes additional results of predictive performance analyses with restriction to randomly selected UKB participants, in subgroup analyses by pre-diabetes status, with lower LASSO bootstrapping selection frequency, and comparisons with protein panels suggested in the literature .

**Table 9. Comparison of the predictive performance of the clinical CDRS with and without proteomic biomarkers for 10-year T2D risk prediction**

| **Metrics** | |
|---|---|
| **Derivation set (70% of UK Biobank)** | |
| **Sample size / incident type 2 diabetes** | **N=14,759 / 569** |
| C-index (Clinical CDRS[a]) | 0.850 (0.833, 0.867) |
| C-index (+15 proteins derived from Olink Explore[b]); *P*-value[c] | **0.881 (0.866, 0.896); *P*<0.001** |
| C-index (+6 proteins derived from Olink Target[d]); *P*-value[c] | **0.869 (0.853, 0.889); *P*<0.001** |
| **Internal validation set (30% of UK Biobank)** | |
| **Sample size / incident type 2 diabetes** | **N=6,339 / 231** |
| C-index (Clinical CDRS[a]) | 0.831 (0.800, 0.861) |
| C-index (+15 proteins derived from Olink Explore[b]); *P*-value[c] | **0.860 (0.833, 0.888); *P*<0.001** |
| Continuous NRI (%); *P*-value[c] | **23.0 (10.0, 36.0); *P*<0.001** |
| IDI; *P*-value[c] | 0.016 (-0.003, 0.034); *P*=0.292 |
| C-index (+6 proteins derived from Olink Target[d]); *P*-value[c] | **0.847 (0.818, 0.876); *P*<0.001** |
| Continuous NRI (%); *P*-value[c] | **29.0 (16.1, 41.9); *P*<0.001** |
| IDI; *P*-value[c] | -0.000 (-0.012, 0.013); *P*=0.964 |
| **External validation set (70% ESTHER)** | |
| **Sample size / incident type 2 diabetes** | **N=3,118 / 290** |
| C-index (Clinical CDRS[a]) | 0.748 (0.720, 0.776) |
| C-index (+6 proteins derived from Olink Target[d]); *P*-value[c] | **0.762 (0.735, 0.789); *P*=0.006** |
| Continuous NRI (%); *P*-value[c] | **19.1 (7.1, 31.1); *P*=0.002** |
| IDI; *P*-value[c] | **0.011 (0.000, 0.022); *P*=0.039** |

**Abbreviations:** CDRS, Cambridge Diabetes Risk Score; IDI, integrated discrimination improvement; NRI, net reclassification improvement

[a] The clinical Cambridge Diabetes Risk Score (CDRS) consists of the following diabetes risk factors: (age, sex, BMI, smoking status, family history of diabetes, prescription of anti-hypertensive medication and steroids, and $HbA_{1c}$).

[b] The clinical CDRS (see above) and the following 15 proteins: ADAMTSL2, ADGRG1, Adiponectin, Apolipoprotein M, CDH2, HGF, IGFBP1, IGSF3, IGSF9, IL18R1, IL1RN, LGALS4, PON3, RTN4R and TWEAK (see legend of Figure 2 for protein abbreviations).

[c] *P*-value for comparison with the clinical CDRS.

[d] The clinical CDRS (see above) and the following 6 proteins: CDCP1, FGF21, HGF, IL18R1, IL6, and TWEAK (see legend of Figure 2 for proteins abbreviations).

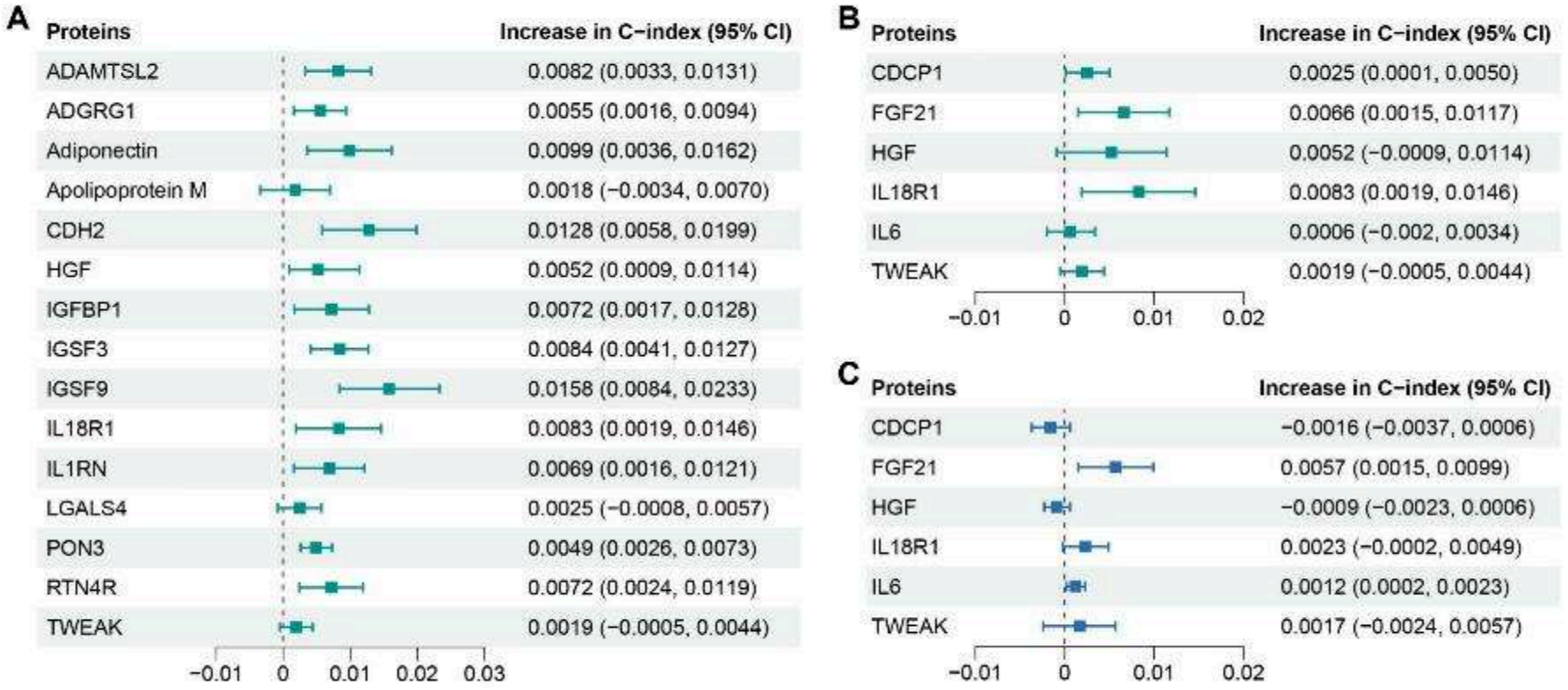



**Figure 13. Increase in C-index by each protein when added to the clinical CDRS for T2D risk prediction**

**Abbreviations:** ADAMTSL2, ADAMTS-like protein 2; ADGRG1, Adhesion G-protein coupled receptor G1; CDH2, Cadherin-2; CDRS, Cambridge Diabetes Risk Score; CI, confidence interval; FDR, false discovery rate; HGF, Hepatocyte growth factor; IGFBP1, Insulin-like growth factor-binding protein 1; IGSF, Immunoglobulin superfamily member; IL1RN, Interleukin-1 receptor antagonist protein; IL18R1, Interleukin-18 receptor 1; LGALS4, Galectin-4; NPX, Normalised Protein eXpression; PON3, Serum paraoxonase/lactonase 3; RTN4R, Reticulon-4 receptor; Serum paraoxonase/lactonase 3; SD, standard deviation; TWEAK, TNF-related weak inducer of apoptosis.

## 3.5 Proteomics for cardiovascular risk prediction in the general population

### 3.5.1 Description of the population included

**Table 10** presents the baseline characteristics of the 47,382 UKB participants included in the analysis. The mean age was 56.4 ± 8.2 years, and 44.1% were male. Over a 10-year follow-up period, 2,163 participants experienced MACE. Compared to participants who did not experience MACE, those who developed MACE were significantly older (mean age, 61.2 ± 8.1 years vs. 56.2 ± 8.2 years), were more frequently male (62.3% vs. 43.2%), had higher SBP (147.0 ± 15.3 mmHg vs. 139.1 ± 13.7 mmHg), lower HDL-C levels (1.4 ± 0.4 mmol/L vs. 1.5 ± 0.4 mmol/L), and were more frequently current smokers (18.9% vs. 10.0%).

**Table 10. Baseline characteristics of selected participants from the UK Biobank**

| Baseline characteristics | Total (N=47,382) | MACE within 10 years | | |
|---|---|---|---|---|
| | | No (N=45,219) | Yes (N=2,163) | *P*-value |
| Male sex, N (%) | 20,900 (44.1) | 19,552 (43.2) | 1,348 (62.3) | <0.001 |
| Age (years) | 56.4 (8.2) | 56.2 (8.2) | 61.2 (6.9) | <0.001 |
| Current smoker, N (%) | 4,909 (10.4) | 4,500 (10.0) | 409 (18.9) | <0.001 |
| Systolic blood pressure (mmHg) | 139.5 (19.8) | 139.1 (19.6) | 147.0 (20.8) | <0.001 |
| Total cholesterol (mmol/L) | 5.8 (1.1) | 5.8 (1.1) | 5.7 (1.2) | 0.005 |
| HDL cholesterol (mmol/L) | 1.5 (0.4) | 1.5 (0.4) | 1.4 (0.4) | <0.001 |

Values are expressed as mean (standard deviation) for continuous variables, and as N (%) for categorical variables. P-values were calculated using t-tests for continuous variables and $\chi^2$ tests for categorical variables.

**Abbreviations:** HDL, high-density lipoprotein; MACE, major adverse cardiovascular event.

### 3.5.2 Associations of selected proteins with MACE

A total of 18 proteins were selected to enhance MACE risk prediction in the SCORE2 model within the derivation set using LASSO analysis and bootstrapping. Among these, 5 proteins were identified in both sexes, 7 were specific to males, and 6 were specific to females (**Appendix 44**).

**Figure 14** illustrates the HRs and CIs for each of the selected proteins in males and females. Among the 18 proteins, 14 were significantly associated with MACE in both sexes. Among the remaining 4 proteins,

ADAMTS13 (a disintegrin and metalloproteinase with thrombospondin motifs 13), BCAN (brevican core protein), and CXCL17 (C-C motif chemokine 7) showed significant associations with MACE in males but not in females, and CRYBB2 (Beta-crystallin B2) was significantly associated with MACE only in females. CRYBB2 also showed a statistically significant interaction with sex (*P* for interaction, 0.045). The only other protein with a statistically significant interaction with sex was NTproBNP (*P* for interaction, 0.038), which showed a slightly more pronounced association with MACE in females than in males.

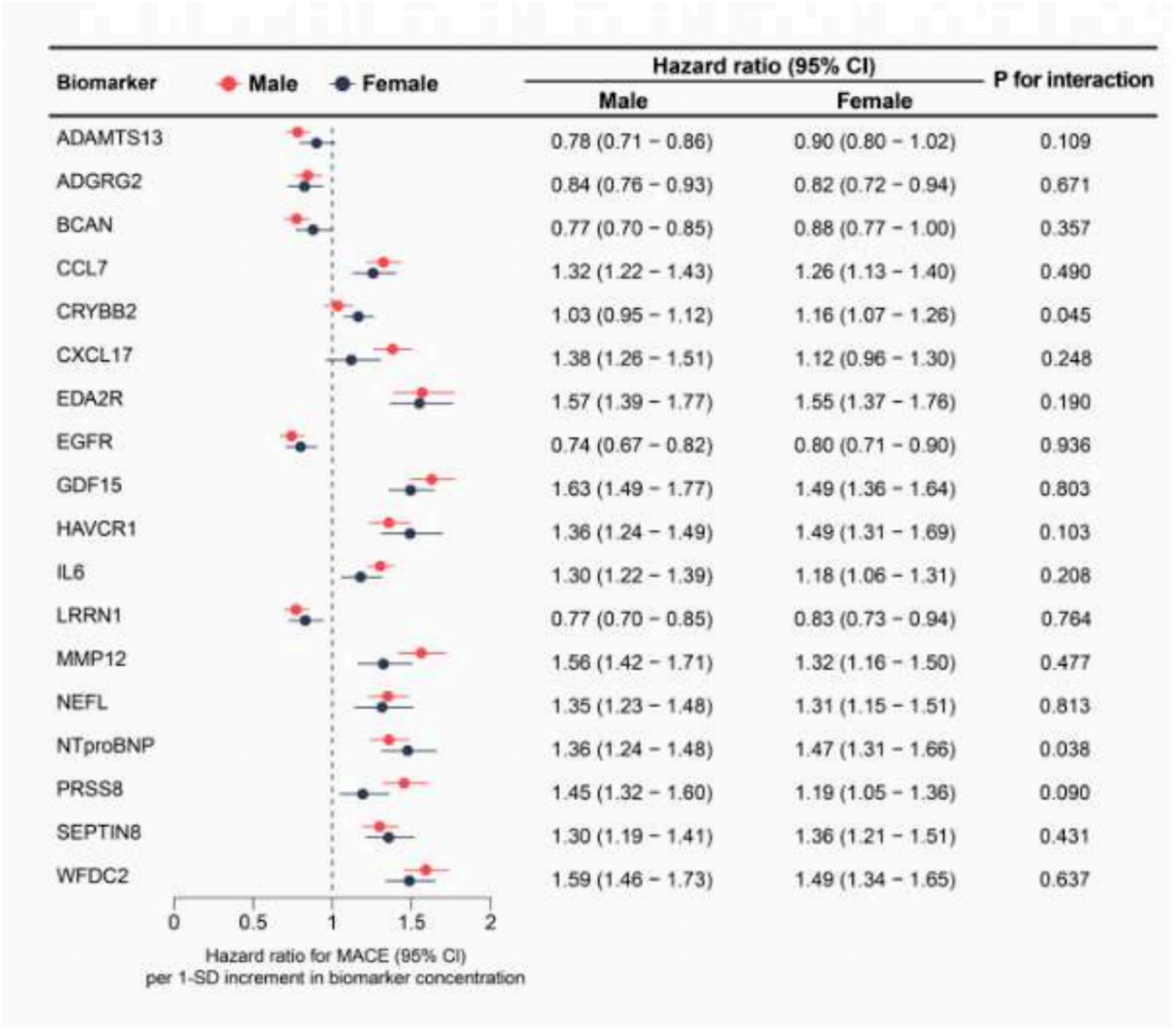



**Figure 14. Associations between selected proteins and major cardiovascular events across sexes in the validation cohort (30% of UK Biobank, N=14,216)**

**Abbreviations:** ADAMTS13, A disintegrin and metalloproteinase with thrombospondin motifs 13; ADGRG2, Adhesion G-protein coupled receptor G2; BCAN, Brevican core protein; CCL7, C-C motif chemokine 7; CI, confidence interval; CRYBB2, Beta-crystallin B2; CXCL17, C-X-C motif chemokine 17; EDA2R, Tumor necrosis factor receptor superfamily member 27; EGFR, Epidermal growth factor receptor; GDF15, Growth/differentiation factor 15; HAVCR1, Hepatitis A virus cellular receptor 1; IL6, Interleukin-6; LRRN1,

Leucine-rich repeat neuronal protein 1; MACE, major adverse cardiovascular events; MMP12, Macrophage metalloelastase; NEFL, Neurofilament light polypeptide; NPX, Normalized Protein eXpresstion; NTproBNP, N-terminal prohormone of brain natriuretic peptide; PRSS8, Prostasin; SEPTIN8, Septin-8; WFDC2, WAP four-disulfide core domain protein 2.

### 3.5.3 MACE risk prediction by proteomic biomarkers

The ß-coefficients for all variables in the extended risk model are detailed in **Appendix 44**. **Table 11** presents the predictive performance metrics of the SCORE2 model for 10-year MACE risk prediction, extended by the selected 18 proteins. In the derivation set, the integration of these proteins significantly improved the C-index of the SCORE2 model from 0.713 (95% CI: 0.701, 0.725) to 0.778 (95% CI: 0.767, 0.790) for the total population. This enhancement was observed in both sexes, with the C-index rising statistically significantly from 0.682 (95% CI: 0.666, 0.699) to 0.763 (95% CI: 0.747, 0.779) in males, and from 0.718 (95% CI: 0.699, 0.738) to 0.780 (95% CI: 0.751, 0.798) in females.

In the validation set, these findings were confirmed, showing similar improvements in the C-index with the C-index increasing from 0.716 (0.698, 0.734) to 0.778 (0.761,0.796) in the total population, from 0.684 (0.659, 0.705) to 0.771 (0.748, 0.795) in males, and from 0.720 (0.690, 0.751) to 0.769 (0.740,0.798) in females ($P$<0.001). **Appendix 45** illustrates the ROC curves comparing the SCORE2 model with and without the inclusion of the proteomic data in the validation set, showing an improved discrimination performance with the addition of the proteins especially in males. In sensitivity analysis, excluding the non-randomly selected participants, the predictive performance of both the SCORE2 model and the proteins-extended model remained essentially unchanged in the total sample (**Appendix 46**).

**Figure 15** illustrates the incremental improvement in the C-statistic achieved by incorporating each of the selected proteomic biomarkers. In males, all 12 selected proteins significantly improved the discriminative ability of the SCORE2 model. Notably, GDF15 (Growth/differentiation factor 15), WFDC2 (WAP four-disulfide core domain protein 2), MMP12 (Macrophage metalloelastase), CXCL17, and IL6 (Interleukin-6) increased the C-index by more than 0.03 index points. In females, out of the 11 selected proteins, 6 significantly improved the SCORE2 model's discriminative ability, while the remaining 5 proteins also contributed to model improvement, albeit without reaching statistical

significance. This difference may be attributable to the already high baseline predictive ability of the SCORE2 model in females (0.720 compared to 0.684 in males). In agreement with the results observed in males, GDF15 and WFDC2 were the two most impactful biomarkers in females, with GDF15 increasing the C-index by 0.026 and WFDC2 by 0.035 index points.

Additionally, adding the proteomic signatures to the SCORE2 model demonstrated significant gains in reclassification statistics, with a notable increase in the NRI for the total population, reaching 19.9% (95% CI: 14.7, 25.2). The NRI was particularly high in males with 36.3% (95% CI: 25.3, 57.1) but the NRI in females was notable as well (18.2% (95% CI: 10.2, 26.5)). In males, more non-events were correctly re-classified than events. In females, it was the opposite, with more events correctly re-classified than non-events. Taken together in the total population, the correct re-classification outweighs the incorrect re-classifications in both events and non-events. This is supported by statistically significant IDIs for the total population, males, and females. **Appendix 47** shows the re-classification in more detail, with re-classification tables for low-risk (≤5%), intermediate risk (5-10%) and high-risk (>10%) categories with respect to MACE.

Calibration curves of both the original and extended SCORE2 models in the internal validation set are depicted in **Appendix 48.** Both are very well calibrated.

**Table 11. Metrics of the predictive performance of the SCORE2 model for 10-year MACE risk without and with extension by proteins**

| Metrics | Male * | Female † | Overall |
|---|---|---|---|
| **Derivation set (70% of UK Biobank)** | | | |
| **Total sample size (N=33,166) / MACE case number (N=1,492)** | | | |
| C-Statistics (SCORE2) | 0.682 (0.666,0.699) | 0.718 (0.699,0.738) | 0.713 (0.701,0.725) |
| C-Statistics (+Proteins) | 0.763 (0.747,0.779) | 0.780 (0.751, 0.798) | 0.778 (0.767,0.790) |
| *P*-values for C-Statistics comparisons | **<0.001** | **<0.001** | **<0.001** |
| **Validation set (30% of UK Biobank)** | | | |
| **Total sample size (N=14,216) / MACE case number (N=671)** | | | |
| C-Statistics (SCORE2) | 0.684 (0.659, 0.705) | 0.720 (0.690, 0.751) | 0.716 (0.698, 0.734) |
| C-Statistics (+Proteins) | 0.771 (0.748,0.795) | 0.769 (0.740,0.798) | 0.778 (0.761,0.796) |
| *P*-values for C-Statistics comparisons | **<0.001** | **<0.001** | **<0.001** |
| NRI total (%) | **36.3 (25.3, 57.1)** | **18.2 (10.2, 26.5)** | **19.9 (14.7, 25.2)** |
| NRI events (%) | -3.6 (-9.2, 5.2) | **19.8 (12.6, 27.5)** | **13.4 (6.2, 17.2)** |
| NRI non-events (%) | **39.9 (26.3, 62.0)** | **-1.6 (-2.7, -0.7)** | **6.5 (4.0, 11.0)** |
| IDI | **0.058 (0.040, 0.075)** | **0.031 (0.016, 0.045)** | **0.045 (0.033, 0.058)** |

**Abbreviations:** IDI, integrated discrimination index; MACE, major cardiovascular events; NRI, net reclassification index.

* Proteins included in the combined model for males were: ADAMTS13, BCAN, CCL7, CXCL17, GDF15, IL6, LRRN1, MMP12, NEFL, NTproBNP, PRSS8, WFDC2.

† Proteins included in the combined model for females were: ADGRG2, BCAN, CRYBB2, EDA2R, EGFR, GDF15, HAVCR1, MMP12, NTproBNP, SEPTIN8, WFDC2.

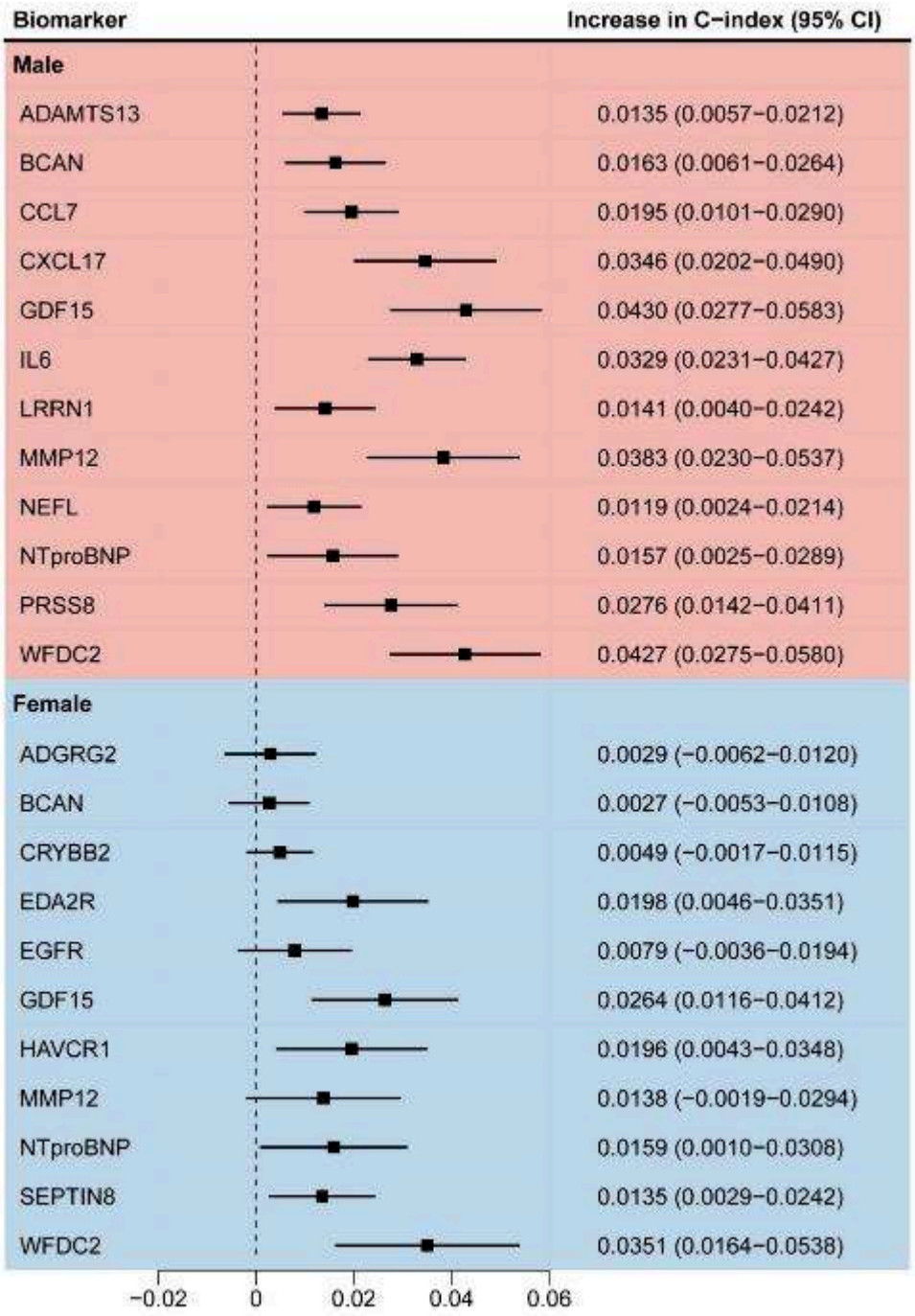



**Figure 15. Contribution of individual proteomic biomarkers to the C-index in MACE risk prediction stratified by sex in the validation cohort (30% of UK Biobank, N=14,216)**

Contribution of each selected proteomic biomarker to the improvement in the C-index, stratified by sex, for predicting 10-year MACE.
**Abbreviations:** see legend of Figure 1.

## 3.6 Inflammation-related proteins for cardiovascular risk prediction

### 3.6.1 Description of the population included

**Table 12** summarizes the baseline characteristics of the 47,382 UKB participants and 6,280 ESTHER participants included in the analysis. The average age of UKB participants was 56.4 years (SD 8.2), with 44.1% being male. Over the 10-year follow-up period, 2,163 MACE and 1,713 competing non-cardiovascular deaths were recorded. In the ESTHER study, the mean age was 61.7 years (SD 6.6), and 41.9% of participants were male, with a total of 602 MACE and 399 competing non-CVD deaths documented during the same follow-up period.

**Table 12. Baseline characteristics of the analyzed individuals from the UK Biobank and the ESTHER study**

| **Baseline characteristics** | UK Biobank (N=47,382) | ESTHER (N=6,280) |
|---|---|---|
| Male sex, N (%) | 20,900 (44.1) | 2,634 (41.9) |
| Age (years), mean (SD) | 56.4 (8.2) | 61.7 (6.6) |
| Current smoker, N (%) | 4,909 (10.4) | 1,030 (16.4) |
| Systolic blood pressure (mmHg), mean (SD) | 139.5 (19.8) | 139.2 (19.1) |
| Total cholesterol (mmol/L), mean (SD) | 5.8 (1.1) | 5.9 (1.2) |
| HDL cholesterol (mmol/L), mean (SD) | 1.5 (0.4) | 1.4 (0.4) |

**Abbreviations:** HDL, high-density lipoprotein; SD, standard deviation.

### 3.6.2 Biomarker Selection and Model Derivation

A total of 7 proteins were selected to enhance MACE risk prediction in the SCORE2 model within the derivation cohort using LASSO analysis and bootstrapping. CRP was not selected as a predictor in the final models. Among the 7 selected proteins, 2 were identified in both sexes, 3 were specific for males, and 2 were specific for females.

**Figure 16** presents the Spearman correlation coefficient (r) matrices for CRP and the 5 selected inflammation-related proteins in males and the 4 selected ones in females. Most of the selected proteins were moderately correlated with each other and with CRP with correlation coefficients in the range from r=0.2-0.49 with a few exceptions of high correlations with correlation coefficients between r=0.5 and

r=0.62. While correlation of all selected proteins with CRP was low among females (all r≤0.29), there was a high correlation between CRP and IL6 among males (r=0.53).

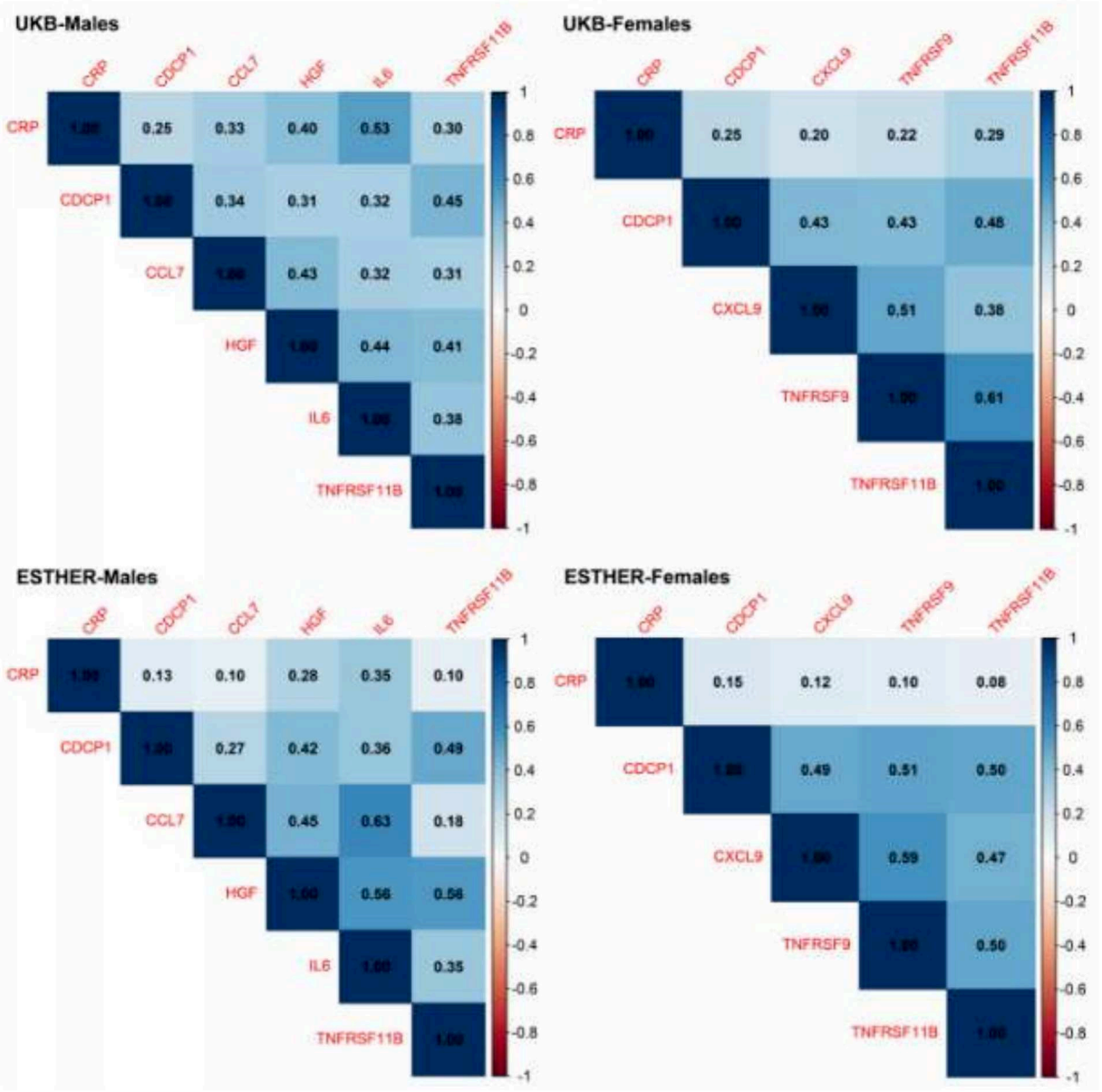



**Figure 16. Spearman correlation coefficients of C-reactive protein and the selected inflammation-related proteins stratified by sex in the UK Biobank and ESTHER study**

**Abbreviations:** CDCP1, CUB domain-containing protein 1; CCL7, C-C motif chemokine 7; CXCL9, C-X-C motif chemokine 9; HGF, Hepatocyte growth factor; IL6, Interleukin-6; TNFRSF9, Tumor necrosis factor receptor superfamily member 9; TNFRSF11, Tumor necrosis factor receptor superfamily member 11.

### 3.6.3 Associations of Selected Proteins with MACE

**Figure 17** provides HRs with 95% CIs for the associations between each selected protein and MACE in males and females in both cohorts. In the UKB, all selected proteins showed significant positive

associations with MACE across sexes. However, none of these associations were observed in the ESTHER study. The observed differences may be partially explained by the fasting status of participants, with non-fasting samples comprising 95.7% of the UKB but only 10.7% of the ESTHER cohort. **Appendix 49** highlights substantial differences in mean protein concentrations for most of the selected proteins between fasting and non-fasting participants in the ESTHER study.

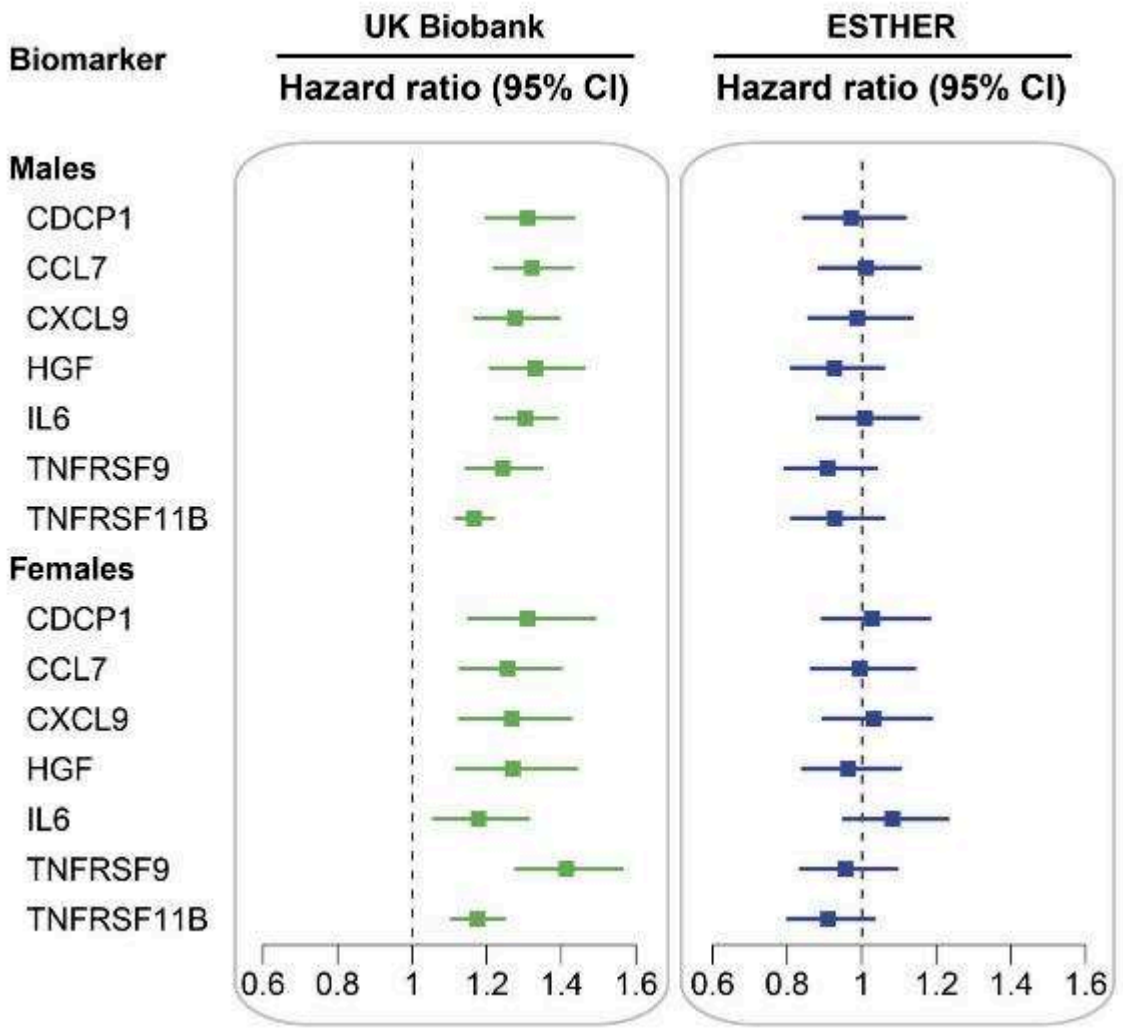



**Figure 17. Associations between selected proteins and MACE across sexes in the internal validation (30% of UK Biobank, N=14,216) and external validation (70% of ESTHER, N=4,397)**

**Abbreviations:** As depicted in **Figure 16**.

### 3.6.4 Model Performance Validation

Despite differences in the HRs, the predictive value of the selected proteins observed in the UKB was validated in the ESTHER study. **Figure 18** illustrates the incremental contributions of each protein to

the C-index in the internal set (UKB) and external validation set (ESTHER). Among males, all selected proteins improved discrimination significantly in the UKB. In the ESTHER study, the increases in the C-index were of similar magnitude but only HGF and TNFRSF11B (Tumor necrosis factor receptor superfamily member11B) reached statistical significance. For females, all selected proteins except CXCL9 (C-X-C motif chemokine 9) significantly improved the C-index in the UKB. In contrast, CXCL9 was the only protein with a statistically significant C-index improvement among female participants of the ESTHER study. Overall, the results in UKB and ESTHER were comparable for females as well.

**Appendix 50-51** show the ß coefficients of the variables of the final prediction model for males and females (SCORE2 variables extended by all selected proteins) in the UK Biobank and ESTHER study, respectively. **Table 13** presents the predictive performance metrics of the original SCORE2 model, the refitted SCORE2 model, and the protein-extended models. In the derivation set, integrating inflammation-related proteins increased the C-index by 0.030 in the total population, with improvements observed in both sexes (+0.044 in males and +0.022 in females). Similar improvements were seen in the internal validation set (+0.034, +0.050, and +0.022 in the total population, males, and females, respectively. External validation in the ESTHER study confirmed the enhanced performance, with the C-index increasing by 0.029 in the total population and by 0.033 in both males and females (**Appendix 52**).

Integration of the selected proteins also enhanced reclassification metrics. In internal validation, the NRI for the total population was 12.4%, with notable improvements among females (NRI: 8.9%) and males (NRI: 12.8%). In external validation, the NRI in the total population was 3.0%, with a much higher NRI observed among females (25.6%) than among males (1.5%).

Calibration curves (**Appendix 53**) demonstrated good alignment between observed and predicted MACE rates for both the refitted SCORE2 and the model extended by proteins in the internal validation and in external validation for females. In males and the total population of the external validation set, the calibration was good for an up to 15% predicted MACE risk and poor for higher predicted risks for both the refitted SCORE2 model and the protein-extended model.

**Table 13. Metrics of the predictive performance of the SCORE2 model for 10-year MACE risk without and with extension by proteins**

| Metrics | Males [a] | Females [b] | Overall |
|---|---|---|---|
| **Derivation set (70% of UK Biobank)** | | | |
| **Total sample size (N=33,166) / MACE case number (N=1,492)** | | | |
| C-Statistics (SCORE2) | 0.678 (0.662, 0.695) | 0.717 (0.698, 0.737) | 0.712 (0.700, 0.724) |
| C-Statistics (Refitted SCORE2) | 0.682 (0.666, 0.699) | 0.718 (0.699, 0.738) | 0.713 (0.701, 0.725) |
| C-Statistics (+Proteins) | 0.726 (0.710, 0.742) | 0.740 (0.721, 0.759) | 0.743 (0.732, 0.755) |
| *P*-values for C-Statistics comparisons[c] | **<0.001** | **<0.001** | **<0.001** |
| **Internal validation set (30% of UK Biobank)** | | | |
| **Total sample size (N=14,216) / MACE case number (N=671)** | | | |
| C-Statistics (SCORE2) | 0.683 (0.658, 0.707) | 0.713 (0.683, 0.743) | 0.711 (0.693, 0.729) |
| C-Statistics (Refitted SCORE2) | 0.684 (0.659, 0.709) | 0.720 (0.690, 0.751) | 0.716 (0.698, 0.734) |
| C-Statistics (+Proteins) | 0.734 (0.710, 0.759) | 0.742 (0.712, 0.772) | 0.750 (0.732, 0.768) |
| *P*-values for C-Statistics comparisons[c] | **<0.001** | **0.001** | **<0.001** |
| NRI categorical total (%)[d] | 12.8 (-6.8, 23.1) | **8.9 (1.7, 15.5)** | **12.4 (5.2, 16.3)** |
| NRI categorical events (%)[d] | 4.0 (-2.7, 10.2) | **8.5 (1.2, 15.2)** | 6.7 (-1.8, 10.3) |
| NRI categorical non-events (%)[d] | 8.8 (-14.2, 22.8) | 0.4 (-0.3, 13.8) | **5.7 (3.6, 8.6)** |
| IDI | **0.025 (0.017, 0.033)** | **0.011 (0.002, 0.021)** | **0.019 (0.012, 0.028)** |
| **External validation set (70% of ESTHER)** | | | |
| **Total sample size (N=4,397) / MACE case number (N=434)** | | | |
| C-Statistics (SCORE2) | 0.627 (0.589, 0.665) | 0.714 (0.679, 0.749) | 0.678 (0.653, 0.704) |
| C-Statistics (Refitted SCORE2) | 0.606 (0.563, 0.649) | 0.712 (0.674, 0.749) | 0.674 (0.647, 0.701) |
| C-Statistics (+Proteins) | 0.639 (0.597, 0.682) | 0.745 (0.707, 0.782) | 0.703 (0.675, 0.730) |
| *P*-values for C-Statistics comparisons[c] | **0.023** | **<0.001** | **<0.001** |
| NRI categorical total (%)[e] | 1.5 (-2.9, 6.7) | **25.6 (1.0, 41.3)** | 3.0 (-0.6, 23.4) |
| NRI categorical events (%)[e] | **4.6 (0.5, 9.2)** | 0.4 (-13.8, 9.2) | 4.8 (-2.0, 8.4) |
| NRI categorical non-events (%)[e] | **-3.0 (-5.0, -1.0)** | 25.1 (-2.0, 46.2) | -1.8 (-2.6, 25.4) |
| IDI | 0.003 (-0.007, 0.014) | 0.012 (-0.005, 0.031) | 0.006 (-0.004, 0.014) |

**Abbreviations:** IDI, integrated discrimination improvement; NRI, net reclassification improvement.

[a] Proteins that were included for males were CCL7, CDCP1, HGF, IL6, and TNFRSF11B.

[b] Proteins that were included for females were CDCP1, CXCL9, TNFRSF9, and TNFRSF11B.

[c] P-values compare the refitted SCORE2 model with the SCORE2 model extended by proteins.

[d] NRI calculated with the pre-specified 10-year MACE risk categories of 0–5%, >5–10%, and >10%.

[e] NRI calculated with the pre-specified 10-year MACE risk categories of 0–7.5%, >7.5–15%, and >15%.

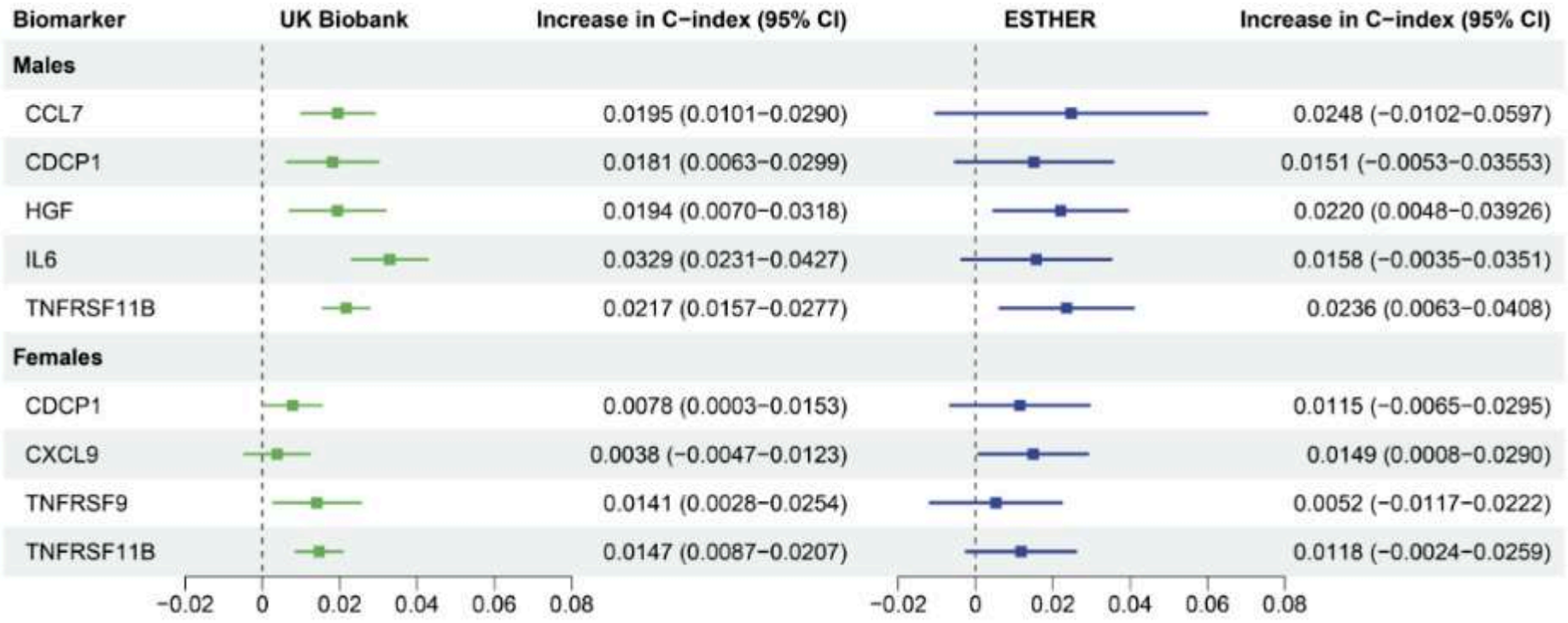



**Figure 18. The incremental discrimination of each protein for the model after the selected proteins were added to SCORE2 separately in different sexes in the internal validation (30% of UK Biobank, N=14,216) and external validation (70% of ESTHER, N=4,397)**

**Abbreviations:** CDCP1, CUB domain-containing protein 1; CCL7, C-C motif chemokine 7; CXCL9, C-X-C motif chemokine 9; HGF, Hepatocyte growth factor; IL6, Interleukin-6; TNFRSF9, Tumor necrosis factor receptor superfamily member 9; TNFRSF11, Tumor necrosis factor receptor superfamily member 11.

### 3.6.5 Sensitivity analyses

We conducted several sensitivity analyses to evaluate the robustness of our findings. First, excluding non-randomly selected participants from both cohorts, demonstrated a slight improvement in the C-index for both the SCORE2 model and the proteins extended model in the total population **(Appendix 54).** Consistent with the main analysis, the protein extended model outperformed the C-index of the refitted SCORE2 model in both the internal and external validation set.

Second, we compared the added predictive value of inflammation-related proteins and cardiac-specific biomarkers (**Appendix 55**). Both cardiac and inflammation-based proteins statistically significantly improved the C-index of the refitted SCORE2 when added to the refitted SCORE2 (all $P > 0.05$). However, the addition of the inflammation-related proteins added more to the C-index than the cardiac biomarkers. While the model including both cardiac biomarkers and inflammation-related proteins showed the highest C-index, the difference compared to the protein-extended model alone was not statistically significant in neither men, women, nor the total population (all $P > 0.05$).

Third, we investigated whether statin use altered the results. At baseline, 11.2% of UKB participants were statin users. After adjustment for SCORE2 variables, all selected proteins except TNFRSF9 were significantly increased among statin users compared to non-users (**Appendix 56**). While C-indices were generally lower among statin users for all models, the inflammation-related protein-extended models consistently outperformed the refitted SCORE2 models in both statin users and non-users (**Appendix 57**).

## 3.7 Proteomics and Multi-omics for cardiovascular risk prediction in individuals with type 2 diabetes

### 3.7.1 Description of the population included

**Table 14** summarizes the baseline characteristics of the 1,751 UKB participants with T2D included in the proteomics-based analysis cohort and the 990 participants in the multi-omics cohort. In the proteomics cohort, the mean age was 60.3 ± 6.6 years and 59.7% of participants were male. The multi-omics cohort was similar, with a mean age of 60.5 ± 6.6 years and 60.3% male participants. The distributions of smoking status, SBP, TC, HDL-C, age at diabetes diagnosis, $HbA_{1c}$, and eGFR were also comparable between the two cohorts, with all *P*-values > 0.05. During the 10-year follow-up, 270 participants (15.4%) in the proteomics cohort experienced a MACE, including 193 males and 77 females. In the multi-omics cohort, 159 participants (16.1%) developed MACE, among whom 114 were males and 44 females.

**Table 14. Baseline characteristics of participants included in the proteomics and multi-omics analyses from the UK Biobank**

| Baseline characteristics | Proteomics cohort (N=1,751) | Multi-omics cohort (N=990) | *P*-value |
|---|---|---|---|
| Male sex, N (%) | 1,045 (59.7%) | 597 (60.3%) | 0.796 |
| Age (years) | 60.3 (6.6) | 60.5 (6.6) | 0.364 |
| Current smoker, N (%) | 187 (10.7%) | 103 (10.4%) | 0.879 |
| Systolic blood pressure (mmHg) | 144.5 (17.7) | 144.3 (17.6) | 0.757 |
| Total cholesterol (mmol/L) | 4.5 (1.0) | 4.5 (1.0) | 0.790 |
| HDL cholesterol (mmol/L) | 1.2 (0.3) | 1.2 (0.3) | 0.671 |
| Age at diabetes diagnosis (years) | 55.0 (6.5) | 55.2 (6.4) | 0.335 |
| $HbA_{1c}$ (mmol/mol) | 51.8 (12.9) | 51.7 (12.4) | 0.833 |
| eGFR (ml/min/1.73m$^2$)[a] | 88.3 (16.1) | 87.4 (16.4) | 0.176 |

[a]The eGFR was calculated with the CKD-EPI 2009 equation

Values are expressed as mean (standard deviation) for continuous variables, and as N (%) for categorical variables. *P*-values were calculated using t-tests for continuous variables and $\chi^2$ tests for categorical variables.

**Abbreviations:** eGFR, estimated Glomerular Filtration Rate; $HbA_{1c}$, glycated hemoglobin; HDL, high-density lipoprotein; MACE, major adverse cardiovascular event.

### 3.7.2 Predictive performance of proteomics and multi-omics extended SCORE2-Diabetes models

**Appendix 58-59** present β-coefficients for variables included in the proteomics-extended and multi-omics-extended SCORE2-Diabetes models, respectively. The incorporation of proteomic biomarkers substantially enhanced the predictive performance of the SCORE2-Diabetes model (**Table 15** and **Appendix 60**). In the total population, the C-index increased from 0.725 (95% CI: 0.697–0.757) to 0.810 (95% CI: 0.782–0.837; ΔC-index = +0.085; $P < 0.001$). In males, the increase was from 0.700 to 0.797 (ΔC-index = +0.097; $P < 0.001$), and in females, from 0.740 to 0.835 (ΔC-index = +0.095; $P < 0.001$). Correspondingly, the total NRI was 24.2% (95% CI: 15.9%, 33.1%) overall, with 51.2% (95% CI: 18.3%, 67.9%) in males and 29.5% (95% CI: 18.4%, 42.1%) in females. The IDI values were also notably improved (0.098 overall, 0.105 in males, and 0.111 in females).

The predictive value was further enhanced by integrating additional omics layers beyond proteomics alone (**Figure 19** and **Appendix 61**). In the total population, integrating all multi-omic biomarkers into the SCORE2-Diabetes model substantially enhanced discrimination, increasing the C-index from 0.766 to 0.848 (ΔC-index = +0.082; $P < 0.001$), with a corresponding NRI of 20.2% (95% CI: 14.3%, 38.4%). In males, the C-index rose from 0.746 to 0.838 (ΔC-index = +0.092; $P < 0.001$), accompanied by a notable NRI of 53.3% (95% CI: 22.0%, 87.2%). In females, the improvement was also significant, with the C-index increasing from 0.784 to 0.860 (ΔC-index = +0.056; $P < 0.001$) and an NRI of 18.9% (95% CI: 1.3%, 37.3%). Among individual omics layers, proteomics yielded the greatest performance gain across all subgroups, followed by metabolomics, while PRS contributed only modest improvements.

**Appendix 62-63** provide reclassification tables, confirming improvements in risk categorization with the addition of proteomics and multi-omics biomarkers, respectively. Calibration curves (**Appendix 64-65**) illustrated good agreement between observed and predicted risks, highlighting the reliable performance of the extended models.

**Figure 20** presents the incremental contributions of individual biomarkers to model performance in males and females. In males, GAST conferred the largest individual increase in the C-index (ΔC-index

= +0.0317), while in females, CTRC (Chymotrypsin-C) was the most influential biomarker (ΔC-index = +0.0389). In males, these included 9 proteins and 1 metabolite; in females, 6 proteins, 3 metabolites, and the CVD-PRS. This pattern further underscores the high predictive ability of proteomic biomarkers.

**Figure 21** evaluates which combination of omics data yields the best model performance for 10-year MACE prediction. Among all single-omics extensions to the SCORE2-Diabetes model, proteomics led to the largest improvement in the total population (C-index: 0.835 vs. 0.766; $P < 0.001$), in males (0.823 vs. 0.746; $P = 0.001$), and in females (0.855 vs. 0.784; $P = 0.027$), and was therefore selected as the reference model for subsequent comparisons. Using the proteomics-extended SCORE2-Diabetes model as the reference, the addition of the CVD-PRS did not significantly improve model performance in the total population (C-index: 0.838 vs. 0.835; $P = 0.080$), in males (0.828 vs. 0.823; $P = 0.078$), or in females (0.857 vs. 0.855; $P = 0.236$). The inclusion of metabolomic biomarkers, however, led to a significant improvement in the total population (C-index: 0.846 vs. 0.835; $P = 0.035$). Further incorporation of the CVD-PRS into the model combining SCORE2-Diabetes variables, proteomics, and metabolomics did not significantly enhance predictive performance in the total population (C-index: 0.848 vs. 0.846; $P = 0.070$), although a small but statistically significant increase was observed in males (ΔC-index = +0.003; $P = 0.047$).

**Table 15. Metrics of the predictive performance of the SCORE2-Diabetes model for 10-year MACE risk without and with extension by proteins (N=1,751)**

| Metrics | Male [a] | Female [b] | Overall |
|---|---|---|---|
| C-Statistics (SCORE2-Diabetes) | 0.700 (0.659, 0.742) | 0.740 (0.682, 0.799) | 0.725 (0.697, 0.757) |
| C-Statistics (+Proteins) | 0.797 (0.760, 0.833) | 0.835 (0.789, 0.881) | 0.810 (0.782, 0.837) |
| *P*-values for C-Statistics comparisons | **<0.001** | **<0.001** | **<0.001** |
| NRI total (%) | **51.2 (18.3, 67.9)** | **29.5 (18.4, 42.1)** | **24.2 (15.9, 33.1)** |
| NRI events (%) | -12.6 (-17.1, 6.8) | **34.9 (24.5, 47.1)** | 19.1 (-2.1, 17.6) |
| NRI non-events (%) | **63.9 (18.6, 78.1)** | **-5.4 (-7.6, -2.4)** | 5.1 (-1.6, 28.3) |
| IDI | **0.105 (0.058, 0.146)** | **0.111 (0.028, 0.110)** | **0.098 (0.052, 0.130)** |

[a] Proteins included in the combined model for males were: ACTA2, CHGA, CRNN, ENPP5, GAST, NT-proBNP, POSTN, TRIM21, WFDC2.

[b] Proteins included in the combined model for females were: CTRC, EFHD1, HS3ST3B1, IL22, SMOC1, TNC, WFDC2.

**Abbreviations:** IDI, integrated discrimination index; MACE, major cardiovascular events; NRI, net reclassification index.

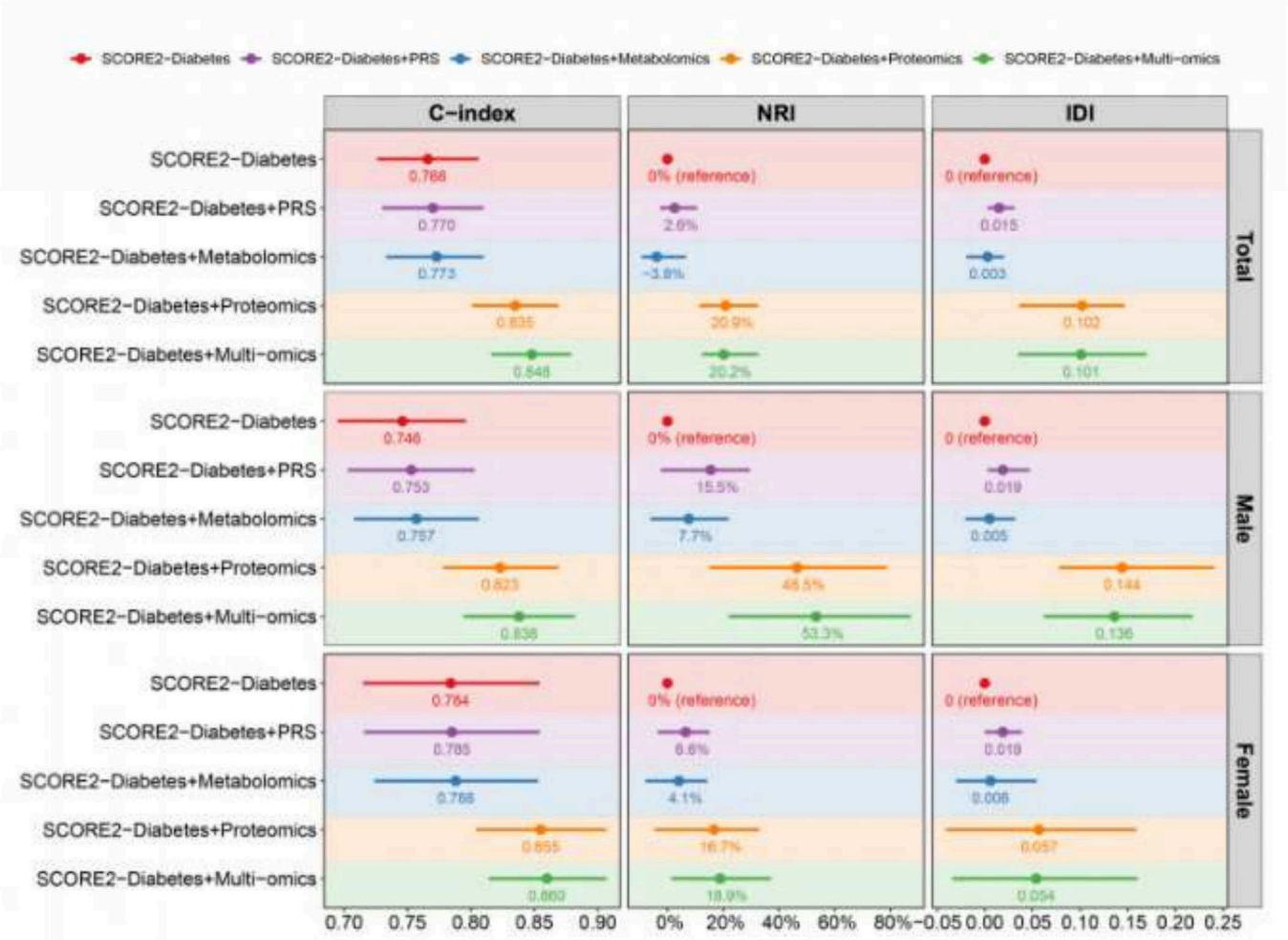



**Figure 19. Predictive performance of the SCORE2-Diabetes model extended with multi-omics biomarkers for 10-year MACE risk (N=990)**

The multi-omics model includes proteomics, metabolomics, and PRS. Performance was evaluated using the C-index, NRI, and IDI.

**Abbreviations:** IDI, Integrated discrimination index; NRI, Net reclassification index.

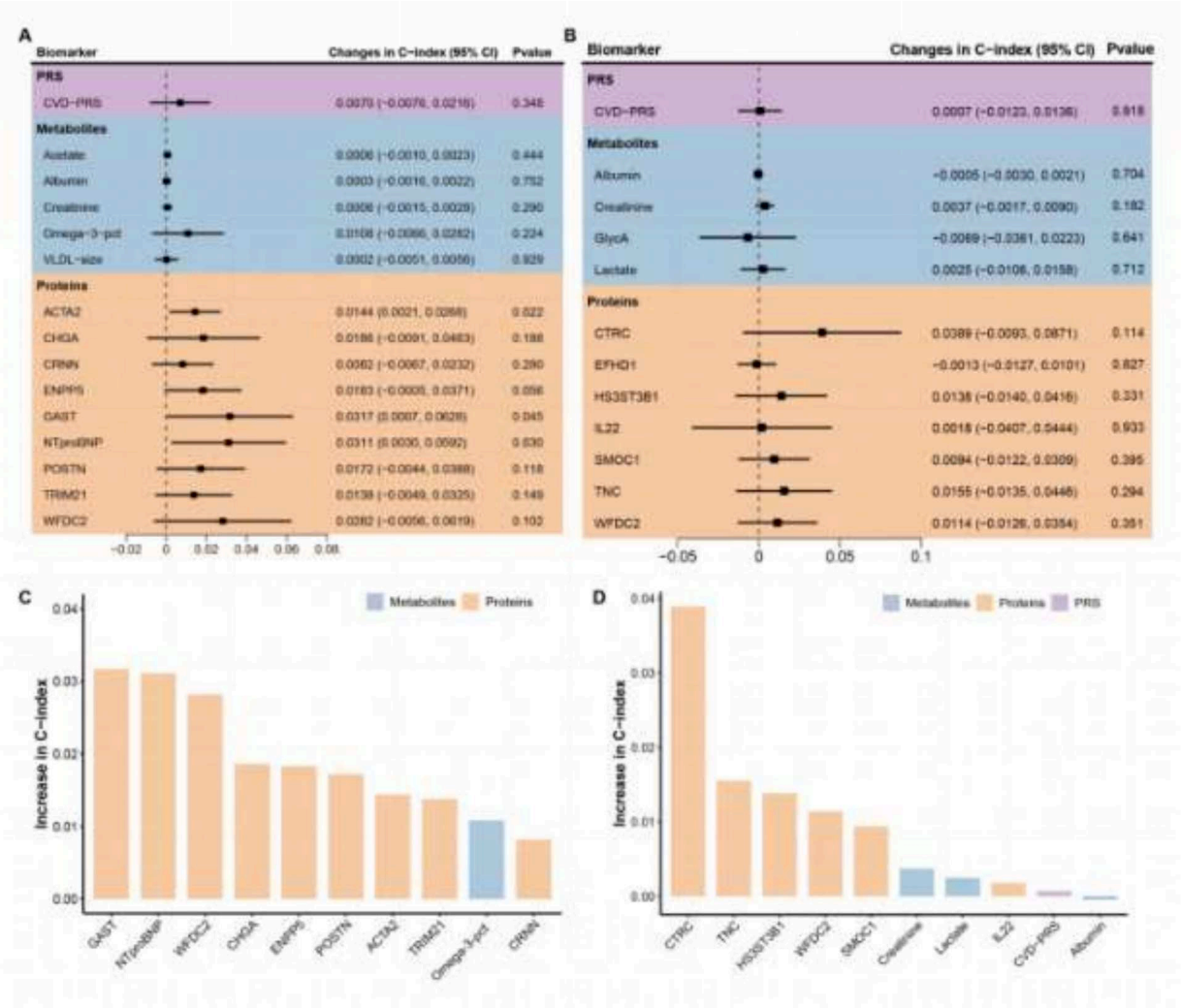
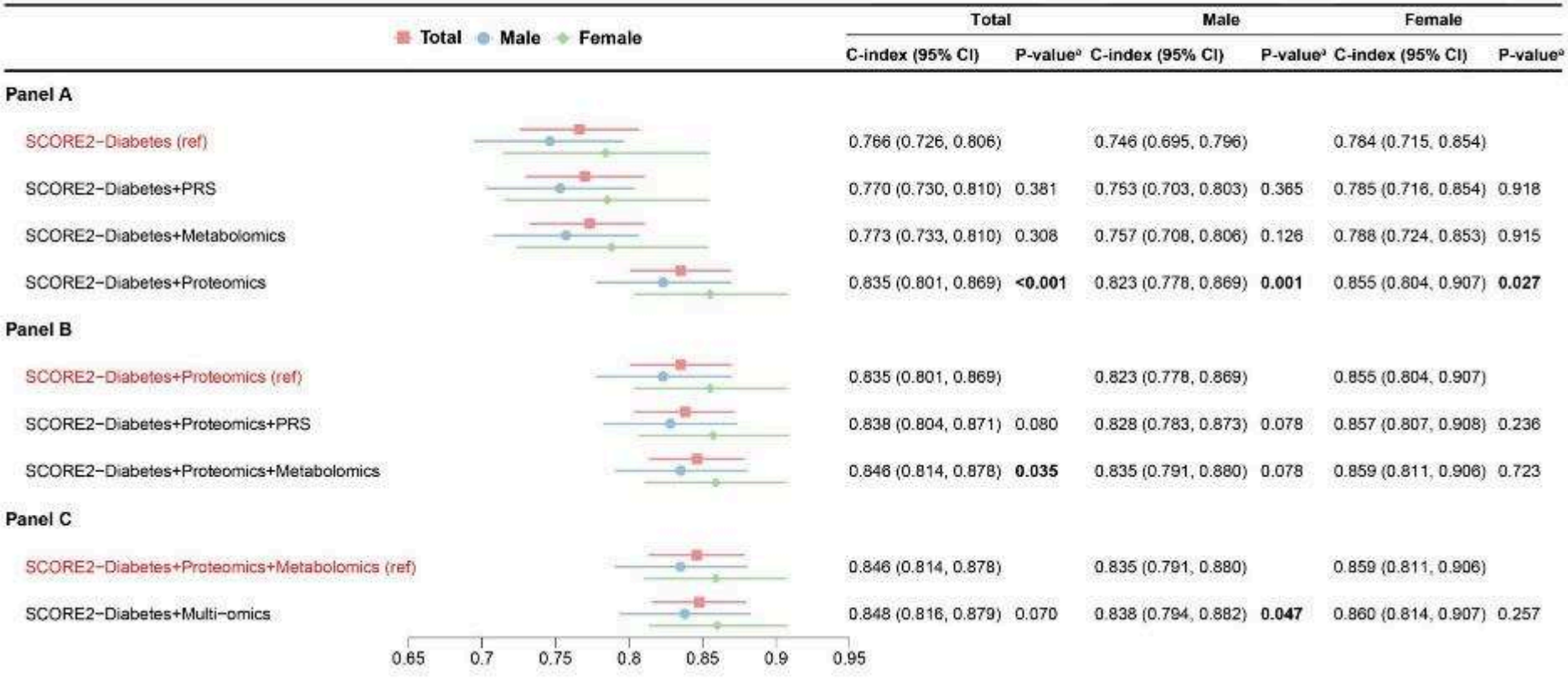



**Figure 20. Incremental predictive value of selected multi-omic biomarkers added to the SCORE2-Diabetes model for 10-year MACE risk prediction (N=990)**

**Panel A** and **B** show the changes in C-index when each biomarker is individually added to the SCORE2-Diabetes model in male and female participants. **Panel C** and **D** illustrate the top 10 biomarkers ranked by individual C-index gains in males and females.

**Abbreviations:** ACTA2, Actin Alpha 2; CHGA, Chromogranin-A; CRNN, Cornulin; CTRC, Chymotrypsin-C; CVD, cardiovascular disease; EFHD1, EF-hand domain-containing protein D1; ENPP5, Ectonucleotide pyrophosphatase/phosphodiesterase family member 5; GAST, Gastrin; HS3ST3B1, Heparan sulfate glucosamine 3-O-sulfotransferase 3B1; IDI, integrated discrimination improvement; IL22, Interleukin-22; MACE, major adverse cardiovascular event; NRI, net reclassification improvement; NT-proBNP, N-terminal prohormone of brain natriuretic peptide; Omega-3-pct, The Omega-3 fatty acids percentage of total fatty acids; PRS, polygenic risk score; SMOC1, SPARC-related modular calcium-binding protein 1; TNC, Tenascin-C; TRIM21; E3 ubiquitin-protein ligase TRIM21; VLDL-size, Average diameter for very-low-density lipoprotein particles; WFDC2, WAP four-disulfide core domain protein.




| | Total | | Male | | Female | |
|---|---|---|---|---|---|---|
| | C-index (95% CI) | P-value[a] | C-index (95% CI) | P-value[a] | C-index (95% CI) | P-value[a] |
| **Panel A** | | | | | | |
| SCORE2-Diabetes (ref) | 0.766 (0.726, 0.806) | | 0.746 (0.695, 0.796) | | 0.784 (0.715, 0.854) | |
| SCORE2-Diabetes+PRS | 0.770 (0.730, 0.810) | 0.381 | 0.753 (0.703, 0.803) | 0.365 | 0.785 (0.716, 0.854) | 0.918 |
| SCORE2-Diabetes+Metabolomics | 0.773 (0.733, 0.810) | 0.308 | 0.757 (0.708, 0.806) | 0.126 | 0.788 (0.724, 0.853) | 0.915 |
| SCORE2-Diabetes+Proteomics | 0.835 (0.801, 0.869) | **<0.001** | 0.823 (0.778, 0.869) | **0.001** | 0.855 (0.804, 0.907) | **0.027** |
| **Panel B** | | | | | | |
| SCORE2-Diabetes+Proteomics (ref) | 0.835 (0.801, 0.869) | | 0.823 (0.778, 0.869) | | 0.855 (0.804, 0.907) | |
| SCORE2-Diabetes+Proteomics+PRS | 0.838 (0.804, 0.871) | 0.080 | 0.828 (0.783, 0.873) | 0.078 | 0.857 (0.807, 0.908) | 0.236 |
| SCORE2-Diabetes+Proteomics+Metabolomics | 0.846 (0.814, 0.878) | **0.035** | 0.835 (0.791, 0.880) | 0.078 | 0.859 (0.811, 0.906) | 0.723 |
| **Panel C** | | | | | | |
| SCORE2-Diabetes+Proteomics+Metabolomics (ref) | 0.846 (0.814, 0.878) | | 0.835 (0.791, 0.880) | | 0.859 (0.811, 0.906) | |
| SCORE2-Diabetes+Multi-omics | 0.848 (0.816, 0.879) | 0.070 | 0.838 (0.794, 0.882) | **0.047** | 0.860 (0.814, 0.907) | 0.257 |

**Figure 19. Discrimination performance of single-omics and multi-omics extensions to the SCORE2-Diabetes in 10-year MACE risk prediction (N=990)**

[a] *P*-values indicate statistical comparisons of C-indices against the corresponding reference model.

**Panel A** shows the C-index with 95% confidence intervals for models extending SCORE2-Diabetes with individual omics layers (PRS, metabolomics, or proteomics). **Panel B** compares multi-omics combinations using SCORE2-Diabetes+proteomics as the reference. **Panel C** contrasts the full multi-omics model with the SCORE2-Diabetes + proteomics + metabolomics model as the reference.
**Abbreviations:** CI, confidence interval; PRS, polygenic risk score; MACE, major adverse cardiovascular event.

### 3.7.3 Correlations among selected biomarkers and associations with MACE

**Figure 22** displays the Spearman correlation matrices for the selected proteins, metabolites, and the CVD-PRS in males and females, respectively. Correlations between biomarkers across omics layers were weak overall. No correlation coefficient exceeded 0.1 between the CVD-PRS and any selected protein or metabolite in either sex. Among male participants, a strong correlation was observed between CHGA (Chromogranin-A) and GAST (Gastrin) (r=0.64). In females, no pair of selected biomarkers exhibited a correlation coefficient ≥ 0.5.

**Figure 23** shows the associations between selected multi-omics biomarkers and 10-year MACE risk in male and female participants, respectively. In males, most selected proteins were significantly associated with MACE, whereas the PRS and the majority of metabolites were not. Similarly, in females, all selected proteins showed significant associations, while neither the PRS nor any selected metabolites reached statistical significance.

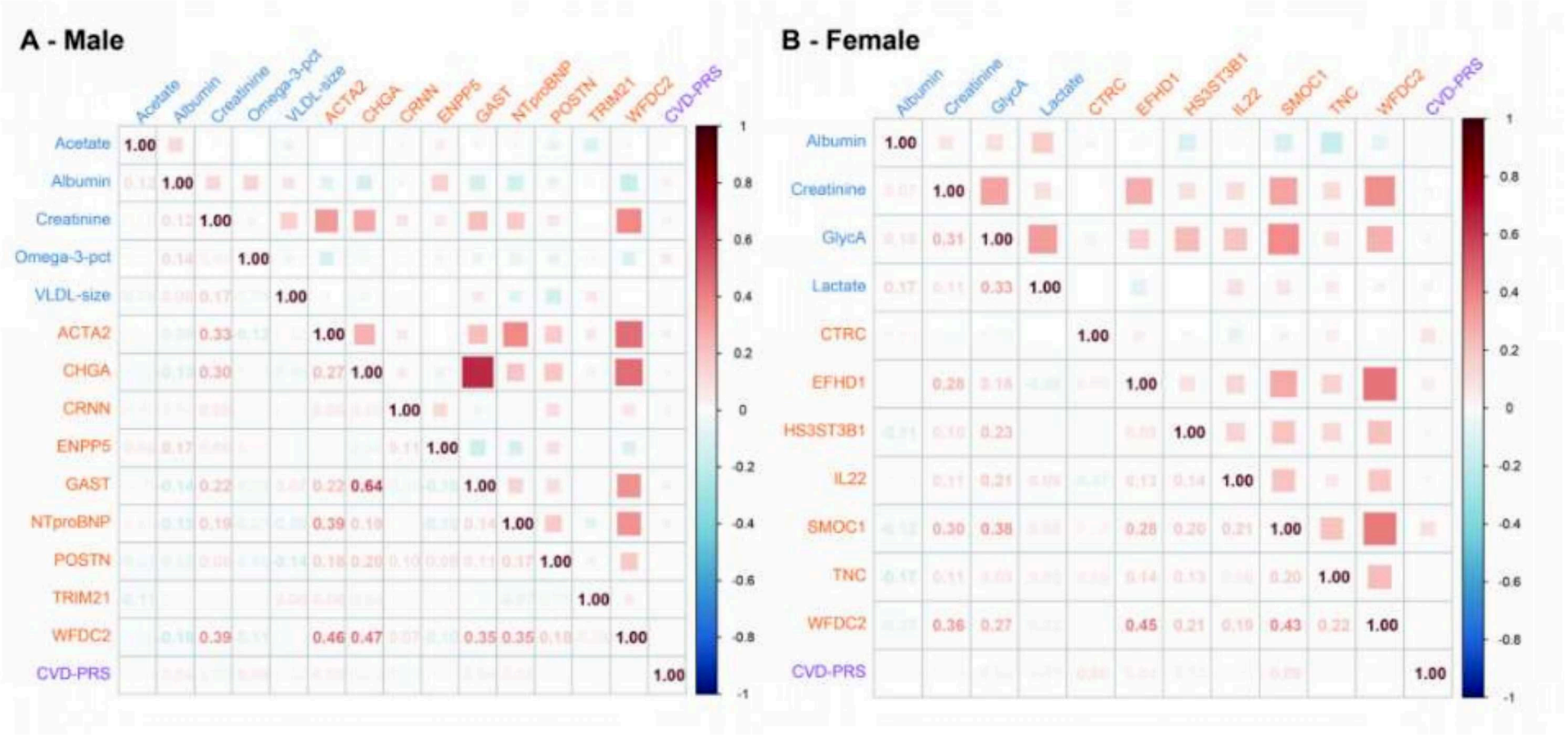



**Figure 22. Spearman correlation matrices for selected multi-omics biomarkers (N=990)**

Correlation plots display the pairwise Spearman correlation coefficients among sex-specific multi-omics biomarkers in male (**Panel A**) and female (**Panel B**) participants
**Abbreviations:** ACTA2, Actin Alpha 2; CHGA, Chromogranin-A; CRNN, Cornulin; CTRC, Chymotrypsin-C; CVD, cardiovascular disease; EFHD1, EF-hand domain-containing protein D1; ENPP5, Ectonucleotide pyrophosphatase/phosphodiesterase family member 5; GAST, Gastrin; HS3ST3B1, Heparan sulfate glucosamine 3-O-sulfotransferase 3B1; IL22, Interleukin-22; NT-proBNP, N-terminal prohormone of brain natriuretic peptide; Omega-3-pct, The Omega-3 fatty acids percentage of total fatty acids; PRS, polygenic risk score; SMOC1, SPARC-related modular calcium-binding protein 1; TNC, Tenascin-C; TRIM21; E3 ubiquitin-protein ligase TRIM21; VLDL-size, Average diameter for very-low-density lipoprotein particles; WFDC2, WAP four-disulfide core domain protein.

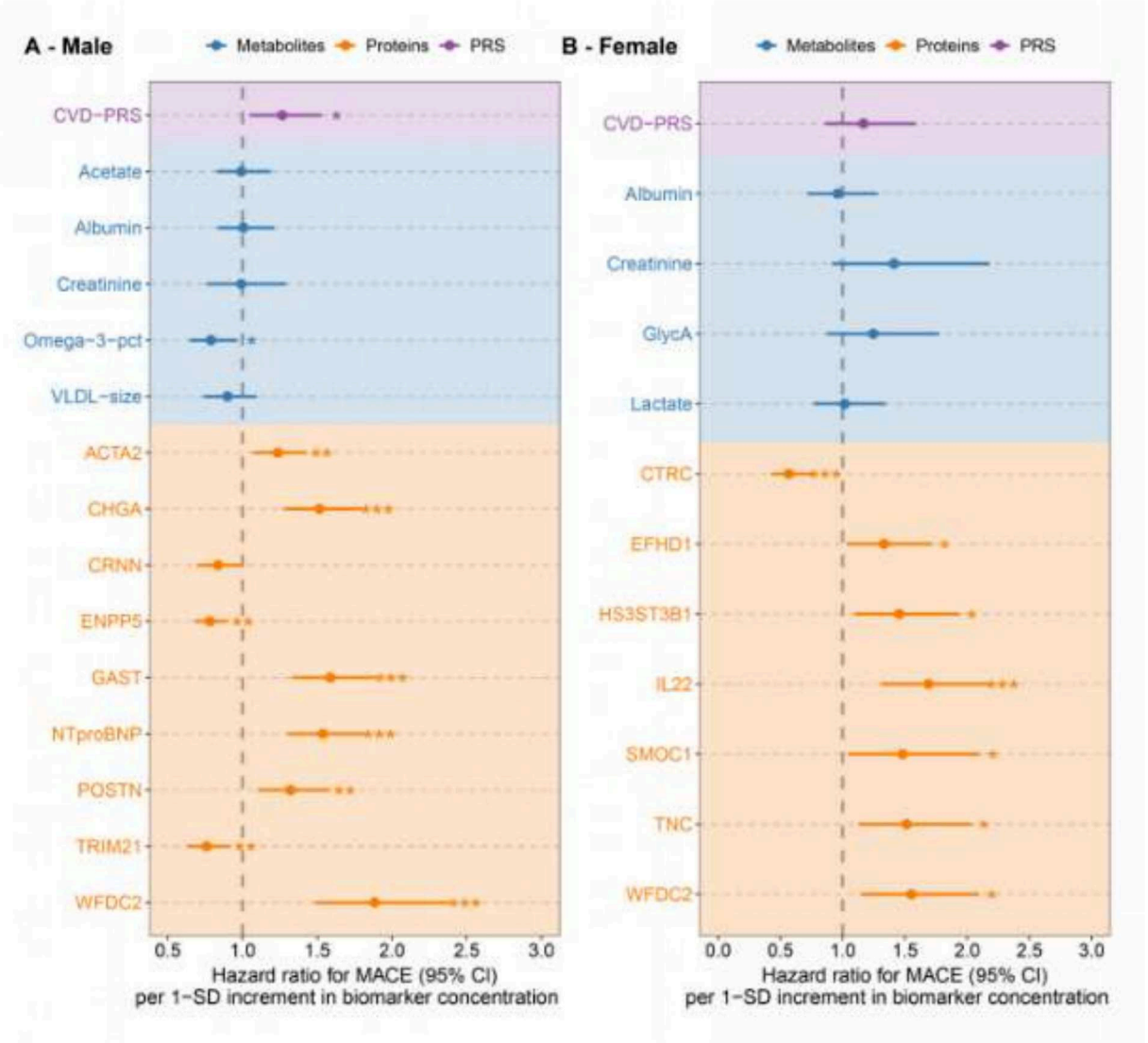



**Figure 20. Associations of selected multi-omics biomarkers and 10-year risk of MACE (N=990)**

Forest plots display hazard ratios (HRs) and 95% confidence intervals (CIs) for incident MACE per 1-standard deviation (SD) increment in the concentration of each sex-specific biomarker. Analyses were conducted separately for male (**Panel A**) and female (**Panel B**) participants, with adjustments for all clinical covariates in the SCORE2-Diabetes model. Statistically significant associations are indicated by asterisks (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$).

**Abbreviations:** HR, hazard ratio; CI, confidence interval; FDR, false discovery rate; other protein abbreviations as in **Figure 19**.

## 3.8 Multi-omics for cardiovascular risk prediction in the general population

### 3.8.1 Description of the population included

**Table 16** presents the baseline characteristics of the 24,042 participants included in the analysis, comprising 10,528 males (43.8%) and 13,514 females (56.2%). The overall mean age was 56.2 ± 8.1

years, with similar age distributions in males (56.1 ± 8.3 years) and females (56.3 ± 8.0 years). Compared to females, males had a higher prevalence of current smoking (12.0% vs. 8.8%), higher SBP (142.8 ± 18.4 mmHg vs. 136.7 ± 20.3 mmHg), and lower HDL-C levels (1.3 ± 0.3 mmol/L vs. 1.6 ± 0.4 mmol/L). In contrast, females had slightly higher total cholesterol levels (5.9 ± 1.1 mmol/L vs. 5.6 ± 1.0 mmol/L). Over a 10-year follow-up period, a total of 1,204 MACE events were recorded, comprising 742 events in males and 462 in females.

**Table 16. Baseline characteristics of the cardiovascular risk cohort**

| Baseline characteristics | Total (N=24,042) | Male (N=10,528) | Female (N=13,514) |
|---|---|---|---|
| Age (years) | 56.2 (8.1) | 56.1 (8.3) | 56.3 (8.0) |
| Current smoker, N (%) | 2,448 (10.2) | 1,262 (12.0) | 1,186 (8.8) |
| Systolic blood pressure (mmHg) | 139.4 (19.7) | 142.8 (18.4) | 136.7 (20.3) |
| Total cholesterol (mmol/L) | 5.8 (1.1) | 5.6 (1.0) | 5.9 (1.1) |
| HDL cholesterol (mmol/L) | 1.5 (0.4) | 1.3 (0.3) | 1.6 (0.4) |

Values are expressed as mean (standard deviation) for continuous variables, and as N (%) for categorical variables.

**Abbreviations:** HDL, high-density lipoprotein; MACE, major adverse cardiovascular event.

### 3.8.2 Correlations among biomarkers and associations with MACE

**Figures 24** illustrate the Spearman correlation matrices among the PRS, the selected metabolites, and proteins in males and females, respectively. Almost no correlations (all $r \leq 0.05$) were observed for the CVD-PRS with metabolites or proteins in both sexes. Correlations between metabolites and proteins were also weak with the exception of GlycA, which is an inflammatory biomarker and is moderately correlated with IL6 (r=0.33 among males) and a few other inflammation related proteins. Within metabolomic and proteomic biomarker panels, a few demonstrated strong correlations (defined as $r \geq 0.6$). In males, WFDC2 was strongly correlated with CXCL17 (r=0.60) and GDF15 (r=0.61). In females, IDL-CE-pct and L-LDL-TG-pct showed a strong correlation (r=0.70).

**Figures 25** show the associations of the selected multi-omics biomarkers with MACE risk in males and females, respectively. Most biomarkers exhibited significant associations with MACE, with overlapping

biomarkers in males and females demonstrating consistent effect directions. In males, five metabolites and two proteins were not significantly associated with MACE. In contrast, all biomarkers except for valine (Val) were significantly associated with MACE risk in females.

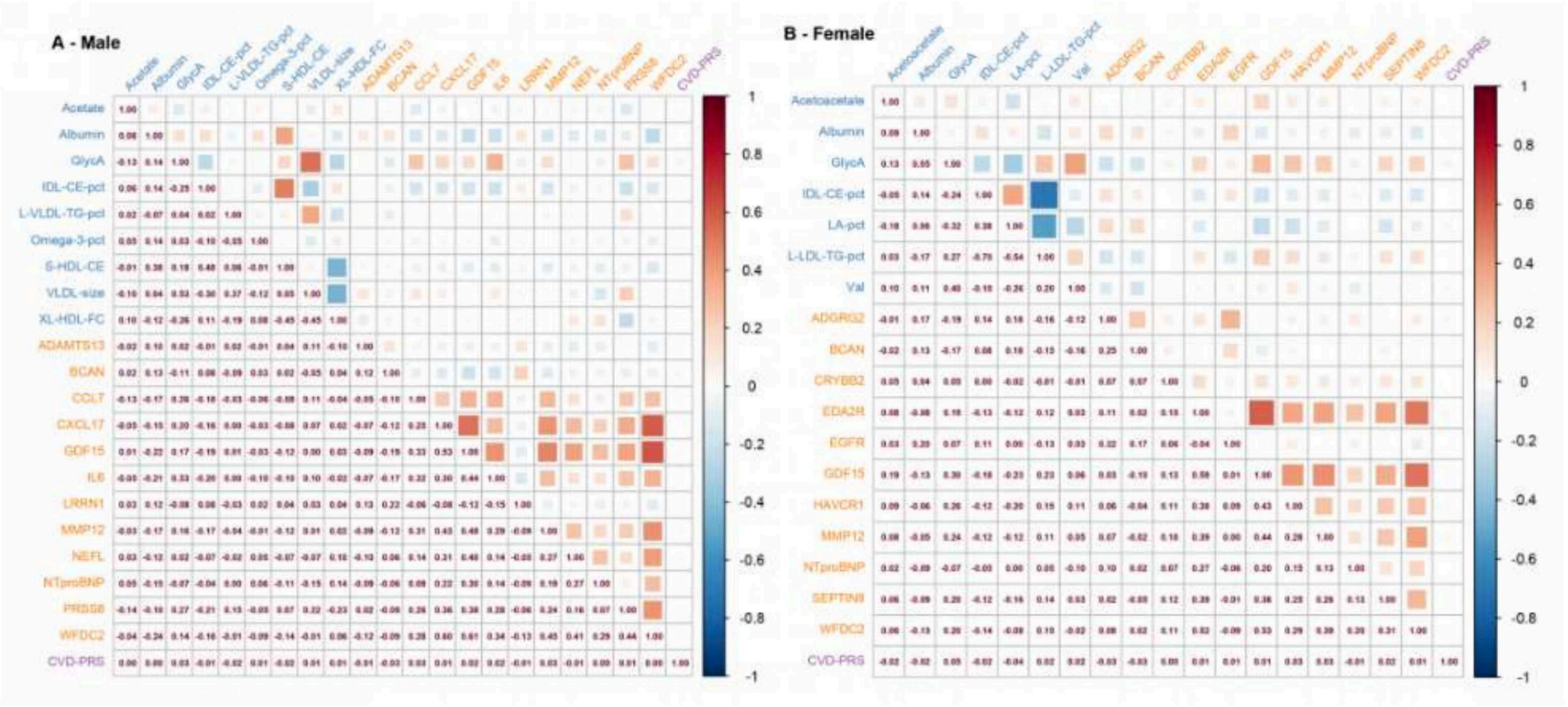



**Figure 21. Sex-specific correlation matrices of selected biomarkers in the validation set (30% of UK Biobank, N=7,214)**

Spearman correlation matrices of selected PRS, metabolites, and proteins are shown for males (Panel A) and females (Panel B).

**Abbreviations:** ADAMTS13, A disintegrin and metalloproteinase with thrombospondin motifs 13; BCAN, Brevican core protein; CCL7, C-C motif chemokine 7; CRYBB2, Beta-crystallin B2; CXCL17, C-X-C motif chemokine 17; EDA2R, Tumor necrosis factor receptor superfamily member 27; EGFR, Epidermal growth factor receptor; GDF15, Growth/differentiation factor 15; GlycA, Glycoprotein acetyls; IDL-CE-pct, The cholesteryl esters percentage of total lipids in intermediate-density lipoprotein; IL6, Interleukin-6; LA-pct, The linoleic acid percentage of total fatty acids; L-LDL-TG-pct, Triglycerides to total lipids in large low-density lipoprotein percentage; L-VLDL-TG-pct, The triglycerides percentage of total lipids in large very-low-density lipoprotein; LRRN1, Leucine-rich repeat neuronal protein 1; MMP12, Macrophage metalloelastase; NEFL, Neurofilament light polypeptide; NTproBNP, N-terminal prohormone of brain natriuretic peptide; Omega-3-pct, The Omega-3 fatty acids percentage of total fatty acids;

PRS, polygenic risk score; PRSS8, Prostasin; S-HDL-CE, Cholesteryl esters in small high-density lipoprotein; SEPTIN8, Septin-8; Val, Valine; VLDL-size, Average diameter for very-low-density lipoprotein particles; WFDC2, WAP four-disulfide core domain protein; XL-HDL-FC, Free cholesterol in very large high-density lipoprotein.

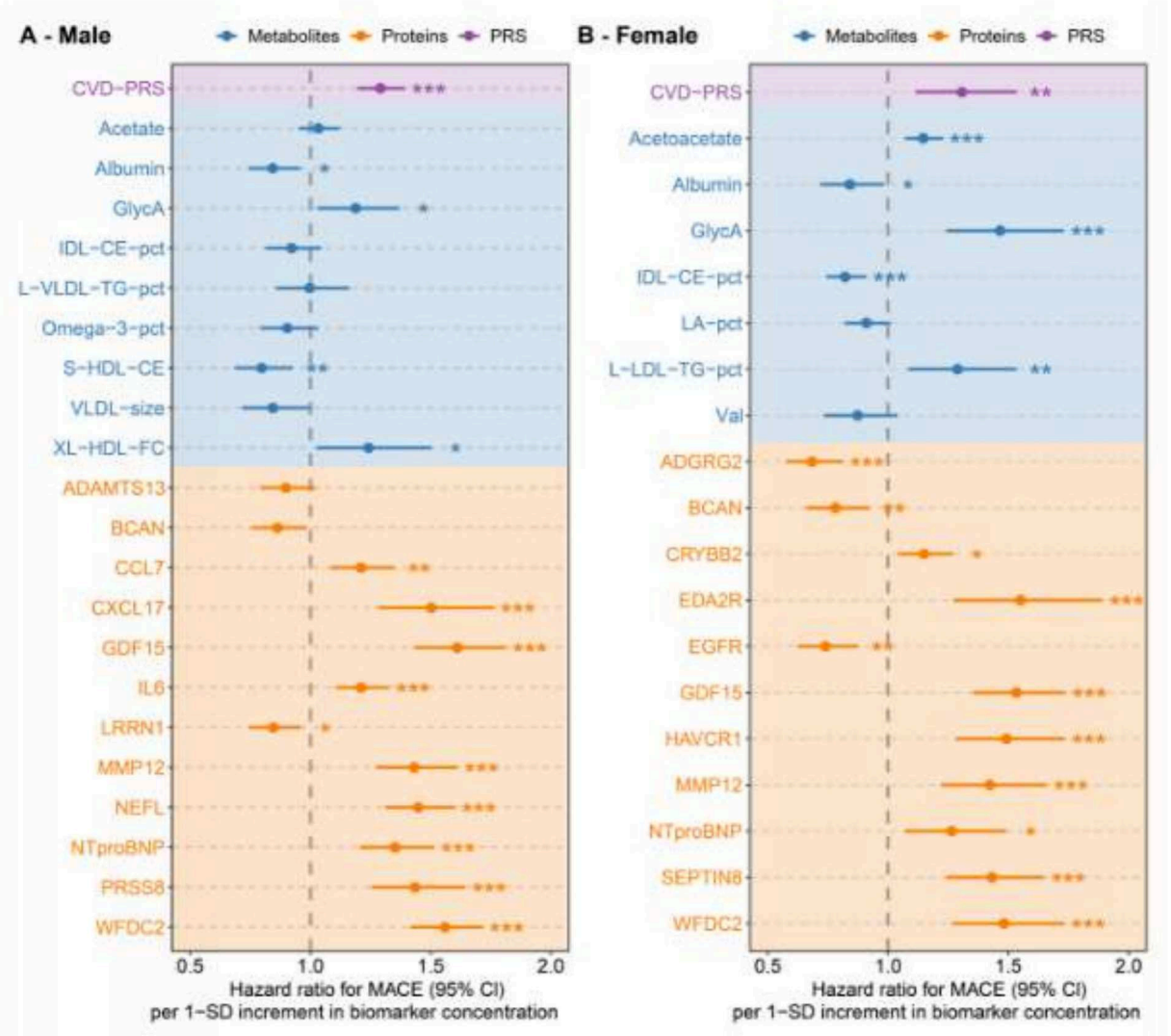



**Figure 22. Associations of selected biomarkers with 10-year MACE risk in males and females in the validation set (N=7,214)**

Hazard ratios (HRs) and 95% confidence intervals for 1-SD increases in biomarker levels are shown for males (Panel A) and females (Panel B).

**Abbreviations:** ADAMTS13, A disintegrin and metalloproteinase with thrombospondin motifs 13; BCAN, Brevican core protein; CCL7, C-C motif chemokine 7; CRYBB2, Beta-crystallin B2; CXCL17, C-X-C motif chemokine 17; EDA2R, Tumor necrosis factor receptor superfamily member 27; EGFR, Epidermal growth factor receptor; GDF15, Growth/differentiation factor 15; GlycA, Glycoprotein acetyls; IDL-CE-pct, The cholesteryl esters percentage of total lipids in intermediate-density lipoprotein; IL6, Interleukin-6; LA-pct, The linoleic acid percentage of total fatty acids; L-LDL-TG-pct, Triglycerides to total lipids in large low-density lipoprotein percentage; L-VLDL-TG-pct, The triglycerides percentage of total lipids in large very-low-density lipoprotein; LRRN1, Leucine-rich repeat neuronal protein 1; MMP12, Macrophage metalloelastase; NEFL, Neurofilament light polypeptide; NTproBNP, N-terminal prohormone of brain natriuretic peptide; Omega-3-pct, The Omega-3 fatty acids percentage of total fatty acids; PRS, polygenic risk score; PRSS8, Prostasin; S-HDL-CE, Cholesteryl esters in small high-density lipoprotein; SD, Standard deviation ; SEPTIN8, Septin-8; Val, Valine; VLDL-size, Average diameter for very-low-density lipoprotein particles; WFDC2, WAP four-disulfide core domain protein; XL-HDL-FC, Free cholesterol in very large high-density lipoprotein.

### 3.8.3 MACE risk prediction by multi-omics

**Figure 26** shows the predictive performance of the SCORE2 model extended with various multi-omics layers for 10-year MACE risk prediction. The addition of PRS, metabolomics, and proteomics each significantly improved model performance. The coefficients for all variables of the multi-omics extended models are detailed in **Appendix 66**.

For the total population, integrating multi-omics data into the SCORE2 model significantly improved predictive performance, with the C-index increasing from 0.708 to 0.769 (ΔC-index = +0.061, $P$<0.001). The NRI was 26.2% (95% CI: 21.2%, 32.5%). In males, the C-index increased from 0.682 to 0.752 (ΔC-index = +0.070, $P$<0.001), with an NRI of 12.4% (95% CI: 3.8%, 22.9%). Similarly, in females, the C-index improved from 0.724 to 0.782 (ΔC-index = +0.058, $P$<0.001), with an NRI of 30.4% (95% CI: 15.1%, 37.6%). It was apparent that proteomics provided the most substantial improvement in both sexes, followed by metabolomics, and the PRS.

**Figure 27** illustrates the incremental changes in C-index attributed to individual biomarkers in males and females. Among all biomarkers, GDF15 contributed the most to risk prediction in both sexes, as reflected by the largest individual increase in the C-index when added to the SCORE2 model. For both sexes, the top 8 biomarkers were proteins, followed by the CVD-PRS and 1 metabolite, which underscores the high predictive ability of proteomic biomarkers.

In **Figure 28,** we evaluate which combination of omics data has the best model performance for MACE prediction. Among all single-omics extensions, proteomics provided the greatest improvement across the total population (C-index: 0.763 vs. 0.708; $P$<0.001), in males (0.740 vs. 0.682; $P$<0.001), and in females (0.779 vs. 0.724; $P$<0.001) and was therefore chosen as the next reference model. Adding the PRS to the proteomics extended SCORE2 model significantly further improved prediction among the total population (C-index: 0.769 vs. 0.763; $P$=0.007) and males (C-index: 0.749 vs. 0.740; $P$=0.017). The C-index also slightly improved among women but this difference was not statistically significant (C-index: 0.782 vs. 0.779; $P$=0.198). The addition of metabolomics yielded no further benefit. Thus, the best model performance was observed for a combination of the variables of the SCORE2 model, proteomics data, and the CVD-PRS. **Appendix 67** presents the sex-specific ß-coefficients for this

optimal multi-omics combination model. Comparing this model with a metabolomics extended version confirmed that metabolomics did not further improve its predictive abilities **(Figure 28**).

**Appendix 68** presents the risk reclassification tables for low-risk (≤7.5%), intermediate-risk (>7.5-15%), and high-risk (>15%) categories. Adding all multi-omics data to the SCORE2 model significantly improved the NRI in males (12.4%) and females (30.4%). The optimal multi-omics combination model, which excludes metabolomics, achieved comparable NRIs of 12.0% in males and 28.6% in females. **Appendix 69** displays the calibration curves for the SCORE2 model and its multi-omics extensions. The optimal multi-omics combination model demonstrated good model calibration.

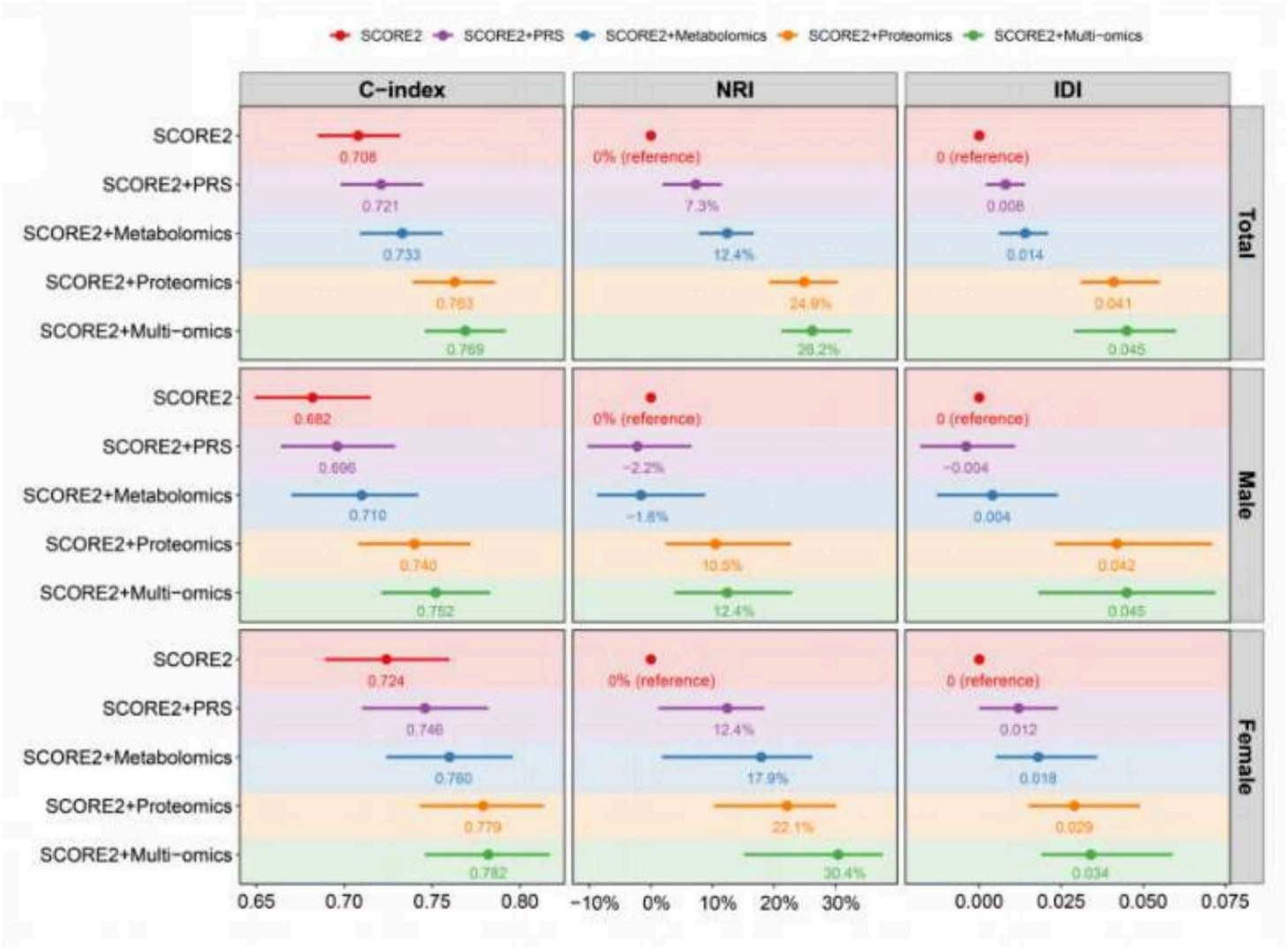



**Figure 23. Performance of SCORE2 with and without omics data extensions for 10-year MACE prediction**

**Abbreviations:** IDI, Integrated discrimination index; NRI, Net reclassification index.

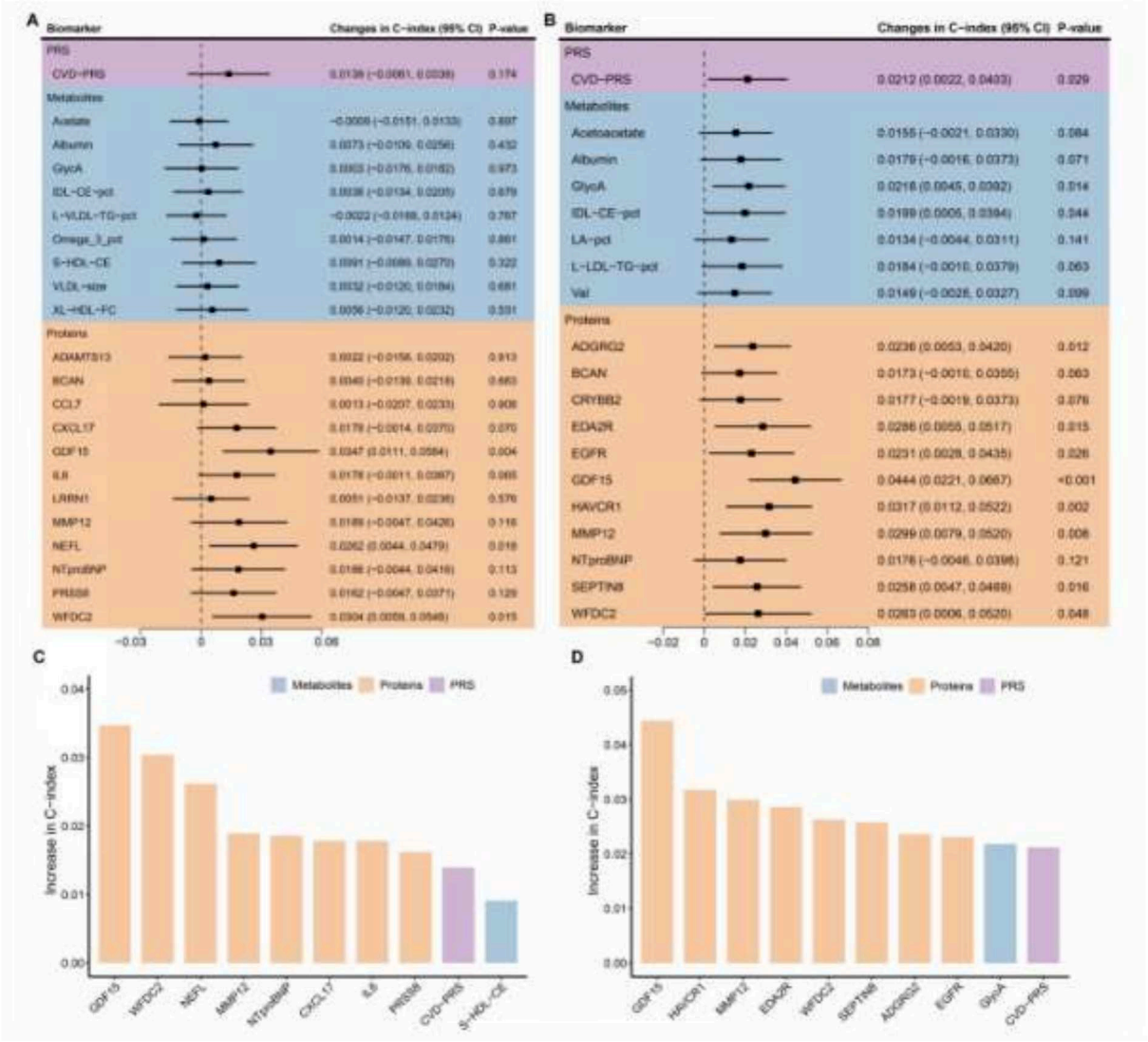
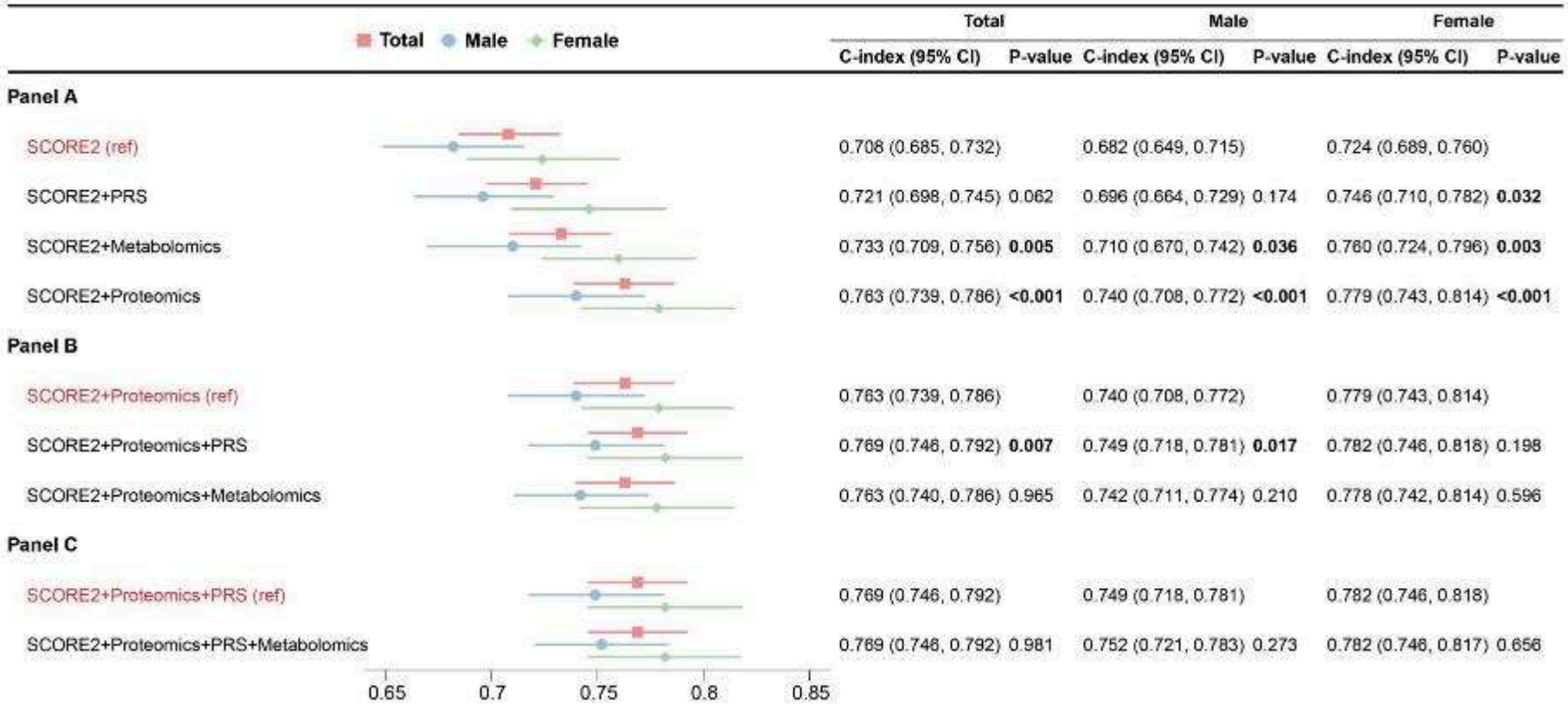



**Figure 24. Incremental changes in C-index attributed to individual biomarkers**

Forest plots of the incremental change in C-index attributed to each biomarker for males (A) and females (B). Panels C and D show the top contributors in males and females, respectively.

**Abbreviations:** ADAMTS13, A disintegrin and metalloproteinase with thrombospondin motifs 13; BCAN, Brevican core protein; CCL7, C-C motif chemokine 7; CRYBB2, Beta-crystallin B2; CXCL17, C-X-C motif chemokine 17; EDA2R, Tumor necrosis factor receptor superfamily member 27; EGFR, Epidermal growth factor receptor; GDF15, Growth/differentiation factor 15; GlycA, Glycoprotein acetyls; IDL-CE-pct, The cholesteryl esters percentage of total lipids in intermediate-density lipoprotein; IL6, Interleukin-6; LA-pct, The linoleic acid percentage of total fatty acids; L-LDL-TG-pct, Triglycerides to total lipids in large low-density lipoprotein percentage; L-VLDL-TG-pct, The triglycerides percentage of total lipids in large very-low-density lipoprotein; LRRN1, Leucine-rich repeat neuronal protein 1; MMP12, Macrophage metalloelastase; NEFL, Neurofilament light polypeptide; NTproBNP, N-terminal prohormone of brain natriuretic peptide; Omega-3-pct, The Omega-3 fatty acids percentage of total fatty acids; PRS, polygenic risk score; PRSS8, Prostasin; S-HDL-CE, Cholesteryl esters in small high-density lipoprotein; SD, standard deviation; SEPTIN8, Septin-8; Val, Valine; VLDL-size, Average diameter for very-low-density lipoprotein particles; WFDC2, WAP four-disulfide core domain protein; XL-HDL-FC, Free cholesterol in very large high-density lipoprotein.




| | Total | | Male | | Female | |
|---|---|---|---|---|---|---|
| | C-index (95% CI) | P-value | C-index (95% CI) | P-value | C-index (95% CI) | P-value |
| **Panel A** | | | | | | |
| SCORE2 (ref) | 0.708 (0.685, 0.732) | | 0.682 (0.649, 0.715) | | 0.724 (0.689, 0.760) | |
| SCORE2+PRS | 0.721 (0.698, 0.745) | 0.062 | 0.696 (0.664, 0.729) | 0.174 | 0.746 (0.710, 0.782) | **0.032** |
| SCORE2+Metabolomics | 0.733 (0.709, 0.756) | **0.005** | 0.710 (0.670, 0.742) | **0.036** | 0.760 (0.724, 0.796) | **0.003** |
| SCORE2+Proteomics | 0.763 (0.739, 0.786) | **<0.001** | 0.740 (0.708, 0.772) | **<0.001** | 0.779 (0.743, 0.814) | **<0.001** |
| **Panel B** | | | | | | |
| SCORE2+Proteomics (ref) | 0.763 (0.739, 0.786) | | 0.740 (0.708, 0.772) | | 0.779 (0.743, 0.814) | |
| SCORE2+Proteomics+PRS | 0.769 (0.746, 0.792) | **0.007** | 0.749 (0.718, 0.781) | **0.017** | 0.782 (0.746, 0.818) | 0.198 |
| SCORE2+Proteomics+Metabolomics | 0.763 (0.740, 0.786) | 0.965 | 0.742 (0.711, 0.774) | 0.210 | 0.778 (0.742, 0.814) | 0.596 |
| **Panel C** | | | | | | |
| SCORE2+Proteomics+PRS (ref) | 0.769 (0.746, 0.792) | | 0.749 (0.718, 0.781) | | 0.782 (0.746, 0.818) | |
| SCORE2+Proteomics+PRS+Metabolomics | 0.769 (0.746, 0.792) | 0.981 | 0.752 (0.721, 0.783) | 0.273 | 0.782 (0.746, 0.817) | 0.656 |

**Figure 25. C-index comparison of single-omics and multi-omics extensions to SCORE2 in MACE prediction**

Panel A shows the C-index with 95% confidence intervals for models extending SCORE2 with individual omics layers (PRS, metabolomics, or proteomics). Panel B compares multi-omics combinations using SCORE2+proteomics as the reference.

## 3.9 Multi-omics for type 2 diabetes risk prediction

### 3.9.1 Description of the population included

**Table 17** presents the baseline characteristics of 23,325 participants included in the study, of whom 719 developed T2D during the 10-year follow-up period. The mean age of the total cohort was 56.4 ± 8.1 years, with 44.5% being male. All features of the clinical CDRS model statistically significantly differed between incident T2D cases and controls. Compared to participants who did not develop T2D in the next 10 years, those who did were more likely to be male (56.9% vs. 44.1%), had higher mean age (59.0 vs. 56.3 years), $HbA_{1c}$ levels (5.8% vs. 5.3%), and BMI (30.9 vs. 27.0 kg/m²). A higher proportion of incident T2D cases also had a family history of diabetes (29.2% vs. 17.4%), were smokers or former smokers (54.4% vs 44.1%) and used antihypertensive (15.9% vs. 9.5%) or steroid medications (2.4% vs. 1.2%) at baseline.

**Table 17. Baseline characteristics of T2D cases and non-cases**

| Baseline characteristics | T2D cases (n=719) | Non-cases (n=22,606) | *P*-value | Total |
|---|---|---|---|---|
| Number of participants | 719 | 22,606 | | 23,325 |
| Male sex, n (%) | 409 (56.9) | 9,970 (44.1) | <0.001 | 10,379 (44.5) |
| Age (years), mean (SD) | 59.0 (7.7) | 56.3 (8.1) | <0.001 | 56.4 (8.1) |
| $HbA_{1c}$ (%), mean (SD) | 5.8 (0.4) | 5.3 (0.3) | <0.001 | 5.4 (0.3) |
| $HbA_{1c}$ (mmol/mol), mean (SD) | 39.9 (4.2) | 34.9 (3.6) | <0.001 | 35.0 (3.7) |
| BMI (kg/m$^2$), mean (SD) | 30.9 (5.2) | 27.0 (4.5) | <0.001 | 27.2 (4.5) |
| BMI category, n (%) | | | <0.001 | |
| <25 kg/m$^2$ | 82 (11.4) | 7,915 (35.0) | | 7,997 (34.3) |
| 25-27.49 kg/m$^2$ | 95 (13.2) | 5,675 (25.1) | | 5,770 (24.7) |
| 27.5-29.99 kg/m$^2$ | 156 (21.7) | 4,268 (18.9) | | 4,424 (19.0) |
| ≥ 30 kg/m$^2$ | 386 (53.7) | 4,748 (21.0) | | 5,134 (22.0) |
| Cigarette smoking, n (%) | | | <0.001 | |
| Never | 328 (45.6) | 12,626 (55.9) | | 12,954 (55.6) |
| Ex-smoker | 301 (41.9) | 7,660 (33.9) | | 7,961 (34.1) |
| Current smoker | 90 (12.5) | 2,320 (10.2) | | 2,410 (10.3) |
| Family history of diabetes, n (%) | | | <0.001 | |
| None | 509 (70.8) | 18,674 (82.6) | | 19,183 (82.2) |
| Parent or sibling | 174 (24.2) | 3,560 (15.8) | | 3,734 (16.0) |
| Parent and sibling | 36 (5.0) | 375 (1.6) | | 408 (1.8) |
| Prescribed medication, n (%) | | | | |
| Anti-hypertensive | 114 (15.9) | 2,144 (9.5) | 0.010 | 2,258 (9.7) |
| Steroid | 17 (2.4) | 274 (1.2) | 0.010 | 291 (1.3) |

**Abbreviations:** BMI, body mass index; $HbA_{1c}$, glycated hemoglobin; SD, standard deviation.
Continuous variables are presented as mean ± standard deviation and compared using the independent t-test. Categorical variables are presented as counts (percentages) and compared using the chi-square test.

### 3.9.2 Correlations among selected omics biomarkers

**Figure 29** illustrates the Spearman correlation matrix among the T2D-PRS, selected metabolites, and proteins. Almost no correlation (all |r| < 0.2) was observed between the T2D-PRS and biomarkers from the other omics layers (metabolites and proteins). Correlations between metabolites and proteins were generally weak to moderate (all |r |< 0.5). Within individual omics layers, only a few biomarkers demonstrated strong correlations, defined as |r| ≥ 0.6.

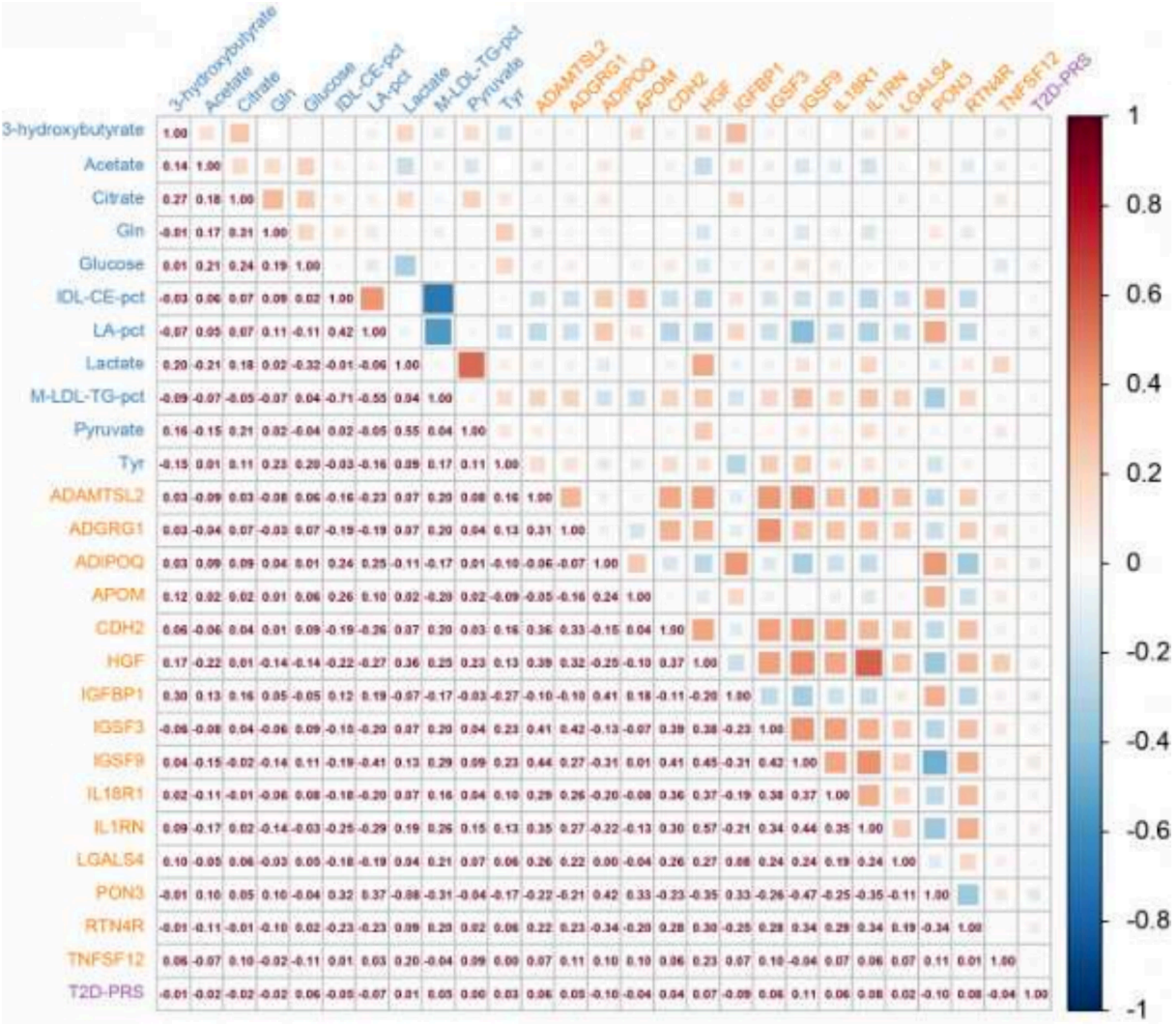



**Figure 26. Correlation matrices of selected biomarkers in the validation set (30% of UK Biobank, N=6,998)**

**Abbreviations:** ADAMTSL2, A disintegrin and metalloproteinase with thrombospondin repeats–like 2; ADGRG1, Adhesion G-protein coupled receptor G1; CDH2, Cadherin-2; CDRS, Cambridge Diabetes Risk Score; HGF, Hepatocyte growth factor; IDL-CE-pct, cholesteryl esters to total lipids in IDL percentage; IGFBP1, Insulin-like growth factor-binding protein 1; IGSF, Immunoglobulin superfamily member; IL1RN, Interleukin-1 receptor antagonist; IL18R1, Interleukin-18 receptor 1; LA-pct, linoleic acid to total fatty acids percentage; M-LDL-TG-

pct, triglycerides to total lipids in medium LDL percentage; PON3, paraoxonase/lactonase 3; RTN4R, Reticulon-4 receptor; T2D-PRS, polygenic risk score for type 2 diabetes; TNFSF12, TNF-related weak inducer of apoptosis.

### 3.9.3 Associations of selected omics biomarkers with type 2 diabetes

**Figure 30** presents the associations of the selected multi-omics biomarkers with incident T2D. Most biomarkers were significantly associated with T2D risk, except for five metabolites that did not reach statistical significance. The T2D-PRS was strongly associated with T2D risk (HR per 1 SD: 1.50, 95% CI: 1.32-1.70). Among the metabolomic biomarkers, glucose demonstrated the strongest association with T2D (HR per 1 SD: 1.50, 95% CI: 1.34–1.67) followed by M-LDL-TG-pct (HR per 1 SD: 1.46, 95% CI: 1.34–1.60). Among the proteomic biomarkers, IGSF9 (Immunoglobulin superfamily member 9) demonstrated the strongest association with T2D (HR [95%CI] per 1 SD): 1.93 [1.72–2.15]).

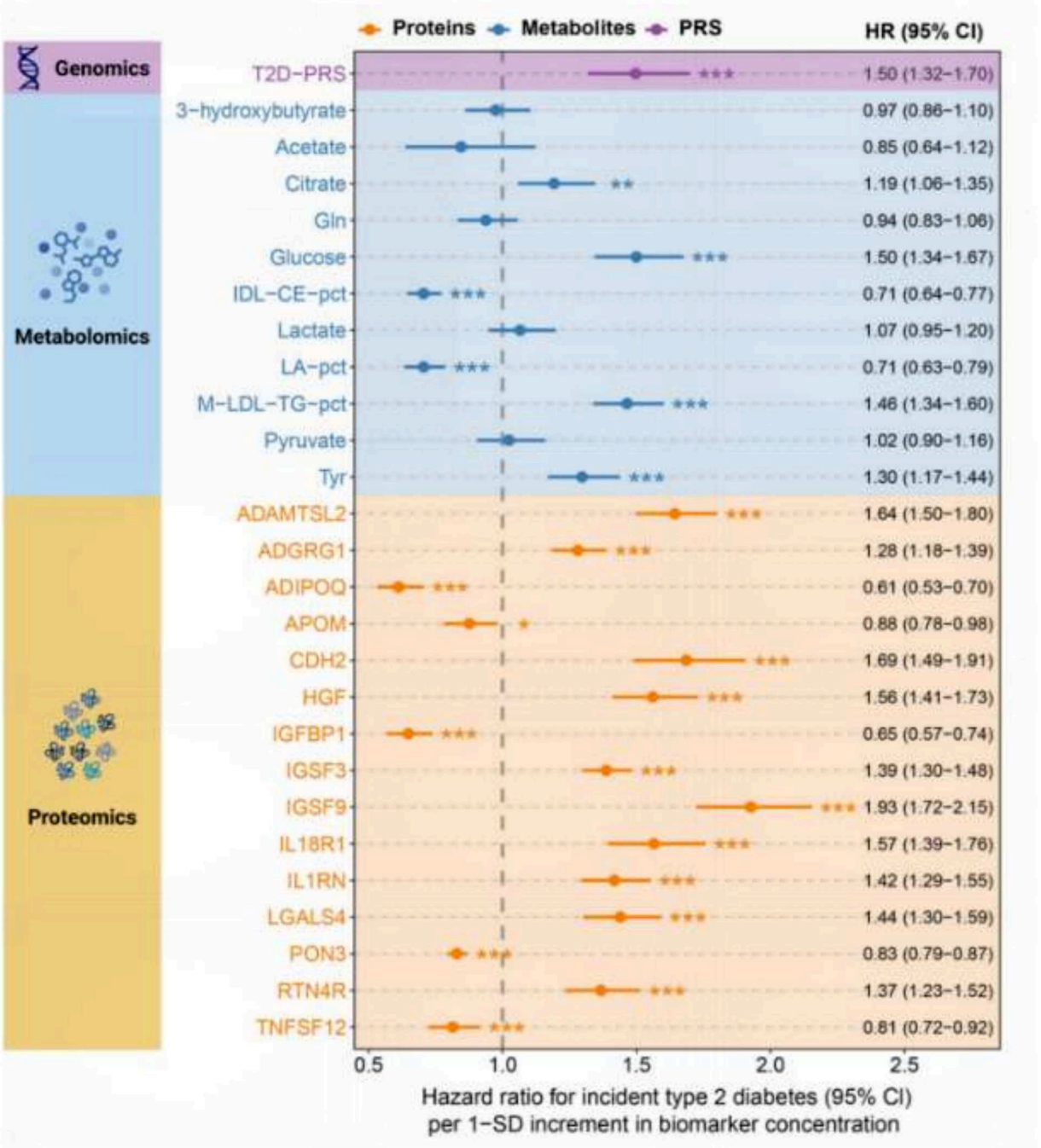



**Figure 27. Associations of selected biomarkers with incident T2D in the test set (30% of UK Biobank, N = 6,998)**

Hazard ratios and 95% confidence intervals for one standard deviation increment in biomarker concentration for type 2 diabetes incidence.
Analyses were adjusted for all variables included in the clinical CDRS model.
Asterisks indicate the level of statistical significance (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$).
**Abbreviations:** ADAMTSL2, A disintegrin and metalloproteinase with thrombospondin repeats–like 2; ADGRG1, Adhesion G-protein coupled receptor G1; CDH2, Cadherin-2; CDRS, Cambridge Diabetes Risk Score; HGF, Hepatocyte growth factor; IDL-CE-pct, cholesteryl esters to total lipids in IDL percentage; IGFBP1, Insulin-like growth factor-binding protein 1; IGSF, Immunoglobulin superfamily member; IL1RN, Interleukin-1 receptor antagonist; IL18R1, Interleukin-18 receptor 1; LA-pct, linoleic acid to total fatty acids percentage; M-LDL-TG-pct, triglycerides to total lipids in medium LDL percentage; PON3, paraoxonase/lactonase 3; RTN4R, Reticulon-4 receptor; SD, standard deviation; T2D-PRS, polygenic risk score for type 2 diabetes; TNFSF12, TNF-related weak inducer of apoptosis.

### 3.9.4 Improvements in Risk Prediction by Multi-Omics Integration

**Figure 31** shows the predictive performance of the clinical CDRS model extended by the three multi-omics layers. The C-index of the clinical CDRS model was 0.857. Adding the T2D-PRS (ΔC-index = +0.006; *P*=0.019) or metabolomics data (ΔC-index = +0.013; *P*=0.001) resulted in modest improvements, while integrating proteomics data provided the most substantial improvement (ΔC-index = +0.023; *P*<0.001), underscored by an especially high NRI (30.0%). Thus, in the next step, the proteomics extended clinical CDRS model was chosen as the reference model. Adding either the T2D-PRS or metabolomics to the proteomics-extended model further improved discrimination slightly, but the differences were not statistically significant. However, adding both the T2D-PRS and metabolomics to the proteomics-extended model significantly enhanced discrimination (ΔC-index = +0.006; *P* = 0.033).

The ß-coefficients of all variables of the main models (CDRS + Proteomics and CDRS + multi-omics) are detailed in **Appendix 70**. **Appendix 71** presents the calibration curves for these models, indicating good agreement between predicted and observed event rates across all models.

**Figure 32** displays the incremental C-index improvements of each selected biomarker when added individually to the clinical CDRS model. Proteomic biomarkers contributed most strongly to model performance, with 7 of the top 10 contributing biomarkers being proteins, followed by 2 metabolites and the T2D-PRS. Among the proteins, IGSF9 provided the greatest improvement in predictive performance, with an increase in C-index of 0.0156. Among metabolites, M-LDL-TG-pct had the highest predictive contribution, increasing the C-index by 0.0107. The T2D-PRS also showed a significant contribution, increasing the C-index by 0.0060.

In subgroup analyses stratified by prediabetes status, the same patterns emerged in the subjects without prediabetes as in the total population (**Appendix 72**). Of the three omics layers, proteomics improved the predictive ability of the clinical CDRS most and the full multi-omics model significantly improved this model further. In individuals with prediabetes, adding of proteomics to the considerably increased the C-index from 0.766 (clinical CDRS) to 0.785 but the C-index increases narrowly missed statistical

significance (p=0.077). Adding metabolomics data or the T2D-PRS to the proteomics extended clinical CDRS did not further improve prediction among people with prediabetes.

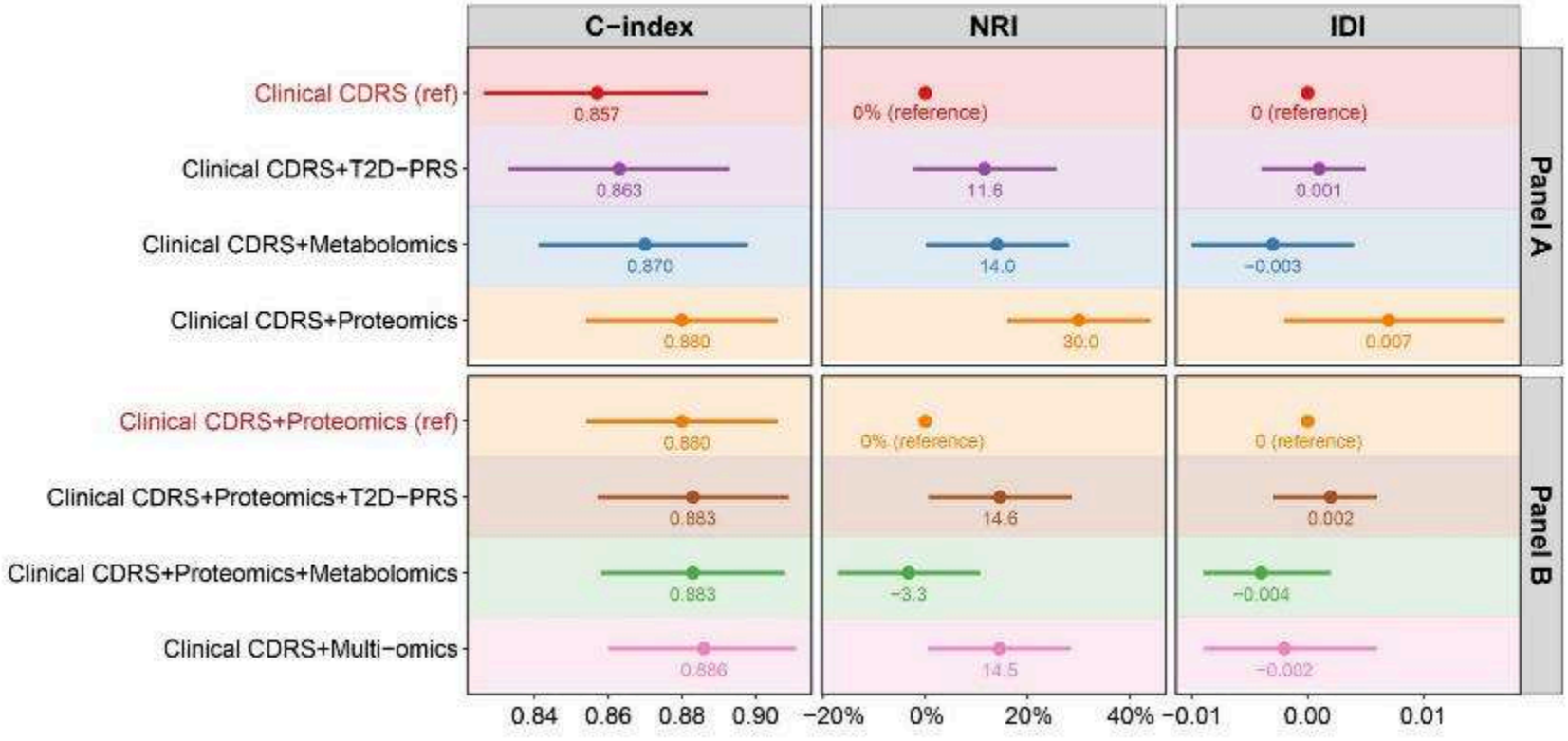



**Figure 28. Improvements in 10-year type 2 diabetes risk prediction by integrating multi-omics data in the test set (30% of UK Biobank, N = 6,998)**

Panel A compares the performance of the clinical CDRS model with its single-omics extensions, including the T2D-PRS, metabolomics, and proteomics. Panel B evaluates the incremental benefit of adding the T2D-PRS or metabolomics to the proteomics-extended model, as well as the full multi-omics model integrating all three layers.
Model performance was assessed using C-index, continuous NRI, and IDI. In Panel A, the clinical CDRS served as the reference model. In Panel B, Clinical CDRS + Proteomics served as the reference.
Asterisks indicate the statistical significance of performance differences compared with the respective reference model (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$).
The multi-omics model refers to the integration of 15 selected proteins, 11 metabolites, and a T2D-PRS.
**Abbreviations:** CDRS, Cambridge Diabetes Risk Score; T2D-PRS, polygenic risk score for type 2 diabetes; NRI, net reclassification index; IDI, integrated discrimination improvement.

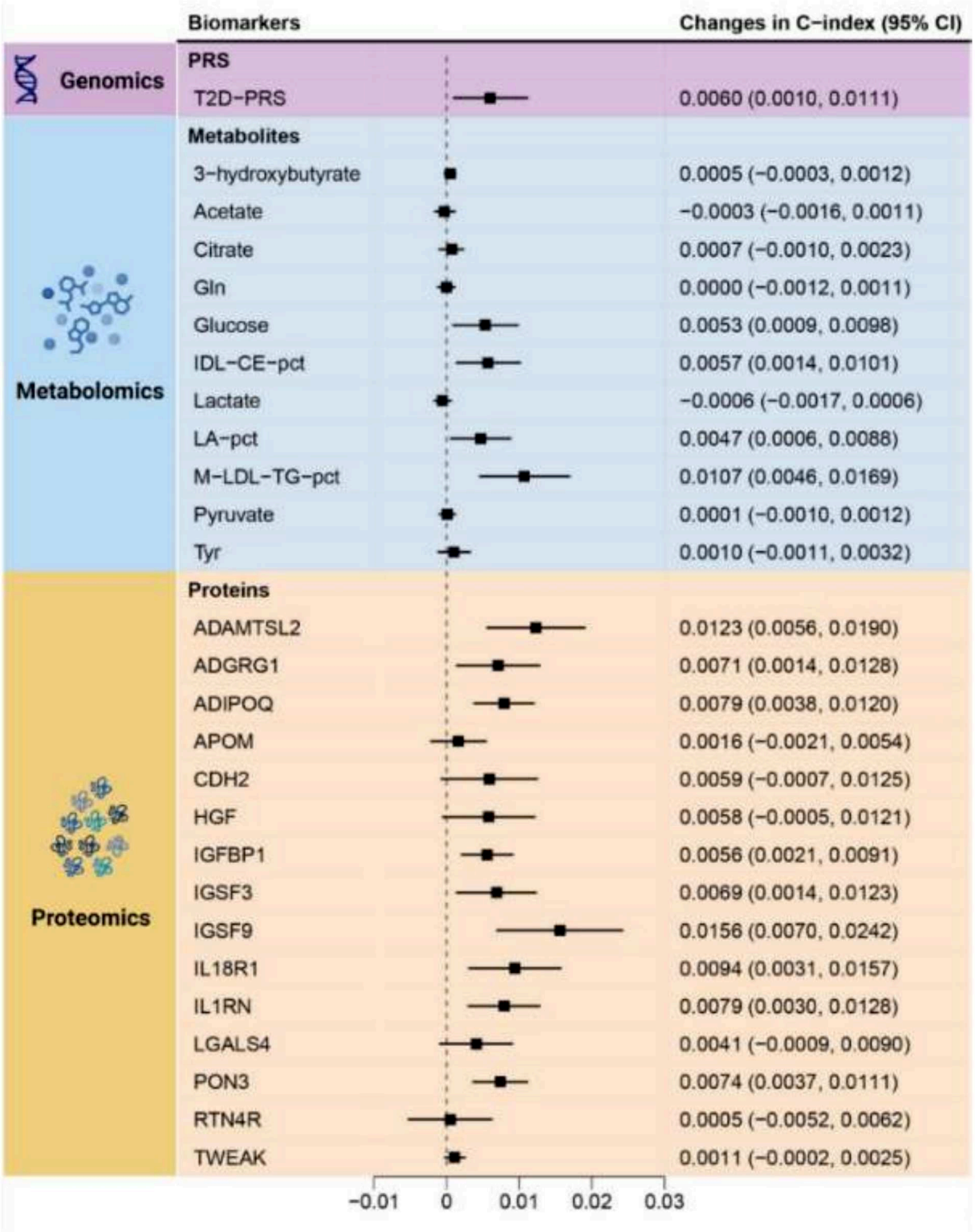



**Figure 29. Incremental contributions of individual biomarkers to risk prediction**

**Abbreviations:** ADAMTSL2, A disintegrin and metalloproteinase with thrombospondin repeats–like 2; ADGRG1, Adhesion G-protein coupled receptor G1; CDH2, Cadherin-2; CDRS, Cambridge Diabetes Risk Score; HGF, Hepatocyte growth factor; IDL-CE-pct, cholesteryl esters to total lipids in IDL percentage; IGFBP1, Insulin-like growth factor-binding protein 1; IGSF, Immunoglobulin superfamily member; IL1RN, Interleukin-1 receptor antagonist; IL18R1, Interleukin-18 receptor 1; LA-pct, linoleic acid to total fatty acids percentage; M-LDL-TG-pct, triglycerides to total lipids in medium LDL percentage; PON3, paraoxonase/lactonase 3; RTN4R, Reticulon-4 receptor; T2D-PRS, polygenic risk score for type 2 diabetes; TNFSF12, TNF-related weak inducer of apoptosis.

# 4 Discussion

## 4.1 Metabolomics for cardiovascular risk prediction in individuals with type 2 diabetes

Using large-scale data from two European cohorts of middle-aged and older patients with T2D, we developed and validated a novel sex-specific competing risk algorithm that combines the SCORE2-Diabetes model with 7 metabolites to improve 10-year cardiovascular risk prediction. By selecting metabolomic biomarkers separately for males and females, we captured sex-specific associations that enhanced the predictive performance of the model. In internal and external validations, the C-index of the SCORE2-Diabetes model significantly increased by 0.018 and 0.043, respectively. Replacing the clinical chemistry-based biomarkers of the original SCORE2-Diabetes model by NMR metabolomics as well was possible and led to comparable, statistically significant increases in the C-index by 0.017 and 0.044 in internal and external validations, respectively.

### 4.1.1 Comparison with previous studies

There are a limited number of studies using metabolites to predict cardiovascular risk in patients with T2D. Three studies based on the ADVANCE trial (sample size: n=3,576-3,779, follow-up: 5 years) investigated the association of specific types of fatty acids, amino acids, and sphingolipids with cardiovascular events and demonstrated the predictive value of these metabolites for cardiovascular risk in patients with T2D (Alshehry et al. 2016; Harris et al. 2020; Welsh et al. 2018). However, the enhancement of the predictive ability of the original models by these metabolomic biomarkers was relatively modest, with improvements in the C-statistic ranging between 0.005 and 0.020. This limited enhancement could be attributed to the investigation focusing only on a selection of a few metabolites. In contrast, the Edinburgh Type 2 Diabetes Study (sample size: n=1,066, follow-up: 10 years) used a broader metabolomics assay with 228 metabolites (Huang et al. 2023). By integrating 12 selected metabolites into the reference model, there was a notable increase in the C-statistic by 0.019; which was comparable to our findings. However, the models derived from the ADVANCE trial and the Edinburgh

Type 2 Diabetes Study lacked external validation and their metabolite profiling measurements were not conducted by a commercial service provider. This limits the application and dissemination of their metabolomics-based risk models in clinical practice (Damen et al. 2016).

A growing body of evidence emphasizes that the associations of some metabolites with MACE differ in men and women, and deriving sex-specific risk models is a potential way to enhance the predictive power (Mittelstrass et al. 2011; Tabassum et al. 2022; Würtz et al. 2015). Sex differences in the association with CVD risk of metabolomic biomarkers are biologically possible due to differences in sex hormones, sex chromosomes, and lifestyle differences (Carrel und Willard 2005; Murphy 2011; Shankar et al. 2007). For example, 141 out of 179 lipid species were found to differ significantly between males and females in the Finnish GeneRISK cohort (Tabassum et al. 2022). Despite this evidence, prior studies have used sex as a covariate rather than selecting biomarkers based on sex specificity (Tabassum et al. 2023; Vaura et al. 2022). Sex-specific biomarker-based algorithms can significantly improve predictive performance, but sex-specific analyses are much more demanding in terms of sample size, which might be the reason why - to our knowledge - there are currently no other studies with diabetic patients, which derived sex-specific, metabolomics-based cardiovascular risk prediction models.

### 4.1.2 Biological mechanisms of selected metabolites

Three metabolites (creatinine, albumin, and GlycA) were chosen to enhance MACE risk prediction for both sexes, with lactate additionally selected for women, and acetate, VLDL-size, and Omega-3-pct specifically selected for men. Creatinine and albumin are indicative of renal function. Patients with moderate to severe chronic kidney disease have an increased MACE risk (Matsushita et al. 2022; Visseren et al. 2022). GlycA reflects the glycosylation status of acute phase proteins and thus a biomarker of inflammation. Inflammatory processes are crucial in the development and progression of cardiovascular diseases (Connelly et al. 2017). Lactate, a key byproduct of glycolysis, is often elevated during metabolic stress. Elevated lactate levels are associated with inflammation, endothelial dysfunction, and oxidative stress (Li et al. 2022). Acetate is the most abundant endogenously produced short-chain fatty acid and has been found to be associated with weight loss, enhanced insulin sensitivity, and a reduced cardiovascular risk reduction (Hernández et al. 2019). The size of VLDL particles is an

independent predictor of atherosclerosis beyond traditional lipid measurements, highlighting its importance in cardiovascular risk stratification in T2D (Amor et al. 2016). Omega-3 fatty acids are known for their beneficial effects on endothelial function and their role in reducing systemic inflammation (Shahidi und Ambigaipalan 2018). If associations of these biomarkers with MACE were not in the expected direction (e.g. for lactate), this might be explained by the comprehensive adjustment for the variables of the SCORE2-Diabetes model.

### 4.1.3 Strengths and limitations

In terms of strengths, our study stands out for its large sample size (n = 10,257 for UKB, n = 1,039 for ESTHER), making it the largest metabolomics study focused on predicting the cardiovascular risk in individuals with T2D so far. Additionally, this study is pioneering in selecting metabolomic biomarkers for cardiovascular risk prediction separately by sex and using an external validation cohort. Furthermore, model calibration was performed for a low risk (UK) and intermediate cardiovascular risk region (Germany).

However, our study has certain limitations. The ESTHER study may not be the ideal external validation cohort for our model due to its differences from the UKB, including a higher mean age (64 vs. 61.5 years), a higher proportion of fasting of at least 4 hours prior to blood sample donation (89.1% vs. 23.7%), which can be relevant for some metabolic biomarkers, and the differences in the ascertainment of non-fatal strokes, which did not allow for the exclusion of hemorrhagic strokes, as well as the regional difference of cardiovascular risk (low in UK and intermediate in Germany). These differences could have negatively influenced the results of the external validation. However, despite these cohort differences, the overall consistency in risk prediction results between the two cohorts provides support for the robustness of the derived model. Another limitation involves the timing of the measurements. In both cohorts, traditional clinical chemistry measurements were performed shortly after sample collection, whereas NMR metabolomics analyses were conducted several years later with stored samples. While this time difference could theoretically introduce metabolite degradation, the quality control statistics provided by Nightingale Health did not justify this concern because the distributions of all metabolites were comparable to previous studies measured by the company, which used fresh

blood samples. Finally, this study may be generalizable primarily to low-to-intermediate risk regions in Europe. Larger validation cohorts from diverse regions, including high and very high-risk areas, are needed to comprehensively validate the new risk model.

### 4.1.4 Potential clinical utility of the derived model

The integration of 7 selected metabolomic biomarkers in the SCORE2-Diabetes model has led to a significant improvement in the discrimination of 10-year cardiovascular risk in patients with T2D. Although statistically significant, the additive improvements in predictive performance are modest. However, they are likely still clinically relevant in the context of a prediction of a severe, and frequent clinical outcome, like MACE, in a high-risk population, like T2D patients, where small individual enhancements in risk stratification can lead to significant improvements on the public health level.

Although the SCORE-2 Diabetes model already includes metabolites, such as TC, HDL-C and creatinine (as part of the eGFR), the 7 additional metabolites improved its predictive power by capturing additional biological information not yet reflected by these traditional clinical measurements. The biomarkers selected reflect additional aspects of inflammation, glucose and lipid metabolism, which are critical in the progression of CVD, especially in high-risk populations like patients with T2D. This augmentation enhances the predictive power of the model without compromising the convenience offered by the conventional model, which requires only a brief physical examination, interview, and blood sample analysis.

Moreover, the NMR metabolomics test could be used as the only laboratory method for MACE prediction in patients with T2D because it measures the clinical chemistry parameters of the SCORE2-Diabetes model as well (TC, HDL-C and creatinine) and can measure glucose as a replacement for the $HbA_{1c}$. The models with these replaced biomarkers performed as good as the original models. Using only NMR metabolomics for the laboratory-based variables of the prediction model could enhance the cost-effectiveness of its use in routine care.

While incorporating metabolomics into clinical practice does incur costs, it is important to note that the NMR method used for metabolomic analysis is generally more cost-effective and standardized compared to the LC-MS/MS (Liquid Chromatography-Mass Spectrometry/Mass Spectrometry) method

(Jacob et al. 2019). Additionally, commercial companies like Nightingale Health are available to handle large sample sizes in a short time, making this approach feasible for widespread clinical application.

However, before recommending routine measurement in clinical practice, it is essential to consider country-specific validation of the model. Currently, our algorithm is primarily validated for the UK and Germany, and extending this validation to other regions is crucial for broader applicability. The differences in cardiovascular risk profiles and healthcare systems across countries necessitate such validation to ensure the effectiveness and relevance of the model in diverse settings. Furthermore, studies are needed to assess the cost-effectiveness and clinical utility of incorporating NMR metabolomics into routine practice.

## 4.2 Metabolomics for cardiovascular risk prediction in the general population

Using large-scale data from two European cohorts of middle-aged and older participants, we derived and validated an algorithm combining the SCORE2 model and 13 metabolites for improving 10-year cardiovascular risk prediction. In internal and external validation, the C-index of the SCORE2 model significantly increased by 0.019 and 0.015, respectively. In addition, the reclassification ability of the SCORE2 model was significantly improved in both internal and external validation. In terms of calibration, the performance of the SCORE2 and its extended model showed good calibration in the internal validation in the UKB, but the external validation in the German ESTHER study revealed an underestimation of risk in both the original and extended model. Thus, calibration of the models to the underlying CVD risk in specific regions is still needed.

### 4.2.1 Comparison with previous studies & novelty

The high-dimensional nature of metabolomics data dictates that sample size is a crucial factor for accurately selecting the most predictive metabolites for health outcomes. Compared to previous studies, this study represents the largest sample size to date in evaluating cardiovascular risk models with metabolomic biomarkers. A recent systematic review and meta-analysis, which included and analyzed

21 studies using metabolomics to assess cardiovascular risk, encompassing a total of 54,337 participants, which is less than one third of the sample size used from the UKB alone in this work (n=187,039). The meta-analysis revealed that incorporating metabolites into the prediction model resulted in an average increase of 0.04 in the C-index (McGranaghan et al. 2020), which was higher than in our results. This might be explained by more metabolites measured. A narrow focus on specific categories of metabolomic biomarkers may limit the accuracy of risk prediction (Würtz et al. 2013; Yang et al. 2019). However, previous studies were constrained by their relatively small sample sizes, with the largest cohort consisting of 8,101 participants (Ganna et al. 2014; Havulinna et al. 2016; Rizza et al. 2014; Shah et al. 2012), which raises concerns regarding to the robustness of their models and, coupled with a widespread absence of external model validation, restricts the generalizability of these models (Stegemann et al. 2014; Wang et al. 2019) .

Recent evidence has highlighted that there are sex differences in both cardiovascular risk and metabolomic biomarkers, suggesting that the derivation of sex-specific risk models is a valuable approach to enhance predictive accuracy (Mittelstrass et al. 2011; Tabassum et al. 2022). Previous studies typically included sex as a covariate in the model derivation process to select metabolites, rather than selecting metabolites based on sex for males and females separately. The primary reason might be that sex-specific analyses are much more demanding in terms of sample size.

### 4.2.2 Biological mechanisms of selected metabolites

Three metabolites (albumin, GlycA and IDL-CE-pct) were chosen to enhance MACE risk prediction for both males and females, four metabolites were additionally selected for females (acetoacetate, LA-pct, L-LDL-TG-pct, and valine), and six metabolites were specifically chosen for males (acetate, L-VLDL-TG-pct, Omega-3-pct, S-HDL-CE, VLDL-size and XL-HDL-FC). Albumin serves as a biomarker of nutritional status and acts as a critical indicator of systemic inflammation and renal health, both of which are pivotal in the assessment and management of cardiovascular risk (Ronit et al. 2020). GlycA reflects the glycosylation status of acute phase proteins and thus is a pro-inflammatory biomarker. Inflammatory processes are crucial in the development and progression of cardiovascular diseases (Connelly et al. 2017). Acetoacetate is a type of ketone body and has been shown to enhance contractile performance

and beta-adrenergic inotropism in stunned myocardium, as indicated by its circulating levels. Clinical research suggests that in non-diabetic populations, a significant association exists between the presence of acetoacetate and the occurrence of heart failure (Lommi et al. 1996). Omega-3-pct and LA-pct may reduce cardiovascular risks by modulating lipid profiles and inflammatory markers (Kleber et al. 2016; Shahidi und Ambigaipalan 2018). Valine is a branched-chain amino acid and has been recognized for its ability to indicate cardiac metabolic risks independently of critical factors such as sex and BMI (Mangge et al. 2016). Acetate is the most abundant endogenously produced short-chain fatty acid, is associated with weight loss, enhanced insulin sensitivity, and reduced cardiovascular risk (Hernández et al. 2019). Lipoprotein components, such as IDL-CE-pct and L-VLDL-TG-pct, represent the intricate dynamics of lipid content and particle size within lipoprotein subclasses, which are crucial in the development and progression of atherosclerosis (Amor et al. 2016; Hicks et al. 2015; Kjellmo et al. 2018; Timmerman et al. 2021).

It was expected that some metabolites were selected for men only and others for women only because of the well-established biological sex differences in lipid metabolism, inflammation, and amino acid utilization. Premenopausal women generally have a more favorable lipid profile than men, with lower levels of atherogenic lipoproteins (VLDL and LDL) and higher levels of HDL cholesterol, largely driven by estrogen-mediated lipid regulation (Robinson et al. 2022; Tabassum et al. 2022). However, metabolic perturbations such as diabetes or menopause can accelerate the loss of this advantage, leading to a disproportionate increase in cardiovascular risk among women (Yoshida et al. 2022).

The male-specific selection of acetate, L-VLDL-TG-pct, Omega-3-pct, S-HDL-CE, VLDL-size, and XL-HDL-FC reflects key sex differences in lipid metabolism. For instance, prior research indicates that men have a higher burden of atherogenic lipids and distinct lipid processing pathways compared to women (Tabassum et al. 2022). Men also tend to have larger VLDL particles, which are strongly associated with increased cardiovascular risk due to their role in promoting remnant lipoproteins that penetrate the arterial wall (Shahidi und Ambigaipalan 2018). Furthermore, Omega-3-pct levels have been reported to be lower in men and associated with a higher cardiovascular risk compared to women (Shahidi und Ambigaipalan 2018). HDL function may also differ between sexes, with men typically exhibiting lower HDL cholesterol and less protective HDL subfractions (Robinson et al. 2022).

Acetoacetate was selected exclusively for females, which could be linked to sex differences in energy metabolism. Prior studies suggest that women rely more on ketone metabolism for energy under fasting and stress conditions, and changes in ketone body levels may have unique implications for cardiovascular risk in women (Yoshida et al. 2022). Additionally, LA-pct was predictive only in women, supporting previous findings that polyunsaturated fatty acids influence lipid metabolism differently between sexes (Shahidi und Ambigaipalan 2018). Notably, valine was also identified as a female-specific risk factor, likely due to its stronger association with insulin resistance and metabolic dysfunction in women than in men (Mangge et al. 2016). Finally, while men generally have higher baseline levels of BCAAs, elevations in BCAAs above a sex-specific threshold appear to be particularly predictive of cardiovascular risk in women, potentially due to differences in amino acid metabolism and muscle mass (Tabassum et al. 2022).

### 4.2.3 Strengths and limitations

Our investigation stands out for its considerable cohort size, comprising 187,039 participants in the UKB and 5,578 in the ESTHER study, rendering it the largest metabolomics analysis dedicated to cardiovascular risk prediction to date. Additionally, this study is pioneering in selecting metabolomic biomarkers for cardiovascular risk prediction separately by sex. Furthermore, model calibration was performed for a low (UK) and intermediate cardiovascular risk region (Germany). Although the results of our model calibration were not satisfactory in external validation, two recent studies using the SCORE2 model in other European cohorts (the EPIC-Norfolk study and a cohort based on a Netherland's population) have also shown poor model calibration, suggesting that recalibration of model risk is warranted in external cohorts (Kist et al. 2023; van Trier et al. 2024).

Our study also has limitations. The ESTHER study may not be the most ideal external validation cohort for our model due to differences with the UKB, particularly in the following aspects: the determination of non-fatal strokes cannot exclude hemorrhagic strokes; and regional differences in cardiovascular risk (low in UK and intermediate in Germany). Most blood samples in the UKB cohort were collected in a non-fasting state (96.6%) in plasma samples, whereas most samples in the ESTHER study were collected in a fasting state in serum samples (89.4%). However, a comparison of the values of the 13

selected metabolites by fasting status in the UK Biobank and ESTHER study shows that the differences in metabolite features between fasting and non-fasting individuals are minimal in both cohorts. This is in agreement with a previous study, which has shown that the duration of fasting has little impact on the variability of these biomarkers (Li-Gao et al. 2019). However, despite these cohort differences, the overall consistency in risk prediction results between the two cohorts provides support for the robustness of the derived model. Additionally, a limitation of the metabolomics data used in this study is that they were obtained via NMR spectroscopy, which predominantly captures high-abundance lipid-related markers and their ratios. Given the lower sensitivity of NMR compared to mass spectrometry, the coverage of low-abundance metabolites is limited. Future studies should integrate MS-based metabolomics to provide a more comprehensive representation of the metabolome and its relevance for cardiovascular risk prediction. Although this study may be generalizable to low-to-intermediate risk regions in Europe, larger validation cohorts are needed, which should originate not only from low to intermediate risk regions in Europe but shall also validate the new risk model in populations from high and very high-risk regions.

### 4.2.4 Potential clinical implications of the derived model

The inclusion of thirteen selected metabolomic biomarkers in the SCORE2 model has significantly improved the discrimination (C-index) and most reclassification measures (continuous NRI, and IDI) for 10-year cardiovascular risk assessment in the total population of the internal and external validation set. Only the categorical NRI, which was statistically significant in the total population of the internal validation set (6.7% [95%CI: 4.8%; 9.1%]), did not reach statistical significance in the external validation set (8% [95%CI: -2%; 14%]).

The addition of metabolomic biomarkers may not only bolster the predictive capacity of the model but would also maintain the simplicity and efficiency of traditional cardiovascular risk assessments, which typically involve only a brief physical examination, patient interview, and blood analysis. A disadvantage of incorporating metabolomics into standard clinical evaluations are the undoubtedly raised expenses. However, it is crucial to recognize that NMR presents a more economical and uniform approach for metabolomic analysis compared to LC-MS/MS (Jacob et al. 2019). The ability of

commercial organizations like Nightingale Health, which can efficiently handle large volumes of samples highlights the feasibility and potential for widespread application in clinical settings.

Prior to recommending routine implementation in clinical practice, it is imperative to conduct model validation specific to particular countries. Currently, our algorithm is primarily validated for the UK and Germany, and extending this validation to other regions is crucial for broader applicability. The differences in cardiovascular risk profiles and healthcare systems across countries necessitate such validation to ensure the model's predictive performance and relevance in diverse settings.

## 4.3 Metabolomics for type 2 diabetes risk prediction

To our knowledge, this study represents the largest-scale investigation to date on the predictive utility of metabolomic biomarkers for T2D risk, leveraging data from two large European cohorts comprising over 90,000 middle-aged and older adults of whom more than 4,000 developed T2D during 10 years of follow-up. We developed and validated the UKB-DRS that integrates an established risk prediction model (clinical CDRS) with 11 metabolites for enhanced predictive performance of 10-year T2D risk. In both internal and external validation, the UKB-DRS demonstrated significant improvements in the C-index compared to the clinical CDRS, with increases of 0.019 and 0.028, respectively. Additionally, we examined the incremental value of each metabolite to the model's predictive capability across both cohorts, finding similar results in internal and external validation for all 11 metabolites. These findings confirm the robustness and generalizability of the new algorithm in enhancing the predictive accuracy for the risk of T2D.

### 4.3.1 Comparison with previous studies & novelty

The association between circulating metabolomic biomarkers and the risk of T2D has been extensively documented. Ahola-Olli et al. investigated four Finnish cohorts and identified that approximately half of the 229 NMR-based metabolomic biomarkers examined were significantly associated with incident T2D (Ahola-Olli et al. 2019). Similarly, Seah et al. observed that multiple metabolites were linked to an increased risk of T2D in cohorts with Asian ethnicity (Seah et al. 2022). Moreover, to enhance the

accuracy of existing prediction models for T2D risk, numerous studies have investigated the incremental value of metabolomic biomarkers (Satheesh et al. 2020). Many traditional prediction models already achieve high predictive accuracy, making significant improvements upon these models challenging (Abbasi et al. 2012; Hippisley-Cox und Coupland 2017). Previous studies often focused on the impact of specific types of metabolomic biomarkers, with several studies finding that incorporating pre-selected metabolomic biomarkers into traditional risk prediction models did not enhance risk identification capabilities (Fall et al. 2016; Tillin et al. 2015). However, other studies have shown that screening among a large number of metabolomic biomarkers tends to improve model discrimination (Bragg et al. 2022; Floegel et al. 2013). The European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam study, which measured 163 metabolites for 800 incident T2D cases (with an average follow-up of 7 years) and 2,282 controls, showed that adding 14 metabolomic biomarkers selected from 163 metabolite candidates to a traditional risk model significantly improved the model's discriminative ability (C-index increased from 0.901 to 0.912) (Floegel et al. 2013). The EPIC-Potsdam study is the second largest study after the UK Biobank using untargeted metabolomics data. Prior to our study, a study from the UKB, which did not have the full number of metabolomics data available had the largest sample size (n=65,684 participants analyzed, including 1,719 T2D cases) (Bragg et al. 2022). The authors conducted internal validation after incorporating metabolomic data into the Framingham Diabetes Risk Score, significantly enhancing the model's C-index (95% CI) from 0.802 (0.791, 0.812) to 0.830 (0.822, 0.841) (Δ C-index=0.028). However, a lack of external validation remains a major barrier to translating metabolomic based models into clinical applications, which we overcame by our joint analysis of the UKB and ESTHER cohort. Furthermore, we included more than twice the number of incident T2D case in our UKB analysis (n=3,536) compared to the previous UKB study by including more study participants with metabolomics data (n=86,232) and by the additional use of primary care records for the outcome assessment.

### 4.3.2 Biological mechanisms of selected metabolites

The ensemble of metabolites identified in our study captures a broad spectrum of biochemical pathways integral to the metabolic dysregulation that precedes T2D. The identified metabolites stem from the

tricarboxylic acid cycle (citrate, glucose, lactate and pyruvate), are ketone bodies (3-hydroxybutyrate and acetate), amino acids (glutamine and tyrosine), lipoproteins (IDL-CE-pct and M-LDL-TG-pct) and a fatty acid (LA-pct). This metabolic signature characterizes an impaired glucose and lipid metabolism that is detectable in blood samples up to 10 years prior diabetes diagnosis and might also give new insights into the mechanisms of T2D pathogenesis.

Citrate is a key metabolite in the tricarboxylic acid cycle and not only facilitates energy production but also plays a vital role in fatty acid synthesis, thereby linking glycolytic processes with lipid metabolism (Jin und Ma 2021). Concurrently, glucose as a direct participant in glycolysis, provides critical insights into the disrupted glucose homeostasis that is characteristic of diabetes (Ortiz-Martínez et al. 2022). The presence of lactate and pyruvate, products of anaerobic and aerobic glycolysis respectively, signals a shift towards inefficient energy utilization, a common observation in insulin-resistant states. These shifts are believed to reflect the underlying mitochondrial adaptations that compromise cellular energy management and exacerbate hyperglycemic conditions (Konrad et al. 1999).

Ketone bodies, including 3-hydroxybutyrate and acetate, accumulate during increased fatty acid oxidation - a compensatory response to impaired glucose utilization in diabetes. Elevated levels of these ketone bodies indicate the body's attempt to maintain energy balance under conditions of insulin resistance (Laffel 1999).

Furthermore, the amino acids glutamine and tyrosine are noted for their dual roles in metabolic and neurotransmitter pathways. Glutamine is involved in gluconeogenesis and plays a crucial role in maintaining glucose homeostasis. Tyrosine is associated with catecholamine synthesis and may influence stress responses that exacerbate beta-cell dysfunction and insulin resistance (Ortiz-Martínez et al. 2022).

Lipoprotein-related metabolites, including IDL-CE-pct and M-LDL-TG-pct, reflect changes in lipoprotein metabolism that contribute to a dyslipidemia pattern commonly associated with insulin resistance (Mackey et al. 2015). Additionally, the percentage of linoleic acid, a crucial component in inflammatory pathways, has been associated with variations in dietary intake and metabolic status, influencing cellular functions pivotal to insulin sensitivity (Zong et al. 2019).

Overall, the identified metabolic profile comprising glycolysis-related metabolites, ketone bodies, amino acids, lipoprotein-related metabolites and fatty acids not only contributes evidence to the biological pathways from early metabolic disturbances to T2D development but also adds predictive power to traditional risk factor models, which will allow to detect those better who are at risk to develop T2D and profit most from early preventive interventions.

### 4.3.3 Strengths and limitations

This study stands out from previous research primarily due to its substantial sample size and the confirmation of our model's robustness through external validation. We employed a well-established, clinically approved targeted NMR metabolomics platform with absolute quantification of 250 biomarkers. This feature does not only facilitate comparative analyses across different populations but also enhances the potential for clinical translation.

However, our study is not without limitations. Although we only included UKB participants with available primary care records to ascertain the incidence of T2D, the IR remained significantly lower than in the ESTHER study (55.0 vs 145.2 per 10,000 PY, respectively). This difference is likely due to an underreporting in the UKB and a more complete ascertainment in the ESTHER study, in which the GPs of the study participants were asked to provide their medical records and these were screened for diabetes diagnoses and glucose-lowering drugs. In addition, there are differences between the two cohorts regarding the blood samples. The UKB used plasma and ESTHER serum samples to determine the metabolites. Furthermore, only a small proportion of blood samples in the UK Biobank cohort were collected under fasting conditions, defined as no consumption of food or drink for at least 8 hours (3.3%), whereas the majority of participants in the ESTHER study were not fasting (90.6%). However, a comparison of the levels of the 11 selected metabolites by fasting status in the UKB and ESTHER study shows that the differences in metabolite concentrations between fasting and non-fasting individuals are small in both cohorts. This is in agreement with a previous study, which has shown that the duration of fasting has little impact on the variability of these metabolites (Li-Gao et al. 2019). Thus, the novel UKB-DRS is robust towards fasting status and use of either serum or plasma samples. However, it is unknown how the model will perform in different populations. As its derivation and

validation are based on British and German populations aged 50 to 69 years, extending its application to ethnically diverse populations or other age groups requires further validation.

### 4.3.4 Potential clinical implications of the derived model

Although the absolute improvement in the C-index for the UKB-DRS over the clinical CDRS alone was modest, this improvement is clinically meaningful for several reasons. First, even small gains in the discriminative power of risk prediction models can have substantial population-level implications when applied to large populations. In the context of T2D, earlier and more accurate identification of individuals at risk can enable timely intervention strategies, potentially preventing thousands of cases annually (Gillies et al. 2008). Second, in a clinical setting, the use of metabolomic biomarkers as an adjunct to traditional risk factors may help to reclassify individuals whose risk of developing diabetes was previously underestimated or overestimated. In this context, the moderately increased continuous NRI of >30% of the UKB-DRS, is more clinically relevant than the modest C-index increase. These reclassifications could refine decision-making processes related to lifestyle interventions or pharmacotherapy (Ortiz-Martínez et al. 2022). Lastly, the metabolomic biomarkers selected in this study were derived from NMR spectroscopy, a high-throughput and cost-effective platform that, compared to liquid chromatography with tandem mass spectrometry , is more suitable to be used in large-scale clinical routine settings (Soininen et al. 2009). In this context, the concise UKB-DRS with just 4 added metabolites is likely the most cost-effective approach.

## 4.4 Proteomics for type 2 diabetes risk prediction

This study utilized data from two large European cohorts comprising 26,352 middle-aged and older adults, among whom 1,227 developed T2D over a 10-year follow-up period. We selected 15 proteins from the Olink Explore and 6 proteins from the Olink Target Inflammation platform to develop and validate two protein-extended risk prediction models based on an established T2D prediction model (clinical CDRS). Both internal and external validation results showed that the model extended by 6 biomarkers of inflammation significantly outperformed the clinical CDRS. The discriminative ability of the model extended by 15 proteins from the comprehensive OLINK Explore proteomics library was

even better in internal validation but external validation of this model is still needed to be done in further studies. These findings suggest that proteomics data have the potential to substantially enhance the predictive accuracy for T2D onset and improve the identification of high-risk individuals. We show that especially biomarkers of inflammation, of which 3 were selected for both models (HGF, TWEAK, and IL18R1), are an important feature of improved risk algorithms, underscoring their potential for clinical application and targeted prevention strategies.

### 4.4.1 Comparison with previous studies & novelty

Previous attempts to enhance the accuracy of T2D risk predictions using proteomic biomarkers have yielded limited success (Carrasco-Zanini et al. 2024c; Elhadad et al. 2020; Huth et al. 2019; Luo et al. 2023; Nowak et al. 2016; Rooney et al. 2023; Yao et al. 2024). Of the few studies that reported improvements, only four demonstrated significant enhancements in model discrimination (ΔAUC [area under the curve]: 0.011-0.022) with the inclusion of proteomic biomarkers (Carrasco-Zanini et al. 2024c; Huth et al. 2019; Luo et al. 2023; Rooney et al. 2023; Yao et al. 2024). This can be explained by the fact that conventional T2D prediction models already achieve high accuracy, making further substantial improvements difficult to attain (Abbasi et al. 2012; Hippisley-Cox und Coupland 2017). In our study, we used the clinical CDRS as the reference model, which is one of the leading T2D risk prediction models worldwide because it has a high C-index, mainly due to integration of $HbA_{1c}$. Nevertheless, we were able to enhance the C-index of the clinical CDRS by 0.029 units by adding 15 proteins in internal validation. In addition, the improved model had a notable NRI of 23% (95%CI: 10-36%). This is a higher improvement than in the studies cited above. Although not statistically significant, only the recently published proteomics enhanced model from the EPIC-Norfolk cohort observed comparable discriminative and reclassification statistics in internal validation (ΔC-index=0.03 for the clinical CDRS from 0.82 to 0.85, NRI=19%) (Carrasco-Zanini et al. 2024c). Interestingly, proteomics data had the highest potential for improving the clinical CDRS in this multi-omics study. Only a model combining proteomics with a polygenic risk score, clinical biomarkers, and metabolites marginally further improved the model (C-index=0.87). As making multi-omics data available in clinical practice is not realistic in the near future due to the high costs, a measurement of a small amount of proteins (e.g., the

15 derived in our study or the 10 from the EPIC-Norfolk study) seems to be a good compromise between model precision and practicality in routine practice. What is needed now are external validation studies. This would either require measurement of the Olink Explore or a newly developed high-throughput assay for the 15 proteins in a large nested case-control. Both are highly cost-intensive endeavors and beyond the scope of this study. As a compromise, we measured the OLINK Target Inflammation panel in our external validation cohort to test whether in principle proteomics-based models for T2D risk prediction can be externally validated. This was successful, as the inflammation-based 6-marker-model showed a comparable and statistically significant incremental C-index and NRI in the external validation set (ΔC-index: 0.016, NRI: 29%) and the internal validation set (ΔC-index: 0.014; NRI: 19%). This incremental C-index is approximately half of the one of the 15-marker model (0.029), although the number of biomarkers to select from was only a small fraction of the number of proteins used for the 15-marker model (73 vs. 2,085 proteins). This finding highlights the association between subclinical inflammation and the pathogenesis of T2D, suggesting that including inflammatory biomarkers could enhance the predictive performance of established risk scores.

Only three other studies included an external validation set (Elhadad et al. 2020; Nowak et al. 2016; Yao et al. 2024). The most recent and largest study utilized 1,896 participants of the China Kadoorie Biobank (CKB) as the derivation cohort and 46,595 participants of the UKB as the validation cohort (Yao et al. 2024). In contrast to the smaller, earlier studies, this study was successful in externally validating a prediction model adding 33 proteins to a traditional risk model. However, as the authors noted, the demographic differences between East Asian and European populations limit its generalizability and applicability. Our study is the first to derive and validate their model in two large Western-European cohorts, which allows the generalizability of its finding to the Western-European population.

### 4.4.2 Biological mechanisms of selected proteins

In this study, we identified a total of 18 proteins for T2D risk prediction. Among these, 17 proteins have been shown to be associated with incident T2D in prospective studies, supporting their role in disease prediction and suggesting a role in disease pathophysiology (Carrasco-Zanini et al. 2024b; Cronjé et al.

2023; Gudmundsdottir et al. 2020; Gummesson et al. 2021; Luo et al. 2023; Ngo et al. 2021; van Heck et al. 2025; Zanini et al. 2020). Only CDH2 (cadherin 2) has not been reported as being associated with T2D in either prospective or cross-sectional studies, previously.

The selected 6 biomarkers of subclinical inflammation, IL6, FGF21, HGF, TWEAK, IL18R1 and the CDCP1, each uniquely reflect processes involved in the exacerbation of insulin resistance and beta-cell dysfunction. CDCP1 is implicated in cellular migration and metastasis, suggesting a role in tissue remodeling and inflammation (Magnusson et al. 2020). IL6 functions as both a pro-inflammatory cytokine and a metabolic regulator, affecting glucose homeostasis and inflammatory pathways (Akbari und Hassan-Zadeh 2018). FGF21 improves insulin sensitivity and supports pancreatic beta-cell viability, serving as an important link between metabolism and cell survival (Gaich et al. 2013). HGF promotes pancreatic beta-cell regeneration and modulates components of the metabolic syndrome (Oliveira et al. 2018), while TWEAK influences both inflammatory responses and cellular apoptosis, with implications for metabolic regulation and insulin resistance (Vendrell und Chacón 2013). IL18R1 plays a crucial role in the immune response and is implicated in autoimmune and inflammatory diseases, including T2D, by modulating IL-18 activity—a cytokine involved in inflammatory processes that can contribute to insulin resistance and impaired glucose metabolism (Thomas et al. 2022; Thorand et al. 2005). Although not part of the Target 96 Inflammation Panel, adiponectin, interleukin-1 receptor antagonist protein (IL1RN), Galectin-4 (LGALS4), and serum paraoxonase/lactonase 3 (PON3) belong to the Inflammation panel I or II of the OLINK Explore library and their primary mechanism of action for T2D development is likely also inflammation (Li et al. 2009; Mohammed et al. 2022; Morosi et al. 2021; Reiner et al. 2008).

Eight of the 15 identified proteins from the OLINK Explore panel are not (primarily) biologically linked to T2D via inflammation-related pathways. Apolipoprotein M plays a crucial role in modulating the lipid metabolism and maintaining high-density lipoprotein function (Hajny et al. 2020; Zhou et al. 2011). The IGFBP1 (insulin-like growth factor-binding protein-1) enhances the bioavailability of insulin-like growth factors, affecting both glucose regulation and cellular growth (Ezzat et al. 2008). IGSF 3 and IGSF9 are involved in cell adhesion and immune response regulation, potentially impacting insulin signaling and glucose metabolism (Liu et al. 2023; Schweitzer et al. 2020). The reticulon-4 receptor

(RTN4R) has been implicated in neuronal outgrowth and may influence cellular plasticity and repair mechanisms in metabolic tissues, offering another layer of complexity to the interplay between cellular structure and metabolic regulation (Wang et al. 2021). Although ADAMTSL2 plays a role in modulating metalloproteinases and extracellular matrix remodeling, suggesting its influence on metabolic processes through effects on tissue structure and inflammation (Corey et al. 2022). Adhesion G-protein coupled receptor G1 (ADGRG1) is a G-protein-coupled receptor involved in cell signaling, immune modulation, and extracellular matrix remodeling (Dunér et al. 2016). While its precise role in glucose metabolism and diabetes is still unclear, its known function in inflammatory and fibrotic pathways suggests a potential link to metabolic dysregulation, which requires further study.

CDH2 has not been directly linked to T2D in previous studies. This protein is primarily involved in cellular adhesion and signaling, playing a crucial role in maintaining structural integrity in insulin-responsive tissues (Mayosi et al. 2017). Although its role in diabetes remains unexplored, alterations in adhesion molecules have been implicated in insulin resistance and metabolic dysfunction, suggesting that CDH2 may contribute to diabetes pathogenesis through mechanisms that warrant further investigation.

### 4.4.3 Strengths and limitations

The main strengths of this study are its large sample size, the robustness of the results confirmed by effective external validation and rigorous case ascertainment. Comprehensive proteomic profiling further strengthens the study by enabling the identification of a wide range of proteins linked to T2D. These features underscore the reliability and generalizability of our findings across diverse populations.

Despite these strengths, the study has several limitations. Although we included only UKB participants with available primary care records to determine the incidence of T2D, the observed IR was significantly lower than in the ESTHER study (41.6 vs. 124.6 per 10,000 PY). This discrepancy may be attributed to the age difference and more comprehensive case ascertainment in the ESTHER study, in which general practitioners were actively approached to provide medical records, including diabetes diagnoses and prescriptions for glucose-lowering medications. Furthermore, the Olink platforms used in this study do

not provide absolute protein concentrations, which implies that the variable coefficients derived from the internal validation set could not be directly applied to external validation cohorts. We addressed this limitation by splitting the ESTHER study into calibration and validation sets to obtain new β-coefficients, although this evidence supports reasonable comparability (Davies et al. 2023), but recalibration may overestimate external performance. Future studies with absolute protein quantification or uniform NPX thresholds are needed to overcome this issue. Additionally, while Olink's PEA has shown high specificity and sensitivity across both EDTA plasma and serum (Assarsson et al. 2014; Wik et al. 2021), minor matrix-based variations cannot be completely ruled out. Finally, our model's performance in non-European or younger populations remains unknown, as both derivation and validation were confined to middle-aged and older British and German adults.

## 4.5 Proteomics for cardiovascular risk prediction in the general population

Our study demonstrates that incorporating sex-specific proteomic biomarkers into the SCORE2 model significantly improves the prediction of 10-year MACE risk in individuals without prior CVD or diabetes. Overall, 18 proteins were selected, with 5 of them identified in both sexes, 7 only in males, and 6 only in females. WFDC2 and GDF15 contributed the highest increase in C-index in both sexes. Furthermore, high NRIs were observed for both sexes, with a particularly high NRI of 36.3% for males.

### 4.5.1 Comparison with previous studies & novelty

Several prior studies have demonstrated the potential of proteomic biomarkers to enhance cardiovascular risk prediction (Bom et al. 2019; Helgason et al. 2023; Hoogeveen et al. 2020; Mazidi et al. 2023; Nurmohamed et al. 2022; Royer et al. 2024a; Wallentin et al. 2021; Williams et al. 2019; Williams et al. 2022). For instance, in the EPIC-Norfolk cohort, incorporating 50 proteins into a cardiovascular risk model resulted in an AUC improvement of 0.024 for predicting MI and 0.071 for predicting cardiovascular events over three years (Hoogeveen et al. 2020). Similarly, in an Icelandic cohort, an increase in the C-index of 0.014 was observed for predicting MACE when a protein risk score was added to a traditional risk model (Helgason et al. 2023).

Additionally, the studies of Royer et al.(Royer et al. 2024a) and Ho et al.(Ho et al. 2025) investigated the enhancement of the predictive capability of the SCORE2 and PREVENT risk models for MACE with proteomic biomarkers in the same data set we used from the UK Biobank. The main difference compared to our analysis is that no sex-specific selection procedure was used. Ho et al. reported that incorporating 85 proteins in the PREVENT risk model improved the C-index for predicting MACE by 0.054 (from 0.745 to 0.799) and resulted in a categorical NRI of 9% (Ho et al. 2025). Royer et al. reported that integrating 114 proteins into the SCORE2 model increased the AUC by 0.031 (from 0.740 to 0.771) and achieved an NRI of 14.0% (Royer et al. 2024a). Compared to the results of Ho et al. and Royer et al, our sex-specific selection procedure achieved a higher improvement of model discrimination (C-index increase of 0.062) from 0.716 to 0.778). However, we achieved this C-index with adding only 12 proteins for males and 11 proteins for females to the SCORE2 model instead of adding 85 (Ho et al.) or 114 (Royer et al.) proteins. In addition, the NRI was higher in our analysis (19.9%) than in the ones of Ho et al. (9%) and Royer et al. (14.0%). For a direct comparison with our sex-specific model, we used the proteins selected Royer at al. and Ho et al. and added them to the SCORE2 model in our study population. Despite the much larger number of biomarkers, both models exhibited comparable (Ho et al.) or substantially lower (Royer et al.) C-index values than our sex-specific protein-extended model. By accounting for biological differences and the distinct baseline cardiovascular risks between males and females, our model more precisely identifies individuals at high risk for MACE in the next 10 years who could benefit from targeted preventive strategies. Finally, measuring up to 12 proteins is not only more cost-effective but also more feasible for clinical implementation than measuring 85 or 114 proteins.

### 4.5.2 Biological mechanisms of selected proteins

GDF15 and WFDC2 gave the largest increase in the discriminative ability of the model for MACE prediction in both sexes. Elevated levels of GDF15 reflect heightened inflammatory activity and are predictive of adverse cardiovascular outcomes (Li et al. 2020). IL6, CCL7, CXCL17, and the tumor necrosis factor receptor superfamily member 27 (EDA2R), proteins involved in inflammation, were also selected. IL6 is a pro-inflammatory cytokine that promotes endothelial dysfunction and plaque formation, leading to increased cardiovascular risk (Libby et al. 2002). CCL7 recruits immune cells to

inflammatory sites, contributing to plaque instability (Chang et al. 2022). CXCL17 is involved in immune cell chemotaxis and may accelerate atherogenesis (Sozzani et al. 2015). EDA2R mediates TNF-related cytokine signaling, exacerbating vascular inflammation and plaque vulnerability (Chiarini et al. 2020).

WFDC2 modulates extracellular matrix degradation, impacting fibrosis and plaque stability (Ressler et al. 2014). Further proteins selected for the MACE risk algorithms which are involved in extracellular matrix remodeling were MMP12, ADAMTS13, and PRSS8 (Prostasin). MMP12 degrades extracellular components, weakening the fibrous cap, which increases rupture risk (Traylor et al. 2014). ADAMTS13 regulates the von Willebrand factor cleavage, affecting thrombosis risk (Zheng 2015). PRSS8 influences vascular stiffness and hypertension through sodium balance regulation (Zhu et al. 2008).

A third large group of selected proteins affect vascular function through neural pathways. These were BCAN, NEFL (Neurofilament light polypeptide), LRRN1 (Leucine-rich repeat neuronal protein 1), CRYBB2, and SEPTIN8 (Septin-8). BCAN may influence vascular integrity via extracellular matrix interactions (Frischknecht und Seidenbecher 2012). NEFL reflects neurodegenerative processes impacting autonomic cardiovascular regulation (Agrawal et al. 2014). LRRN1 is involved in neuronal signaling affecting vascular tone (Tossell et al. 2011). CRYBB2 may be involved in oxidative stress pathways or hormonal regulation affecting cardiovascular health in women (Zhang et al. 2008). SEPTIN8 is involved in cytoskeletal organization and contributes to atherosclerosis by influencing cell migration and proliferation within the vessel wall (Werner und Yadav 2023).

The remaining selected proteins have, other, specific functions. NTproBNP (N-terminal prohormone of brain natriuretic peptide) is a well-established biomarker reflecting cardiac wall stress and dysfunction (Panagopoulou et al. 2013). EGFR (Epidermal growth factor receptor) plays a role in vascular smooth muscle cell proliferation and neointimal formation, and its activation may lead to vascular remodeling and plaque progression (Dreux et al. 2006). ADGRG2 (Adhesion G-protein coupled receptor G2) is involved in cell adhesion and signaling, potentially affecting endothelial function and vascular integrity (Peeters et al. 2015). HAVCR1 (Hepatitis A virus cellular receptor 1) is implicated in immune regulation

and may contribute to atherosclerosis by modulating immune cell activity and endothelial interactions (Liu et al. 2022).

### 4.5.3 Strengths and limitations

The strengths of this study include its large sample size and the comprehensive evaluation of a wide array of proteomic biomarkers. The sex-specific biomarker selection led to better risk prediction models and supports personalized risk assessment. By focusing on proteins with high selection frequency in bootstrap samples, we minimized overfitting and improved model generalizability.

However, there are limitations to consider. We did not test the robustness of our model in external validation cohorts due to the lack of other publicly available cohorts with Olink Explore proteomics data and the substantial costs associated with measuring them in large-scale cohorts. This should be addressed in future research. Furthermore, future studies should calibrate and validate the model in other populations because the findings can currently only be generalized to the population of the UK and other Western European countries with low cardiovascular risk.

## 4.6 Inflammation-related proteins for cardiovascular risk prediction

In this study involving large-scale data from the UKB and the ESTHER study, we demonstrated that incorporating selected inflammation-related proteins into the SCORE2 model significantly improves the prediction of 10-year MACE in individuals without prior CVD or diabetes. By employing a sex-specific selection approach using LASSO regression and bootstrapping, we selected 7 inflammation-related proteins, of which 2 were selected for both sexes, 3 only for males, and 2 only for females. The inclusion of these proteins led to significant increases in the C-index in both internal and external validations, as well as improvements in reclassification metrics.

### 4.6.1 Comparison with previous studies & novelty

Inflammation is a key mechanism in the development of atherosclerosis and the triggering of major cardiovascular events (Ross 1999). CRP as a non-specific marker of inflammation, has been extensively

investigated over the past two decades for its value in cardiovascular risk prediction (Lloyd-Jones et al. 2006). Some studies have shown that adding CRP to risk models can improve the C-index (Danesh et al. 2004; Koenig et al. 2004; Ridker et al. 2002). For example, in a cohort of 3,435 men aged 45 to 74 years from southern Germany, adding CRP to the Framingham risk score increased the C-index from 0.735 to 0.750 (Koenig et al. 2004). However, many other studies have found that incorporating CRP does not significantly enhance the model's discriminatory ability (Lloyd-Jones et al. 2006). In our study, we included CRP in the protein selection process, but it was not selected as a predictor. This may be due to the limited additional information CRP provides compared to other inflammation-related proteins. Especially IL-6 is highly correlated with CRP (which we also observed in male participants of the UKB) and is more strongly related to MACE than CRP (Georgakis et al. 2022; Held et al. 2017). We presume that this is the reason why IL6 and not CRP was selected for the risk prediction model for males in our derivation set.

The recent development of large-scale targeted proteomics has offered unprecedented opportunities to identify novel circulating biomarkers for MACE (Cui et al. 2022). For instance, Helgason et al. found in an Icelandic cohort that adding a protein risk score to traditional risk models increased the C-index by 0.014 (Helgason et al. 2023). Similarly, Hoogeveen et al. included 50 proteins in a cardiovascular risk model in the EPIC-Norfolk cohort and ranked the importance of these proteins (Hoogeveen et al. 2020). Notably, most of these proteins were inflammation-related, and 5 of the 7 proteins selected in our study — CXCL9, IL6, CDCP1 (CUB domain-containing protein 1), TNFRSF11B, and TNFRSF9 — were ranked 6$^{th}$, 8$^{th}$, 12$^{th}$, 18$^{th}$, and 23$^{rd}$, respectively, in their work. Additionally, Royer et al. investigated the enhancement of the SCORE2 model's predictive ability with proteomic biomarkers in approximately 38,000 participants from the UKB (Royer et al. 2024b). They found that integrating 114 proteins (selected from 2,919 proteins of the Olink Explore panel) into the model increased the AUC by 0.031 (from 0.740 to 0.771). To our knowledge, this study is the first to focus specifically on inflammation-related proteomic biomarkers for improving MACE risk prediction. Compared to Royer et al., we selected far fewer number of proteins (7 vs. 114) and observed a similar improvement in the C-index of 0.034 (from 0.716 to 0.750) in the UKB population.

These findings highlight the important role of subclinical inflammation in the prediction of MACE, which was not detected or underestimated by previous studies, which only evaluated CRP. Measuring a small number of more specific proteins for different aspects of subclinical inflammation (similar to the 5 proteins for males and the 4 proteins for females derived from our study) may offer a practical balance between improving model accuracy and maintaining feasibility for routine clinical practice.

### 4.6.2 Biological mechanisms of selected proteins

The 7 inflammation-related proteins selected in our study — IL6, HGF, CDCP1, TNFRSF11B, CCL7, TNFRSF9, and CXCL9 — reflect processes involved in the initiation and progression of CVD, including inflammation, immune responses, and vascular remodeling.

IL6 is a pro-inflammatory cytokine central to the acute-phase response and immune activation (Kopf et al. 1994). By promoting endothelial dysfunction and facilitating immune cell recruitment, it plays a key role in atherogenesis and the destabilization of plaques (Schieffer et al. 2004). Elevated IL6 levels have consistently been associated with cardiovascular outcomes, reflecting chronic low-grade inflammation that underlies many CVD processes (Zhang et al. 2018).

HGF serves as a dual mediator, promoting tissue repair and angiogenesis, while also being implicated in pathological vascular remodeling (Chen et al. 2023b). Elevated HGF levels are thought to indicate vascular injury and reparative efforts in response to endothelial damage (Libetta et al. 2016).

Proteins involved in cell adhesion and apoptosis include CDCP1 and TNFRSF11B. CDCP1 contributes to vascular endothelial integrity and repair mechanisms but may also indicate ongoing endothelial damage in pathological conditions (Zhang et al. 2022). TNFRSF11B functions as a decoy receptor for RANKL and modulates vascular calcification and inflammation (Kiechl et al. 2004). Elevated TNFRSF11B levels have been associated with arterial stiffness, vascular calcification, and an increased risk of cardiovascular events, potentially reflecting maladaptive remodeling in atherosclerosis (Tschiderer et al. 2018).

Chemokines, such as CCL7, CXCL9, and TNFRSF9, reflect immune cell recruitment and inflammation in the vascular microenvironment. CCL7 recruits monocytes and macrophages, which are critical for

foam cell formation and plaque development (Chang et al. 2022). CXCL9 supports T-cell recruitment and promotes chronic inflammatory states within atherosclerotic plaques (Shamsi et al. 2023). TNFRSF9 is expressed on T-cells and endothelial cells and amplifies inflammatory signaling, exacerbating immune cell infiltration and plaque instability (Hashimoto 2021).

### 4.6.3 Strengths and limitations

The main strengths of this study are its large sample size, the robustness of the results confirmed by effective external validation, and rigorous case ascertainment. Additionally, this study is pioneering in deriving sex-specific predictive models for cardiovascular risk by selecting proteomic biomarkers separately for males and females, which reflects biological differences between sexes and supports more precise and personalized risk stratification.

However, several limitations should be acknowledged. First, the study cohorts differed in baseline cardiovascular risk, with the UK Biobank representing a low-risk population and the ESTHER study representing an intermediate-risk population. Second, the Olink platforms used in this study differed and provided cohort-specific relative NPX values, which implied that the variable coefficients derived from the UKB derivation set could not be directly applied to the ESTHER external validation set. We addressed this limitation by splitting the validation cohort into two parts, with one part used for calibration to obtain new ß-coefficients. However, this approach may have led to an overestimation of the model performance in external validation. Third, the fasting status of participants varied markedly between cohorts, with the UK Biobank predominantly comprising non-fasting participants (95.7%) compared to only 10.7% in the ESTHER study. As most of the protein concentrations also varied by fasting status, this limitation was another reason for re-calibrating the model in ESTHER study. Future studies should use absolute protein quantification and study populations with 100% fasting status to obtain coefficients that are more generalizable across populations. Finally, although this study may be applicable to low-to-intermediate risk regions in Europe, larger validation cohorts from high and very high-risk regions are required to extend the generalizability of these findings.

### 4.6.4 Potential clinical implications of the derived model

Regarding practical implementation, we acknowledge that reproducibility, availability, and costs are key considerations. All of the 5 selected proteins for males and the 4 selected ones for women are in the library of the OLINK Flex assay, which can measure up to 30 proteins with absolute concentrations. Thus, the selected proteins could be measured reproducibly and simultaneously at relatively low costs because the OLINK Flex measurements are more affordable compared to large-scale proteomics assays. The last price we got for a scientific project was 70 € per sample for the simultaneous measurement of 21 proteins with the OLINK Flex kit measurement. The availability of this technology is given because the OLINK technology can measure large sample volumes in very short time and the company has contracts with multiple labs in several countries by now that offer these measurements. However, before the developed prediction model can be used in clinical routine, the algorithm needs to be recalibrated with the proteins measured in absolute concentrations with OLINK Flex kits in cohorts with only fasting blood samples. This should ideally be done in different European countries with varying underlying cardiovascular risk, as done before for the calibration of the SCORE2 algorithm (Collaboration und Group 2021).

While the combination of the selected proteins with the cardiac biomarkers NT-proBNP and cTnI achieved an even higher C-index in our study, the added benefit over the "protein-only+SCORE2" model was not statistically significant in neither women, men nor the total population (all $P > 0.05$), suggesting that inflammation-based markers provide a higher predictive value than NT-proBNP and cTnI and could even replace them in a prediction model without much loss in predictive information. These findings underscore the high value of inflammation-related proteins in cardiovascular risk prediction.

While we acknowledge that inflammation-related proteins can fluctuate due to factors such as recent food intake or acute illness (Hruby und Jacques 2019), they seem to be stable enough over time on the population-level for the long-term prediction of MACE. Otherwise they would not have been selected in our study as improvements of the SCORE2 model. Nevertheless, on the individual level, the

prediction model should only be used in practice with fasting blood samples and not in patients with acute infections or acute phases of illnesses known to cause acute, clinically relevant inflammation.

## 4.7 Proteomics and Multi-omics for cardiovascular risk prediction in individuals with type 2 diabetes

In this study of individuals with T2D from the UKB, we developed sex-specific risk models that integrated proteomic biomarkers into the established SCORE2-Diabetes algorithm to improve 10-year prediction of MACE. A total of 15 proteins were selected using a bootstrap-LASSO approach, including one protein (WFDC2) shared across sexes, 8 specific to males, and 6 specific to females. The inclusion of these proteomic signatures markedly enhanced model discrimination for both sexes. Additional gains in model performance by further adding metabolites and a CVD-PRS were only statistically significant among males but not among females. This sex difference was supported by a lack of statistically significant associations of all selected metabolites and the CVD-PRS with MACE incidence in females, whereas the CVD-PRS and one metabolite showed statistically significance in males.

### 4.7.1 Comparison with previous studies & novelty

To our knowledge, this is the first study to develop and validate sex-specific proteomic signatures for cardiovascular risk prediction in individuals with T2D. Most previous studies have investigated proteomic biomarkers in general populations, with limited attention to individuals with T2D (Helgason et al. 2023; Hoogeveen et al. 2020; Nurmohamed et al. 2022; Royer et al. 2024a; Williams et al. 2022). These studies often overlooked the distinct cardiovascular risk profiles and pathophysiological mechanisms unique to diabetic populations. In our prior analysis using the UKB cohort without diabetes (N = 47,382), we applied the same sex-specific biomarker selection framework and identified 18 proteins predictive of MACE (Xie et al. 2025b). Interestingly, only two proteins overlapped with the current diabetes-specific analysis NT-proBNP and WFDC2 in males; WFDC2 alone in females)]. This limited overlap in selected proteins underscores the need to develop proteomic risk models tailored to individuals with T2D.

Prior proteomics-based cardiovascular risk prediction studies in diabetic populations have utilized less extensive protein panels without sex-specific biomarker selection (Ferreira et al. 2021; Li et al. 2025; Looker et al. 2015; Luo et al. 2024; Nowak et al. 2018; Yu et al. 2025). For example, the SUMMIT study employed 6 plasma proteins across five European T2D cohorts (N = 2,310) and observed a notable improvement in AUC from 0.59 to 0.72 for 5-year cardiovascular risk prediction (Looker et al. 2015). Similarly, Nowak et al. utilized 20 proteins measured via the Olink platform in pooled Swedish T2D cohorts (N = 1,211), and achieved a C-index increase from 0.686 to 0.747 (Nowak et al. 2018). The EXAMINE trial (N = 5,131), which focused on T2D patients with recent MI, demonstrated that adding a protein panel including NT-proBNP and Galectin-9 improved the C-index by over 0.05 and significantly enhanced reclassification metrics (Ferreira et al. 2021). Although these studies support the utility of proteomics in T2D populations, their lack of sex-specific biomarker selection potentially limits their predictive precision.

Our study is the first to specifically develop and validate sex-specific proteomic signatures for cardiovascular risk prediction among individuals with T2D. Unlike previous studies that required extensive protein panels, we achieved comparable or superior predictive accuracy with a concise set of biomarkers (9 proteins for males, 7 for females). Additionally, by integrating metabolomics and genomic data, our analysis systematically evaluated the incremental predictive value of multi-omics combinations.

### 4.7.2 Clinical Translation and Feasibility

Previous large-scale proteomics studies often selected a large number of proteins, significantly increasing measurement complexity, cost, and limiting clinical feasibility (Ruiz‐Canela et al. 2017; Yin et al. 2014). Our approach substantially reduces this burden by utilizing concise, sex-specific protein panels (9 proteins for males and 7 for females). For example, it would be possible to design an Olink Focus panel, which allows simultaneous measurement of up to 21 proteins with absolute concentration (https://olink.com/products/olink-focus). After relatively high investment costs once to develop an OLINK Focus panel, the scaling up of measuring it in clinical routine is feasible at low costs. As our

results need to be externally validated anyway before they can be translated in clinical routine, it would be ideal if further studies would use an OLINK Focus panel or a comparable technology to measure the selected proteins with absolute concentration.

Our results suggest that integrating further multi-omics layers (metabolomics and CVD-PRS) would improve the predictive accuracy in males, but the clinical translation of these additional biomarker platforms is challenging in regards of costs and timely measurement as further labs would need to get involved (Lewis und Vassos 2020). Future studies should carefully weigh the incremental predictive gains by additional omics biomarker measurements in men (which were marginal in our study) against the associated measurement costs and feasibility. In women, the marginal improvements in model accuracy did not reach statistical significance in our study, which suggests that a protein-extended SCORE2-Diabetes model is sufficient for MACE risk assessment in women with T2D.

### 4.7.3 Biological mechanisms of selected proteins

The 15 proteins identified in this study are involved in a range of biological processes relevant to cardiovascular risk in individuals with T2D, including pathways related to cardiac stress, fibrosis, inflammation, and metabolic dysfunction. Several proteins, including NT-proBNP, CHGA, and EFHD1, are associated with cardiac stress and neurohumoral signaling (Bayes‐Genis et al. 2023; Gayen et al. 2010; Hou et al. 2016). NT-proBNP is a well-established biomarker of myocardial wall stress and subclinical heart dysfunction, extensively validated in both diabetic and non-diabetic populations (Bibbins-Domingo et al. 2007; Malachias et al. 2022). CHGA is a prohormone co-released with catecholamines that influences autonomic regulation and vascular tone through active peptides such as catestatin (Gayen et al. 2010). EFHD1 is a mitochondrial calcium-binding protein linked to oxidative stress and mitochondrial dysfunction, which may contribute to cardiomyocyte injury in metabolically impaired states such as T2D (Hou et al. 2016).

Several other proteins are involved in fibrosis and extracellular matrix remodeling, processes known to drive plaque instability and adverse ventricular remodeling. WFDC2, the sole protein selected in both sexes in our study, is associated with myocardial and renal fibrosis, potentially via mechanisms

involving collagen deposition regulation and protease inhibition (LeBleu et al. 2013; Yamamoto et al. 2021). TNC (Tenascin-C), POSTN (Periostin), ACTA2 (Actin Alpha 2), and SMOC1 (SPARC-related modular calcium-binding protein 1) are matricellular proteins or contractile regulators that influence fibroblast activation, vascular smooth muscle cell proliferation, and extracellular matrix turnover (Awwad et al. 2015; Guo et al. 2009; Jones et al. 1997; Li et al. 2010). These processes contribute to vascular stiffening, neointimal hyperplasia, and fibrotic remodeling, particularly relevant in T2D patients who experience accelerated atherosclerosis and impaired tissue repair (Dhulekar und Simionescu 2018).

Inflammatory and immune-regulatory mechanisms are also reflected among the selected proteins. IL22 (Interleukin-22) is a pleiotropic cytokine involved in mucosal immunity and tissue regeneration; however, in chronic inflammatory states such as T2D, it has been associated with endothelial dysfunction and atherogenesis (Herder et al. 2017). TRIM21 (E3 ubiquitin-protein ligase TRIM21) functions as an intracellular Fc receptor and E3 ligase, regulating immune signaling and oxidative stress pathways associated with vascular inflammation (Gong et al. 2025). CRNN (Cornulin) is an epithelial stress-response protein associated with inflammation, potentially reflecting systemic inflammatory burden or metabolic stress relevant to cardiovascular risk in T2D (Sundkvist et al. 2018).

Several proteins are involved in metabolic-vascular regulation and endothelial homeostasis. GAST is primarily known for its role in gastrointestinal function but has also been linked to hypertension and subclinical atherosclerosis, possibly through stimulation of vascular smooth muscle (Shan et al. 2018). CTRC is a pancreatic serine protease that may reflect systemic proteolytic activity and metabolic disturbances in diabetes, and has recently been linked to coronary artery disease in a sex-specific manner (Schuermans et al. 2024; Whitcomb 2013). This may reflect a broader disruption of protease-antiprotease balance in T2D, potentially affecting endothelial integrity or vascular remodeling, although the precise mechanisms remain unclear. ENPP5 (Ectonucleotide pyrophosphatase/phosphodiesterase family member 5) is part of the ectonucleotide pyrophosphatase family and has been implicated in cellular senescence, vascular stiffening, and low-grade inflammation (Takaya und Kishi 2024). HS3ST3B1 (Heparan sulfate glucosamine 3-O-sulfotransferase 3B1) encodes a 3-O-sulfotransferase

that modifies the fine structure of heparan sulfate chains (Zhang et al. 2015). By altering sulfation patterns on endothelial surfaces, it may indirectly modulate leukocyte adhesion and inflammatory signaling, as heparan sulfate structure is known to influence cytokine binding and vascular homeostasis. These mechanisms may underlie early metabolic-endothelial dysfunction in T2D and help explain the elevated cardiovascular risk observed in this population.

### 4.7.4 Strengths and limitations

This study's strengths include the availability of multi-omics data from proteomics, metabolomics, and genomics in a large sample of people with T2D with 10-year MACE follow-up. The use of sex-specific modeling accounted for biological differences between males and females and represents a first step toward personalized cardiovascular disease prevention.

However, several limitations must also be considered. First, the lack of external validation means the reported improvements in predictive performance may be overly optimistic, and replication in independent cohorts is necessary to confirm generalizability. Second, the metabolomics data derived from the Nightingale Health platform were limited to approximately 250 biomarkers, and more comprehensive, untargeted metabolomics approaches based on mass spectrometry might yield superior predictive performance. Third, UK Biobank participants were predominantly adults of European ancestry, aged 40-69 years , who are generally healthier than the general diabetic population, potentially limiting the generalizability to other populations. Fourth, although beta coefficients are provided for the proteomics-extended and multi-omics data extended SCORE2-Diabetes model for external validation, recalibration may be necessary when these models are applied to populations with different baseline CVD risk or if alternative omics measurement platforms were used (Schoeler et al. 2023). Fourth, although beta coefficients are provided for the selected biomarkers, recalibration may be necessary when these models are applied to other populations or using alternative omics measurement platforms. Finally, an in-depth cost-effectiveness analysis of implementing these multi-omics risk models into routine clinical practice was beyond the scope of this study but remains essential for evaluating feasibility and cost-effectiveness for large-scale clinical implementation.

## 4.8 Multi-omics for cardiovascular risk prediction in the general population

This study investigated the integration of multi-omics data into the SCORE2 model to improve 10-year risk prediction of MACE. By applying sex-specific biomarker selection, we developed novel risk algorithms for women and men, incorporating a CVD-PRS, metabolites, and proteins, which yielded significant improvements in predictive performance. The optimal model was the combination of proteomics and the CVD-PRS with the SCORE2 model, which achieved predictive accuracy comparable to that of the full multi-omics model in both males and females. These findings highlight the importance of integrating multi-omics data and sex-specific modeling in enhancing precision cardiovascular risk prediction.

### 4.8.1 Comparison with previous studies & novelty

To our knowledge, this is the first study to investigate the predictive value of multi-omics data (PRS, metabolomics, and proteomics) integrated into an established cardiovascular risk model. Previous studies have reported improvements in cardiovascular risk prediction using single omics layers (Helgason et al. 2023; Ho et al. 2024; Huang et al. 2023; Li et al. 2024; McGranaghan et al. 2020; O'Sullivan et al. 2022; Royer et al. 2024b; Samani et al. 2024; Xie et al. 2025a). For example, Li et al. demonstrated that combining a CVD-PRS with the SCORE2 model significantly enhanced risk stratification in the UK Biobank population (Li et al. 2024). A systematic review and meta-analysis analyzing 21 studies with 54,337 participants reported an average increase of 0.04 in the C-index when incorporating metabolomics into cardiovascular risk models (McGranaghan et al. 2020). Similarly, in an Icelandic cohort, an increase in the C-index by 0.014 was observed for predicting MACE when a protein risk score was added to a traditional risk model (Helgason et al. 2023).

Our study showed for the first time that integrating both proteomics and CVD-PRS data is useful to enhance cardiovascular risk prediction by a traditional risk factor model, while metabolomics data add too little C-index improvement to be of clinical relevance after proteomics data were added. Among the individual omics layers, proteomics contributed the greatest predictive value, outperforming both the CVD-PRS and metabolomics in both males and females. Even the addition of the CVD-PRS to a

proteomics extended SCORE2 model added little incremental predictive value (although it was statistically significant), which likely makes the proteomics extended SCORE2 model the model with the highest translation potential because it is almost as good as the model additionally including the CVD-PRS and does not require cost-intensive genetic testing and the labor of using software to compute the CVD-PRS from several million SNPs. While several studies have tested large-scale omics features comprising hundreds of proteins or metabolites, the costs and logistical demands of validating and harmonizing such extensive assays limit their feasibility for clinical translation (Ruiz‐Canela et al. 2017; Yin et al. 2014). Our protein-panel only includes 12 proteins for men and 11 proteins for women, which could be measured at lower costs than full proteomics. However, to measure these few proteins simultaneously with absolute concentrations, the assays need to be developed first with, for example, the OLINK Focus technology, which can measure up to 21 proteins from the whole OLINK Explore protein library (https://olink.com/products/olink-focus). Subsequently, the predictive performance of this OLINK Focus panel for 10-year MACE risk should be validated in external cohorts before this sex-specific prediction algorithm could be used in the clinical routine.

Sex differences in cardiovascular risk and multi-omics biomarkers have been well-documented (Chandramouli et al. 2022; Lau et al. 2019; Tabassum et al. 2023; Titova et al. 2024). Previous studies typically included sex as a covariate in the model derivation process rather than selecting biomarkers separately for males and females. We addressed these limitations by developing sex-specific sparse models, restricting the tested omics features to a small number of highly informative biomarkers. Our previous work has validated the effectiveness of this approach in different populations (Xie et al. 2025a; Xie et al. 2025b). For instance, Ho et al. reported that incorporating 85 proteins into the PREVENT risk model improved the C-index for predicting MACE by 0.054 (from 0.745 to 0.799) and achieved a categorical NRI of 9% (Ho et al. 2024). Similarly, Royer et al. found that adding 114 proteins to the SCORE2 model increased the AUC by 0.031 (from 0.740 to 0.771) and resulted in an NRI of 14.0% (Royer et al. 2024b). Compared to these studies, our sex-specific sparse model achieved a greater improvement in model discrimination (C-index increased from 0.716 to 0.778) (Xie et al. 2025b). Notably, this enhancement was achieved by incorporating only 12 proteins for males and 11 for females into the SCORE2 model, rather than the 85 proteins used by Ho et al. or the 114 proteins used by Royer

et al. Furthermore, the NRI observed in our analysis (19.9%) was higher than that reported by Ho et al. (9%) and Royer et al. (14.0%), highlighting the effectiveness of a sex-specific approach in improving predictive accuracy. By integrating sex-specific biomarker selection while accounting for baseline differences in cardiovascular risk profiles, our model facilitates more personalized and clinically actionable risk stratification. In addition, measuring up to 12 proteins is more cost-effective and practical for clinical implementation compared to measuring 114 proteins, facilitating the translation of our approach into routine clinical practice.

### 4.8.2 Strengths and limitations

This study's strengths include the availability of multi-omics data from all 3 layers (genetics, proteomics, and metabolomics) in a large cohort study with 10-year MACE follow-up. The use of sex-specific modelling considered the biological differences between males and females and is a first step towards personalized CVD prevention. However, several limitations should be noted. First, the metabolomics platform of Nightingale Health is limited to 250 metabolites and more comprehensive, untargeted metabolomics data measured with mass spectrometry methods might have yielded better results. Second, our analyses were primarily based on middle-aged and older adults of European ancestry, necessitating validation in more diverse populations with varying baseline cardiovascular risks and ethnic backgrounds, to ensure the generalizability of our findings. Third, although we provided beta coefficients for selected biomarkers, these coefficients may require recalibration when applied to populations with varying demographic and clinical profiles or when using different measurement platforms. Finally, the study did not include an in-depth cost-effectiveness analysis of integrating multi-omics data into clinical practice, which is crucial for assessing the feasibility of large-scale implementation.

## 4.9 Multi-omics for type 2 diabetes risk prediction

This study investigated whether integrating multi-omics data into a clinical risk model could improve 10-year prediction of T2D in a large, population-based cohort of 23,325 adults without previously diagnosed diabetes. We combined a T2D-PRS, 11 metabolites, and 15 proteins with the clinical CDRS

and observed significant gains in predictive performance in terms of model discrimination and reclassification. While the full multi-omics model achieved the highest discrimination, its incremental benefit over the proteomics-extended model was modest.

### 4.9.1 Comparison with previous studies & novelty

To our knowledge, this is the first large-scale study to systematically evaluate whether integrating genomics, metabolomics, and proteomics improves T2D risk prediction beyond single-omics extensions and established clinical models. Previous studies have primarily examined the predictive value of individual omics layers (Chen und Gerszten 2020; Läll et al. 2017; Satheesh et al. 2020; Zanini et al. 2020). T2D-PRSs have been investigated as early predictors of T2D, but their incremental value beyond clinical models has generally been limited (Meigs 2019; Meigs et al. 2008). Metabolomics and proteomics have revealed biological pathways relevant to insulin resistance and metabolic dysfunction, and several studies have demonstrated considerable abilities in diabetes risk prediction for these omics data (Bragg et al. 2022; Chen und Gerszten 2020; Floegel et al. 2013; Luo et al. 2023; Morze et al. 2022; Rooney et al. 2023; Yao et al. 2024). However, these studies focused on single omics layers and did not assess their joint predictive value. In the only prior multi-omics study, conducted in the EPIC-Norfolk cohort (N = 1,105; 375 incident cases), Zanini et al. reported that integration of top features from genomics, proteomics, and metabolomics significantly improved T2D prediction compared to the clinical CDRS model (C-index: 0.82 to 0.87; $P$ = 0.045), whereas single-omics models showed no significant benefit (Carrasco-Zanini et al. 2024c). Our findings confirm and extend these results in a substantially larger cohort (N = 23,325; 719 incident cases), demonstrating that multi-omics integration significantly enhanced risk prediction (C-index: 0.857 to 0.886). The larger sample size of our cohort allowed more in-depth analyses and elucidated that each omics layer alone significantly improved the predictive abilities of the clinical CDRS, and that proteomics data contributed the largest C-index increase.

Beyond improving overall prediction, multi-omics integration may help identifying individuals who fall below current clinical thresholds for intervention. Existing guidelines recommend preventive action for individuals classified as having prediabetes, defined as $HbA_{1c}$ levels ≥5.7% (≥39 mmol/mol) according

to ADA criteria (Echouffo-Tcheugui und Selvin 2021). However, this threshold may fail to identify all individuals at elevated risk for T2D and its complications. Our results demonstrate that especially among participants without prediabetes ($HbA_{1c}$ <5.7% [<39 mmol/mol]), the multi-omics model significantly improved risk prediction compared to the clinical CDRS. Thus, multi-omics data may allow early identification of high-risk individuals before prediabetes is clinically detectable. This may be explained by the lower discrimination of the clinical model in individuals without prediabetes, allowing greater room for improvement by biomarkers. In contrast, in those with prediabetes, $HbA_{1c}$ and glucose levels are already strongly predictive, which may limit the added value of omics data.

Recent studies in translational omics have similarly emphasized the need for simplified, interpretable biomarker sets to enable real-world application (Carrasco-Zanini et al. 2024a; Carrasco-Zanini et al. 2024b). While the full multi-omics model achieved the highest overall performance in the total population as well as in people with or without prediabetes, adding proteomics alone achieved almost as good C-index increases in all populations. Thus, the proteomics-extended clinical CDRS model may represent a pragmatic balance between accuracy and feasibility for clinical translation considering the high costs and operational complexity of measuring multi-omics data in clinical settings. In fact, the plasma concentration measurements of only 15 proteins would be needed for the suggested model in this study. The OLINK Focus platform enables simultaneous measurement of up to 21 selected proteins from the broad OLINK Explore library with around 3,000 proteins (https://olink.com/products/olink-focus). A developed 15-protein-panel could offer a cost-effective and scalable solution for implementation in clinical workflows.

### 4.9.2 Strengths and limitations

This study has several strengths. This is the largest population-based study to evaluate the integration of proteomics, metabolomics, and a T2D-PRS into an established clinical model for T2D risk prediction. The large samples size allowed testing each omics layer separately and to conduct a sub-group analysis by prediabetes status. Nonetheless, several limitations should be considered. First, the metabolomics data were derived from a targeted nuclear magnetic resonance platform limited to 250 metabolites, which may not capture the full metabolic complexity relevant to T2D. Second, our analyses cannot be

generalized to other populations than those with European ancestry aged 37-73 years. Third, external validation in independent cohorts is essential before adoption in routine care. Fourth, although we provide regression coefficients for the selected biomarkers, these may require recalibration when applied to other populations or measurement platforms.

## 4.10 Conclusion

This dissertation systematically addressed the overarching aim of enhancing the prediction of CVD and T2D by integrating multi-omics biomarkers into established clinical risk models. Leveraging large-scale data from the UK Biobank and ESTHER cohorts, this work consistently demonstrated that molecular biomarkers, when selected and applied with methodological rigor, can significantly improve risk stratification beyond traditional clinical factors. The findings provide a clear path toward more precise, personalized, and sex-specific preventive medicine for cardiometabolic diseases.

Across multiple analyses, this research established that both metabolomics and proteomics offer substantial value in refining risk prediction for both CVD and T2D. For cardiovascular risk prediction, metabolomic signatures were successfully developed and externally validated to enhance the performance of the SCORE2 and SCORE2-Diabetes models in both the general population and in individuals with T2D. For T2D prediction, a novel and externally validated metabolomics-based risk score, the UKB-DRS, was developed, with a concise four-metabolite version offering a highly pragmatic tool for clinical use.

A central and recurring theme throughout this dissertation is the primacy of proteomics. In every multi-omics comparison, whether for CVD or T2D, proteomics consistently emerged as the single most powerful molecular layer, providing the largest incremental improvement in predictive accuracy. This underscores the unique value of circulating proteins in capturing the dynamic pathophysiology that precedes clinical disease onset. Furthermore, this work pioneered the use of sex-specific biomarker selection, a strategy that proved highly effective in improving model performance by accounting for the well-established biological differences in cardiometabolic risk between men and women.

By systematically dissecting the contributions of each omics layer, this dissertation also provides crucial insights into the practicalities of clinical translation. While full multi-omics models consistently yielded the highest statistical performance, the results repeatedly demonstrated that more parsimonious models, particularly those extended with a concise, sex-specific protein panel, achieved nearly equivalent predictive accuracy. For cardiovascular risk prediction in the general population, the optimal combination was proteomics and a CVD-PRS, but a proteomics-only extension of the SCORE2 model represented a powerful and more feasible alternative. For cardiovascular risk prediction in individuals with T2D, the optimal multi-omics extension of the SCORE2-Diabetes model differed by sex, combining proteomics and metabolomics for men, while proteomics alone was sufficient for women. For T2D, the proteomics-extended clinical CDRS similarly offered a pragmatic balance between maximal accuracy and clinical feasibility. These findings are critical, as they suggest that the substantial benefits of omics-based prediction can be realized without the prohibitive cost and complexity of full multi-omics profiling in a routine clinical setting.

In conclusion, this dissertation provides comprehensive evidence that the integration of multi-omics, especially the use of sex-specific proteomics, offers a powerful pathway to enhance cardiometabolic risk stratification. The developed and validated models have the potential to reclassify a significant number of individuals into more accurate risk categories, thereby improving the targeting of preventive interventions. Future research should focus on the external validation of these novel multi-omics models in diverse, non-European populations, assessing their cost-effectiveness, and developing strategies for their implementation to bridge the gap between high-dimensional research and routine clinical care. Ultimately, this work paves the way for a new generation of risk prediction tools that can usher in a new era of precision preventive medicine.

# 5 Summary

## 5.1 English summary

Cardiovascular disease (CVD) and type 2 diabetes (T2D) constitute two of the most significant public health challenges globally, with traditional risk models demonstrating limited accuracy for effective population screening. This dissertation aimed to address this critical gap by systematically investigating the incremental predictive value of metabolomic, proteomic, and genomic biomarkers. Leveraging large-scale data from the UK Biobank (UKB) and the German ESTHER cohorts, this work sought to enhance established clinical risk models and thereby pave the way for more precise, personalized preventive medicine.

The findings for cardiovascular risk prediction consistently affirmed the utility of molecular biomarkers. In the domain of metabolomics, this research established that sex-specific biomarker signatures could significantly augment existing risk scores. For the general population without type 2 diabetes (T2D), a 13-metabolite panel enhanced the SCORE2 model's discrimination (ΔC-index = +0.019), while in the T2D population, a distinct 7-metabolite panel provided a substantial improvement to the SCORE2-Diabetes model (ΔC-index = +0.018). Proteomics offered even more profound predictive power. In the general population without T2D, a parsimonious, sex-specific 18-protein panel markedly improved the SCORE2 model (ΔC-index = +0.062). In individuals with T2D, sex-specific protein signatures yielded a particularly strong predictive gain, increasing the C-index of the SCORE2-Diabetes model by 0.085. The subsequent multi-omics analyses combined the metabolomic and proteomic signatures with polygenic risk scores (PRS). For the general population without T2D, the optimal model for extending SCORE2 combined proteomics and a CVD PRS, increasing the C-index from 0.716 to 0.778. For individuals with T2D, the optimal multi-omics extension of the SCORE2-Diabetes model was sex-dependent: in men, the optimal model combined proteomics and metabolomics (C-index: 0.835), whereas in women, proteomics alone achieved the best model performance (C-index: 0.855).

In the context of T2D risk prediction, a similar stepwise improvement was observed. Metabolomics analyses culminated in the development and successful external validation of a novel score, the UKB-DRS, which significantly improved the clinical Cambridge Diabetes Risk Score (CDRS) (ΔC-index =

+0.019). Notably, a concise 4-metabolite version of this score was developed, offering a highly pragmatic tool for clinical use. Proteomics also substantially enhanced T2D prediction; and a 15-protein panel increased the C-index of the clinical CDRS by 0.029, and a 6-protein inflammation panel proved robust in validation. Finally, the multi-omics integration of metabolomics, proteomics and a T2D PRS yielded the highest overall accuracy, increasing the C-index from 0.857 to 0.886. A key finding from this analysis was that the multi-omics model was particularly effective in improving risk stratification for individuals not yet classified as having prediabetes, highlighting its potential for very early risk detection.

In conclusion, this dissertation provides comprehensive evidence that the integration of multi-omics in established risk prediction models offers a powerful avenue to enhance cardiometabolic risk stratification. A central and recurring theme is the primacy of proteomics, which consistently emerged as the most impactful molecular layer, coupled with the finding that sex-specific models systematically outperform ones developed for the total population. The results strongly suggest that concise, protein-based biomarker panels represent the most pragmatic and effective approach for future clinical translation. These advanced models have the potential to enable earlier and more accurate identification of high-risk individuals, paving the way for a new era of targeted and personalized preventive medicine for cardiometabolic diseases.

## 5.2 Deutsche Zusammenfassung

Herz-Kreislauf-Erkrankungen (CVD) und Typ-2-Diabetes (T2D) stellen weltweit zwei der größten Herausforderungen für die öffentliche Gesundheit dar, und traditionelle Risikomodelle zeigen eine begrenzte Genauigkeit für ein effektives bevölkerungsweites Screening. Ziel dieser Dissertation war es, dieses Problem zu addressieren, indem der inkrementelle Vorhersagewert von metabolomischen, proteomischen und genomischen Biomarkern systematisch untersucht wurde. Unter Nutzung von umfangreichen Daten der UK Biobank (UKB) und der deutschen ESTHER-Kohorte zielte diese Arbeit darauf ab, etablierte klinische Risikomodelle zu verbessern und damit den Weg für eine präzisere, personalisierte Präventivmedizin zu ebnen.

Die Ergebnisse für die kardiovaskuläre Risikovorhersage bestätigten durchweg den Nutzen molekularer Biomarker. Im Bereich der Metabolomik konnte diese Arbeit zeigen, dass geschlechtsspezifische Biomarker-Signaturen bestehende Risikomodelle signifikant verbessern können. Für die Allgemeinbevölkerung ohne T2D verbesserte ein Panel aus 13 Metaboliten die prognostischen Wert des SCORE2-Modells (ΔC-Index = +0,019), während bei T2D-Patienten ein 7-Metaboliten-Panel eine wesentliche Verbesserung des SCORE2-Diabetes-Modells bewirkte (ΔC-Index = +0,018). Die Proteomik bot eine noch bessere Vorhersagekraft. Für die Allgemeinbevölkerung ohne T2D verbesserte ein geschlechtsspezifisches 18-Protein-Panel das SCORE2-Modell deutlich (ΔC-Index = +0,062). Bei Personen mit T2D erzielten geschlechtsspezifische Proteinsignaturen eine besonders starke Vorhersageverbesserung und erhöhten den C-Index des SCORE2-Diabetes-Modells um 0,085 Punkte. Die nachfolgenden Multi-Omics-Analysen kombinierten diese Ergebnisse zur Metabolomik und Proteomik mit polygenetischen Risikoscores (PRS). Für die Allgemeinbevölkerung ohne T2D kombinierte das optimale Modell zur Erweiterung von SCORE2 Proteomik mit einem PRS für CVD, wodurch der C-Index von 0,716 auf 0,778 anstieg. Bei Personen mit T2D war die optimale Multi-Omics-Erweiterung des SCORE2-Diabetes-Modells geschlechtsabhängig: Bei Männern kombinierte das optimale Modell Proteomik und Metabolomik (C-Index: 0,835), während bei Frauen die Proteomik allein die beste Vorhersageleistung erzielte (C-Index: 0,855).

Im Kontext der T2D-Risikovorhersage wurde eine ähnliche schrittweise Verbesserung beobachtet. Metabolomik-Analysen führten zu der Entwicklung und erfolgreichen externen Validierung eines neuen

Scores, des UKB-DRS, der den klinischen Cambridge Diabetes Risk Score (CDRS) signifikant verbesserte (ΔC-Index = +0,019). Bemerkenswerterweise konnte eine reduzierte Version dieses Scores mit vier Metaboliten entwickelt werden, die ein äußerst pragmatisches Werkzeug für die klinische Anwendung darstellt. Die Proteomik verbesserte die T2D-Vorhersage ebenfalls erheblich; ein 15-Protein-Panel erhöhte den C-Index des klinischen CDRS um 0,029, und ein 6-Protein-Panel aus Entzündungsfaktoren erwies sich in der Validierung als robust. Schließlich erzielte die Multi-Omics-Variante aus Metabolomik, Proteomik und einem Diabetes-PRS die höchste Gesamtgenauigkeit und steigerte den C-Index von 0,857 auf 0,886. Ein zentrales Ergebnis dieser Analyse war, dass das Multi-Omics-Modell besonders wirksam bei der Verbesserung der Risikostratifizierung von Personen war, die noch nicht als Prädiabetiker eingestuft waren, was sein Potenzial für eine sehr frühe Risikoerkennung unterstreicht.

Zusammenfassend liefert diese Dissertation umfassende Belege dafür, dass die Integration von Multi-Omics-Daten in etablierte Vorhersagemodelle einen wirkungsvollen Weg zur Verbesserung der kardiometabolischen Risikostratifizierung bietet. Ein zentrales und wiederkehrendes Thema ist die Vormachtstellung der Proteomik, die sich durchweg als die einflussreichste molekulare Ebene erwies, gekoppelt mit der Erkenntnis, dass geschlechtsspezifische Modelle stets Modelle übertreffen, die für die Gesamtbevölkerung entwickelt wurden. Die Ergebnisse deuten stark darauf hin, dass prägnante, auf Proteinen basierende Biomarker-Panels den pragmatischsten und effektivsten Ansatz für die zukünftige klinische Anwendung darstellen. Diese fortschrittlichen Modelle haben das Potenzial, eine frühere und genauere Identifizierung von Hochrisikopersonen zu ermöglichen und ebnen so den Weg für eine neue Ära der gezielten und personalisierten Präventivmedizin für kardiometabolische Erkrankungen.

# 6 References

Abbasi, A., Peelen, L. M., Corpeleijn, E., van der Schouw, Y. T., Stolk, R. P., Spijkerman, A. M., van der, A. D., Moons, K. G., Navis, G., Bakker, S. J. und Beulens, J. W. (2012). **Prediction models for risk of developing type 2 diabetes: systematic literature search and independent external validation study**. BMJ *345*, e5900, doi: 10.1136/bmj.e5900.

Agrawal, P. B., Joshi, M., Marinakis, N. S., Schmitz-Abe, K., Ciarlini, P. D., Sargent, J. C., Markianos, K., De Girolami, U., Chad, D. A. und Beggs, A. H. (2014). **Expanding the phenotype associated with the NEFL mutation: neuromuscular disease in a family with overlapping myopathic and neurogenic findings**. JAMA neurology *71 (11)*, 1413-1420.

Ahmad, E., Lim, S., Lamptey, R., Webb, D. R. und Davies, M. J. (2022). **Type 2 diabetes**. Lancet *400 (10365)*, 1803-1820, doi: 10.1016/s0140-6736(22)01655-5.

Ahola-Olli, A. V., Mustelin, L., Kalimeri, M., Kettunen, J., Jokelainen, J., Auvinen, J., Puukka, K., Havulinna, A. S., Lehtimäki, T. und Kähönen, M. (2019). **Circulating metabolites and the risk of type 2 diabetes: a prospective study of 11,896 young adults from four Finnish cohorts**. Diabetologia *62*, 2298-2309.

Akbari, M. und Hassan-Zadeh, V. (2018). **IL-6 signalling pathways and the development of type 2 diabetes**. Inflammopharmacology *26*, 685-698.

Alshehry, Z. H., Mundra, P. A., Barlow, C. K., Mellett, N. A., Wong, G., McConville, M. J., Simes, J., Tonkin, A. M., Sullivan, D. R., Barnes, E. H., Nestel, P. J., Kingwell, B. A., Marre, M., Neal, B., Poulter, N. R., Rodgers, A., Williams, B., Zoungas, S., Hillis, G. S., Chalmers, J., Woodward, M. und Meikle, P. J. (2016). **Plasma Lipidomic Profiles Improve on Traditional Risk Factors for the Prediction of Cardiovascular Events in Type 2 Diabetes Mellitus**. Circulation *134 (21)*, 1637-1650, doi: 10.1161/circulationaha.116.023233.

Amor, A. J., Catalan, M., Pérez, A., Herreras, Z., Pinyol, M., Sala-Vila, A., Cofán, M., Gilabert, R., Ros, E. und Ortega, E. (2016). **Nuclear magnetic resonance lipoprotein abnormalities in newly-diagnosed type 2 diabetes and their association with preclinical carotid atherosclerosis**. Atherosclerosis *247*, 161-169, doi: 10.1016/j.atherosclerosis.2016.02.014.

Anagnostakis, F., Ko, S., Saadatinia, M., Wang, J., Davatzikos, C. und Wen, J. (2025). **Multi-organ metabolome biological age implicates cardiometabolic conditions and mortality risk**. Nat Commun *16 (1)*, 4871, doi: 10.1038/s41467-025-59964-z.

Anderson, N. L. und Anderson, N. G. (2002). **The human plasma proteome: history, character, and diagnostic prospects**. Mol Cell Proteomics *1 (11)*, 845-867, doi: 10.1074/mcp.r200007-mcp200.

Assarsson, E., Lundberg, M., Holmquist, G., Björkesten, J., Bucht Thorsen, S., Ekman, D., Eriksson, A., Rennel Dickens, E., Ohlsson, S. und Edfeldt, G. (2014). **Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability**. PLoS One *9 (4)*, e95192.

Awwad, K., Hu, J., Shi, L., Mangels, N., Abdel Malik, R., Zippel, N., Fisslthaler, B., Eble, J. A., Pfeilschifter, J. und Popp, R. (2015). **Role of secreted modular calcium-binding protein 1 (SMOC1) in transforming growth factor β signalling and angiogenesis**. Cardiovasc Res *106 (2)*, 284-294.

Bach, F. R. (2008). **Bolasso: model consistent lasso estimation through the bootstrap**. Beitrag präsentiert bei: Proceedings of the 25th international conference on Machine learning.

Balakumar, P., Maung-U, K. und Jagadeesh, G. (2016). **Prevalence and prevention of cardiovascular disease and diabetes mellitus**. Pharmacol Res *113*, 600-609.

Bayes-Genis, A., Docherty, K. F., Petrie, M. C., Januzzi, J. L., Mueller, C., Anderson, L., Bozkurt, B., Butler, J., Chioncel, O. und Cleland, J. G. (2023). **Practical algorithms for early diagnosis of heart failure and heart stress using NT-proBNP: A clinical consensus statement from the Heart Failure Association of the ESC**. Eur J Heart Fail *25 (11)*, 1891-1898.

Bibbins-Domingo, K., Gupta, R., Na, B., Wu, A. H., Schiller, N. B. und Whooley, M. A. (2007). **N-terminal fragment of the prohormone brain-type natriuretic peptide (NT-proBNP), cardiovascular events, and mortality in patients with stable coronary heart disease**. JAMA *297 (2)*, 169-176.

Binuya, M., Engelhardt, E., Schats, W., Schmidt, M. und Steyerberg, E. (2022). **Methodological guidance for the evaluation and updating of clinical prediction models: a systematic review**. BMC Med Res Methodol *22 (1)*, 316.

Bom, M. J., Levin, E., Driessen, R. S., Danad, I., Van Kuijk, C. C., van Rossum, A. C., Narula, J., Min, J. K., Leipsic, J. A. und Pereira, J. P. B. (2019). **Predictive value of targeted proteomics for coronary plaque morphology in patients with suspected coronary artery disease**. EBioMedicine *39*, 109-117.

Bragg, F., Trichia, E., Aguilar-Ramirez, D., Bešević, J., Lewington, S. und Emberson, J. (2022). **Predictive value of circulating NMR metabolic biomarkers for type 2 diabetes risk in the UK Biobank study**. BMC Med *20 (1)*, 159, doi: 10.1186/s12916-022-02354-9.

Brindle, P. M., McConnachie, A., Upton, M. N., Hart, C. L., Davey Smith, G. und Watt, G. C. (2005). **The accuracy of the Framingham risk-score in different socioeconomic groups: a prospective study**. Br J Gen Pract *55 (520)*, 838-845.

Buijsse, B., Simmons, R. K., Griffin, S. J. und Schulze, M. B. (2011). **Risk assessment tools for identifying individuals at risk of developing type 2 diabetes**. Epidemiol Rev *33 (1)*, 46-62, doi: 10.1093/epirev/mxq019.

Carrasco-Zanini, J., Pietzner, M., Davitte, J., Surendran, P., Croteau-Chonka, D. C., Robins, C., Torralbo, A., Tomlinson, C., Grünschläger, F., Fitzpatrick, N., Ytsma, C., Kanno, T., Gade, S., Freitag, D., Ziebell, F., Haas, S., Denaxas, S., Betts, J. C., Wareham, N. J., Hemingway, H., Scott, R. A. und Langenberg, C. (2024a). **Proteomic signatures improve risk prediction for common and rare diseases**. Nat Med *30 (9)*, 2489-2498, doi: 10.1038/s41591-024-03142-z.

Carrasco-Zanini, J., Pietzner, M., Koprulu, M., Wheeler, E., Kerrison, N. D., Wareham, N. J. und Langenberg, C. (2024b). **Proteomic prediction of diverse incident diseases: a machine learning-guided biomarker discovery study using data from a prospective cohort study**. Lancet Digit Health *6 (7)*, e470-e479, doi: 10.1016/s2589-7500(24)00087-6.

Carrasco-Zanini, J., Pietzner, M., Lindbohm, J. V., Wheeler, E., Oerton, E., Kerrison, N., Simpson, M., Westacott, M., Drolet, D. und Kivimaki, M. (2022). **Proteomic signatures for identification of impaired glucose tolerance**. Nat Med *28 (11)*, 2293-2300.

Carrasco-Zanini, J., Pietzner, M., Wheeler, E., Kerrison, N. D., Langenberg, C. und Wareham, N. J. (2024c). **Multi-omic prediction of incident type 2 diabetes**. Diabetologia *67 (1)*, 102-112, doi: 10.1007/s00125-023-06027-x.

Carrel, L. und Willard, H. F. (2005). **X-inactivation profile reveals extensive variability in X-linked gene expression in females**. Nature *434 (7031)*, 400-404, doi: 10.1038/nature03479.

Chandramouli, C., Ting, T. W., Tromp, J., Agarwal, A., Svedlund, S., Saraste, A., Hage, C., Tan, R. S., Beussink-Nelson, L. und Lagerström Fermer, M. (2022). **Sex differences in proteomic correlates of coronary microvascular dysfunction among patients with heart failure and preserved ejection fraction**. Eur J Heart Fail *24 (4)*, 681-684.

Chang, T.-T., Chen, C. und Chen, J.-W. (2022). **CCL7 as a novel inflammatory mediator in cardiovascular disease, diabetes mellitus, and kidney disease**. Cardiovasc Diabetol *21 (1)*, 185.

Chen, C., Wang, J., Pan, D., Wang, X., Xu, Y., Yan, J., Wang, L., Yang, X., Yang, M. und Liu, G. P. (2023a). **Applications of multi-omics analysis in human diseases**. MedComm (2020) *4 (4)*, e315, doi: 10.1002/mco2.315.

Chen, J., Ruan, X., Sun, Y., Lu, S., Hu, S., Yuan, S. und Li, X. (2024). **Multi-omic insight into the molecular networks of mitochondrial dysfunction in the pathogenesis of inflammatory bowel disease**. EBioMedicine *99*, 104934, doi: 10.1016/j.ebiom.2023.104934.

Chen, L., Magliano, D. J. und Zimmet, P. Z. (2012). **The worldwide epidemiology of type 2 diabetes mellitus—present and future perspectives**. Nature reviews endocrinology *8 (4)*, 228-236.

Chen, X., Yu, W., Zhang, J., Fan, X., Liu, X., Liu, Q., Pan, S., Dixon, R. A., Li, P. und Yu, P. (2023b). **Therapeutic angiogenesis and tissue revascularization in ischemic vascular disease**. J Biol Eng *17 (1)*, 13.

Chen, Z. Z. und Gerszten, R. E. (2020). **Metabolomics and Proteomics in Type 2 Diabetes**. Circ Res *126 (11)*, 1613-1627, doi: 10.1161/circresaha.120.315898.

Chiarini, A., Dal Prà, I., Faggian, G., Armato, U. und Luciani, G. B. (2020). **Maladaptive remodeling of pulmonary artery root autografts after Ross procedure: a proteomic study**. The Journal of Thoracic and Cardiovascular Surgery *159 (2)*, 621-632. e623.

Collaboration, E. C. R. und Group, S. W. (2021). **SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe**. Eur Heart J *42 (25)*, 2439-2454.

Connelly, M. A., Otvos, J. D., Shalaurova, I., Playford, M. P. und Mehta, N. N. (2017). **GlycA, a novel biomarker of systemic inflammation and cardiovascular disease risk**. J Transl Med *15 (1)*, 219, doi: 10.1186/s12967-017-1321-6.

Conroy, R. M., Pyörälä, K., Fitzgerald, A. e., Sans, S., Menotti, A., De Backer, G., De Bacquer, D., Ducimetiere, P., Jousilahti, P. und Keil, U. a. (2003). **Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project**. Eur Heart J *24 (11)*, 987-1003.

Cooney, M.-T., Dudina, A., Whincup, P., Capewell, S., Menotti, A., Jousilahti, P., Njølstad, I., Oganov, R., Thomsen, T. und Tverdal, A. (2009a). **Re-evaluating the Rose approach: comparative benefits of the population and high-risk preventive strategies**. European Journal of Preventive Cardiology *16 (5)*, 541-549.

Cooney, M. T., Dudina, A. L. und Graham, I. M. (2009b). **Value and limitations of existing scores for the assessment of cardiovascular risk: a review for clinicians**. J Am Coll Cardiol *54 (14)*, 1209-1227, doi: 10.1016/j.jacc.2009.07.020.

Corey, K. E., Pitts, R., Lai, M., Loureiro, J., Masia, R., Osganian, S. A., Gustafson, J. L., Hutter, M. M., Gee, D. W. und Meireles, O. R. (2022). **ADAMTSL2 protein and a soluble biomarker signature identify at-risk non-alcoholic steatohepatitis and fibrosis in adults with NAFLD**. J Hepatol *76 (1)*, 25-33.

Cronjé, H. T., Mi, M. Y., Austin, T. R., Biggs, M. L., Siscovick, D. S., Lemaitre, R. N., Psaty, B. M., Tracy, R. P., Djoussé, L. und Kizer, J. R. (2023). **Plasma proteomic risk markers of incident type 2 diabetes reflect physiologically distinct components of glucose-insulin homeostasis**. Diabetes *72 (5)*, 666-673.

Cui, M., Cheng, C. und Zhang, L. (2022). **High-throughput proteomics: a methodological mini-review**. Lab Invest *102 (11)*, 1170-1181, doi: 10.1038/s41374-022-00830-7.

Damen, J. A., Hooft, L., Schuit, E., Debray, T. P., Collins, G. S., Tzoulaki, I., Lassale, C. M., Siontis, G. C., Chiocchia, V., Roberts, C., Schlüssel, M. M., Gerry, S., Black, J. A., Heus, P., van der Schouw, Y. T., Peelen, L. M. und Moons, K. G. (2016). **Prediction models for cardiovascular disease risk in the general population: systematic review**. BMJ *353*, i2416, doi: 10.1136/bmj.i2416.

Danesh, J., Wheeler, J. G., Hirschfield, G. M., Eda, S., Eiriksdottir, G., Rumley, A., Lowe, G. D., Pepys, M. B. und Gudnason, V. (2004). **C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease**. N Engl J Med *350 (14)*, 1387-1397, doi: 10.1056/NEJMoa032804.

Davies, M. P. A., Sato, T., Ashoor, H., Hou, L., Liloglou, T., Yang, R. und Field, J. K. (2023). **Plasma protein biomarkers for early prediction of lung cancer**. EBioMedicine *93*, 104686, doi: 10.1016/j.ebiom.2023.104686.

Davyson, E., Shen, X., Gadd, D. A., Bernabeu, E., Hillary, R. F., McCartney, D. L., Adams, M., Marioni, R. und McIntosh, A. M. (2023). **Metabolomic Investigation of Major Depressive Disorder Identifies a Potentially Causal Association With Polyunsaturated Fatty Acids**. Biol Psychiatry, doi: 10.1016/j.biopsych.2023.01.027.

De Rosa, S., Arcidiacono, B., Chiefari, E., Brunetti, A., Indolfi, C. und Foti, D. P. (2018). **Type 2 diabetes mellitus and cardiovascular disease: genetic and epigenetic links**. Front Endocrinol (Lausanne) *9*, 2.

DeFronzo, R. A. und Abdul-Ghani, M. (2011). **Type 2 diabetes can be prevented with early pharmacological intervention**. Diabetes Care *34 (Supplement_2)*, S202-S209.

Dhulekar, J. und Simionescu, A. (2018). **Challenges in vascular tissue engineering for diabetic patients**. Acta Biomater *70*, 25-34, doi: 10.1016/j.actbio.2018.01.008.

Di Cesare, M., Perel, P., Taylor, S., Kabudula, C., Bixby, H., Gaziano, T. A., McGhie, D. V., Mwangi, J., Pervan, B., Narula, J., Pineiro, D. und Pinto, F. J. (2024). **The Heart of the World**. Glob Heart *19 (1)*, 11, doi: 10.5334/gh.1288.

Dreux, A. C., Lamb, D. J., Modjtahedi, H. und Ferns, G. A. (2006). **The epidermal growth factor receptors and their family of ligands: their putative role in atherogenesis**. Atherosclerosis *186 (1)*, 38-53.

Dunér, P., Al-Amily, I. M., Soni, A., Asplund, O., Safi, F., Storm, P., Groop, L., Amisten, S. und Salehi, A. (2016). **Adhesion G protein-coupled receptor G1 (ADGRG1/GPR56) and pancreatic β-cell function**. The Journal of Clinical Endocrinology & Metabolism *101 (12)*, 4637-4645.

Echouffo-Tcheugui, J. B. und Selvin, E. (2021). **Prediabetes and What It Means: The Epidemiological Evidence**. Annu Rev Public Health *42*, 59-77, doi: 10.1146/annurev-publhealth-090419-102644.

Einarson, T. R., Acs, A., Ludwig, C. und Panton, U. H. (2018). **Prevalence of cardiovascular disease in type 2 diabetes: a systematic literature review of scientific evidence from across the world in 2007–2017**. Cardiovasc Diabetol *17*, 1-19.

Elhadad, M. A., Jonasson, C., Huth, C., Wilson, R., Gieger, C., Matias, P., Grallert, H., Graumann, J., Gailus-Durner, V., Rathmann, W., von Toerne, C., Hauck, S. M., Koenig, W., Sinner, M. F., Oprea, T. I., Suhre, K., Thorand, B., Hveem, K., Peters, A. und Waldenberger, M. (2020). **Deciphering the Plasma Proteome of Type 2 Diabetes**. Diabetes *69 (12)*, 2766-2778, doi: 10.2337/db20-0296.

Ezzat, V. A., Duncan, E. R., Wheatcroft, S. B. und Kearney, M. T. (2008). **The role of IGF-I and its binding proteins in the development of type 2 diabetes and cardiovascular disease**. Diabetes, Obesity and Metabolism *10 (3)*, 198-211.

Fall, T., Salihovic, S., Brandmaier, S., Nowak, C., Ganna, A., Gustafsson, S., Broeckling, C. D., Prenni, J. E., Kastenmüller, G. und Peters, A. (2016). **Non-targeted metabolomics combined with genetic analyses identifies bile acid synthesis and phospholipid metabolism as being associated with incident type 2 diabetes**. Diabetologia *59 (10)*, 2114-2124.

Farzadfar, F. (2019). **Cardiovascular disease risk prediction models: challenges and perspectives**. The Lancet Global Health *7 (10)*, e1288-e1289.

Ferreira, J. P., Sharma, A., Mehta, C., Bakris, G., Rossignol, P., White, W. B. und Zannad, F. (2021). **Multi-proteomic approach to predict specific cardiovascular events in patients with diabetes and myocardial infarction: findings from the EXAMINE trial**. Clin Res Cardiol *110 (7)*, 1006-1019, doi: 10.1007/s00392-020-01729-3.

Floegel, A., Stefan, N., Yu, Z., Mühlenbruch, K., Drogan, D., Joost, H. G., Fritsche, A., Häring, H. U., Hrabě de Angelis, M., Peters, A., Roden, M., Prehn, C., Wang-Sattler, R., Illig, T., Schulze, M. B., Adamski, J., Boeing, H. und Pischon, T. (2013). **Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach**. Diabetes *62 (2)*, 639-648, doi: 10.2337/db12-0495.

Folsom, A. R. (2013). **Classical and novel biomarkers for cardiovascular risk prediction in the United States**. J Epidemiol *23 (3)*, 158-162, doi: 10.2188/jea.je20120157.

Frischknecht, R. und Seidenbecher, C. I. (2012). **Brevican: a key proteoglycan in the perisynaptic extracellular matrix of the brain**. The international journal of biochemistry & cell biology *44 (7)*, 1051-1054.

Gaich, G., Chien, J. Y., Fu, H., Glass, L. C., Deeg, M. A., Holland, W. L., Kharitonenkov, A., Bumol, T., Schilske, H. K. und Moller, D. E. (2013). **The effects of LY2405319, an FGF21 analog, in obese human subjects with type 2 diabetes**. Cell Metab *18 (3)*, 333-340.

Ganna, A., Salihovic, S., Sundström, J., Broeckling, C. D., Hedman, A. K., Magnusson, P. K., Pedersen, N. L., Larsson, A., Siegbahn, A., Zilmer, M., Prenni, J., Arnlöv, J., Lind, L., Fall, T. und Ingelsson, E. (2014). **Large-scale metabolomic profiling identifies novel biomarkers for incident coronary heart disease**. PLoS Genet *10 (12)*, e1004801, doi: 10.1371/journal.pgen.1004801.

Ganz, P., Heidecker, B., Hveem, K., Jonasson, C., Kato, S., Segal, M. R., Sterling, D. G. und Williams, S. A. (2016). **Development and Validation of a Protein-Based Risk Score for Cardiovascular Outcomes Among Patients With Stable Coronary Heart Disease**. JAMA *315 (23)*, 2532-2541, doi: 10.1001/jama.2016.5951.

Garg, M., Karpinski, M., Matelska, D., Middleton, L., Burren, O. S., Hu, F., Wheeler, E., Smith, K. R., Fabre, M. A., Mitchell, J., O'Neill, A., Ashley, E. A., Harper, A. R., Wang, Q., Dhindsa, R. S., Petrovski, S. und Vitsios, D. (2024). **Disease prediction with multi-omics and biomarkers empowers case-control genetic discoveries in the UK Biobank**. Nat Genet *56 (9)*, 1821-1831, doi: 10.1038/s41588-024-01898-1.

Gayen, J. R., Zhang, K., RamachandraRao, S. P., Mahata, M., Chen, Y., Kim, H.-S., Naviaux, R. K., Sharma, K., Mahata, S. K. und O'Connor, D. T. (2010). **Role of reactive oxygen species in hyperadrenergic hypertension: biochemical, physiological, and pharmacological evidence from targeted ablation of the chromogranin a (Chga) gene**. Circ Cardiovasc Genet *3 (5)*, 414-425.

Georgakis, M. K., Malik, R., Richardson, T. G., Howson, J. M., Anderson, C. D., Burgess, S., Hovingh, G. K., Dichgans, M. und Gill, D. (2022). **Associations of genetically predicted IL-6 signaling with cardiovascular disease risk across population subgroups**. BMC Med *20 (1)*, 245.

Gillies, C. L., Lambert, P. C., Abrams, K. R., Sutton, A. J., Cooper, N. J., Hsu, R. T., Davies, M. J. und Khunti, K. (2008). **Different strategies for screening and prevention of type 2 diabetes in adults: cost effectiveness analysis**. BMJ *336 (7654)*, 1180-1185, doi: 10.1136/bmj.39545.585289.25.

Golovaty, I., Ritchie, N. D., Tuomilehto, J., Mohan, V., Ali, M. K., Gregg, E. W., Bergman, M. und Moin, T. (2023). **Two decades of diabetes prevention efforts: A call to innovate and revitalize our approach to lifestyle change**. Diabetes Res Clin Pract *198*, 110195, doi: 10.1016/j.diabres.2022.110195.

Gong, X., Xu, L. und Cai, P. (2025). **Friend or foe of tripartite motif-containing protein 21 in cardiovascular disease: A review**. Int J Biol Macromol *308 (Pt 4)*, 142682, doi: 10.1016/j.ijbiomac.2025.142682.

Gore, M. O., Ayers, C. R., Khera, A., deFilippi, C. R., Wang, T. J., Seliger, S. L., Nambi, V., Selvin, E., Berry, J. D., Hundley, W. G., Budoff, M., Greenland, P., Drazner, M. H., Ballantyne, C. M., Levine, B. D. und de Lemos, J. A. (2020). **Combining Biomarkers and Imaging for Short-Term Assessment of Cardiovascular Disease Risk in Apparently Healthy Adults**. J Am Heart Assoc *9 (15)*, e015410, doi: 10.1161/jaha.119.015410.

Griffin, S. J., Little, P. S., Hales, C. N., Kinmonth, A. L. und Wareham, N. J. (2000). **Diabetes risk score: towards earlier detection of type 2 diabetes in general practice**. Diabetes Metab Res Rev *16 (3)*, 164-171, doi: 10.1002/1520-7560(200005/06)16:3<164::aid-dmrr103>3.0.co;2-r.

Grundy, S. M., Brewer, H. B., Jr., Cleeman, J. I., Smith, S. C., Jr. und Lenfant, C. (2004). **Definition of metabolic syndrome: Report of the National Heart, Lung, and Blood Institute/American Heart Association conference on scientific issues related to definition**. Circulation *109 (3)*, 433-438, doi: 10.1161/01.Cir.0000111245.75752.C6.

Gudmundsdottir, V., Zaghlool, S. B., Emilsson, V., Aspelund, T., Ilkov, M., Gudmundsson, E. F., Jonsson, S. M., Zilhão, N. R., Lamb, J. R., Suhre, K., Jennings, L. L. und Gudnason, V. (2020). **Circulating Protein Signatures and Causal Candidates for Type 2 Diabetes**. Diabetes *69 (8)*, 1843-1853, doi: 10.2337/db19-1070.

Gummesson, A., Björnson, E., Fagerberg, L., Zhong, W., Tebani, A., Edfors, F., Schmidt, C., Lundqvist, A., Adiels, M. und Bäckhed, F. (2021). **Longitudinal plasma protein profiling of newly diagnosed type 2 diabetes**. EBioMedicine *63*.

Guo, D.-C., Papke, C. L., Tran-Fadulu, V., Regalado, E. S., Avidan, N., Johnson, R. J., Kim, D. H., Pannu, H., Willing, M. C. und Sparks, E. (2009). **Mutations in smooth muscle alpha-actin (ACTA2) cause coronary artery disease, stroke, and Moyamoya disease, along with thoracic aortic disease**. The American Journal of Human Genetics *84 (5)*, 617-627.

Haffner, S. M., Lehto, S., Rönnemaa, T., Pyörälä, K. und Laakso, M. (1998). **Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction**. N Engl J Med *339 (4)*, 229-234.

Hajny, S., Christoffersen, M., Dalila, N., Nielsen, L. B., Tybjærg-Hansen, A. und Christoffersen, C. (2020). **Apolipoprotein M and risk of type 2 diabetes**. The Journal of Clinical Endocrinology & Metabolism *105 (9)*, 3046-3057.

Harris, K., Oshima, M., Sattar, N., Würtz, P., Jun, M., Welsh, P., Hamet, P., Harrap, S., Poulter, N., Chalmers, J. und Woodward, M. (2020). **Plasma fatty acids and the risk of vascular disease and mortality outcomes in individuals with type 2 diabetes: results from the ADVANCE study**. Diabetologia *63 (8)*, 1637-1647, doi: 10.1007/s00125-020-05162-z.

Hashimoto, K. (2021). **CD137 as an attractive T cell co-stimulatory target in the TNFRSF for immuno-oncology drug development**. Cancers (Basel) *13 (10)*, 2288.

Hasin, Y., Seldin, M. und Lusis, A. (2017). **Multi-omics approaches to disease**. Genome Biol *18 (1)*, 83, doi: 10.1186/s13059-017-1215-1.

Havulinna, A. S., Sysi-Aho, M., Hilvo, M., Kauhanen, D., Hurme, R., Ekroos, K., Salomaa, V. und Laaksonen, R. (2016). **Circulating Ceramides Predict Cardiovascular Outcomes in the Population-Based FINRISK 2002 Cohort**. Arterioscler Thromb Vasc Biol *36 (12)*, 2424-2430, doi: 10.1161/atvbaha.116.307497.

Held, C., White, H. D., Stewart, R. A., Budaj, A., Cannon, C. P., Hochman, J. S., Koenig, W., Siegbahn, A., Steg, P. G. und Soffer, J. (2017). **Inflammatory biomarkers interleukin-6 and C-reactive protein and outcomes in stable coronary heart disease: experiences from the STABILITY (stabilization of atherosclerotic plaque by initiation of darapladib therapy) trial**. Journal of the American Heart Association *6 (10)*, e005077.

Helgason, H., Eiriksdottir, T., Ulfarsson, M. O., Choudhary, A., Lund, S. H., Ivarsdottir, E. V., Hjorleifsson Eldjarn, G., Einarsson, G., Ferkingstad, E., Moore, K. H. S., Honarpour, N., Liu, T., Wang, H., Hucko, T., Sabatine, M. S., Morrow, D. A., Giugliano, R. P., Ostrowski, S. R., Pedersen, O. B., Bundgaard, H., Erikstrup, C., Arnar, D. O., Thorgeirsson, G., Masson, G., Magnusson, O. T., Saemundsdottir, J., Gretarsdottir, S., Steinthorsdottir, V., Thorleifsson, G., Helgadottir, A., Sulem, P., Thorsteinsdottir, U., Holm, H., Gudbjartsson, D. und Stefansson, K. (2023). **Evaluation of Large-Scale Proteomics for Prediction of Cardiovascular Events**. JAMA *330 (8)*, 725-735, doi: 10.1001/jama.2023.13258.

Herder, C., Kannenberg, J. M., Carstensen-Kirberg, M., Huth, C., Meisinger, C., Koenig, W., Peters, A., Rathmann, W., Roden, M. und Thorand, B. (2017). **Serum levels of interleukin-22, cardiometabolic risk factors and incident type 2 diabetes: KORA F4/FF4 study**. Cardiovasc Diabetol *16*, 1-8.

Hernández, M. A. G., Canfora, E. E., Jocken, J. W. E. und Blaak, E. E. (2019). **The Short-Chain Fatty Acid Acetate in Body Weight Control and Insulin Sensitivity**. Nutrients *11 (8)*, doi: 10.3390/nu11081943.

Hicks, K., Walzem, R., Carroll, R. und Turner, N. (2015). **A polyphenol rich sumac sorghum cereal alters lipoprotein subfractions resulting in a more cardioprotective lipoprotein profile**. The FASEB Journal *29*, 923.921.

Hippisley-Cox, J. und Coupland, C. (2017). **Development and validation of QDiabetes-2018 risk prediction algorithm to estimate future risk of type 2 diabetes: cohort study**. BMJ *359*, j5019, doi: 10.1136/bmj.j5019.

Ho, F. K., Mark, P. B., Lees, J. S., Pell, J. P., Strawbridge, R. J., Kimenai, D. M., Mills, N. L., Woodward, M., McMurray, J. J. und Sattar, N. (2025). **A proteomics-based approach for prediction of different cardiovascular diseases and dementia**. Circulation *151 (5)*, 277-287.

Ho, F. K., Mark, P. B., Lees, J. S., Pell, J. P., Strawbridge, R. J., Kimenai, D. M., Mills, N. L., Woodward, M., McMurray, J. J. V., Sattar, N. und Welsh, P. (2024). **A Proteomics-Based Approach for Prediction of Different Cardiovascular Diseases and Dementia**. Circulation, doi: 10.1161/circulationaha.124.070454.

Hong, J., Medzikovic, L., Sun, W., Wong, B., Ruffenach, G., Rhodes, C. J., Brownstein, A., Liang, L. L., Aryan, L., Li, M., Vadgama, A., Kurt, Z., Schwantes-An, T. H., Mickler, E. A., Gräf, S., Eyries, M., Lutz, K. A., Pauciulo, M. W., Trembath, R. C., Perros, F., Montani, D., Morrell, N. W., Soubrier, F., Wilkins, M. R., Nichols, W. C., Aldred, M. A., Desai, A. A., Trégouët, D. A., Umar, S., Saggar, R., Channick, R., Tuder, R. M., Geraci, M. W., Stearman, R. S., Yang, X. und Eghbali, M. (2024). **Integrative Multiomics in the Lung Reveals a Protective Role of Asporin in Pulmonary Arterial Hypertension**. Circulation *150 (16)*, 1268-1287, doi: 10.1161/circulationaha.124.069864.

Hoogeveen, R. M., Pereira, J. P. B., Nurmohamed, N. S., Zampoleri, V., Bom, M. J., Baragetti, A., Boekholdt, S. M., Knaapen, P., Khaw, K. T., Wareham, N. J., Groen, A. K., Catapano, A. L., Koenig, W., Levin, E. und Stroes, E. S. G. (2020). **Improved cardiovascular risk prediction using targeted plasma proteomics in primary prevention**. Eur Heart J *41 (41)*, 3998-4007, doi: 10.1093/eurheartj/ehaa648.

Hou, T., Jian, C., Xu, J., Huang, A. Y., Xi, J., Hu, K., Wei, L., Cheng, H. und Wang, X. (2016). **Identification of EFHD1 as a novel Ca2+ sensor for mitoflash activation**. Cell Calcium *59 (5)*, 262-270.

Hruby, A. und Jacques, P. F. (2019). **Dietary Protein and Changes in Biomarkers of Inflammation and Oxidative Stress in the Framingham Heart Study Offspring Cohort**. Curr Dev Nutr *3 (5)*, nzz019, doi: 10.1093/cdn/nzz019.

Huang, Z., Klaric, L., Krasauskaite, J., Khalid, W., Strachan, M. W. J., Wilson, J. F. und Price, J. F. (2023). **Combining serum metabolomic profiles with traditional risk factors improves 10-year cardiovascular risk prediction in people with type 2 diabetes**. Eur J Prev Cardiol *30 (12)*, 1255-1262, doi: 10.1093/eurjpc/zwad160.

Huth, C., von Toerne, C., Schederecker, F., de Las Heras Gala, T., Herder, C., Kronenberg, F., Meisinger, C., Rathmann, W., Koenig, W., Waldenberger, M., Roden, M., Peters, A., Hauck, S. M. und Thorand, B. (2019). **Protein markers and risk of type 2 diabetes and prediabetes: a targeted proteomics approach in the KORA F4/FF4 study**. Eur J Epidemiol *34 (4)*, 409-422, doi: 10.1007/s10654-018-0475-8.

Jacob, M., Lopata, A. L., Dasouki, M. und Abdel Rahman, A. M. (2019). **Metabolomics toward personalized medicine**. Mass Spectrom Rev *38 (3)*, 221-238, doi: 10.1002/mas.21548.

Jin, Q. und Ma, R. C. W. (2021). **Metabolomics in Diabetes and Diabetic Complications: Insights from Epidemiological Studies**. Cells *10 (11)*, doi: 10.3390/cells10112832.

Jones, P. L., Crack, J. und Rabinovitch, M. (1997). **Regulation of tenascin-C, a vascular smooth muscle cell survival factor that interacts with the αvβ3 integrin to promote epidermal growth factor receptor phosphorylation and growth**. J Cell Biol *139 (1)*, 279-293.

Kahn, R., Robertson, R. M., Smith, R. und Eddy, D. (2008). **The impact of prevention on reducing the burden of cardiovascular disease**. Circulation *118 (5)*, 576-585.

Karczewski, K. J. und Snyder, M. P. (2018). **Integrative omics for health and disease**. Nat Rev Genet *19 (5)*, 299-310, doi: 10.1038/nrg.2018.4.

Kasartzian, D. I. und Tsiampalis, T. (2025). **Transforming Cardiovascular Risk Prediction: A Review of Machine Learning and Artificial Intelligence Innovations**. Life (Basel) *15 (1)*, doi: 10.3390/life15010094.

Kerr, K. F., Wang, Z., Janes, H., McClelland, R. L., Psaty, B. M. und Pepe, M. S. (2014). **Net reclassification indices for evaluating risk-prediction instruments: a critical review**. Epidemiology (Cambridge, Mass.) *25 (1)*, 114.

Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., Natarajan, P., Lander, E. S., Lubitz, S. A., Ellinor, P. T. und Kathiresan, S. (2018). **Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations**. Nat Genet *50 (9)*, 1219-1224, doi: 10.1038/s41588-018-0183-z.

Kiechl, S., Schett, G., Wenning, G., Redlich, K., Oberhollenzer, M., Mayr, A., Santer, P., Smolen, J., Poewe, W. und Willeit, J. (2004). **Osteoprotegerin is a risk factor for progressive atherosclerosis and cardiovascular disease**. Circulation *109 (18)*, 2175-2180.

Kist, J. M., Vos, R. C., Mairuhu, A. T. A., Struijs, J. N., van Peet, P. G., Vos, H. M. M., van Os, H. J. A., Beishuizen, E. D., Sijpkens, Y. W. J., Faiq, M. A., Numans, M. E. und Groenwold, R. H. H. (2023). **SCORE2 cardiovascular risk prediction models in an ethnic and socioeconomic diverse population in the Netherlands: an external validation study**. EClinicalMedicine *57*, 101862, doi: 10.1016/j.eclinm.2023.101862.

Kjellmo, C. A., Hovland, A. und Lappegård, K. T. (2018). **CVD Risk Stratification in the PCSK9 Era: Is There a Role for LDL Subfractions?** Diseases *6 (2)*, doi: 10.3390/diseases6020045.

Kleber, M. E., Delgado, G. E., Lorkowski, S., März, W. und von Schacky, C. (2016). **Omega-3 fatty acids and mortality in patients referred for coronary angiography. The Ludwigshafen Risk and Cardiovascular Health Study**. Atherosclerosis *252*, 175-181, doi: 10.1016/j.atherosclerosis.2016.06.049.

Knowles, J. W. und Ashley, E. A. (2018). **Cardiovascular disease: The rise of the genetic risk score**. PLoS Med *15 (3)*, e1002546, doi: 10.1371/journal.pmed.1002546.

Koenig, W., Löwel, H., Baumert, J. und Meisinger, C. (2004). **C-reactive protein modulates risk prediction based on the Framingham Score: implications for future risk assessment: results from a large cohort study in southern Germany**. Circulation *109 (11)*, 1349-1353, doi: 10.1161/01.Cir.0000120707.98922.E3.

Konrad, T., Vicini, P., Kusterer, K., Höflich, A., Assadkhani, A., Böhles, H. J., Sewell, A., Tritschler, H. J., Cobelli, C. und Usadel, K. H. (1999). **alpha-Lipoic acid treatment decreases serum lactate and pyruvate concentrations and improves glucose effectiveness in lean and obese patients with type 2 diabetes**. Diabetes Care *22 (2)*, 280-287.

Kopf, M., Baumann, H., Freer, G., Freudenberg, M., Lamers, M., Kishimoto, T., Zinkernagel, R., Bluethmann, H. und Köhler, G. (1994). **Impaired immune and acute-phase responses in interleukin-6-deficient mice**. Nature *368 (6469)*, 339-342.

Laakso, M. und Kuusisto, J. (2014). **Insulin resistance and hyperglycaemia in cardiovascular disease development**. Nature reviews endocrinology *10 (5)*, 293-302.

Laffel, L. (1999). **Ketone bodies: a review of physiology, pathophysiology and application of monitoring to diabetes**. Diabetes Metab Res Rev *15 (6)*, 412-426.

Läll, K., Mägi, R., Morris, A., Metspalu, A. und Fischer, K. (2017). **Personalized risk prediction for type 2 diabetes: the potential of genetic risk scores**. Genet Med *19 (3)*, 322-329.

Lau, E. S., Paniagua, S. M., Guseh, J. S., Bhambhani, V., Zanni, M. V., Courchesne, P., Lyass, A., Larson, M. G., Levy, D. und Ho, J. E. (2019). **Sex Differences in Circulating Biomarkers of Cardiovascular Disease**. J Am Coll Cardiol *74 (12)*, 1543-1553, doi: 10.1016/j.jacc.2019.06.077.

Lauber, C., Gerl, M. J., Klose, C., Ottosson, F., Melander, O. und Simons, K. (2022). **Lipidomic risk scores are independent of polygenic risk scores and can predict incidence of diabetes and cardiovascular disease in a large population cohort**. PLoS Biol *20 (3)*, e3001561, doi: 10.1371/journal.pbio.3001561.

LeBleu, V. S., Teng, Y., O'Connell, J. T., Charytan, D., Müller, G. A., Müller, C. A., Sugimoto, H. und Kalluri, R. (2013). **Identification of human epididymis protein-4 as a fibroblast-derived mediator of fibrosis**. Nat Med *19 (2)*, 227-231.

Levey, A. S., Stevens, L. A., Schmid, C. H., Zhang, Y., Castro III, A. F., Feldman, H. I., Kusek, J. W., Eggers, P., Van Lente, F. und Greene, T. (2009). **A new equation to estimate glomerular filtration rate**. Ann Intern Med *150 (9)*, 604-612.

Lewis, C. M. und Vassos, E. (2020). **Polygenic risk scores: from research tools to clinical instruments**. Genome Med *12 (1)*, 44, doi: 10.1186/s13073-020-00742-5.

Li-Gao, R., Hughes, D. A., le Cessie, S., de Mutsert, R., den Heijer, M., Rosendaal, F. R., Willems van Dijk, K., Timpson, N. J. und Mook-Kanamori, D. O. (2019). **Assessment of reproducibility and biological variability of fasting and postprandial plasma metabolite concentrations using 1H NMR spectroscopy**. PLoS One *14 (6)*, e0218549, doi: 10.1371/journal.pone.0218549.

Li, G., Jin, R., Norris, R. A., Zhang, L., Yu, S., Wu, F., Markwald, R. R., Nanda, A., Conway, S. J. und Smyth, S. S. (2010). **Periostin mediates vascular smooth muscle cell migration through the integrins αvβ3 and αvβ5 and focal adhesion kinase (FAK) pathway**. Atherosclerosis *208 (2)*, 358-365.

Li, L., Pang, S., Starnecker, F., Mueller-Myhsok, B. und Schunkert, H. (2024). **Integration of a polygenic score into guideline-recommended prediction of cardiovascular disease**. Eur Heart J *45 (20)*, 1843-1852, doi: 10.1093/eurheartj/ehae048.

Li, M., Duan, L., Cai, Y.-L., Li, H.-Y., Hao, B.-C., Chen, J.-Q. und Liu, H.-B. (2020). **Growth differentiation factor-15 is associated with cardiovascular outcomes in patients with coronary artery disease**. Cardiovasc Diabetol *19*, 1-12.

Li, S., Shin, H. J., Ding, E. L. und van Dam, R. M. (2009). **Adiponectin levels and risk of type 2 diabetes: a systematic review and meta-analysis**. JAMA *302 (2)*, 179-188.

Li, X., Yang, Y., Zhang, B., Lin, X., Fu, X., An, Y., Zou, Y., Wang, J. X., Wang, Z. und Yu, T. (2022). **Lactate metabolism in human health and disease**. Signal Transduct Target Ther *7 (1)*, 305, doi: 10.1038/s41392-022-01151-3.

Li, Y., Li, D., Lin, J., Zhou, L., Yang, W., Yin, X., Xu, C., Cao, Z. und Wang, Y. (2025). **Proteomic signatures of type 2 diabetes predict the incidence of coronary heart disease**. Cardiovasc Diabetol *24 (1)*, 120, doi: 10.1186/s12933-025-02670-3.

Libby, P., Ridker, P. M. und Maseri, A. (2002). **Inflammation and atherosclerosis**. Circulation *105 (9)*, 1135-1143.

Libetta, C., Esposito, P., Martinelli, C., Grosjean, F., Gregorini, M., Rampino, T. und Dal Canton, A. (2016). **Hepatocyte growth factor (HGF) and hemodialysis: physiopathology and clinical implications**. Clin Exp Nephrol *20*, 371-378.

Lin, M., Guo, J., Gu, Z., Tang, W., Tao, H., You, S., Jia, D., Sun, Y. und Jia, P. (2025). **Machine learning and multi-omics integration: advancing cardiovascular translational research and clinical practice**. J Transl Med *23 (1)*, 388.

Lindstrom, J. und Tuomilehto, J. (2003). **The diabetes risk score: a practical tool to predict type 2 diabetes risk**. Diabetes Care *26 (3)*, 725-731.

Liu, S., Tang, W., Cao, J., Shang, M., Sun, H., Gong, J. und Hu, B. (2022). **A comprehensive analysis of HAVCR1 as a prognostic and diagnostic marker for pan-cancer**. Frontiers in Genetics *13*, 904114.

Liu, Y., Wang, H., Zhao, X., Zhang, J., Zhao, Z., Lian, X., Zhang, J., Kong, F., Hu, T. und Wang, T. (2023). **Targeting the Immunoglobulin IGSF9 Enhances Antitumor T-cell Activity and Sensitivity to Anti–PD-1 Immunotherapy**. Cancer Res *83 (20)*, 3385-3399.

Lloyd-Jones, D. M., Liu, K., Tian, L. und Greenland, P. (2006). **Narrative review: Assessment of C-reactive protein in risk prediction for cardiovascular disease**. Ann Intern Med *145 (1)*, 35-42, doi: 10.7326/0003-4819-145-1-200607040-00129.

Lommi, J., Kupari, M., Koskinen, P., Näveri, H., Leinonen, H., Pulkki, K. und Härkönen, M. (1996). **Blood ketone bodies in congestive heart failure**. J Am Coll Cardiol *28 (3)*, 665-672.

Looker, H. C., Colombo, M., Agakov, F., Zeller, T., Groop, L., Thorand, B., Palmer, C. N., Hamsten, A., de Faire, U., Nogoceke, E., Livingstone, S. J., Salomaa, V., Leander, K., Barbarini, N., Bellazzi, R., van Zuydam, N., McKeigue, P. M. und Colhoun, H. M. (2015). **Protein biomarkers for the prediction of cardiovascular disease in type 2 diabetes**. Diabetologia *58 (6)*, 1363-1371, doi: 10.1007/s00125-015-3535-6.

Low Wang, C. C., Hess, C. N., Hiatt, W. R. und Goldfine, A. B. (2016). **Clinical update: cardiovascular disease in diabetes mellitus: atherosclerotic cardiovascular disease and heart failure in type 2 diabetes mellitus–mechanisms, management, and clinical considerations**. Circulation *133 (24)*, 2459-2502.

Lu, X., Xie, Q., Pan, X., Zhang, R., Zhang, X., Peng, G., Zhang, Y., Shen, S. und Tong, N. (2024). **Type 2 diabetes mellitus in adults: pathogenesis, prevention and therapy**. Signal transduction and targeted therapy *9 (1)*, 262.

Luo, H., Bauer, A., Nano, J., Petrera, A., Rathmann, W., Herder, C., Hauck, S. M., Sun, B. B., Hoyer, A., Peters, A. und Thorand, B. (2023). **Associations of plasma proteomics with type 2 diabetes and related traits: results from the longitudinal KORA S4/F4/FF4 Study**. Diabetologia *66 (9)*, 1655-1668, doi: 10.1007/s00125-023-05943-2.

Luo, H., Huemer, M. T., Petrera, A., Hauck, S. M., Rathmann, W., Herder, C., Koenig, W., Hoyer, A., Peters, A. und Thorand, B. (2024). **Association of plasma proteomics with incident coronary heart disease in individuals with and without type 2 diabetes: results from the population-based KORA study**. Cardiovasc Diabetol *23 (1)*, 53, doi: 10.1186/s12933-024-02143-z.

Mackey, R. H., Mora, S., Bertoni, A. G., Wassel, C. L., Carnethon, M. R., Sibley, C. T. und Goff Jr, D. C. (2015). **Lipoprotein particles and incident type 2 diabetes in the multi-ethnic study of atherosclerosis**. Diabetes Care *38 (4)*, 628-636.

Magnusson, L., Espes, D., Casas, R. und Carlsson, P.-O. (2020). **Increased plasma levels of the co-stimulatory proteins CDCP1 and SLAMF1 in patients with autoimmune endocrine diseases**. Front Immunol *11*, 1916.

Malachias, M. V. B., Wijkman, M. O. und Bertoluci, M. C. (2022). **NT-proBNP as a predictor of death and cardiovascular events in patients with type 2 diabetes**. Diabetol Metab Syndr *14 (1)*, 64.

Mangge, H., Zelzer, S., Prüller, F., Schnedl, W. J., Weghuber, D., Enko, D., Bergsten, P., Haybaeck, J. und Meinitzer, A. (2016). **Branched-chain amino acids are associated with cardiometabolic risk profiles found already in lean, overweight and obese young**. J Nutr Biochem *32*, 123-127, doi: 10.1016/j.jnutbio.2016.02.007.

Matsushita, K., Kaptoge, S., Hageman, S. H. J., Sang, Y., Ballew, S. H., Grams, M. E., Surapaneni, A., Sun, L., Arnlov, J., Bozic, M., Brenner, H., Brunskill, N. J., Chang, A. R., Chinnadurai, R., Cirillo, M., Correa, A., Ebert, N., Eckardt, K.-U., Gansevoort, R. T., Gutierrez, O., Hadaegh, F., He, J., Hwang, S.-J., Jafar, T. H., Jassal, S. K., Kayama, T., Kovesdy, C. P., Landman, G. W., Levey, A. S., Lloyd-Jones, D. M., Major, R. W., Miura, K., Muntner, P., Nadkarni, G. N., Nowak, C., Ohkubo, T., Pena, M. J., Polkinghorne, K. R., Sairenchi, T., Schaeffner, E., Schneider, M. P., Shalev, V., Shlipak, M. G., Solbu, M. D., Stempniewicz, N., Tollitt, J., Valdivielso, J. M., van der Leeuw, J., Wang, A. Y.-M., Wen, C.-P., Woodward, M., Yamagishi, K., Yatsuya, H., Zhang, L., Dorresteijn, J. A. N., Di Angelantonio, E., Visseren, F. L. J., Pennells, L. und Coresh, J. (2022). **Including measures of chronic kidney disease to improve cardiovascular risk prediction by SCORE2 and SCORE2-OP**. European Journal of Preventive Cardiology *30 (1)*, 8-16, doi: 10.1093/eurjpc/zwac176.

Mayosi, B. M., Fish, M., Shaboodien, G., Mastantuono, E., Kraus, S., Wieland, T., Kotta, M.-C., Chin, A., Laing, N. und Ntusi, N. B. (2017). **Identification of cadherin 2 (CDH2) mutations in arrhythmogenic right ventricular cardiomyopathy**. Circ Cardiovasc Genet *10 (2)*, e001605.

Mazidi, M., Wright, N., Yao, P., Kartsonaki, C., Millwood, I. Y., Fry, H., Said, S., Pozarickij, A., Pei, P. und Chen, Y. (2023). **Plasma proteomics to identify drug targets for ischemic heart disease**. J Am Coll Cardiol *82 (20)*, 1906-1920.

McGranaghan, P., Saxena, A., Rubens, M., Radenkovic, J., Bach, D., Schleussner, L., Pieske, B., Edelmann, F. und Trippel, T. D. (2020). **Predictive value of metabolomic biomarkers for cardiovascular disease risk: A systematic review and meta-analysis**. Biomarkers *25 (2)*, 101-111.

Meigs, J. B. (2019). **The Genetic Epidemiology of Type 2 Diabetes: Opportunities for Health Translation**. Curr Diab Rep *19 (8)*, 62, doi: 10.1007/s11892-019-1173-y.

Meigs, J. B., Shrader, P., Sullivan, L. M., McAteer, J. B., Fox, C. S., Dupuis, J., Manning, A. K., Florez, J. C., Wilson, P. W. und D'Agostino Sr, R. B. (2008). **Genotype score in addition to common risk factors for prediction of type 2 diabetes**. N Engl J Med *359 (21)*, 2208-2219.

Mittelstrass, K., Ried, J. S., Yu, Z., Krumsiek, J., Gieger, C., Prehn, C., Roemisch-Margl, W., Polonikov, A., Peters, A., Theis, F. J., Meitinger, T., Kronenberg, F., Weidinger, S., Wichmann, H. E., Suhre, K., Wang-Sattler, R., Adamski, J. und Illig, T. (2011). **Discovery of sexual dimorphisms in metabolic and genetic biomarkers**. PLoS Genet *7 (8)*, e1002215, doi: 10.1371/journal.pgen.1002215.

Mohammed, C. J., Lamichhane, S., Connolly, J. A., Soehnlen, S. M., Khalaf, F. K., Malhotra, D., Haller, S. T., Isailovic, D. und Kennedy, D. J. (2022). **A PON for all seasons: comparing paraoxonase enzyme substrates, activity and action including the role of PON3 in health and disease**. Antioxidants *11 (3)*, 590.

Morosi, L. G., Cutine, A. M., Cagnoni, A. J., Manselle-Cocco, M. N., Croci, D. O., Merlo, J. P., Morales, R. M., May, M., Pérez-Sáez, J. M. und Girotti, M. R. (2021). **Control of intestinal inflammation by glycosylation-dependent lectin-driven immunoregulatory circuits**. Science advances *7 (25)*, eabf8630.

Morze, J., Wittenbecher, C., Schwingshackl, L., Danielewicz, A., Rynkiewicz, A., Hu, F. B. und Guasch-Ferré, M. (2022). **Metabolomics and type 2 diabetes risk: an updated systematic review and meta-analysis of prospective cohort studies**. Diabetes Care *45 (4)*, 1013-1024.

Murphy, E. (2011). **Estrogen signaling and cardiovascular disease**. Circ Res *109 (6)*, 687-696, doi: 10.1161/circresaha.110.236687.

Ngo, D., Benson, M. D., Long, J. Z., Chen, Z.-Z., Wang, R., Nath, A. K., Keyes, M. J., Shen, D., Sinha, S. und Kuhn, E. (2021). **Proteomic profiling reveals biomarkers and pathways in type 2 diabetes risk**. JCI insight *6 (5)*, e144392.

Niersmann, C., Carstensen-Kirberg, M., Maalmi, H., Holleczek, B., Roden, M., Brenner, H., Herder, C. und Schöttker, B. (2020). **Higher circulating omentin is associated with increased risk of primary cardiovascular events in individuals with diabetes**. Diabetologia *63 (2)*, 410-418, doi: 10.1007/s00125-019-05017-2.

Nowak, C., Carlsson, A. C., Östgren, C. J., Nyström, F. H., Alam, M., Feldreich, T., Sundström, J., Carrero, J. J., Leppert, J., Hedberg, P., Henriksen, E., Cordeiro, A. C., Giedraitis, V., Lind, L., Ingelsson, E., Fall, T. und Ärnlöv, J. (2018). **Multiplex proteomics for prediction of major cardiovascular events in type 2 diabetes**. Diabetologia *61 (8)*, 1748-1757, doi: 10.1007/s00125-018-4641-z.

Nowak, C., Sundström, J., Gustafsson, S., Giedraitis, V., Lind, L., Ingelsson, E. und Fall, T. (2016). **Protein Biomarkers for Insulin Resistance and Type 2 Diabetes Risk in Two Large Community Cohorts**. Diabetes *65 (1)*, 276-284, doi: 10.2337/db15-0881.

Nurmohamed, N. S., Belo Pereira, J. P., Hoogeveen, R. M., Kroon, J., Kraaijenhof, J. M., Waissi, F., Timmerman, N., Bom, M. J., Hoefer, I. E., Knaapen, P., Catapano, A. L., Koenig, W., de Kleijn, D., Visseren, F. L. J., Levin, E. und Stroes, E. S. G. (2022). **Targeted proteomics improves cardiovascular risk prediction in secondary prevention**. Eur Heart J *43 (16)*, 1569-1577, doi: 10.1093/eurheartj/ehac055.

O'Sullivan, J. W., Raghavan, S., Marquez-Luna, C., Luzum, J. A., Damrauer, S. M., Ashley, E. A., O'Donnell, C. J., Willer, C. J. und Natarajan, P. (2022). **Polygenic risk scores for cardiovascular disease: a scientific statement from the American Heart Association**. Circulation *146 (8)*, e93-e118.

Oliveira, A. G., Araújo, T. G., Carvalho, B. d. M., Rocha, G. Z., Santos, A. und Saad, M. J. (2018). **The role of hepatocyte growth factor (HGF) in insulin resistance and diabetes**. Front Endocrinol (Lausanne) *9*, 370744.

Ortiz-Martínez, M., González-González, M., Martagón, A. J., Hlavinka, V., Willson, R. C. und Rito-Palomares, M. (2022). **Recent developments in biomarkers for diagnosis and screening of type 2 diabetes mellitus**. Curr Diab Rep *22 (3)*, 95-115.

Panagopoulou, V., Deftereos, S., Kossyvakis, C., Raisakis, K., Giannopoulos, G., Bouras, G., Pyrgakis, V. und W Cleman, M. (2013). **NTproBNP: an important biomarker in cardiac diseases**. Curr Top Med Chem *13 (2)*, 82-94.

Patti, G. J., Yanes, O. und Siuzdak, G. (2012). **Innovation: Metabolomics: the apogee of the omics trilogy**. Nat Rev Mol Cell Biol *13 (4)*, 263-269, doi: 10.1038/nrm3314.

Peeters, M. C., Fokkelman, M., Boogaard, B., Egerod, K. L., van de Water, B., Ijzerman, A. P. und Schwartz, T. W. (2015). **The adhesion G protein-coupled receptor G2 (ADGRG2/GPR64) constitutively activates SRE and NFκB and is involved in cell adhesion and migration**. Cell Signal *27 (12)*, 2579-2588.

Pennells, L., Kaptoge, S., Ostergaard, H. B., Read, S. H., Carinci, F., Franch-Nadal, J., Petitjean, C., Taylor, O., Hageman, S. H. und Xu, Z. (2023). **SCORE2-Diabetes: 10-year cardiovascular risk estimation in type 2 diabetes in Europe**. Eur Heart J *44 (28)*, 2544-2556.

Rahman, M., Simmons, R. K., Harding, A. H., Wareham, N. J. und Griffin, S. J. (2008). **A simple risk score identifies individuals at high risk of developing Type 2 diabetes: a prospective cohort study**. Fam Pract *25 (3)*, 191-196, doi: 10.1093/fampra/cmn024.

Reiner, A. P., Wurfel, M. M., Lange, L. A., Carlson, C. S., Nord, A. S., Carty, C. L., Rieder, M. J., Desmarais, C., Jenny, N. S. und Iribarren, C. (2008). **Polymorphisms of the IL1-receptor antagonist gene (IL1RN) are associated with multiple markers of systemic inflammation**. Arterioscler Thromb Vasc Biol *28 (7)*, 1407-1412.

Ressler, S. J., Dang, T. D., Wu, S. M., Dennis, Y. T., Gilbert, B. E., Vyakarnam, A., Yang, F., Schauer, I. G., Barron, D. A. und Rowley, D. R. (2014). **WFDC1 is a key modulator of inflammatory and wound repair responses**. The American Journal of Pathology *184 (11)*, 2951-2964.

Rezaee, M., Putrenko, I., Takeh, A., Ganna, A. und Ingelsson, E. (2020). **Development and validation of risk prediction models for multiple cardiovascular diseases and Type 2 diabetes**. PLoS One *15 (7)*, e0235758.

Ridker, P. M., Rifai, N., Rose, L., Buring, J. E. und Cook, N. R. (2002). **Comparison of C-reactive protein and low-density lipoprotein cholesterol levels in the prediction of first cardiovascular events**. N Engl J Med *347 (20)*, 1557-1565, doi: 10.1056/NEJMoa021993.

Rizza, S., Copetti, M., Rossi, C., Cianfarani, M. A., Zucchelli, M., Luzi, A., Pecchioli, C., Porzio, O., Di Cola, G., Urbani, A., Pellegrini, F. und Federici, M. (2014). **Metabolomics signature improves the prediction of cardiovascular events in elderly subjects**. Atherosclerosis *232 (2)*, 260-264, doi: 10.1016/j.atherosclerosis.2013.10.029.

Robinson, G. A., Pineda-Torra, I., Ciurtin, C. und Jury, E. C. (2022). **Sex differences in lipid metabolism: implications for systemic lupus erythematosus and cardiovascular disease risk**. Frontiers in Medicine *9*, 914016.

Roglic, G. (2016). **Global report on diabetes**, World Health Organization.

Ronit, A., Kirkegaard-Klitbo, D. M., Dohlmann, T. L., Lundgren, J., Sabin, C. A., Phillips, A. N., Nordestgaard, B. G. und Afzal, S. (2020). **Plasma Albumin and Incident Cardiovascular Disease: Results From the CGPS and an Updated Meta-Analysis**. Arterioscler Thromb Vasc Biol *40 (2)*, 473-482, doi: 10.1161/atvbaha.119.313681.

Rooney, M. R., Chen, J., Echouffo-Tcheugui, J. B., Walker, K. A., Schlosser, P., Surapaneni, A., Tang, O., Chen, J., Ballantyne, C. M., Boerwinkle, E., Ndumele, C. E., Demmer, R. T., Pankow, J. S., Lutsey, P. L., Wagenknecht, L. E., Liang, Y., Sim, X., van Dam, R., Tai, E. S., Grams, M. E., Selvin, E. und Coresh, J. (2023). **Proteomic Predictors of Incident Diabetes: Results From the Atherosclerosis Risk in Communities (ARIC) Study**. Diabetes Care *46 (4)*, 733-741, doi: 10.2337/dc22-1830.

Ross, R. (1999). **Atherosclerosis--an inflammatory disease**. N Engl J Med *340 (2)*, 115-126, doi: 10.1056/nejm199901143400207.

Roth, G. A., Mensah, G. A., Johnson, C. O., Addolorato, G., Ammirati, E., Baddour, L. M., Barengo, N. C., Beaton, A. Z., Benjamin, E. J. und Benziger, C. P. (2020a). **Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study**. J Am Coll Cardiol *76 (25)*, 2982-3021.

Roth, G. A., Mensah, G. A., Johnson, C. O., Addolorato, G., Ammirati, E., Baddour, L. M., Barengo, N. C., Beaton, A. Z., Benjamin, E. J., Benziger, C. P., Bonny, A., Brauer, M., Brodmann, M., Cahill, T. J., Carapetis, J., Catapano, A. L., Chugh, S. S., Cooper, L. T., Coresh, J., Criqui, M., DeCleene, N., Eagle, K. A., Emmons-Bell, S., Feigin, V. L., Fernández-Solà, J., Fowkes, G., Gakidou, E., Grundy, S. M., He, F. J., Howard, G., Hu, F., Inker, L., Karthikeyan, G., Kassebaum, N., Koroshetz, W., Lavie, C., Lloyd-Jones, D., Lu, H. S., Mirijello, A., Temesgen, A. M., Mokdad, A., Moran, A. E., Muntner, P., Narula, J., Neal, B., Ntsekhe, M., Moraes de Oliveira, G., Otto, C., Owolabi, M., Pratt, M., Rajagopalan, S., Reitsma, M., Ribeiro, A. L. P., Rigotti, N., Rodgers, A., Sable, C., Shakil, S., Sliwa-Hahnle, K., Stark, B., Sundström, J., Timpel, P., Tleyjeh, I. M., Valgimigli, M., Vos, T., Whelton, P. K., Yacoub, M., Zuhlke, L., Murray, C. und Fuster, V. (2020b). **Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019: Update From the GBD 2019 Study**. J Am Coll Cardiol *76 (25)*, 2982-3021, doi: 10.1016/j.jacc.2020.11.010.

Royer, P., Björnson, E., Adiels, M., Josefson, R., Hagberg, E., Gummesson, A. und Bergström, G. (2024a). **Large-scale plasma proteomics in the UK Biobank modestly improves prediction of major cardiovascular events in a population without previous cardiovascular disease**. European Journal of Preventive Cardiology *31 (14)*, 1681-1689.

Royer, P., Björnson, E., Adiels, M., Josefson, R., Hagberg, E., Gummesson, A. und Bergström, G. (2024b). **Large-scale plasma proteomics in the UK Biobank modestly improves prediction of major cardiovascular events in a population without previous cardiovascular disease**. Eur J Prev Cardiol, doi: 10.1093/eurjpc/zwae124.

Ruiz-Canela, M., Hruby, A., Clish, C. B., Liang, L., Martínez-González, M. A. und Hu, F. B. (2017). **Comprehensive metabolomic profiling and incident cardiovascular disease: a systematic review**. Journal of the American heart association *6 (10)*, e005705.

Samani, N. J., Beeston, E., Greengrass, C., Riveros-McKay, F., Debiec, R., Lawday, D., Wang, Q., Budgeon, C. A., Braund, P. S., Bramley, R., Kharodia, S., Newton, M., Marshall, A., Krzeminski, A., Zafar, A., Chahal, A., Heer, A., Khunti, K., Joshi, N., Lakhani, M., Farooqi, A., Plagnol, V., Donnelly, P., Weale, M. E. und Nelson, C. P. (2024). **Polygenic risk score adds to a clinical risk score in the prediction of cardiovascular disease in a clinical setting**. Eur Heart J *45 (34)*, 3152-3160, doi: 10.1093/eurheartj/ehae342.

Sarwar, N., Gao, P., Seshasai, S. R., Gobin, R., Kaptoge, S., Di Angelantonio, E., Ingelsson, E., Lawlor, D. A., Selvin, E., Stampfer, M., Stehouwer, C. D., Lewington, S., Pennells, L., Thompson, A., Sattar, N., White, I. R., Ray, K. K. und Danesh, J. (2010). **Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies**. Lancet *375 (9733)*, 2215-2222, doi: 10.1016/s0140-6736(10)60484-9.

Satheesh, G., Ramachandran, S. und Jaleel, A. (2020). **Metabolomics-based prospective studies and prediction of type 2 diabetes mellitus risks**. Metab Syndr Relat Disord *18 (1)*, 1-9.

Schieffer, B., Selle, T., Hilfiker, A., Hilfiker-Kleiner, D., Grote, K., Tietge, U. J., Trautwein, C., Luchtefeld, M., Schmittkamp, C. und Heeneman, S. (2004). **Impact of interleukin-6 on plaque development and morphology in experimental atherosclerosis**. Circulation *110 (22)*, 3493-3500.

Schoeler, T., Speed, D., Porcu, E., Pirastu, N., Pingault, J.-B. und Kutalik, Z. (2023). **Participation bias in the UK Biobank distorts genetic associations and downstream analyses**. Nature Human Behaviour *7 (7)*, 1216-1227.

Schöttker, B., Herder, C., Rothenbacher, D., Roden, M., Kolb, H., Müller, H. und Brenner, H. (2013). **Proinflammatory cytokines, adiponectin, and increased risk of primary cardiovascular events in diabetic patients with or without renal dysfunction: results from the ESTHER study**. Diabetes Care *36 (6)*, 1703-1711, doi: 10.2337/dc12-1416.

Schuermans, A., Pournamdari, A. B., Lee, J., Bhukar, R., Ganesh, S., Darosa, N., Small, A. M., Yu, Z., Hornsby, W., Koyama, S., Kooperberg, C., Reiner, A. P., Januzzi, J. L., Honigberg, M. C. und Natarajan, P. (2024). **Integrative proteomic analyses across common cardiac diseases yield mechanistic insights and enhanced prediction**. Nat Cardiovasc Res *3 (12)*, 1516-1530, doi: 10.1038/s44161-024-00567-0.

Schulze, M. B., Hoffmann, K., Boeing, H., Linseisen, J., Rohrmann, S., Möhlig, M., Pfeiffer, A. F., Spranger, J., Thamer, C., Häring, H. U., Fritsche, A. und Joost, H. G. (2007). **An accurate risk score based on anthropometric, dietary, and lifestyle factors to predict the development of type 2 diabetes**. Diabetes Care *30 (3)*, 510-515, doi: 10.2337/dc06-2089.

Schweitzer, K. S., Jinawath, N., Yonescu, R., Ni, K., Rush, N., Charoensawan, V., Bronova, I., Berdyshev, E., Leach, S. M. und Gillenwater, L. A. (2020). **IGSF3 mutation identified in patient with severe COPD alters cell function and motility**. JCI insight *5 (14)*.

Seah, J. Y. H., Hong, Y., Cichońska, A., Sabanayagam, C., Nusinovici, S., Wong, T. Y., Cheng, C.-Y., Jousilahti, P., Lundqvist, A. und Perola, M. (2022). **Circulating metabolic biomarkers are consistently associated with type 2 diabetes risk in Asian and European populations**. The Journal of Clinical Endocrinology & Metabolism *107 (7)*, e2751-e2761.

Shah Ebrahim, S. E. und Smith, G. (2001). **Exporting failure? Coronary heart disease and stroke in developing countries**.

Shah, S. H., Sun, J. L., Stevens, R. D., Bain, J. R., Muehlbauer, M. J., Pieper, K. S., Haynes, C., Hauser, E. R., Kraus, W. E., Granger, C. B., Newgard, C. B., Califf, R. M. und Newby, L. K. (2012). **Baseline metabolomic profiles predict cardiovascular events in patients at risk for coronary artery disease**. Am Heart J *163 (5)*, 844-850.e841, doi: 10.1016/j.ahj.2012.02.005.

Shahidi, F. und Ambigaipalan, P. (2018). **Omega-3 Polyunsaturated Fatty Acids and Their Health Benefits**. Annu Rev Food Sci Technol *9*, 345-381, doi: 10.1146/annurev-food-111317-095850.

Shamsi, A., Roghani, S. A., Abdan, Z., Soufivand, P., Pournazari, M., Bahrehmand, F., Vafaei, A., Salari, N., Soroush, M. G. und Taghadosi, M. (2023). **CXCL9 and its receptor CXCR3, an important link between inflammation and cardiovascular risks in RA patients**. Inflammation *46 (6)*, 2374-2385.

Shan, J., Bai, X., Han, L., Yuan, Y., Yang, J. und Sun, X. (2018). **Association between atherosclerosis and gastric biomarkers concerning Helicobacter pylori infection in a Chinese healthy population**. Exp Gerontol *112*, 97-102, doi: 10.1016/j.exger.2018.09.009.

Shankar, R. R., Charchar, F. J., Eckert, G. J., Saha, C., Tu, W., Dominiczak, A. F. und Pratt, J. H. (2007). **Studies of an association in boys of blood pressure and the Y chromosome**. Am J Hypertens *20 (1)*, 27-31, doi: 10.1016/j.amjhyper.2006.06.013.

Siam, N. H., Snigdha, N. N., Tabasumma, N. und Parvin, I. (2024). **Diabetes Mellitus and Cardiovascular Disease: Exploring Epidemiology, Pathophysiology, and Treatment Strategies**. Rev Cardiovasc Med *25 (12)*, 436, doi: 10.31083/j.rcm2512436.

Singh, H., Mhasawade, V. und Chunara, R. (2022). **Generalizability challenges of mortality risk prediction models: A retrospective analysis on a multi-center database**. PLOS Digit Health *1 (4)*, e0000023, doi: 10.1371/journal.pdig.0000023.

Soininen, P., Kangas, A. J., Würtz, P., Suna, T. und Ala-Korpela, M. (2015). **Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics**. Circ Cardiovasc Genet *8 (1)*, 192-206.

Soininen, P., Kangas, A. J., Würtz, P., Tukiainen, T., Tynkkynen, T., Laatikainen, R., Järvelin, M. R., Kähönen, M., Lehtimäki, T., Viikari, J., Raitakari, O. T., Savolainen, M. J. und Ala-Korpela, M. (2009). **High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism**. Analyst *134 (9)*, 1781-1785, doi: 10.1039/b910205a.

Sozzani, S., Del Prete, A., Bonecchi, R. und Locati, M. (2015). **Chemokines as effector and target molecules in vascular biology**. Cardiovasc Res *107 (3)*, 364-372.

Spitz, M. R., Wu, X. und Mills, G. (2005). **Integrative epidemiology: from risk assessment to outcome prediction**. J Clin Oncol *23 (2)*, 267-275.

Stegemann, C., Pechlaner, R., Willeit, P., Langley, S. R., Mangino, M., Mayr, U., Menni, C., Moayyeri, A., Santer, P., Rungger, G., Spector, T. D., Willeit, J., Kiechl, S. und Mayr, M. (2014). **Lipidomics profiling and risk of cardiovascular disease in the prospective population-based Bruneck study**. Circulation *129 (18)*, 1821-1831, doi: 10.1161/circulationaha.113.002500.

Stekhoven, D. J. und Bühlmann, P. (2012). **MissForest--non-parametric missing value imputation for mixed-type data**. Bioinformatics *28 (1)*, 112-118, doi: 10.1093/bioinformatics/btr597.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T. und Collins, R. (2015). **UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age**. PLoS Med *12 (3)*, e1001779, doi: 10.1371/journal.pmed.1001779.

Sun, B. B., Chiou, J., Traylor, M., Benner, C., Hsu, Y.-H., Richardson, T. G., Surendran, P., Mahajan, A., Robins, C. und Vasquez-Grinnell, S. G. (2023). **Plasma proteomic associations with genetics and health in the UK Biobank**. Nature *622 (7982)*, 329-338.

Sun, C., Cheng, X., Xu, J., Chen, H., Tao, J., Dong, Y., Wei, S., Chen, R., Meng, X., Ma, Y., Tian, H., Guo, X., Bi, S., Zhang, C., Kang, J., Zhang, M., Lv, H., Shang, Z., Lv, W., Zhang, R. und Jiang, Y. (2024). **A review of disease risk prediction methods and applications in the omics era**. Proteomics *24 (18)*, e2300359, doi: 10.1002/pmic.202300359.

Sun, H., Saeedi, P., Karuranga, S., Pinkepank, M., Ogurtsova, K., Duncan, B. B., Stein, C., Basit, A., Chan, J. C. N., Mbanya, J. C., Pavkov, M. E., Ramachandaran, A., Wild, S. H., James, S., Herman, W. H., Zhang, P., Bommer, C., Kuo, S., Boyko, E. J. und Magliano, D. J. (2022). **IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045**. Diabetes Res Clin Pract *183*, 109119, doi: 10.1016/j.diabres.2021.109119.

Sundkvist, A., Myte, R., Bodén, S., Enroth, S., Gyllensten, U., Harlid, S. und van Guelpen, B. (2018). **Targeted plasma proteomics identifies a novel, robust association between cornulin and Swedish moist snuff**. Sci Rep *8 (1)*, 2320.

Tabassum, R., Ruotsalainen, S., Ottensmann, L., Gerl, M. J., Klose, C., Tukiainen, T., Pirinen, M., Simons, K., Widén, E. und Ripatti, S. (2022). **Lipidome- and Genome-Wide Study to Understand Sex Differences in Circulatory Lipids**. J Am Heart Assoc *11 (19)*, e027103, doi: 10.1161/jaha.122.027103.

Tabassum, R., Widén, E. und Ripatti, S. (2023). **Effect of biological sex on human circulating lipidome: An overview of the literature**. Atherosclerosis, 117274, doi: 10.1016/j.atherosclerosis.2023.117274.

Takaya, K. und Kishi, K. (2024). **Regulation of ENPP5, a senescence-associated secretory phenotype factor, prevents skin aging**. Biogerontology *25 (3)*, 529-542, doi: 10.1007/s10522-024-10096-9.

Thomas, J. M., Huuskes, B. M., Sobey, C. G., Drummond, G. R. und Vinh, A. (2022). **The IL-18/IL-18R1 signalling axis: Diagnostic and therapeutic potential in hypertension and chronic kidney disease**. Pharmacol Ther *239*, 108191.

Thompson, D. J., Wells, D., Selzam, S., Peneva, I., Moore, R., Sharp, K., Tarran, W. A., Beard, E. J., Riveros-Mckay, F., Giner-Delgado, C., Palmer, D., Seth, P., Harrison, J., Futema, M., McVean, G., Plagnol, V., Donnelly, P. und Weale, M. E. (2024). **A systematic evaluation of the performance and properties of the UK Biobank Polygenic Risk Score (PRS) Release**. PLoS One *19 (9)*, e0307270, doi: 10.1371/journal.pone.0307270.

Thorand, B., Kolb, H., Baumert, J., Koenig, W., Chambless, L., Meisinger, C., Illig, T., Martin, S. und Herder, C. (2005). **Elevated levels of interleukin-18 predict the development of type 2 diabetes: results from the MONICA/KORA Augsburg Study, 1984-2002**. Diabetes *54 (10)*, 2932-2938, doi: 10.2337/diabetes.54.10.2932.

Tibshirani, R. (1996). **Regression shrinkage and selection via the lasso**. Journal of the Royal Statistical Society Series B: Statistical Methodology *58 (1)*, 267-288.

Tillin, T., Hughes, A. D., Wang, Q., Würtz, P., Ala-Korpela, M., Sattar, N., Forouhi, N. G., Godsland, I. F., Eastwood, S. V. und McKeigue, P. M. (2015). **Diabetes risk and amino acid profiles: cross-sectional and prospective analyses of ethnicity, amino acids and diabetes in a South Asian and European cohort from the SABRE (Southall And Brent REvisited) Study**. Diabetologia *58*, 968-979.

Timmerman, N., Waissi, F., Dekker, M., van de Pol, Q. Y., van Bennekom, J., Schoneveld, A., Klein Avink, M. J. M., de Winter, R. J., Pasterkamp, G., de Borst, G. J. und de Kleijn, D. P. V. (2021). **Pre-Operative Plasma Extracellular Vesicle Proteins are Associated with a High Risk of Long Term Secondary Major Cardiovascular Events in Patients Undergoing Carotid Endarterectomy**. Eur J Vasc Endovasc Surg *62 (5)*, 705-715, doi: 10.1016/j.ejvs.2021.06.039.

Titova, O. E., Yuan, S., Byberg, L., Baron, J. A., Lind, L., Michaëlsson, K. und Larsson, S. C. (2024). **Plasma proteome and incident myocardial infarction: sex-specific differences**. Eur Heart J *45 (43)*, 4647-4657, doi: 10.1093/eurheartj/ehae658.

Tokgozoglu, L. und Torp-Pedersen, C. (2021). **Redefining cardiovascular risk prediction: is the crystal ball clearer now?** Eur Heart J *42 (25)*, 2468-2471, doi: 10.1093/eurheartj/ehab310.

Torkamani, A., Wineinger, N. E. und Topol, E. J. (2018). **The personal and clinical utility of polygenic risk scores**. Nat Rev Genet *19 (9)*, 581-590, doi: 10.1038/s41576-018-0018-x.

Tossell, K., Andreae, L. C., Cudmore, C., Lang, E., Muthukrishnan, U., Lumsden, A., Gilthorpe, J. D. und Irving, C. (2011). **Lrrn1 is required for formation of the midbrain–hindbrain boundary and organiser through regulation of affinity differences between midbrain and hindbrain cells in chick**. Dev Biol *352 (2)*, 341-352.

Trares, K., Bhardwaj, M., Perna, L., Stocker, H., Petrera, A., Hauck, S. M., Beyreuther, K., Brenner, H. und Schöttker, B. (2022). **Association of the inflammation-related proteome with dementia development at older age: results from a large, prospective, population-based cohort study**. Alzheimers Res Ther *14 (1)*, 128, doi: 10.1186/s13195-022-01063-y.

Trares, K., Wiesenfarth, M., Stocker, H., Perna, L., Petrera, A., Hauck, S. M., Beyreuther, K., Brenner, H. und Schöttker, B. (2024). **Addition of inflammation-related biomarkers to the CAIDE model for risk prediction of all-cause dementia, Alzheimer's disease and vascular dementia in a prospective study**. Immun Ageing *21 (1)*, 23, doi: 10.1186/s12979-024-00427-2.

Traylor, M., Mäkelä, K.-M., Kilarski, L. L., Holliday, E. G., Devan, W. J., Nalls, M. A., Wiggins, K. L., Zhao, W., Cheng, Y.-C. und Achterberg, S. (2014). **A novel MMP12 locus is associated with large artery atherosclerotic stroke using a genome-wide age-at-onset informed approach**. PLoS Genet *10 (7)*, e1004469.

Tschiderer, L., Klingenschmid, G., Nagrani, R., Willeit, J., Laukkanen, J. A., Schett, G., Kiechl, S. und Willeit, P. (2018). **Osteoprotegerin and Cardiovascular Events in High-Risk Populations: Meta-Analysis of 19 Prospective Studies Involving 27 450 Participants**. Journal of the American Heart Association *7 (16)*, e009012.

Vaduganathan, M., Mensah, G. A., Turco, J. V., Fuster, V. und Roth, G. A. (2022). **The global burden of cardiovascular diseases and risk: a compass for future health** (American College of Cardiology Foundation Washington DC2361-2371.

van Geloven, N., Giardiello, D., Bonneville, E. F., Teece, L., Ramspek, C. L., van Smeden, M., Snell, K. I. E., van Calster, B., Pohar-Perme, M., Riley, R. D., Putter, H. und Steyerberg, E. (2022). **Validation of prediction models in the presence of competing risks: a guide through modern methods**. BMJ *377*, e069249, doi: 10.1136/bmj-2021-069249.

van Heck, J. I., Ajie, M., Joosten, L. A., Tack, C. J. und Stienstra, R. (2025). **Circulating inflammatory proteins are elevated in type 1 and type 2 diabetes and associated to complications**. Diabetes, Obesity and Metabolism *27 (2)*, 719-728.

van Trier, T. J., Snaterse, M., Boekholdt, S. M., Scholte Op Reimer, W. J. M., Hageman, S. H. J., Visseren, F. L. J., Dorresteijn, J. A. N., Peters, R. J. G. und Jørstad, H. T. (2024). **Validation of Systematic Coronary Risk Evaluation 2 (SCORE2) and SCORE2-Older Persons in the EPIC-Norfolk prospective population cohort**. Eur J Prev Cardiol *31 (2)*, 182-189, doi: 10.1093/eurjpc/zwad318.

Vaura, F., Palmu, J., Aittokallio, J., Kauko, A. und Niiranen, T. (2022). **Genetic, Molecular, and Cellular Determinants of Sex-Specific Cardiovascular Traits**. Circ Res *130 (4)*, 611-631, doi: 10.1161/circresaha.121.319891.

Vendrell, J. und Chacón, M. R. (2013). **TWEAK: a new player in obesity and diabetes**. Front Immunol *4*, 488.

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A. und Yang, J. (2017). **10 Years of GWAS Discovery: Biology, Function, and Translation**. Am J Hum Genet *101 (1)*, 5-22, doi: 10.1016/j.ajhg.2017.06.005.

Visseren, F. L., Mach, F., Smulders, Y. M., Carballo, D., Koskinas, K. C., Bäck, M., Benetos, A., Biffi, A., Boavida, J.-M. und Capodanno, D. (2022). **2021 ESC Guidelines on cardiovascular disease prevention in clinical practice: Developed by the Task Force for cardiovascular disease prevention in clinical practice with representatives of the European Society of Cardiology and 12 medical societies With the special contribution of the European Association of Preventive Cardiology (EAPC)**. European journal of preventive cardiology *29 (1)*, 5-115.

Wallentin, L., Eriksson, N., Olszowka, M., Grammer, T. B., Hagström, E., Held, C., Kleber, M. E., Koenig, W., März, W., Stewart, R. A. H., White, H. D., Åberg, M. und Siegbahn, A. (2021). **Plasma proteins associated with cardiovascular death in patients with chronic coronary heart disease: A retrospective study**. PLoS Med *18 (1)*, e1003513, doi: 10.1371/journal.pmed.1003513.

Wang, J., Miao, Y., Wicklein, R., Sun, Z., Wang, J., Jude, K. M., Fernandes, R. A., Merrill, S. A., Wernig, M. und Garcia, K. C. (2021). **RTN4/NoGo-receptor binding to BAI adhesion-GPCRs regulates neuronal development**. Cell *184 (24)*, 5869-5885. e5825.

Wang, T. J., Larson, M. G., Vasan, R. S., Cheng, S., Rhee, E. P., McCabe, E., Lewis, G. D., Fox, C. S., Jacques, P. F. und Fernandez, C. (2011). **Metabolite profiles and the risk of developing diabetes**. Nat Med *17 (4)*, 448-453.

Wang, Z., Zhu, C., Nambi, V., Morrison, A. C., Folsom, A. R., Ballantyne, C. M., Boerwinkle, E. und Yu, B. (2019). **Metabolomic Pattern Predicts Incident Coronary Heart Disease**. Arterioscler Thromb Vasc Biol *39 (7)*, 1475-1482, doi: 10.1161/atvbaha.118.312236.

Welsh, P., Rankin, N., Li, Q., Mark, P. B., Würtz, P., Ala-Korpela, M., Marre, M., Poulter, N., Hamet, P., Chalmers, J., Woodward, M. und Sattar, N. (2018). **Circulating amino acids and the risk of macrovascular, microvascular and mortality outcomes in individuals with type 2 diabetes: results from the ADVANCE trial**. Diabetologia *61 (7)*, 1581-1591, doi: 10.1007/s00125-018-4619-x.

Werner, B. und Yadav, S. (2023). **Phosphoregulation of the septin cytoskeleton in neuronal development and disease**. Cytoskeleton *80 (7-8)*, 275-289.

Whitcomb, D. C. (2013). **Genetic risk factors for pancreatic disorders**. Gastroenterology *144 (6)*, 1292-1302, doi: 10.1053/j.gastro.2013.01.069.

Wik, L., Nordberg, N., Broberg, J., Björkesten, J., Assarsson, E., Henriksson, S., Grundberg, I., Pettersson, E., Westerberg, C. und Liljeroth, E. (2021). **Proximity extension assay in combination with next-generation sequencing for high-throughput proteome-wide analysis**. Mol Cell Proteomics *20*.

Williams, S. A., Kivimaki, M., Langenberg, C., Hingorani, A. D., Casas, J. P., Bouchard, C., Jonasson, C., Sarzynski, M. A., Shipley, M. J. und Alexander, L. (2019). **Plasma protein patterns as comprehensive indicators of health**. Nat Med *25 (12)*, 1851-1857.

Williams, S. A., Ostroff, R., Hinterberg, M. A., Coresh, J., Ballantyne, C. M., Matsushita, K., Mueller, C. E., Walter, J., Jonasson, C., Holman, R. R., Shah, S. H., Sattar, N., Taylor, R., Lean, M. E., Kato, S., Shimokawa, H., Sakata, Y., Nochioka, K., Parikh, C. R., Coca, S. G., Omland, T., Chadwick, J., Astling, D., Hagar, Y., Kureshi, N., Loupy, K., Paterson, C., Primus, J., Simpson, M., Trujillo, N. P. und Ganz, P. (2022). **A proteomic surrogate for cardiovascular outcomes that is sensitive to multiple mechanisms of change in risk**. Sci Transl Med *14 (639)*, eabj9625, doi: 10.1126/scitranslmed.abj9625.

Wilson, P. W., D'Agostino, R. B., Levy, D., Belanger, A. M., Silbershatz, H. und Kannel, W. B. (1998). **Prediction of coronary heart disease using risk factor categories**. Circulation *97 (18)*, 1837-1847.

Wörheide, M. A., Krumsiek, J., Kastenmüller, G. und Arnold, M. (2021). **Multi-omics integration in biomedical research - A metabolomics-centric review**. Anal Chim Acta *1141*, 144-162, doi: 10.1016/j.aca.2020.10.038.

Würtz, P., Havulinna, A. S., Soininen, P., Tynkkynen, T., Prieto-Merino, D., Tillin, T., Ghorbani, A., Artati, A., Wang, Q., Tiainen, M., Kangas, A. J., Kettunen, J., Kaikkonen, J., Mikkilä, V., Jula, A., Kähönen, M., Lehtimäki, T., Lawlor, D. A., Gaunt, T. R., Hughes, A. D., Sattar, N., Illig, T., Adamski, J., Wang, T. J., Perola, M., Ripatti, S., Vasan, R. S., Raitakari, O. T., Gerszten, R. E., Casas, J. P., Chaturvedi, N., Ala-Korpela, M. und Salomaa, V. (2015). **Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts**. Circulation *131 (9)*, 774-785, doi: 10.1161/circulationaha.114.013116.

Würtz, P., Soininen, P., Kangas, A. J., Rönnemaa, T., Lehtimäki, T., Kähönen, M., Viikari, J. S., Raitakari, O. T. und Ala-Korpela, M. (2013). **Branched-chain and aromatic amino acids are predictors of insulin resistance in young adults**. Diabetes Care *36 (3)*, 648-655, doi: 10.2337/dc12-0895.

Xie, R., Sha, S., Peng, L., Holleczek, B., Brenner, H. und Schöttker, B. (2025a). **Metabolomics data improve 10-year cardiovascular risk prediction with the SCORE2 algorithm for the general population without cardiovascular disease or diabetes**. Eur J Prev Cardiol, doi: 10.1093/eurjpc/zwaf254.

Xie, R., Vlaski, T., Sha, S., Brenner, H. und Schöttker, B. (2025b). **Sex-specific proteomic signatures improve cardiovascular risk prediction for the general population without cardiovascular disease or diabetes**. J Adv Res, doi: 10.1016/j.jare.2025.03.034.

Xuan, Y., Gào, X., Anusruti, A., Holleczek, B., Jansen, E., Muhlack, D. C., Brenner, H. und Schöttker, B. (2019). **Association of Serum Markers of Oxidative Stress With Incident Major Cardiovascular Events, Cancer Incidence, and All-Cause Mortality in Type 2 Diabetes Patients: Pooled Results From Two Cohort Studies**. Diabetes Care *42 (8)*, 1436-1445, doi: 10.2337/dc19-0292.

Yamamoto, M., Hanatani, S., Araki, S., Izumiya, Y., Yamada, T., Nakanishi, N., Ishida, T., Yamamura, S., Kimura, Y., Arima, Y., Nakamura, T., Takashio, S., Yamamoto, E., Sakamoto, K., Kaikita, K., Matsushita, K., Morimoto, S., Ito, T. und Tsujita, K. (2021). **HE4 Predicts Progressive Fibrosis and Cardiovascular Events in Patients With Dilated Cardiomyopathy**. J Am Heart Assoc *10 (15)*, e021069, doi: 10.1161/jaha.120.021069.

Yang, W. S., Chen, Y. Y., Chen, P. C., Hsu, H. C., Su, T. C., Lin, H. J., Chen, M. F., Lee, Y. T. und Chien, K. L. (2019). **Association between Plasma N-6 Polyunsaturated Fatty Acids Levels and the Risk of Cardiovascular Disease in a Community-based Cohort Study**. Sci Rep *9 (1)*, 19298, doi: 10.1038/s41598-019-55686-7.

Yao, P., Iona, A., Pozarickij, A., Said, S., Wright, N., Lin, K., Millwood, I., Fry, H., Kartsonaki, C. und Mazidi, M. (2024). **Proteomic analyses in diverse populations improved risk prediction and identified new drug targets for type 2 diabetes**. Diabetes Care *47 (6)*, 1012-1019.

Yin, X., Subramanian, S., Hwang, S.-J., O'Donnell, C. J., Fox, C. S., Courchesne, P., Muntendam, P., Gordon, N., Adourian, A. und Juhasz, P. (2014). **Protein biomarkers of new-onset cardiovascular disease: prospective study from the systems approach to biomarker research in cardiovascular disease initiative**. Arterioscler Thromb Vasc Biol *34 (4)*, 939-945.

Yoshida, Y., Chen, Z. und Baudier, R. (2022). **Sex differences in the progression of metabolic risk factors in diabetes development. JAMA Netw Open. 2022; 5 (7): e2222070** (

Yu, H., Zhang, J., Qian, F., Yao, P., Xu, K., Wu, P., Li, R., Qiu, Z., Li, R., Zhu, K., Li, L., Geng, T., Yu, X., Li, D., Liao, Y., Pan, A. und Liu, G. (2025). **Large-Scale Plasma Proteomics Improves Prediction of Peripheral Artery Disease in Individuals With Type 2 Diabetes: A Prospective Cohort Study**. Diabetes Care *48 (3)*, 381-389, doi: 10.2337/dc24-1696.

Zanini, J. C., Pietzner, M. und Langenberg, C. (2020). **Integrating Genetics and the Plasma Proteome to Predict the Risk of Type 2 Diabetes**. Curr Diab Rep *20 (11)*, 60, doi: 10.1007/s11892-020-01340-w.

Zhang, B., Li, X.-L., Zhao, C.-R., Pan, C.-L. und Zhang, Z. (2018). **Interleukin-6 as a predictor of the risk of cardiovascular disease: a meta-analysis of prospective epidemiological studies**. Immunol Invest *47 (7)*, 689-699.

Zhang, J., Li, J., Huang, C., Xue, L., Peng, Y., Fu, Q., Gao, L., Zhang, J. und Li, W. (2008). **Targeted knockout of the mouse βB2-crystallin gene (Crybb2) induces age-related cataract**. Invest Ophthalmol Vis Sci *49 (12)*, 5476-5483.

Zhang, L., Borjini, N., Lun, Y., Parab, S., Asonye, G., Singh, R., Bell, B. A., Bonilha, V. L., Ivanov, A. und Fox, D. A. (2022). **CDCP1 regulates retinal pigmented epithelial barrier integrity for the development of experimental autoimmune uveitis**. JCI insight *7 (18)*.

Zhang, L., Song, K., Zhou, L., Xie, Z., Zhou, P., Zhao, Y., Han, Y., Xu, X. und Li, P. (2015). **Heparan sulfate D-glucosaminyl 3-O-sulfotransferase-3B1 (HS3ST3B1) promotes angiogenesis and proliferation by induction of VEGF in acute myeloid leukemia cells**. J Cell Biochem *116 (6)*, 1101-1112, doi: 10.1002/jcb.25066.

Zheng, X. L. (2015). **ADAMTS13 and von Willebrand factor in thrombotic thrombocytopenic purpura**. Annu Rev Med *66 (1)*, 211-225.

Zhou, J.-W., Tsui, S. K., Ng, M. C., Geng, H., Li, S.-K., So, W.-Y., Ma, R. C., Wang, Y., Tao, Q. und Chen, Z.-Y. (2011). **Apolipoprotein M gene (APOM) polymorphism modifies metabolic and disease traits in type 2 diabetes**. PLoS One *6 (2)*, e17324.

Zhu, H., Guo, D., Li, K., Yan, W., Tan, Y., Wang, X., Treiber, F. A., Chao, J., Snieder, H. und Dong, Y. (2008). **Prostasin: a possible candidate gene for human hypertension**. Am J Hypertens *21 (9)*, 1028-1033.

Zong, G., Liu, G., Willett, W. C., Wanders, A. J., Alssema, M., Zock, P. L., Hu, F. B. und Sun, Q. (2019). **Associations between linoleic acid intake and incident type 2 diabetes among US men and women**. Diabetes Care *42 (8)*, 1406-1413.

# 7 Own publications and contributions

This dissertation uses data from the ESTHER study and the UK Biobank. Six accepted publications and three submitted papers resulted from this dissertation. In these nine papers, my involvement spanned the literature review, comprehensive data analyses, interpretation of results, and the writing and refinement of the manuscripts associated with this dissertation. Substantial contributions from my supervisor were significantly noted in the management of the original databases and extensive revisions of the manuscripts prior to publication. Additionally, co-authors provided meticulous comments on the methodology, the interpretation of the results, and the discussion in each of the publications.

**All the findings, including texts, tables, figures, and appendix presented in this dissertation, have been or will be published in peer-reviewed journals with high impact factors through the following nine papers:**

1. Xie, R., Seum, T., Sha, S., Trares, K., Holleczek, B., Brenner, H. und Schöttker, B. (2025). **Improving 10-year cardiovascular risk prediction in patients with type 2 diabetes with metabolomics**. Cardiovasc Diabetol *24 (1)*, 18, doi: 10.1186/s12933-025-02581-3.

2. Xie, R., Sha, S., Peng, L., Holleczek, B., Brenner, H. und Schöttker, B. (2025). **Metabolomics data improve 10-year cardiovascular risk prediction with the SCORE2 algorithm for the general population without cardiovascular disease or diabetes**. Eur J Prev Cardiol, doi: 10.1093/eurjpc/zwaf254.

3. Xie, R., Herder, C., Sha, S., Peng, L., Brenner, H. und Schoettker, B. (2025). **Novel type 2 diabetes prediction score based on traditional risk factors and circulating metabolites: Model derivation and validation in two large cohort studies**. eClinicalMedicine *79*.

4. Xie, R., Vlaski, T., Trares, K., Herder, C., Holleczek, B., Brenner, H. und Schöttker, B. (2025). **Large-Scale Proteomics Improve Risk Prediction for Type 2 Diabetes**. Diabetes Care, doi: 10.2337/dc24-2478.

5. Xie, R., Vlaski, T., Sha, S., Brenner, H. und Schöttker, B. (2025). **Sex-specific proteomic signatures improve cardiovascular risk prediction for the general population without cardiovascular disease or diabetes**. J Adv Res, doi: 10.1016/j.jare.2025.03.034.

6. Xie, R., Sha, S., Brenner, H. und Schöttker, B. (2025). **Circulating inflammation-related proteome improves cardiovascular risk prediction. Results from two large European cohort studies**. Eur J Epidemiol, doi: 10.1007/s10654-025-01285-y.

7. Xie, R., Herder, C., Sha, S., Brenner, H. und Schöttker, B. (202X). **Sex-specific proteomic signatures for 10-year cardiovascular risk prediction in people with type 2 diabetes**. [submitted to journal]

8. Xie, R., Bhardwaj, M., Sha, S., Peng, L., Vlaski, T., Brenner, H. und Schöttker, B. (202X). **Sex-specific prediction of major cardiovascular events in apparently healthy individuals with multi-omics data** [under review]

9. Xie, R., Herder, C., und Schöttker, B. (202X). **Large-scale multi-omics enhance risk prediction for type 2 diabetes** [under review]

In this dissertation, Chapter 2.1, 3.1, 4.1, part of the Chapter 1 and 4.10 are derived from **publication 1**. My own contributions to this publication include the literature review, performing the data analyses, interpretation of the results, and writing the manuscript. My supervisor, Prof. Dr. Ben Schöttker, made significant contributions to the revision of the manuscript and played a crucial role in proposing the methodology for variable selection.

Chapters 2.2, 3.2, 4.2, as well as part of the Chapter 1 and 4.10 in this dissertation originate from **publication 2**. For this paper, I conducted the literature review, data analysis, and interpretation of results, as well as drafted and refined the manuscript. The manuscript was revised by my supervisor Prof. Dr. Ben Schöttker.

Chapters 2.3, 3.3, 4.3, as well as part of the Chapter 1 and 4.10 of this dissertation stem from **publication 3**. My contributions to this publication encompass the literature review, the data analyses, the interpretation of results, and the writing of the manuscript, which was revised by my supervisor Prof. Dr. Ben Schöttker.

Chapters 2.4, 3.4, 4.4, and part of the Chapter 1 and 4.10 of this dissertation are grounded on **publication 4**. My own contributions to this publication include the literature review, data analyses, interpretation

of the results, and writing and refinement of the manuscript, which was revised by my supervisor Prof. Dr. Ben Schöttker.

Chapters 2.5, 3.5, 4.5, and part of the Chapter 1 and 4.10 of this dissertation are based on **publication 5**. For this publication, I was responsible for the literature review, all data analyses, the interpretation of the findings, and the drafting and revision of the manuscript. The manuscript was revised by my supervisor Prof. Dr. Ben Schöttker.

The work presented in chapters 2.6, 3.6, 4.6, and part of the Chapter 1 and 4.10 originates from **publication 6**. My own contributions to this publication include the literature review, performing the data analyses, interpretation of the results, and writing the manuscript, which was revised by my supervisor Prof. Dr. Ben Schöttker.

Chapters 2.7, 3.7, 4.7, and part of the Chapter 1 and 4.10 of this dissertation are derived from the work in **publication 7**, which is currently submitted. My contributions to this work encompass the literature review, data analyses, interpretation of results, and the writing of the manuscript, which was revised by my supervisor Prof. Dr. Ben Schöttker.

Chapters 2.8, 3.8, 4.8, and part of the Chapter 1 and 4.10 are grounded on **publication 8**, which is in submission. My contributions to this publication include the literature review, the implementation of data analysis, interpretation of the results, and writing the manuscript, which was revised by my supervisor Prof. Dr. Ben Schöttker.

The research in chapters 2.9, 3.9, 4.9, and part of the Chapter 1 and 4.10 is based on **publication 9**, which is also submitted. For this paper, I conducted the literature review, data analyses, and interpretation of results, and I was responsible for drafting and refining the manuscript, which was revised by my supervisor Prof. Dr. Ben Schöttker.

### Further own and co-authored publications:

Chen, L. J., Herder, C., Xie, R., Brenner, H. und Schöttker, B. (2025). **Comparison of the metabolic profiles and their cardiovascular event risks of metformin users versus insulin users. A cohort study of people with type 2 diabetes from the UK Biobank**. Diabetes Res Clin Pract *222*, 112108, doi: 10.1016/j.diabres.2025.112108.

Cui, Z., Xie, R., Lu, X., Schomburg, L., Brenner, H. und Schöttker, B. (2025). **Associations of selenium status with all-cause and cause-specific mortality: a systematic review and meta-analysis of cohort studies**. Redox Biol *85*, 103755, doi: 10.1016/j.redox.2025.103755.

Sha, S., Xie, R., Gwenzi, T., Wang, Y., Brenner, H. und Schöttker, B. (2025). **Real-world evidence for an association of vitamin D supplementation with atherosclerotic cardiovascular disease in the UK Biobank**. Clin Nutr *49*, 118-127, doi: 10.1016/j.clnu.2025.04.017.

Xie, R., Liu, X., Wu, H., Liu, M. und Zhang, Y. (2023). **Associations between systemic immune-inflammation index and abdominal aortic calcification: Results of a nationwide survey**. Nutr Metab Cardiovasc Dis *33 (7)*, 1437-1443, doi: 10.1016/j.numecd.2023.04.015.

Xie, R., Ning, Z., Xiao, M., Li, L., Liu, M. und Zhang, Y. (2023). **Dietary inflammatory potential and biological aging among US adults: a population-based study**. Aging Clin Exp Res *35 (6)*, 1273-1281, doi: 10.1007/s40520-023-02410-1.

Xie, R. und Zhang, Y. (2023). **Associations between dietary flavonoid intake with hepatic steatosis and fibrosis quantified by VCTE: Evidence from NHANES and FNDDS**. Nutr Metab Cardiovasc Dis *33 (6)*, 1179-1189, doi: 10.1016/j.numecd.2023.03.005.

**Moderated poster presentation at scientific conference:**

- **The European Society of Cardiology (ESC) Congress 2025**

  29th August-1stSeptember, 2025, Madrid, Spain

  "Sex-specific prediction of major cardiovascular events in apparently healthy individuals with multi-omics data"

# Appendix

Appendix 1. Flow chart for participant inclusion and exclusion ..... 197

Appendix 2. Laboratory methods applied in the UK Biobank (UKB) and ESTHER study ..... 198

Appendix 3. Definition of endpoint major cardiovascular event (MACE) ..... 199

Appendix 4. Flow chart for participant inclusion and exclusion ..... 2004

Appendix 5. Flow chart for participant inclusion and exclusion ..... 2015

Appendix 6. Diagnosis and medication codes for assessment of type 2 diabetes status in primary and secondary healthcare and death registry records and verbal interview ..... 2026

Appendix 7. Flow chart for participant inclusion and exclusion ..... 2037

Appendix 8. Flow chart for participant inclusion and exclusion ..... 2048

Appendix 9. Flow chart for participant inclusion and exclusion ..... 2059

Appendix 10. Flow chart for participant inclusion and exclusion ..... 206

Appendix 11. Flow chart for participant inclusion and exclusion ..... 207

Appendix 12. Flow chart for participant inclusion and exclusion ..... 208

Appendix 13. ß-coefficients of the variables of model 2 “Original SCORE2-Diabetes extended by NMR metabolites” for 10-year prediction of MACE in the derivation set of the UK Biobank ..... 209

Appendix 14. Associations between selected metabolites and MACE across sexes in the internal validation cohort (30% of UK Biobank, N=3,077) ..... 2104

Appendix 15. Associations between selected metabolites and MACE across sexes in the external validation cohort (ESTHER study, N=1,039) ..... 2115

Appendix 16. ß-coefficients of the variables of model 2 “Original SCORE2-Diabetes extended by NMR metabolites” for 10-year prediction of MACE in the derivation set of the UK Biobank ... 2126

Appendix 17. Calibration curves of the SCORE2-Diabetes model and its combination with the metabolites for 10-year cardiovascular risk prediction in the internal validation set (30% of UK Biobank, N=3,077) ..... 2137

Appendix 18. Calibration curves of the SCORE2-Diabetes model and its combination with the metabolites for 10-year cardiovascular risk prediction in the external validation (ESTHER study, N=1,039) ..... 2148

Appendix 19. ß-coefficients of the variables of model 4 “Age, smoking, SBP, diabetes age at diagnosis, and NMR metabolites” for 10-year prediction of MACE in the derivation set of the UK Biobank ..... 2159

Appendix 20. Comparison of the predictive performance of model 1 “Original SCORE2-Diabetes” and model 4 “Age, smoking, SBP, diabetes age at diagnosis, and NMR metabolites” ..... 216

Appendix 21. ß-coefficients of the variables of the SCORE2 model extended by 13 metabolites for 10-year prediction of major cardiovascular events in the derivation set of the UK Biobank ..... 217

Appendix 22. Associations between selected metabolites and MACE across sexes in the external validation cohort (ESTHER study, N=5,578) ..... 218

Appendix 23. ß-coefficients of the variables of the SCORE2 model extended by 13 metabolites for 10-year prediction of MACE in the derivation set of the UK Biobank ..... 220

Appendix 24. Calibration curves of the SCORE2 model and its combination with the metabolites for 10-year cardiovascular risk prediction in the internal validation cohort (30% of UK Biobank, N=56,112) ..... 2215

Appendix 25. Calibration curves of the SCORE2 model and its combination with the metabolites for 10-year cardiovascular risk prediction in the external validation cohort (ESTHER study, N=5,578) ..... 2226

Appendix 26. The incremental discrimination of each metabolite for the model after the selected metabolites were added to SCORE2 separately in different sexes in the external validation cohort (ESTHER study, N=5,578) ..... 2237

Appendix 27. Correlation matrix of Pearson correlation coefficients for the 11 selected metabolites ..... 2248

Appendix 28. Hazard ratios per standard deviation increments with 95% confidence intervals and p-values for the associations of the selected metabolites for incident diabetes prediction in the test set (30% of UK Biobank, N=25,870) and external validation set (ESTHER, N=4,383) ..... 2259

Appendix 29. ß-coefficients of the variables of the UKB-DRS for 10-year prediction of T2D ..... 226

Appendix 30. ROC curves for the UKB-DRS, concise UKB-DRS, and clinical CDRS in internal validation (30% UK Biobank, N=25,870) and external validation (ESTHER, N=4,383) ..... 227

Appendix 31. Calibration curves for the UKB-DRS, Concise UKB-DRS, and Clinical CDRS in internal validation (30% UK Biobank, N=25,870) and external validation (ESTHER, N=4,383) . 229

Appendix 32. Comparison of the predictive performance of the clinical CDRS plus glucose and the UKB-DRS for 10-year type 2 diabetes risk prediction ..... 230

Appendix 33. Subgroup analyses comparing the C-index of the UKB-DRS and the clinical CDRS in internal validation (30% UK Biobank, N=25,870) and external validation (ESTHER, N=4,383) ..... 2314

Appendix 34. NRI and IDI if each of the 11 selected metabolite is added one by one to the clinical CDRS in the test set (30% of UK Biobank, N=25,870) and external validation set (ESTHER, N=4,383) ..... 2325

Appendix 35. Correlation matrix showing Spearman correlation coefficients between BMI, $HbA_{1c}$, and the selected proteins ..... 2336

Appendix 36. ROC curves of the clinical CDRS with and without proteomics data for 10-year type 2 diabetes risk prediction in the internal validation (30% of UK Biobank, N=6,339) and external validation (70% of ESTHER, N=3,118) ..... 2357

Appendix 37. ß-coefficients of the variables of the 15 proteins extended clinical CDRS for 10-year prediction of type 2 diabetes ..... 2368

Appendix 38. ß-coefficients of the variables of the 6 proteins extended clinical CDRS for 10-year prediction of type 2 diabetes in UK Biobank and ESTHER study ..... 2379

Appendix 39. Calibration curves of the clinical CDRS with and without proteomics data for 10-year type 2 diabetes risk prediction in the internal validation (30% of UK Biobank, N=6,339) and external validation (70% of ESTHER, N=3,118) ..... 238

Appendix 40. Comparison of 6 selected proteins levels by fasting status in the ESTHER study .. 239

Appendix 41. Comparison of the predictive performance of the clinical Cambridge Diabetes Risk Score with and without proteomic biomarkers for 10-year type 2 diabetes risk prediction in UK Biobank without non-random selected participants ..... 240

Appendix 42. Comparison of the incremental improvements in the C-index by proteomic biomarkers added to the clinical CDRS in the total population, subjects with and without pre-diabetes status in the UKB internal validation and external validation with the ESTHER study ..... 241

Appendix 43. Comparison of the C-index of the 6- and 15-protein-extended models derived in this study with different protein selection approaches with LASSO bootstrapping and suggested protein panels from the literature in the internal validation set (30% of UK Biobank, N=6,339) ..... 2424

Appendix 44. ß-coefficients of the variables of the SCORE2 model extended by 18 proteins for 10-year prediction of MACE ..... 2435

Appendix 45. Receiver operating characteristic (ROC) curves for the SCORE2 model with and without proteomic biomarkers in the validation cohort (30% of UK Biobank, N=14,216) ..... 2456

Appendix 46. Metrics of the predictive performance of the (protein extended) SCORE2 model for 10-year MACE risk including only randomly selected participants ..... 2467

Appendix 47. Reclassification of the SCORE2 model with and without proteomic biomarkers in the validation set (30% of UK Biobank, N=14,216) ..... 2478

Appendix 48. Calibration curves of the SCORE2 model with and without proteomics data for 10-year MACE risk prediction in the validation set (30% of UK Biobank, N=14,216) ..... 2489

Appendix 49. Comparison of the standardized circulating concentration of the 7 selected proteins by fasting status in the ESTHER study ..... 249

Appendix 50. ß-coefficients of the variables of the SCORE2 model extended by 7 proteins for 10-year prediction of MACE in the derivation set of the UK Biobank ..... 250

Appendix 51. ß-coefficients of the variables of the SCORE2 model extended by 7 proteins for 10-year prediction of MACE in the calibration set of the ESTHER study ..... 251

Appendix 52. Receiver operating characteristic (ROC) curves of the SCORE2 model with and without proteomics data for 10-year MACE risk prediction in the internal validation (30% of UK Biobank, N=14,216) and external validation (70% of ESTHER, N=4,397) ..... 252

Appendix 53. Calibration curves of the SCORE2 model with and without proteomics data for 10-year MACE risk prediction in the internal validation (30% of UK Biobank, N=14,216) and external validation (70% of ESTHER, N=4,397) ..... 2534

Appendix 54. Metrics of the predictive performance of the SCORE2 model for 10-year MACE risk without and with extension by proteins in random selected participants ..... 2545

Appendix 55. Comparison of C-index for five risk prediction models based on SCORE2 and its extensions in the internal validation (30% of UK Biobank, N=14,216) ..... 256

Appendix 56. Comparison of adjusted standardized circulating protein concentrations by statin use in the UK Biobank ..... 257

Appendix 57. Performance of the SCORE2 model and its extension with inflammation-related proteins stratified by statin use in the internal validation (30% of UK Biobank, N=14,216) ..... 258

Appendix 58. ß-coefficients of the variables of the SCORE2-Diabetes model extended by 15 proteins for 10-year prediction of MACE ..... 259

Appendix 59. ß-coefficients of the variables of the SCORE2-Diabetes model extended by selected multi-omics biomarkers for 10-year prediction of MACE ..... 261

Appendix 60. ROC curves comparing the discrimination of the SCORE2-Diabetes model with and without proteomics extendsion for 10-year MACE risk (N=1,751) ..... 263

Appendix 61. ROC curves comparing the discrimination of the SCORE2-Diabetes model with and without multi-omics extendsion for 10-year MACE risk (N=990)..... 264

Appendix 62. Reclassification of cardiovascular risk by adding proteomics to the SCORE2-Diabetes model (N=1,751)..... 265

Appendix 63. Reclassification of cardiovascular risk by adding multi-omics to the SCORE2-Diabetes model (N=990)..... 266

Appendix 64. Calibration curves of the SCORE2-Diabetes model with and without proteomics extension for 10-year MACE risk prediction (N=1,751)..... 267

Appendix 65. Calibration curves of the SCORE2-Diabetes model with and without multi-omics extension for 10-year MACE risk prediction (N=990)..... 268

Appendix 66. ß-coefficients of the variables of the SCORE2 model extended by multi-omics for 10-year prediction of MACE..... 269

Appendix 67. ß-coefficients of the variables of the optimal multi-omics combination models for 10-year prediction of MACE..... 271

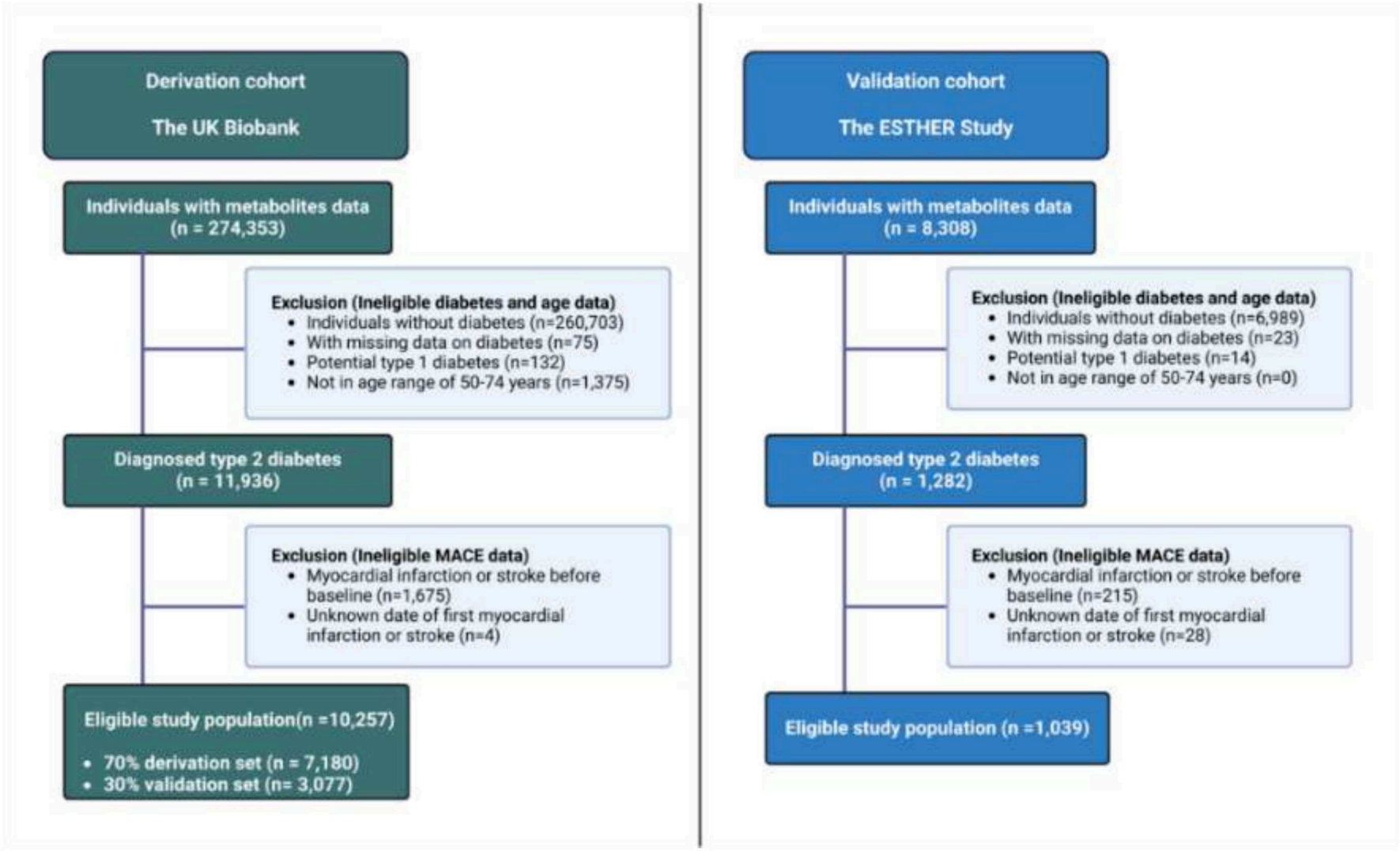
Appendix 68. Reclassification table of the SCORE2 model with Proteomics + PRS + Metabolomics or optimal multi-omics combination in the validation set (30% of UK Biobank, N=7,214)..... 273

Appendix 69. Calibration curves of the SCORE2 model with and without Proteomics + PRS + Metabolomics or optimal multi-omics combination extended for 10-year MACE risk prediction in the validation set (30% of UK Biobank, N=7,214)..... 274

Appendix 70. ß-coefficients of the different extensions of the clinical CDRS model with multi-omics data for 10-year prediction of T2D..... 276

Appendix 71. Calibration curves of the different extensions of the clinical CDRS model with multi-omics data for 10-year T2D risk prediction in the test set (30% of UK Biobank, N=6,998)..... 278

Appendix 72. Subgroup analysis of predictive performance by prediabetes status (30% of UK Biobank, N=6,998)..... 279




**Appendix 1. Flow chart for participant inclusion and exclusion**

**Appendix 2. Laboratory methods applied in the UK Biobank (UKB) and ESTHER study**

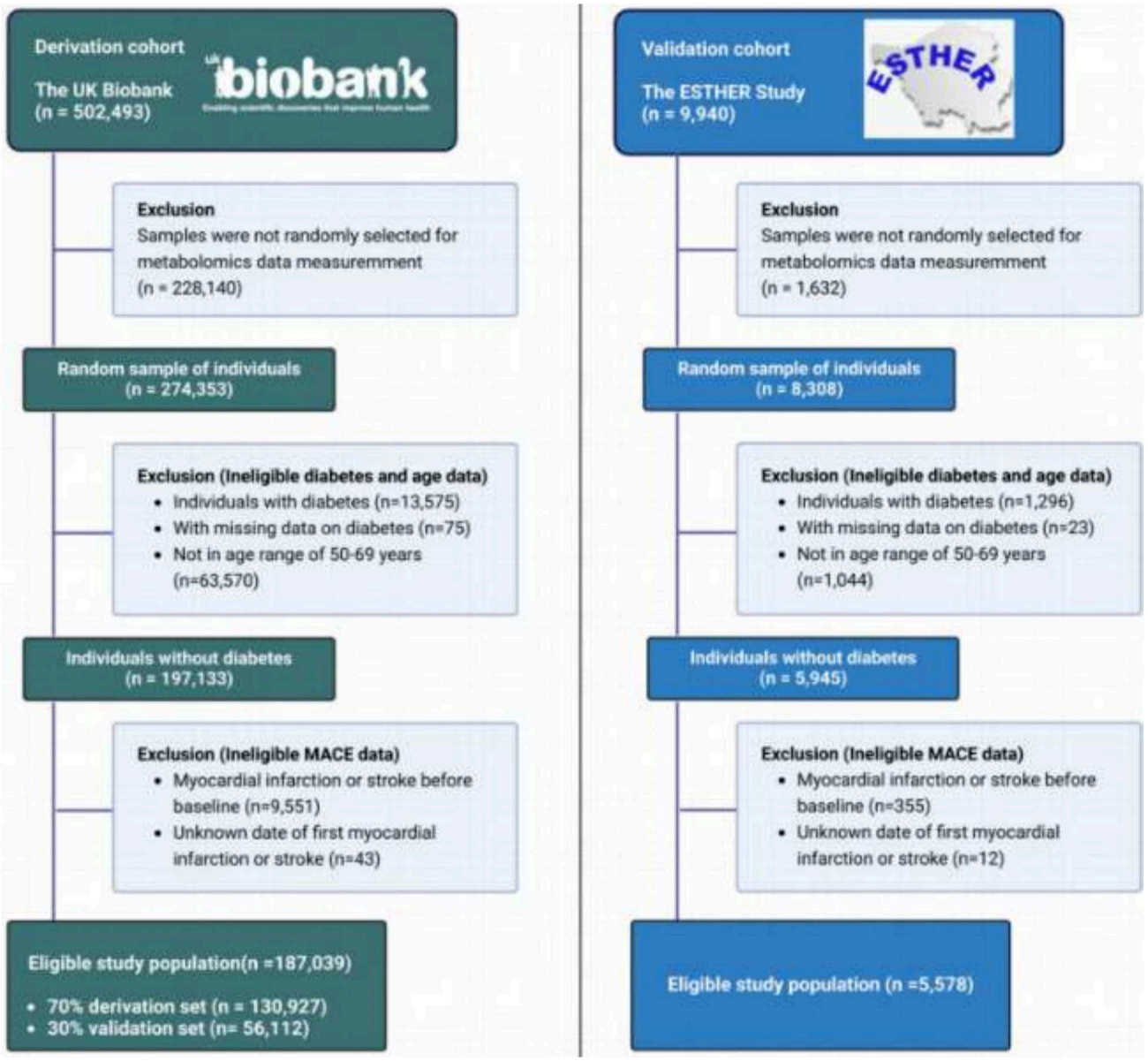
| Biomarker | Study | Analytical method | Instrument | Supplier |
|---|---|---|---|---|
| HDL-C | UKB | Enzymatic | Beckman Coulter AU5800 | Beckman Coulter (UK), Ltd |
| HDL-C | ESTHER | Enzymatic | Cobas 8000 C701 | Roche Diagnostics, Mannheim, Germany |
| Total cholesterol | UKB | Enzymatic | Beckman Coulter AU5800 | Beckman Coulter (UK), Ltd |
| Total cholesterol | ESTHER | Enzymatic | Cobas 8000 C701 | Roche Diagnostics, Mannheim, Germany |
| $HbA_{1c}$ | UKB | HPLC | Bio-Rad Variant II | Bio-Rad Laboratories, Hercules, CA, USA |
| $HbA_{1c}$ | ESTHER | HPLC | Bio-Rad Variant II | Bio-Rad Laboratories, Hercules, CA, USA |
| Creatinine | UKB | Enzymatic | Beckman Coulter AU5800 | Beckman Coulter (UK), Ltd |
| Creatinine | ESTHER | Kinetic Jaffé method | Cobas 8000 C701 | Roche Diagnostics, Mannheim, Germany |

**Abbreviations:** $HbA_{1c}$, glycated hemoglobin; HDL-C, high-density lipoprotein cholesterol; HPLC, high pressure liquid chromatography.

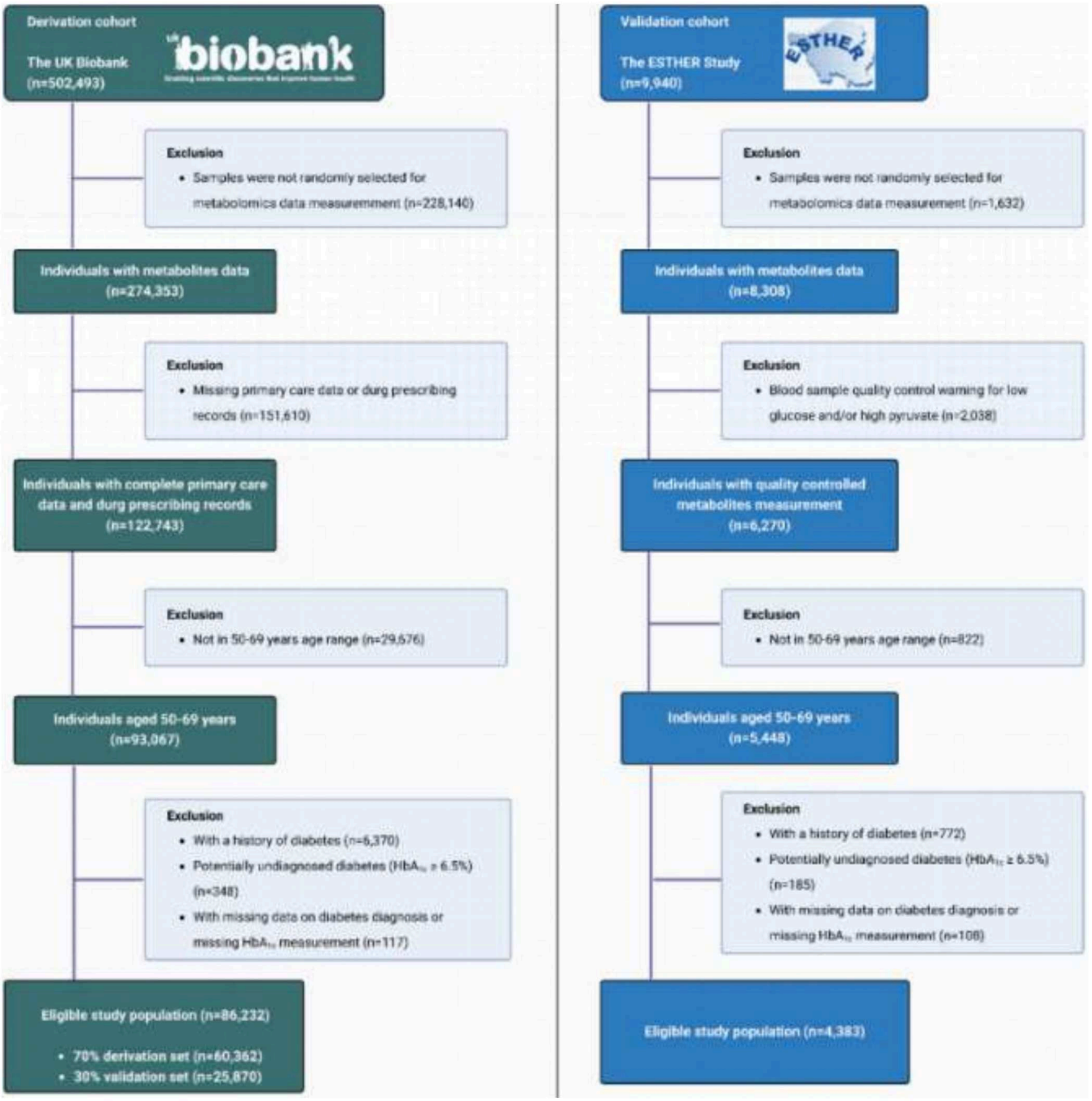
**Appendix 3. Definition of endpoint major cardiovascular event (MACE)**

| **Fatal MACE – cause-specific mortality due to any of the following:** | |
|---|---|
| *Endpoints included* | *ICD10-codes* |
| Hypertensive disease | I10-16 |
| Ischemic heart disease | I20-25 |
| Arrhythmias, heart failure | I46-52 |
| Cerebrovascular disease | I60-69 |
| Atherosclerosis/aortic aneurysm | I70-73 |
| Sudden death and death within 24 hours of symptom onset | R96.0-96.1 |
| | |
| *Endpoints excluded from the above endpoint:* | *ICD10-codes* |
| Myocarditis, unspecified | I51.4 |
| Subarachnoid haemorrhage | I60 |
| Subdural hemorrhage | I62 |
| Cerebral aneurysm | I67.1 |
| Cerebral arteritis | I68.2 |
| Moyamoya | I67.5 |
| | |
| **Non-fatal MACE** | *ICD10-codes* |
| Non-fatal myocardial infarction | I21-I23 |
| Non-fatal stroke | I60-69* |

***** In UK Biobank: I61, I63-I66, I69; in ESTHER: I60-I69




**Appendix 4. Flow chart for participant inclusion and exclusion**



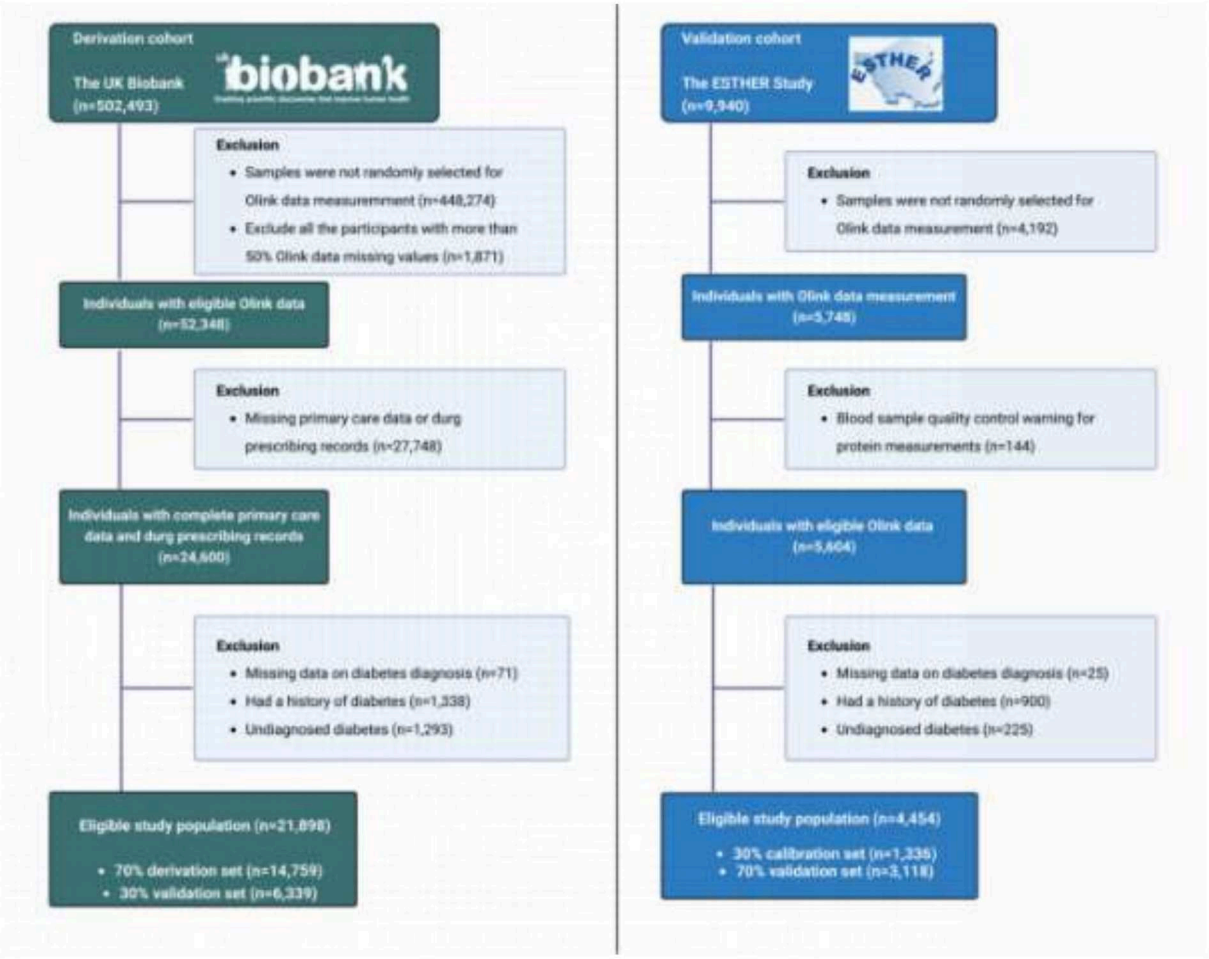

**Appendix 5. Flow chart for participant inclusion and exclusion**

**Appendix 6. Diagnosis and medication codes for assessment of type 2 diabetes status in primary and secondary healthcare and death registry records and verbal interview**

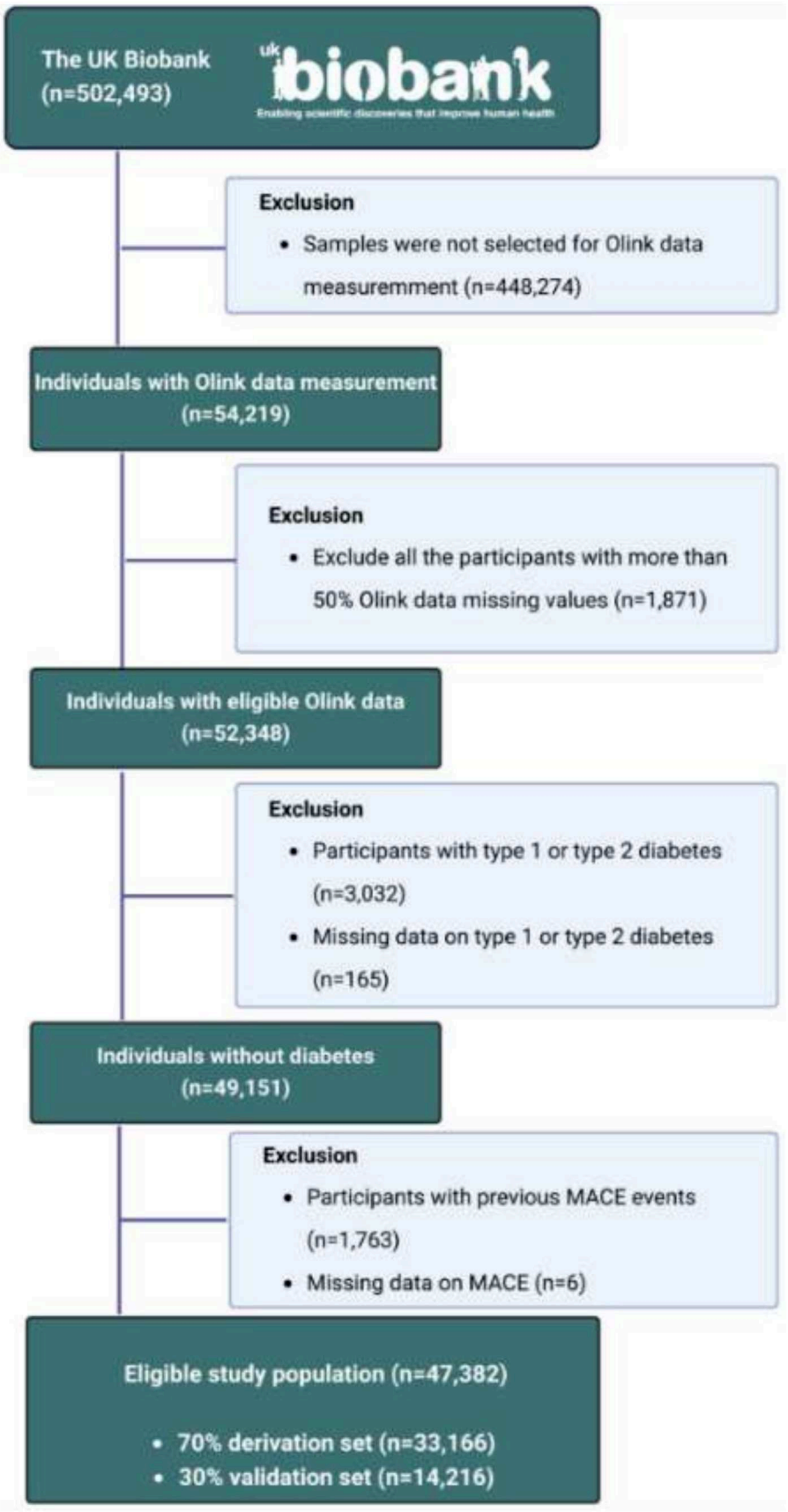
| Code format | Codes |
|---|---|
| ***Diagnosis codes*** | |
| **ICD-10** | E10-E14 |
| ***Medication codes*** | |
| **ATC codes** | A10AB01, A10AB02, A10AB03, A10AB04, A10AB05, A10AB06, A10AB30, A10AC01, A10AC02, A10AC03, A10AC04, A10AC30, A10AD01, A10AD02, A10AD03, A10AD04, A10AD05, A10AD30, A10AE01, A10AE02, A10AE03, A10AE04, A10AE05, A10AE30, A10AF01, A10BA01, A10BA02, A10BA03, A10BB01, A10BB02, A10BB03, A10BB04, A10BB05, A10BB06, A10BB07, A10BB08, A10BB09, A10BB10, A10BB11, A10BB12, A10BB31, A10BC01, A10BD01, A10BD02, A10BD03, A10BD04, A10BD05, A10BD06, A10BD07, A10BD08, A10BD09, A10BD10, A10BD11, A10BF01, A10BF02, A10BF03, A10BG01, A10BG02, A10BG03, A10BH01, A10BH02, A10BH03, A10BH04, A10BH05, A10BJ01, A10BJ02, A10BJ03, A10BJ04, A10BJ05, A10BJ06, A10BK01, A10BK02, A10BK03, A10BX01, A10BX02, A10BX03, A10BX05, A10BX06, A10BX08, A10XA01 |
| **Verbal interview**[a] | 1140857494, 1140857496, 1140857500, 1140857502, 1140857506, 1140857584, 1140857586, 1140857590, 1140868902, 1140868908, 1140874646, 1140874650, 1140874652, 1140874658, 1140874660, 1140874664, 1140874666, 1140874674, 1140874678, 1140874680, 1140874686, 1140874690, 1140874706, 1140874712, 1140874716, 1140874718, 1140874724, 1140874726, 1140874728, 1140874732, 1140874736, 1140874740, 1140874744, 1140874746, 1140882964, 1140883066, 1140884600, 1140910564, 1140910566, 1140910818, 1140921964, 1141152590, 1141153254, 1141153262, 1141156984, 1141157284, 1141168660, 1141168668, 1141169504, 1141171508, 1141171646, 1141171652, 1141173786, 1141173882, 1141177600, 1141177606, 1141189090, 1141189094 |

[a] Only assessed in UK Biobank. The medication codes for the verbal interview in UKB and details of the corresponding medications can be viewed at: https://biobank.ndph.ox.ac.uk/ukb/field.cgi?id=20003

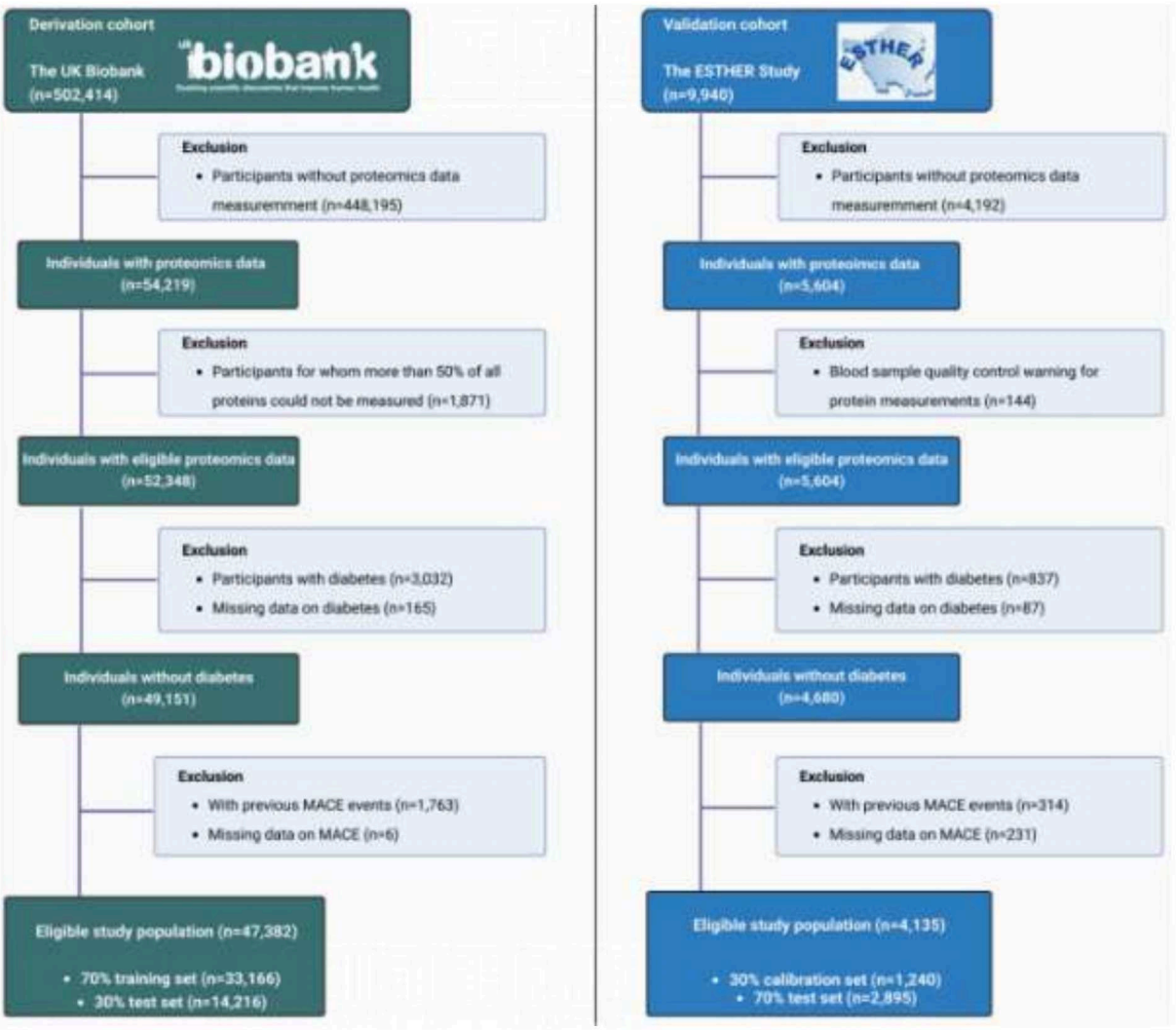
**Abbreviations:** ATC, Anatomical Therapeutic Chemical; ICD-10, the International Statistical Classification of Diseases and Related Health Problems 10th Revision.



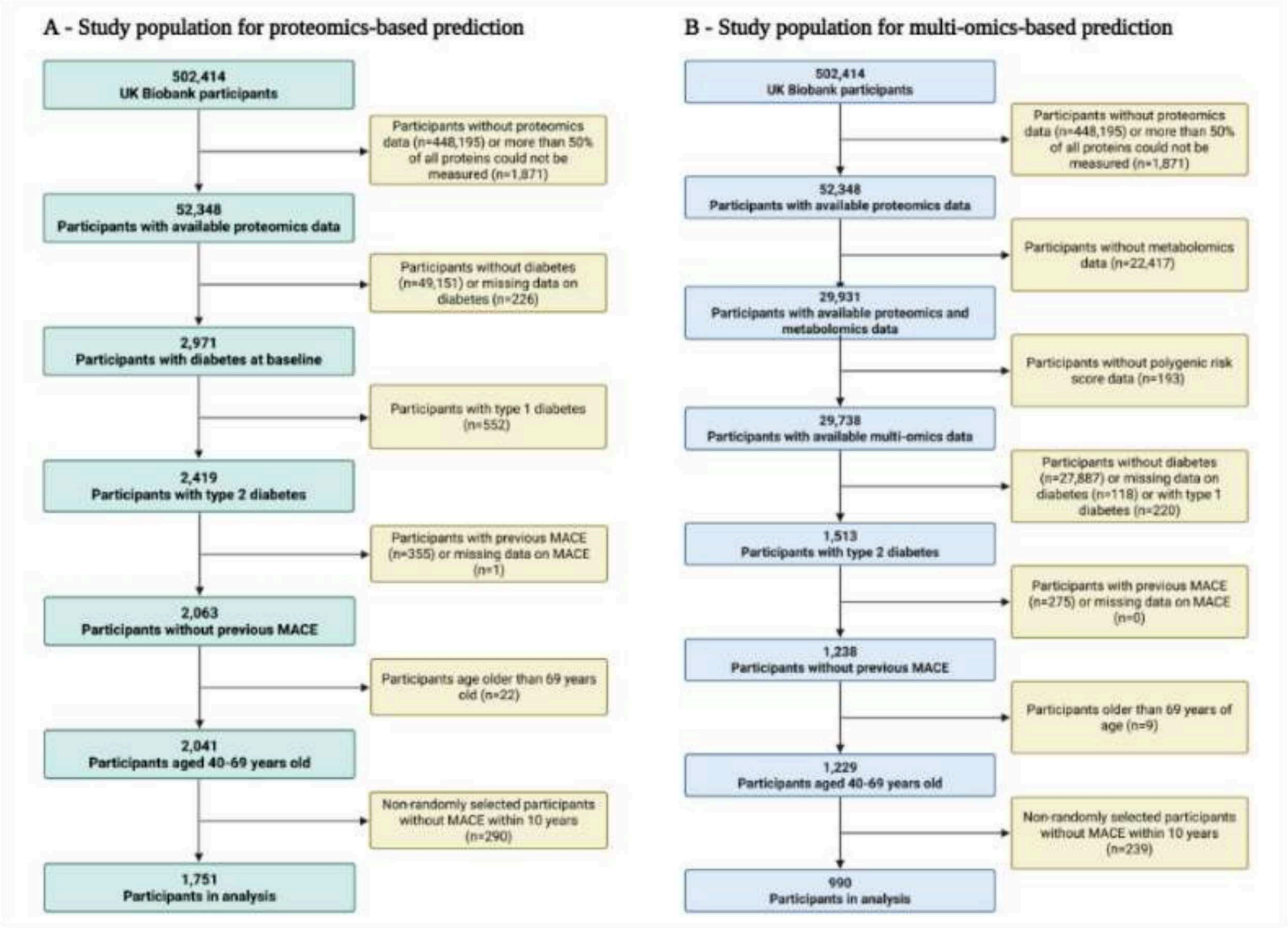

**Appendix 7. Flow chart for participant inclusion and exclusion**



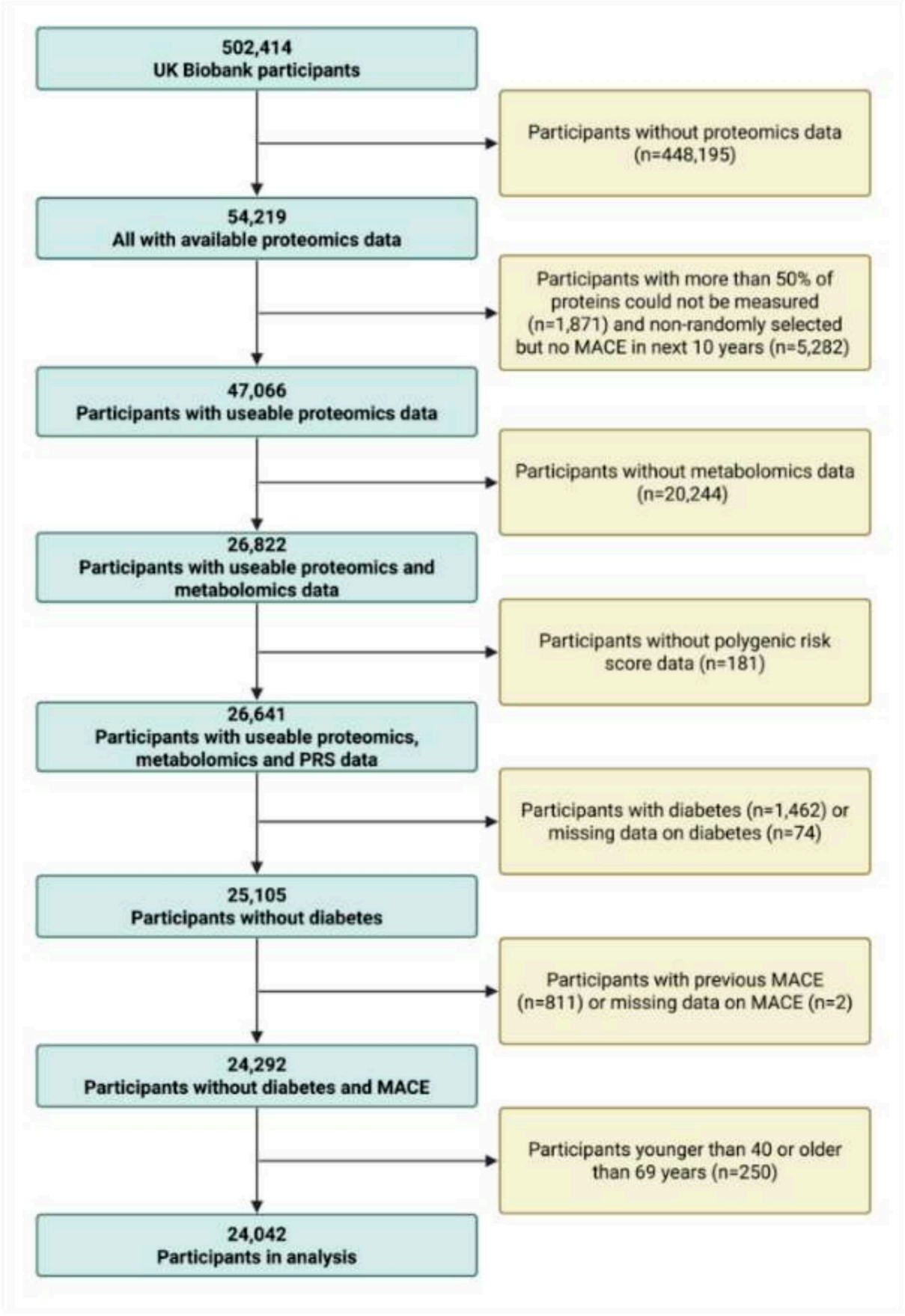

**Appendix 8. Flow chart for participant inclusion and exclusion**



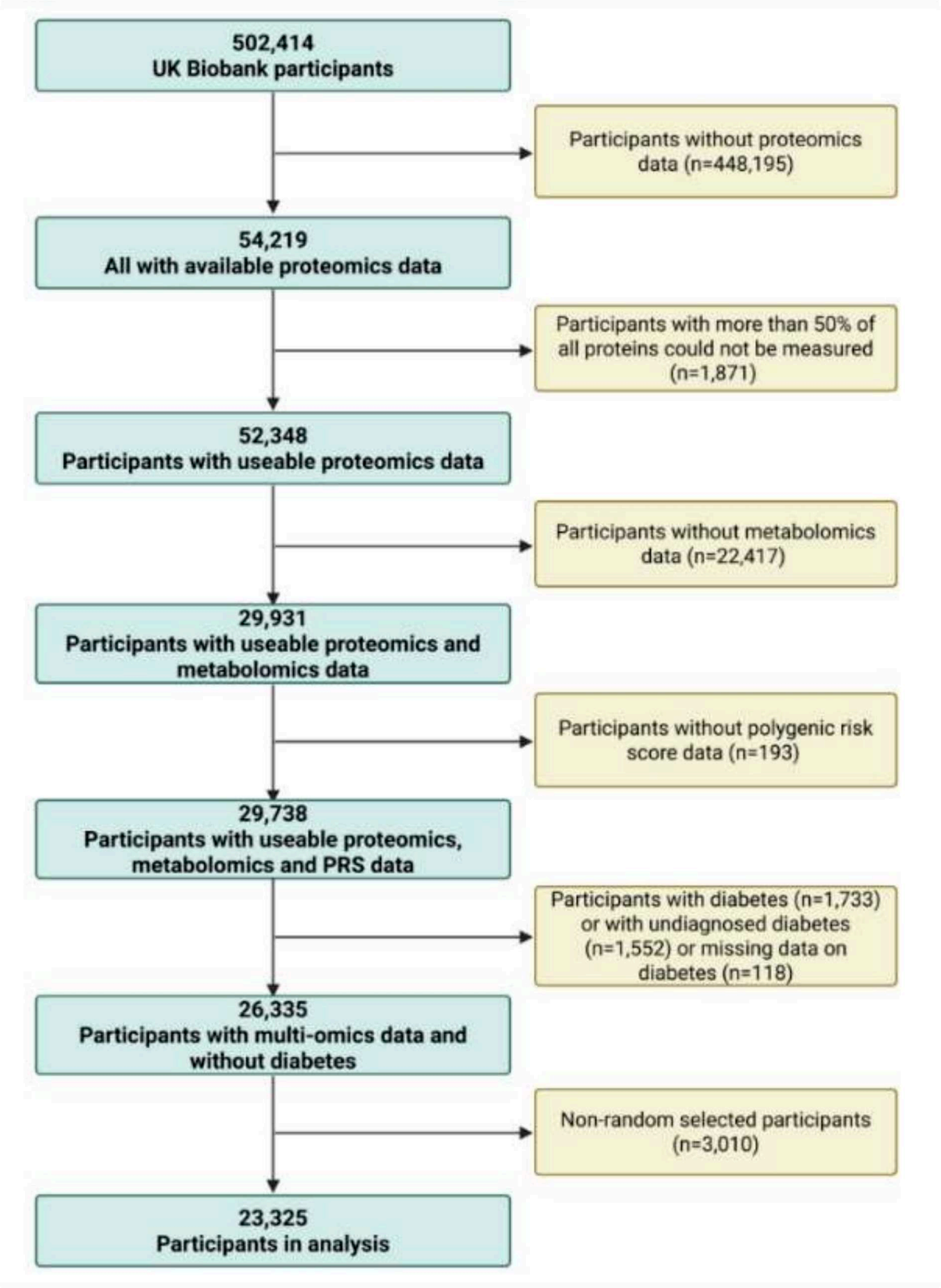

**Appendix 9. Flow chart for participant inclusion and exclusion**




**Appendix 10. Flow chart for participant inclusion and exclusion**




**Appendix 11. Flow chart for participant inclusion and exclusion**




**Appendix 12. Flow chart for participant inclusion and exclusion**

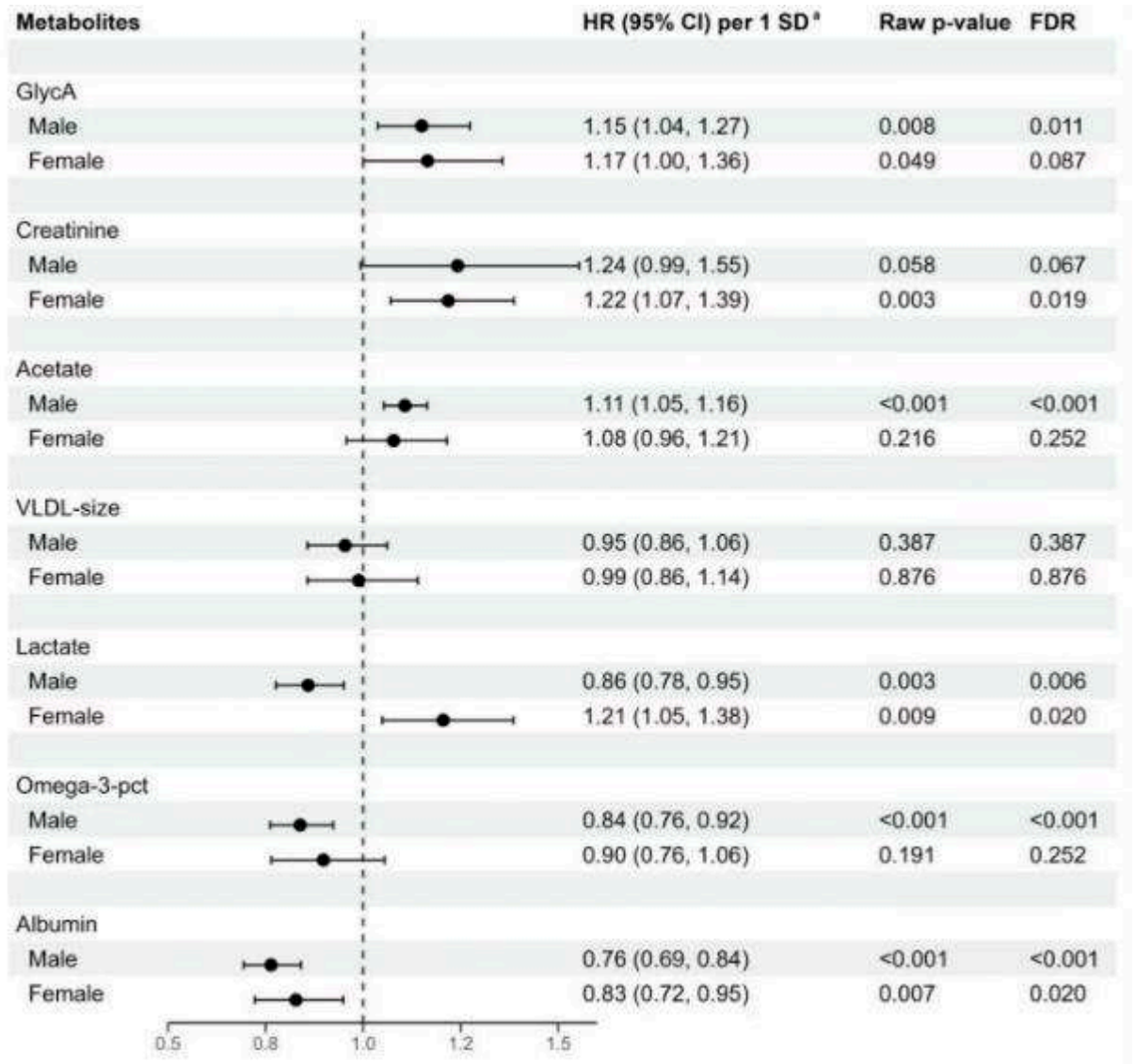
**Appendix 13. ß-coefficients of the variables of model 2 "Original SCORE2-Diabetes extended by NMR metabolites" for 10-year prediction of MACE in the derivation set of the UK Biobank**

| Risk factor (units) | ß-coefficients | |
|---|---|---|
| | **Men** | **Women** |
| **SCORE2-Diabetes variables**[a] | | |
| Age (per 5 years) | 0.3649 | 0.4972 |
| Current smoking | 0.1815 | 0.7378 |
| SBP (per 20 mmHg) | 0.0831 | 0.2404 |
| Total cholesterol (per 1 mmol/L) | 0.0060 | 0.1019 |
| HDL cholesterol (per 0.5 mmol/L) | 0.2256 | -0.1160 |
| Smoking interaction with age | -0.2768 | -0.1317 |
| SBP interaction with age | -0.0260 | -0.0815 |
| Total cholesterol interaction with age | -0.0363 | -0.0102 |
| HDL interaction with age | 0.0509 | -0.0408 |
| Diabetes age at diagnosis (per 5 years) | -0.1146 | -0.1317 |
| $HbA_{1c}$ (per 9.34 mmol/mol) | 0.1368 | 0.1098 |
| Ln eGFR (per 0.15 ml/min/1.73$m^2$) | -0.1393 | -0.0947 |
| Ln $eGFR^2$ (quadratic term) | -0.0067 | -0.0051 |
| $HbA_{1c}$ interaction with age | 0.0363 | 0.0069 |
| Ln eGFR interaction with age | 0.0182 | -0.0208 |
| **Added NMR metabolites** | | |
| Creatinine (per 1 SD[b]) | 0.0341 | 0.0652 |
| Albumin (per 1 SD[b]) | -0.0491 | -0.2865 |
| GlycA (per 1 SD[b]) | 0.1704 | 0.1283 |
| Acetate (per 1 SD[b]) | 0.0395 | - |
| Omega-3-pct (per 1 SD[b]) | -0.0682 | - |
| VLDL-size (per 1 SD[b]) | -0.1076 | - |
| Lactate (per 1 SD[b]) | - | 0.1282 |

**Abbreviations:** CI, confidence interval; GlycA, glycoprotein acetyls; $HbA_{1c}$, glycated hemoglobin; SBP, systolic blood pressure; SHR, subdistribution hazard ratio; Omega-3-pct, the Omega-3 fatty acids percentage of total fatty acids; SD, standard deviation; VLDL-size, average diameter for very-low-density lipoprotein particles.

[a] The ß-coefficients of all variables from SCORE2-Diabtes were re-fitted and the variables "diabetes" and "diabetes-age-interaction" were removed because all included study participants are diabetes patients.

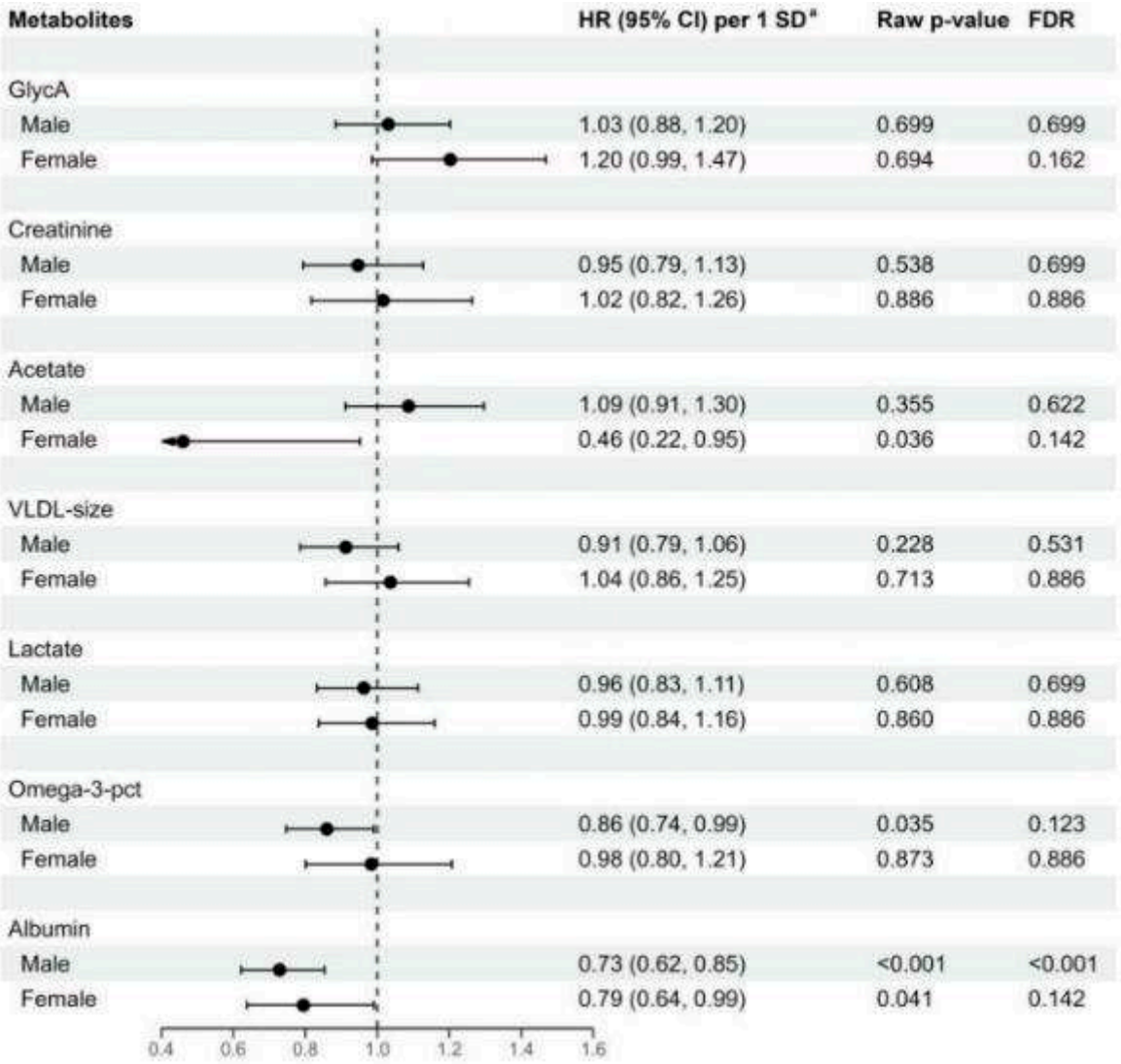
[b] The derivation set (n=7,180) standard deviations of the metabolites creatinine, albumin, GlycA, acetate, omega-3-pct, VLDL-size, and lactate were 0.02 mmol/L, 3.60 mmol/L, 0.13 mmol/L, 0.04 mmol/L, 1.51%, 1.41 nm, and 1.23 mmol/L, respectively.




**Appendix 14. Associations between selected metabolites and MACE across sexes in the internal validation cohort (30% of UK Biobank, N=3,077)**

**Abbreviations:** CI, confidence interval; FDR, false discovery rate; GlycA, glycoprotein acetyls; HR, hazard ratio; Omega-3-pct, the Omega-3 fatty acids percentage of total fatty acids; SD, standard deviation; VLDL-size, average diameter for very-low-density lipoprotein particles.

[a] Hazard ratios are expressed per 1 standard deviation of the respective metabolite concentration and are adjusted for age, systolic blood pressure, smoking status, diabetes age at diagnosis, glycated hemoglobin, and the estimated glomerular filtration rate. The standard deviations of creatinine, albumin, GlycA, acetate, omega-3-pct, VLDL-size, and lactate were 0.02 mmol/L, 3.60 mmol/L, 0.13 mmol/L, 0.04 mmol/L, 1.51%, 1.41 mmol/L, and 1.23 mmol/L, respectively.




**Appendix 15. Associations between selected metabolites and MACE across sexes in the external validation cohort (ESTHER study, N=1,039)**

**Abbreviations:** CI, confidence interval; GlycA, glycoprotein acetyls; HR, hazard ratio; Omega-3-pct, the Omega-3 fatty acids percentage of total fatty acids; VLDL-size, average diameter for very-low-density lipoprotein particles.

[a] Hazard ratios are expressed per 1 standard deviation of the respective metabolite concentration and are adjusted for age, systolic blood pressure, smoking status, diabetes age at diagnosis, glycated hemoglobin, and estimated glomerular filtration rate. The standard deviations of creatinine, albumin, GlycA, acetate, omega-3-pct, VLDL-size, and lactate were 0.03 mmol/L, 4.08 mmol/L, 0.14 mmol/L, 0.04 mmol/L, 1.19%, 1.08 mmol/L, and 5.36 mmol/L, respectively.

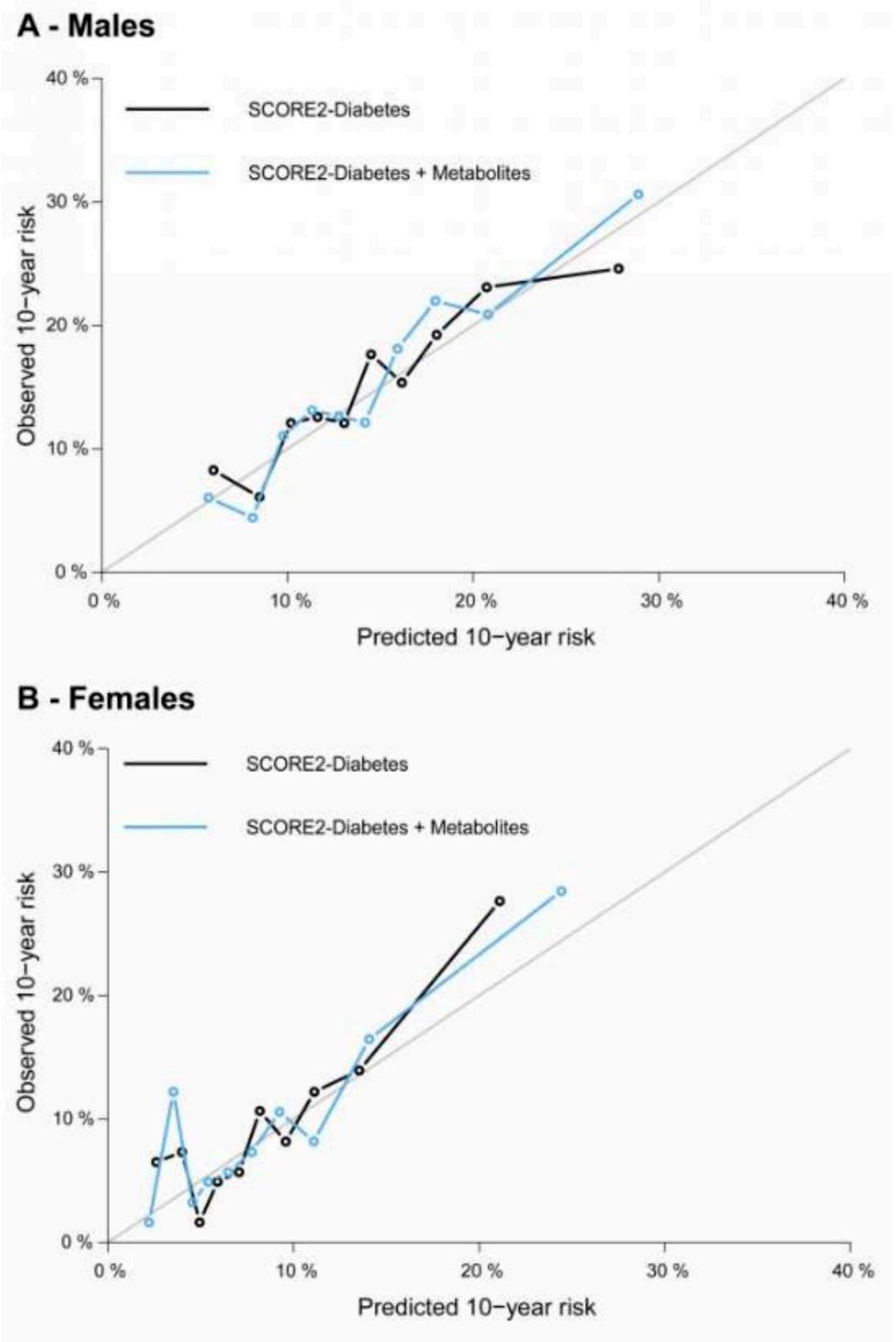
**Appendix 16. ß-coefficients of the variables of model 2 "Original SCORE2-Diabetes extended by NMR metabolites" for 10-year prediction of MACE in the derivation set of the UK Biobank**

| Risk factor (units) | ß-coefficients | |
|---|---|---|
| | **Men** | **Women** |
| **SCORE2-Diabetes variables**[a] | | |
| Age (per 5 years) | 0.3649 | 0.4972 |
| Current smoking | 0.1815 | 0.7378 |
| SBP (per 20 mmHg) | 0.0831 | 0.2404 |
| Total cholesterol (per 1 mmol/L) | 0.0060 | 0.1019 |
| HDL cholesterol (per 0.5 mmol/L) | 0.2256 | -0.1160 |
| Smoking interaction with age | -0.2768 | -0.1317 |
| SBP interaction with age | -0.0260 | -0.0815 |
| Total cholesterol interaction with age | -0.0363 | -0.0102 |
| HDL interaction with age | 0.0509 | -0.0408 |
| Diabetes age at diagnosis (per 5 years) | -0.1146 | -0.1317 |
| $HbA_{1c}$ (per 9.34 mmol/mol) | 0.1368 | 0.1098 |
| Ln eGFR (per 0.15 ml/min/1.73m$^2$) | -0.1393 | -0.0947 |
| Ln eGFR$^2$ (quadratic term) | -0.0067 | -0.0051 |
| $HbA_{1c}$ interaction with age | 0.0363 | 0.0069 |
| Ln eGFR interaction with age | 0.0182 | -0.0208 |
| **Added NMR metabolites** | | |
| Creatinine (per 1 SD[b]) | 0.0341 | 0.0652 |
| Albumin (per 1 SD[b]) | -0.0491 | -0.2865 |
| GlycA (per 1 SD[b]) | 0.1704 | 0.1283 |
| Acetate (per 1 SD[b]) | 0.0395 | - |
| Omega-3-pct (per 1 SD[b]) | -0.0682 | - |
| VLDL-size (per 1 SD[b]) | -0.1076 | - |
| Lactate (per 1 SD[b]) | - | 0.1282 |

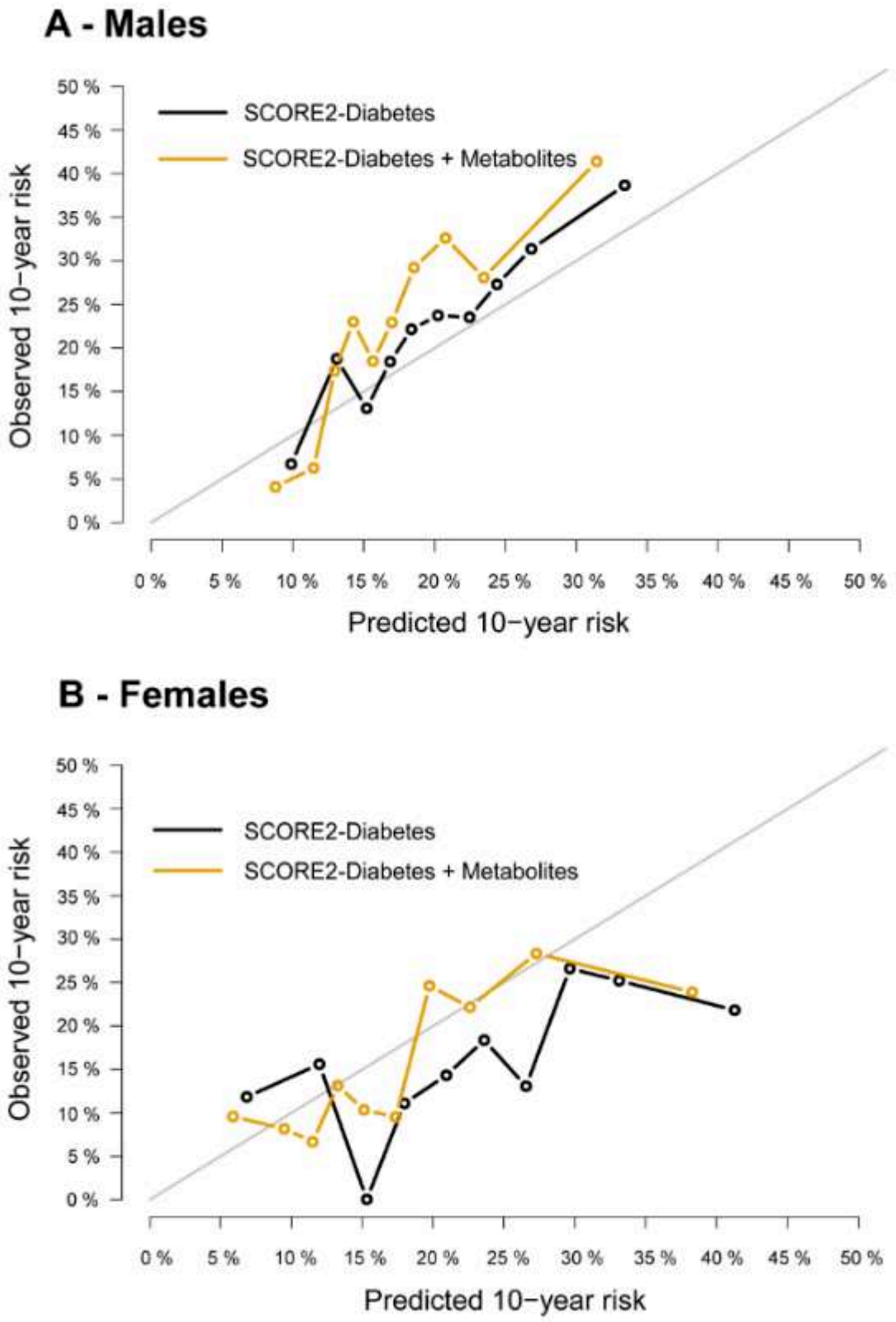
**Abbreviations:** CI, confidence interval; GlycA, glycoprotein acetyls; $HbA_{1c}$, glycated hemoglobin; SBP, systolic blood pressure; SHR, subdistribution hazard ratio; Omega-3-pct, the Omega-3 fatty acids percentage of total fatty acids; SD, standard deviation; VLDL-size, average diameter for very-low-density lipoprotein particles.

[a] The ß-coefficients of all variables from SCORE2-Diabtes were re-fitted and the variables "diabetes" and "diabetes-age-interaction" were removed because all included study participants are diabetes patients.

[b] The derivation set (n=7,180) standard deviations of the metabolites creatinine, albumin, GlycA, acetate, omega-3-pct, VLDL-size, and lactate were 0.02 mmol/L, 3.60 mmol/L, 0.13 mmol/L, 0.04 mmol/L, 1.51%, 1.41 nm, and 1.23 mmol/L, respectively.




**Appendix 17. Calibration curves of the SCORE2-Diabetes model and its combination with the metabolites for 10-year cardiovascular risk prediction in the internal validation set (30% of UK Biobank, N=3,077)**




**Appendix 18. Calibration curves of the SCORE2-Diabetes model and its combination with the metabolites for 10-year cardiovascular risk prediction in the external validation (ESTHER study, N=1,039)**

**Appendix 19. ß-coefficients of the variables of model 4 "Age, smoking, SBP, diabetes age at diagnosis, and NMR metabolites" for 10-year prediction of MACE in the derivation set of the UK Biobank**

| Risk factor (units) | ß-coefficients | |
|---|---|---|
| | Men | Women |
| **Non-Chemistry variables from SCORE2-Diabetes** | | |
| Age (per 5 years) | 0.3706 | 0.4945 |
| Current smoking | 0.1819 | 0.7327 |
| Smoking interaction with age | -0.2385 | -0.1277 |
| SBP (per 20 mmHg) | 0.0822 | 0.2503 |
| SBP interaction with age | -0.0245 | -0.0819 |
| Diabetes age at diagnosis (per 5 years) | -0.1191 | -0.1310 |
| **NMR metabolites** | | |
| Total cholesterol (per 1 mmol/L) | 0.0086 | 0.0959 |
| Total cholesterol interaction with age | -0.0432 | -0.0117 |
| HDL cholesterol (per 0.5 mmol/L) | 0.2321 | -0.1156 |
| HDL interaction with age | 0.0512 | -0.0480 |
| Ln eGFR (per 0.15 ml/min/1.73m$^2$) | -0.1395 | -0.0919 |
| Ln eGFR$^2$ (quadratic term) | -0.0059 | -0.0060 |
| Ln eGFR interaction with age | 0.0185 | -0.0214 |
| Creatinine (per 1 SD[b]) | 0.0348 | 0.0623 |
| Albumin (per 1 SD[b]) | -0.0490 | -0.2841 |
| GlycA (per 1 SD[b]) | 0.1592 | 0.1276 |
| Acetate (per 1 SD[b]) | 0.0378 | - |
| Omega-3-pct (per 1 SD[b]) | -0.0681 | - |
| VLDL-size (per 1 SD[b]) | -0.1067 | - |
| Lactate (per 1 SD[b]) | - | 0.1220 |
| Glucose (per 1 SD[b]) | 0.1254 | 0.1028 |
| Glucose interaction with age | -0.0108 | -0.0094 |

**Abbreviations:** CI, confidence interval; GlycA, glycoprotein acetyls; $HbA_{1c}$, glycated hemoglobin; SBP, systolic blood pressure; Omega-3-pct, the Omega-3 fatty acids percentage of total fatty acids; SD, standard deviation; VLDL-size, average diameter for very-low-density lipoprotein particles.

[a] The ß-coefficients of all variables from SCORE2-Diabtes were re-fitted and the variables "diabetes" and "diabetes-age-interaction" were removed because all included study participants are diabetes patients.

[b] The derivation set (n=7,180) standard deviations of the metabolites creatinine, albumin, GlycA, acetate, omega-3-pct, VLDL-size, lactate and glucose were 0.02 mmol/L, 3.60 mmol/L, 0.13 mmol/L, 0.04 mmol/L, 1.51%, 1.41 nm, 1.23 mmol/L, and 1.19 mmol/L, respectively.

**Appendix 20. Comparison of the predictive performance of model 1 "Original SCORE2-Diabetes" and model 4 "Age, smoking, SBP, diabetes age at diagnosis, and NMR metabolites"**

| Metrics | Men | Women | Overall |
|---|---|---|---|
| **Derivation set (70% of UK Biobank)** | | | |
| Total sample size / MACE cases | 4,298/691 | 2,882/249 | 7,180/940 |
| C-index (Model 1) | 0.630 (0.606, 0.653) | 0.673 (0.638, 0.707) | 0.662 (0.644, 0.681) |
| C-index (Model 4[a]) | 0.645 (0.622, 0.668) | 0.678 (0.643, 0.712) | 0.676 (0.664, 0.700) |
| *P*-values for C-index comparisons[b] | **0.026** | 0.072 | **0.010** |
| **Internal validation set (30% of UK Biobank)** | | | |
| Total sample size / MACE cases | 1,842/313 | 1,235/113 | 3,077/426 |
| C-index (Model 1) | 0.632 (0.597, 0.668) | 0.670 (0.617, 0.722) | 0.660 (0.632, 0.688) |
| C-index (Model 4[a]) | 0.652 (0.618, 0.687) | 0.678 (0.624, 0.733) | 0.677 (0.649, 0.705) |
| *P*-values for C-index comparisons[b] | **0.030** | 0.731 | **0.041** |
| NRI total (%)[c] | 0.8 (-7.3, 9.5) | 6.5 (-9.3, 24.0) | 5.3 (-2.8, 13.5) |
| NRI events (%)[c] | 3.5 (-4.0, 7.4) | 2.3 (-13.9, 26.6) | 0.6 (-3.6, 6.8) |
| NRI non-events (%)[c] | -2.7 (-7.5, 3.8) | 4.2 (-4.9, 12.5) | 4.7 (-1.9, 9.1) |
| IDI | 0.000 (-0.015, 0.015) | **0.019 (0.001, 0.043)** | **0.010 (0.000, 0.020)** |
| **External validation set (ESTHER)** | | | |
| Total sample size / MACE cases | 528/107 | 511/68 | 1,039/175 |
| C-index (Model 1) | 0.632 (0.577, 0.687) | 0.618 (0.545, 0.691) | 0.612 (0.566, 0.657) |
| C-index (Model 4[a]) | 0.660 (0.605, 0.716) | 0.651 (0.578, 0.724) | 0.656 (0.612, 0.700) |
| *P*-values for C-index comparisons[b] | 0.201 | 0.168 | **0.009** |
| NRI total (%)[c] | 7.3 (-6.8, 15.7) | -6.0 (-22.5, 17.9) | 4.5 (-5.5, 13.1) |
| NRI events (%)[c] | **9.5 (0.8, 16.0)** | -2.5 (-15.5, 16.3) | 4.3 (-3.8, 11.5) |
| NRI non-events (%)[c] | -2.2 (-16.7, 10.5) | -3.5 (-17.0, 10.7) | 0.2 (-7.6, 8.7) |
| IDI | 0.004 (-0.026, 0.035) | 0.002 (-0.028, 0.040) | 0.013 (-0.007, 0.033) |

**Abbreviations:** IDI, integrated discrimination improvement; MACE, major cardiovascular events; NRI, net reclassification improvement.

[a] Metabolites that were included in model 4 were total cholesterol, HDL cholesterol, glucose, creatinine, albumin, GlycA, acetate, Omega-3-pct, and VLDL-size for men, and total cholesterol, HDL cholesterol, glucose, lactate, creatinine, albumin, and GlycA for women (see Table 1, main document).

[b] The statistical method for comparing differences in C-index is based on U-statistics estimation, using asymptotic normal approximation for hypothesis testing.

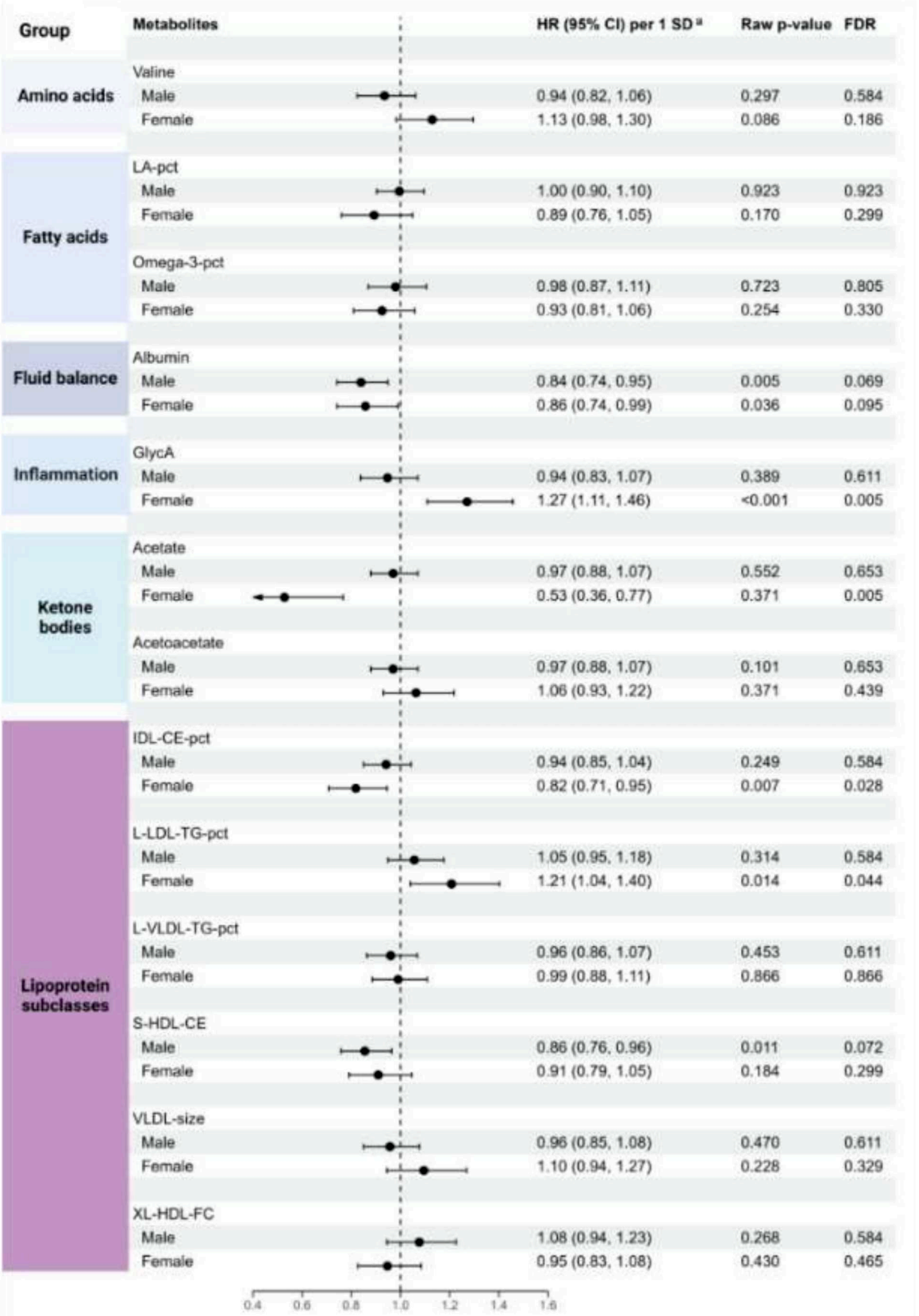
[c] NRI calculated with the pre-specified 10-year MACE risk categories of 0–10%, >10–20%, and >20%.

**Appendix 21. ß-coefficients of the variables of the SCORE2 model extended by 13 metabolites for 10-year prediction of major cardiovascular events in the derivation set of the UK Biobank**

| Risk factor (units) | Coefficient | |
|---|---|---|
| | Men | Women |
| **SCORE2-Diabetes variables** | | |
| Age (per 5 years) | 0.2940 | 0.4691 |
| Current smoking | 0.5926 | 0.7999 |
| Systolic blood pressure (per 20mmHg) | 0.1388 | 0.1834 |
| Total cholesterol (per 1 mmol/L) | 0.0401 | 0.0557 |
| HDL cholesterol (per 0.5 mmol/L) | -0.4797 | -0.1400 |
| Smoking interaction with age | -0.1546 | -0.1701 |
| SBP interaction with age | -0.0205 | -0.0360 |
| Total cholesterol interaction with age | -0.0836 | -0.0574 |
| HDL interaction with age | 0.0863 | 0.0529 |
| **Additional metabolic variables** | | |
| Acetate (per 1 SD[a]) | 0.0303 | - |
| Albumin (per 1 SD[a]) | -0.0912 | -0.1148 |
| GlycA (per 1 SD[a]) | 0.1792 | 0.1758 |
| IDL-CE-pct (per 1 SD[a]) | -0.1134 | -0.0084 |
| L-VLDL-TG-pct (per 1 SD[a]) | 0.0667 | - |
| Omega-3-pct (per 1 SD[a]) | -0.0835 | - |
| S-HDL-CE (per 1 SD[a]) | 0.1585 | - |
| VLDL-size (per 1 SD[a]) | -0.1911 | - |
| XL-HDL-FC (per 1 SD[a]) | 0.1706 | - |
| Acetoacetate (per 1 SD[a]) | - | 0.0413 |
| LA-pct (per 1 SD[a]) | - | -0.0669 |
| L-LDL-TG-pct (per 1 SD[a]) | - | 0.0843 |
| Valine (per 1 SD[a]) | - | -0.0886 |

**Abbreviations:** GlycA, glycoprotein acetyls; IDL-CE-pct, the cholesteryl esters percentage of total lipids in intermediate-density lipoprotein; LA-pct, the linoleic acid percentage of total fatty acids; L-LDL-TG-pct, triglycerides to total lipids in large low-density lipoprotein percentage; L-VLDL-TG-pct, the triglycerides percentage of total lipids in large very-low-density lipoprotein; Omega-3-pct, the Omega-3 fatty acids percentage of total fatty acids; SD, standard deviation; S-HDL-CE, cholesteryl esters in small high-density lipoprotein; VLDL-size, average diameter for very-low-density lipoprotein particles; XL-HDL-FC, free cholesterol in very large high-density lipoprotein.

[a] The standard deviations of the metabolites acetate, albumin, GlycA, IDL-CE-pct, L-VLDL-TG-pct, Omega-3-pct, S-HDL-CE, VLDL-size, XL-HDL-FC, acetoacetate, LA-pct, L-LDL-TG-pct, and valine were 0.03 mmol/L, 3.35 mmol/L, 0.12 mmol/L, 2.31%, 5.84%, 1.58nm, 0.05 mmol/L, 1.21 nm, 0.01 mmol/L, 0.01 mmol/L, 3.33mmol/L, 1.68%, and 0.04 mmol/L, respectively.




**Appendix 22. Associations between selected metabolites and MACE across sexes in the external validation cohort (ESTHER study, N=5,578)**

**Abbreviations:** CI, confidence interval; FDR, false discovery rate; GlycA, glycoprotein acetyls; IDL-CE-pct, the cholesteryl esters percentage of total lipids in intermediate-density lipoprotein; LA-pct, the linoleic acid percentage of total fatty acids; L-LDL-TG-pct, triglycerides to total lipids in large low-density lipoprotein percentage; L-VLDL-TG-pct, the triglycerides percentage of total lipids in large very-low-density lipoprotein; Omega-3-pct, the Omega-3 fatty acids percentage of total fatty acids; SD, standard deviation; S-HDL-CE, cholesteryl esters in small high-density lipoprotein; VLDL-size, average diameter for very-low-density lipoprotein particles; XL-HDL-FC, free cholesterol in very large high-density lipoprotein.

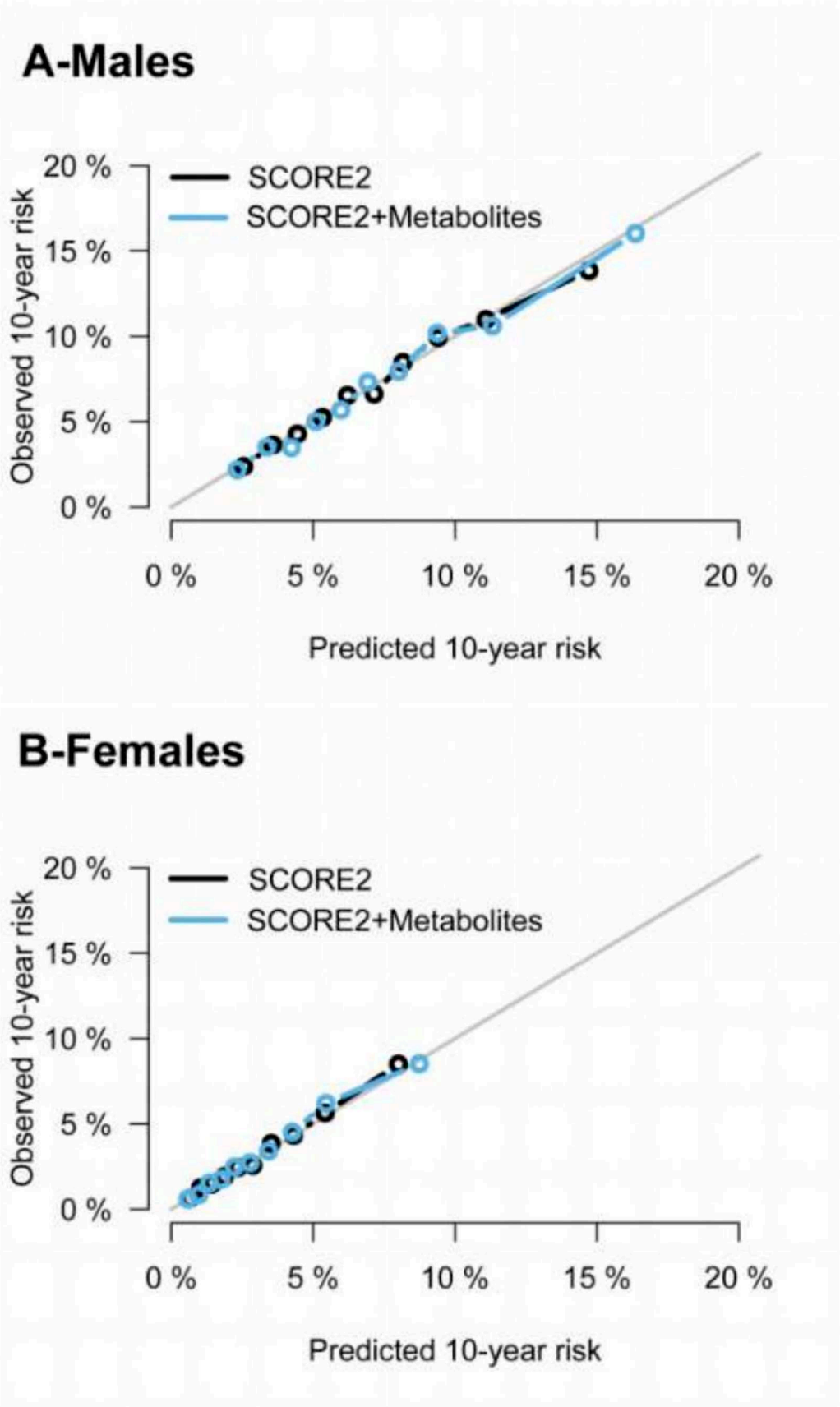
[a] The standard deviations of the metabolites acetate, albumin, GlycA, IDL-CE-pct, L-VLDL-TG-pct, Omega-3-pct, S-HDL-CE, VLDL-size, XL-HDL-FC, acetoacetate, LA-pct, L-LDL-TG-pct, and valine were 0.03 mmol/L, 3.35 mmol/L, 0.12 mmol/L, 2.31%, 5.84%, 1.58nm, 0.05 mmol/L, 1.21 mmol/L, 0.01 mmol/L, 0.01 mmol/L, 3.33mmol/L, 1.68%, and 0.04 mmol/L, respectively.

**Appendix 23. ß-coefficients of the variables of the SCORE2 model extended by 13 metabolites for 10-year prediction of MACE in the derivation set of the UK Biobank**

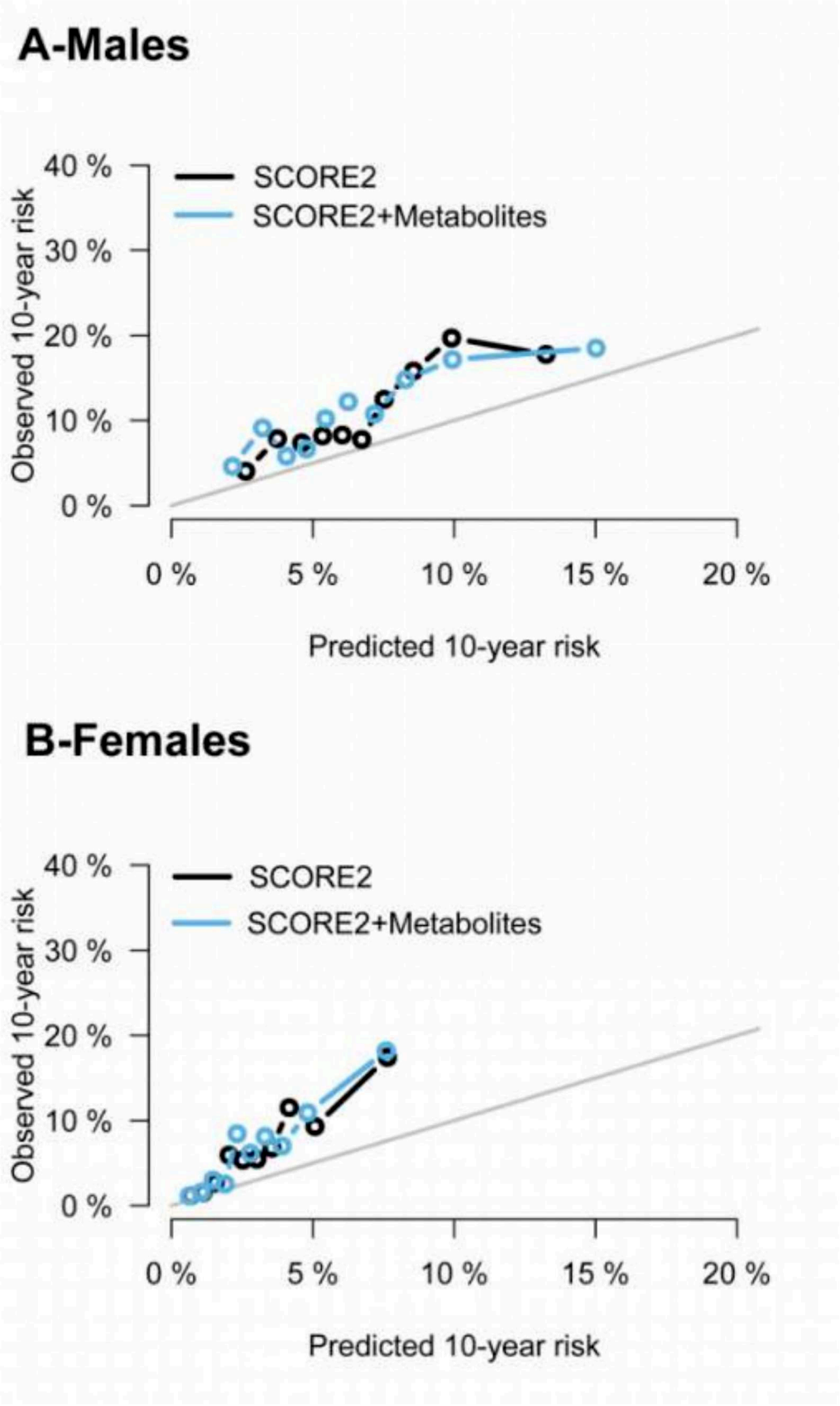
| Risk factor (units) | Coefficient | |
|---|---|---|
| | Men | Women |
| **SCORE2-Diabetes variables** | | |
| Age (per 5 years) | 0.2940 | 0.4691 |
| Current smoking | 0.5926 | 0.7999 |
| Systolic blood pressure (per 20mmHg) | 0.1388 | 0.1834 |
| Total cholesterol (per 1 mmol/L) | 0.0401 | 0.0557 |
| HDL cholesterol (per 0.5 mmol/L) | -0.4797 | -0.1400 |
| Smoking interaction with age | -0.1546 | -0.1701 |
| SBP interaction with age | -0.0205 | -0.0360 |
| Total cholesterol interaction with age | -0.0836 | -0.0574 |
| HDL interaction with age | 0.0863 | 0.0529 |
| **Additional metabolic variables** | | |
| Acetate (per 1 SD[a]) | 0.0303 | - |
| Albumin (per 1 SD[a]) | -0.0912 | -0.1148 |
| GlycA (per 1 SD[a]) | 0.1792 | 0.1758 |
| IDL-CE-pct (per 1 SD[a]) | -0.1134 | -0.0084 |
| L-VLDL-TG-pct (per 1 SD[a]) | 0.0667 | - |
| Omega-3-pct (per 1 SD[a]) | -0.0835 | - |
| S-HDL-CE (per 1 SD[a]) | 0.1585 | - |
| VLDL-size (per 1 SD[a]) | -0.1911 | - |
| XL-HDL-FC (per 1 SD[a]) | 0.1706 | - |
| Acetoacetate (per 1 SD[a]) | - | 0.0413 |
| LA-pct (per 1 SD[a]) | - | -0.0669 |
| L-LDL-TG-pct (per 1 SD[a]) | - | 0.0843 |
| Valine (per 1 SD[a]) | - | -0.0886 |

**Abbreviations:** GlycA, glycoprotein acetyls; IDL-CE-pct, the cholesteryl esters percentage of total lipids in intermediate-density lipoprotein; LA-pct, the linoleic acid percentage of total fatty acids; L-LDL-TG-pct, triglycerides to total lipids in large low-density lipoprotein percentage; L-VLDL-TG-pct, the triglycerides percentage of total lipids in large very-low-density lipoprotein; Omega-3-pct, the Omega-3 fatty acids percentage of total fatty acids; SD, standard deviation; S-HDL-CE, cholesteryl esters in small high-density lipoprotein; VLDL-size, average diameter for very-low-density lipoprotein particles; XL-HDL-FC, free cholesterol in very large high-density lipoprotein.

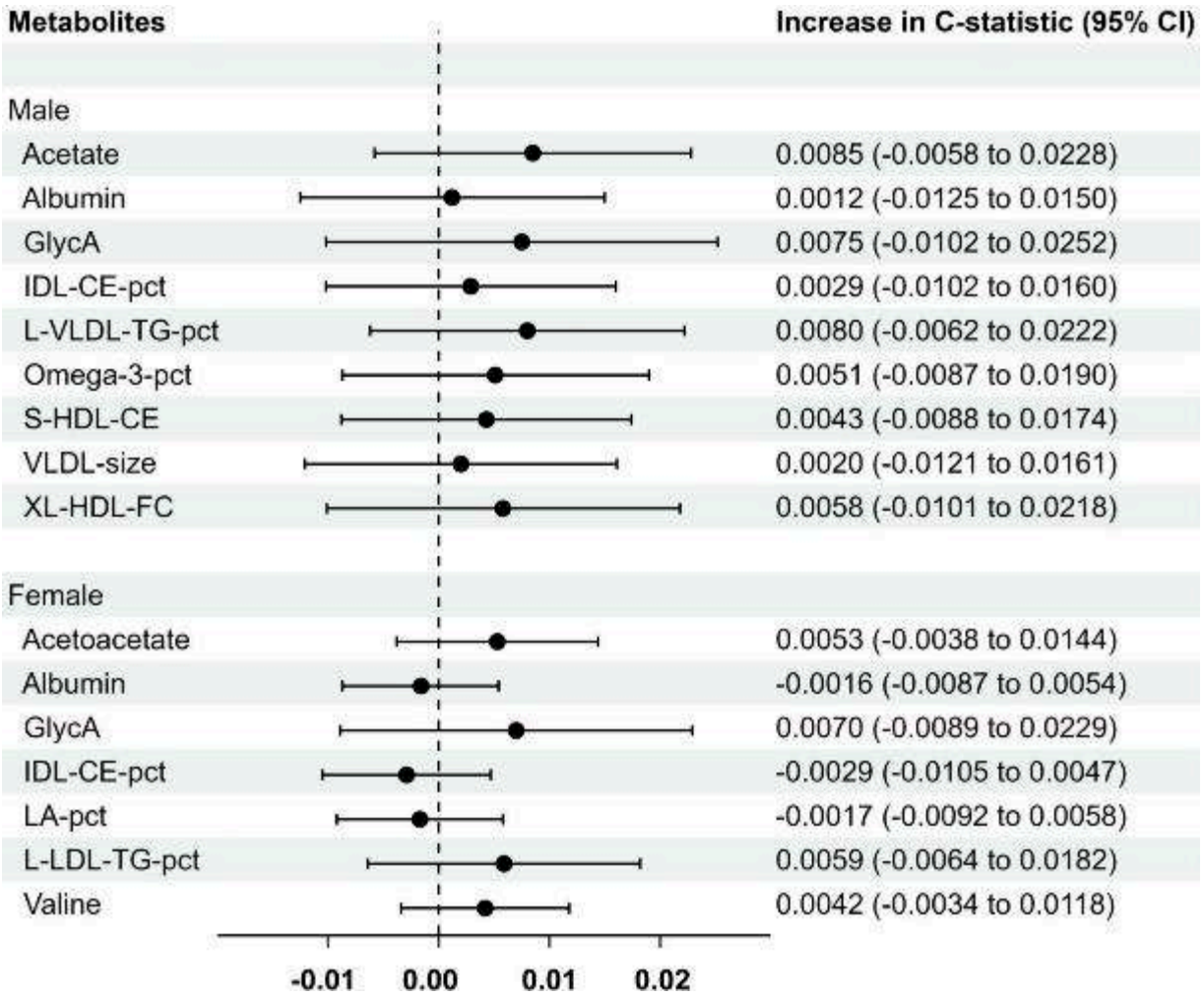
[a] The standard deviations of the metabolites acetate, albumin, GlycA, IDL-CE-pct, L-VLDL-TG-pct, Omega-3-pct, S-HDL-CE, VLDL-size, XL-HDL-FC, acetoacetate, LA-pct, L-LDL-TG-pct, and valine were 0.03 mmol/L, 3.35 mmol/L, 0.12 mmol/L, 2.31%, 5.84%, 1.58nm, 0.05 mmol/L, 1.21 nm, 0.01 mmol/L, 0.01 mmol/L, 3.33mmol/L, 1.68%, and 0.04 mmol/L, respectively.




**Appendix 24. Calibration curves of the SCORE2 model and its combination with the metabolites for 10-year cardiovascular risk prediction in the internal validation cohort (30% of UK Biobank, N=56,112)**

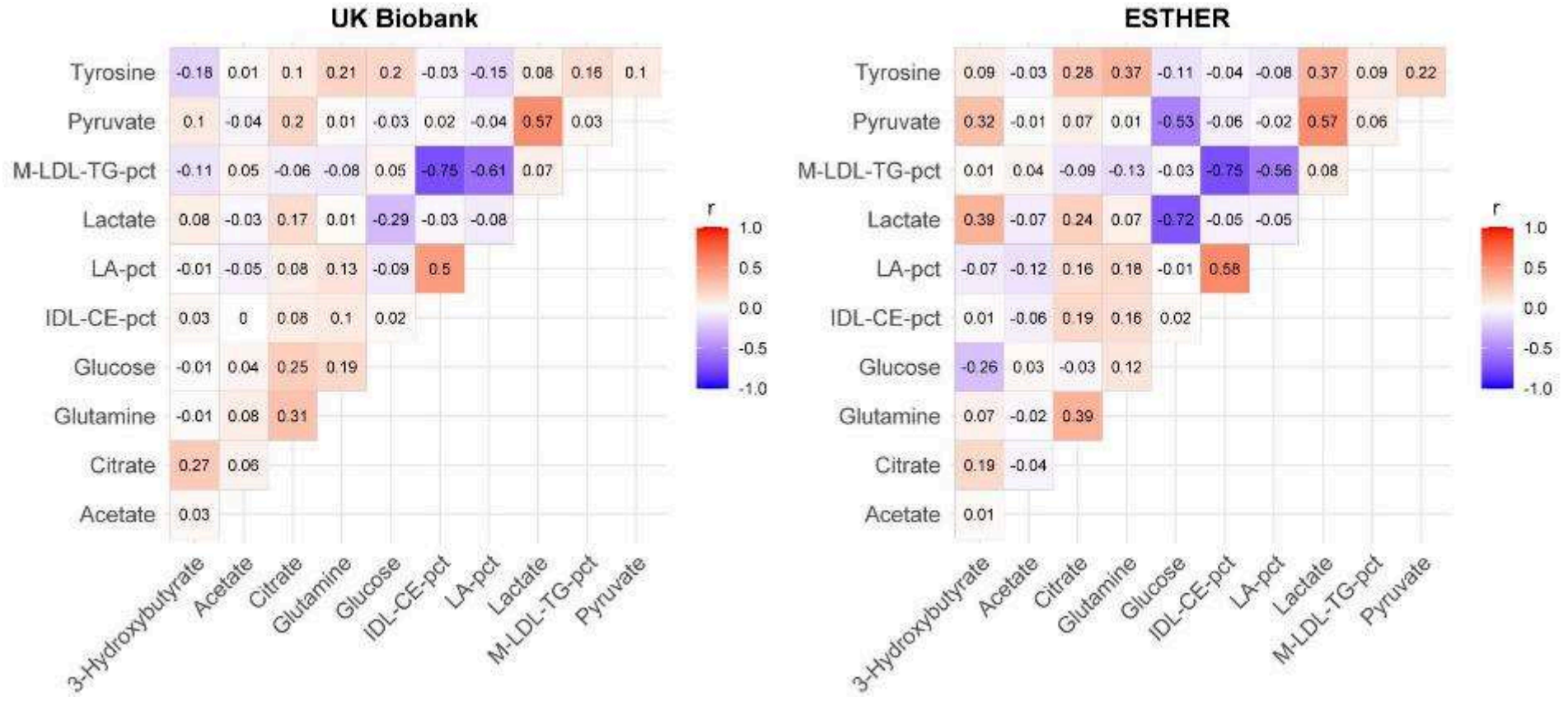



**Appendix 25. Calibration curves of the SCORE2 model and its combination with the metabolites for 10-year cardiovascular risk prediction in the external validation cohort (ESTHER study, N=5,578)**




**Appendix 26. The incremental discrimination of each metabolite for the model after the selected metabolites were added to SCORE2 separately in different sexes in the external validation cohort (ESTHER study, N=5,578)**

**Abbreviations:** CI, confidence interval; GlycA, glycoprotein acetyls; IDL-CE-pct, the cholesteryl esters percentage of total lipids in intermediate-density lipoprotein; LA-pct, the linoleic acid percentage of total fatty acids; L-LDL-TG-pct, triglycerides to total lipids in large low-density lipoprotein percentage; L-VLDL-TG-pct, the triglycerides percentage of total lipids in large very-low-density lipoprotein; Omega-3-pct, the Omega-3 fatty acids percentage of total fatty acids; S-HDL-CE, cholesteryl esters in small high-density lipoprotein; VLDL-size, average diameter for very-low-density lipoprotein particles; XL-HDL-FC, free cholesterol in very large high-density lipoprotein.




**Appendix 27. Correlation matrix of Pearson correlation coefficients for the 11 selected metabolites**

**Abbreviations:** IDL-CE-pct, cholesteryl esters to total lipids in IDL percentage; LA-pct, linoleic acid to total fatty acids percentage; M-LDL-TG-pct, triglycerides to total lipids in medium LDL percentage.

**Appendix 28. Hazard ratios per standard deviation increments with 95% confidence intervals and p-values for the associations of the selected metabolites for incident diabetes prediction in the test set (30% of UK Biobank, N=25,870) and external validation set (ESTHER, N=4,383)**

| Metabolites | Test set | | | External validation set | | |
|---|---|---|---|---|---|---|
| | SD | HR (95%CI) | *P*-value | SD | HR (95%CI) | *P*-value |
| 3-Hydroxybutyrate | 0.063 mmol/L | 0.99 (0.92, 1.05) | 0.671 | 0.059 mmol/L | 1.22 (1.13, 1.30) | <0.001 |
| Acetate | 0.035 mmol/L | 1.01 (0.95, 1.06) | 0.798 | 0.062 mmol/L | 0.93 (0.82, 1.05) | 0.247 |
| Citrate | 0.013 mmol/L | 1.10 (1.03, 1.17) | 0.003 | 0.016 mmol/L | 1.04 (0.95, 1.14) | 0.425 |
| Glutamine | 0.083 mmol/L | 0.93 (0.87, 0.99) | 0.017 | 0.106 mmol/L | 0.87 (0.80, 0.95) | 0.002 |
| Glucose | 0.863 mmol/L | 1.49 (1.41, 1.57) | <0.001 | 1.478 mmol/L | 1.63 (1.44, 1.85) | <0.001 |
| IDL-CE-pct | 2.369 % | 0.76 (0.72, 0.80) | <0.001 | 2.548 % | 0.78 (0.72, 0.84) | <0.001 |
| LA-pct | 3.325 % | 0.76 (0.72, 0.80) | <0.001 | 3.574 % | 0.77 (0.71, 0.83) | <0.001 |
| Lactate | 1.072 mmol/L | 1.21 (1.14, 1.29) | <0.001 | 3.351 mmol/L | 1.25 (1.13, 1.38) | <0.001 |
| M-LDL-TG-pct | 1.581 % | 1.34 (1.27, 1.40) | <0.001 | 2.075 % | 1.38 (1.25, 1.53) | <0.001 |
| Pyruvate | 0.028 mmol/L | 1.15 (1.07, 1.23) | <0.001 | 0.068 mmol/L | 1.21 (0.94, 1.56) | 0.139 |
| Tyrosine | 0.014 mmol/L | 1.19 (1.12, 1.26) | <0.001 | 0.015 mmol/L | 1.24 (1.14, 1.36) | <0.001 |

Note: Each metabolite was included individually in a Cox proportional hazards regression model adjusted for the variables of the clinical CDRS.

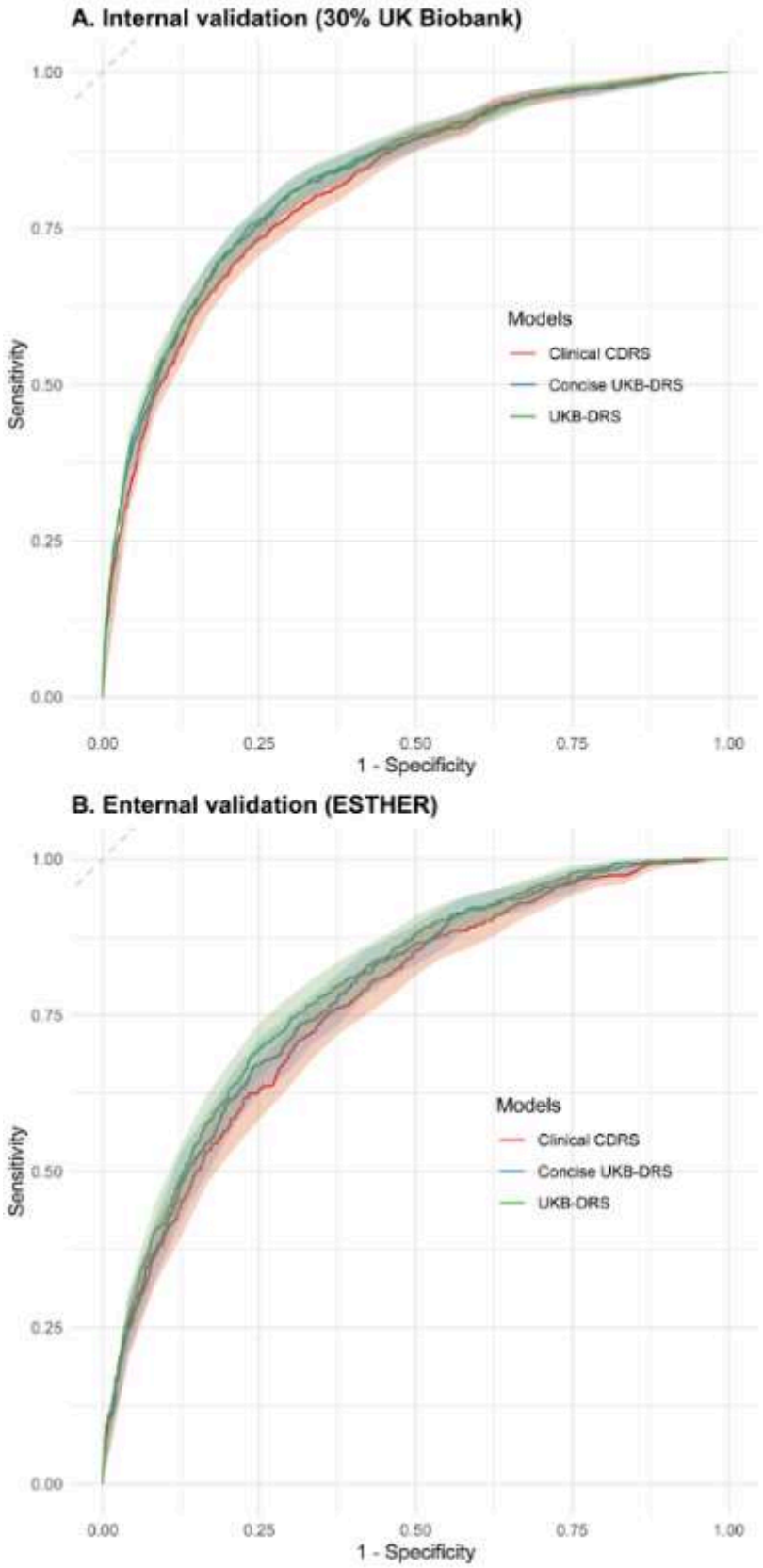
**Abbreviations:** CI, confidence interval; IDL-CE-pct, cholesteryl esters to total lipids in IDL percentage; LA-pct, linoleic acid to total fatty acids percentage; M-LDL-TG-pct, triglycerides to total lipids in medium LDL percentage; SD, standard deviation.

**Appendix 29. ß-coefficients of the variables of the UKB-DRS for 10-year prediction of T2D**

| Risk factor (units) | ß-coefficient | SE (coef) |
|---|---|---|
| **Clinical CDRS variables** | | |
| Female | -0.2878 | 0.0473 |
| Prescribed anti-hypertension medication | 0.1046 | 0.0536 |
| Prescribed steroids | 0.3950 | 0.1359 |
| Age (per years) | 0.1018 | 0.0040 |
| Body mass index <25 kg/m$^2$ | Ref | Ref |
| Body mass index 25-27.49 kg/m$^2$ | 0.3632 | 0.0819 |
| Body mass index 27.5-29.99 kg/m$^2$ | 0.6876 | 0.0783 |
| Body mass index $\geq$ 30 kg/m$^2$ | 1.2019 | 0.0722 |
| No first degree relative with diabetes | Ref | Ref |
| Parent or sibling with diabetes | 0.5114 | 0.0480 |
| Parent and sibling with diabetes | 0.7203 | 0.1033 |
| None-smoker | Ref | Ref |
| Ex-smoker | 0.1297 | 0.0447 |
| Current smoker | 0.3409 | 0.0654 |
| $HbA_{1c}$ (per mmol/mol) | 0.1014 | 0.0018 |
| **Added metabolites** | | |
| 3-Hydroxybutyrate (per 1 SD[a]) | 0.0493 | 0.0247 |
| Acetate (per 1 SD[a]) | -0.0214 | 0.0206 |
| Citrate (per 1 SD[a]) | -0.0286 | 0.0246 |
| Glutamine (per 1 SD[a]) | -0.1055 | 0.0216 |
| Glucose (per 1 SD[a]) | 0.3880 | 0.0217 |
| IDL-CE-pct (per 1 SD[a]) | -0.1233 | 0.0277 |
| LA-pct (per 1 SD[a]) | -0.0691 | 0.0251 |
| Lactate (per 1 SD[a]) | 0.1707 | 0.0271 |
| M-LDL-TG-pct (per 1 SD[a]) | 0.1300 | 0.0326 |
| Pyruvate (per 1 SD[a]) | -0.0454 | 0.0291 |
| Tyrosine (per 1 SD[a]) | 0.0458 | 0.0216 |

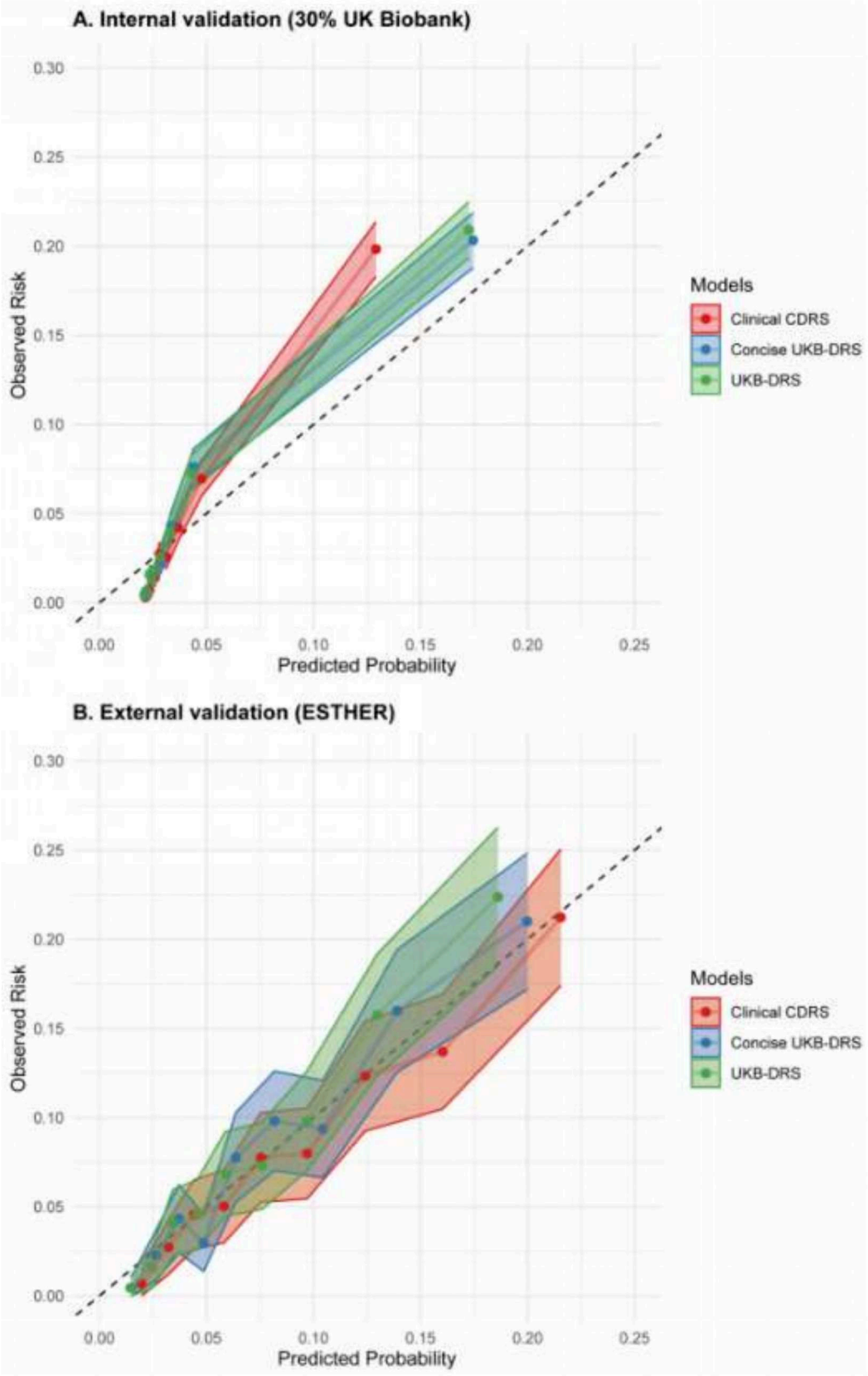
[a] The standard deviations of the 11 selected metabolites can be found in **Appendix 28**.

**Abbreviations:** CDRS, Cambridge Diabetes Risk Score; IDL-CE-pct, cholesteryl esters to total lipids in IDL percentage; LA-pct, linoleic acid to total fatty acids percentage; M-LDL-TG-pct, triglycerides to total lipids in medium LDL percentage; Ref, reference; SD, standard deviation; SE, standard error.




**Appendix 30. ROC curves for the UKB-DRS, concise UKB-DRS, and clinical CDRS in internal validation (30% UK Biobank, N=25,870) and external validation (ESTHER, N=4,383)**

**Abbreviations:** ROC, receiver operating characteristic; CDRS, Cambridge Diabetes Risk Score; UKB-DRS, UK Biobank Diabetes Risk Score




**Appendix 31. Calibration curves for the UKB-DRS, Concise UKB-DRS, and Clinical CDRS in internal validation (30% UK Biobank, N=25,870) and external validation (ESTHER, N=4,383)**

Notes: Confidence intervals are represented by shaded regions for each model.
**Abbreviations:** CDRS, Cambridge Diabetes Risk Score; UKB-DRS, UK Biobank Diabetes Risk Score.

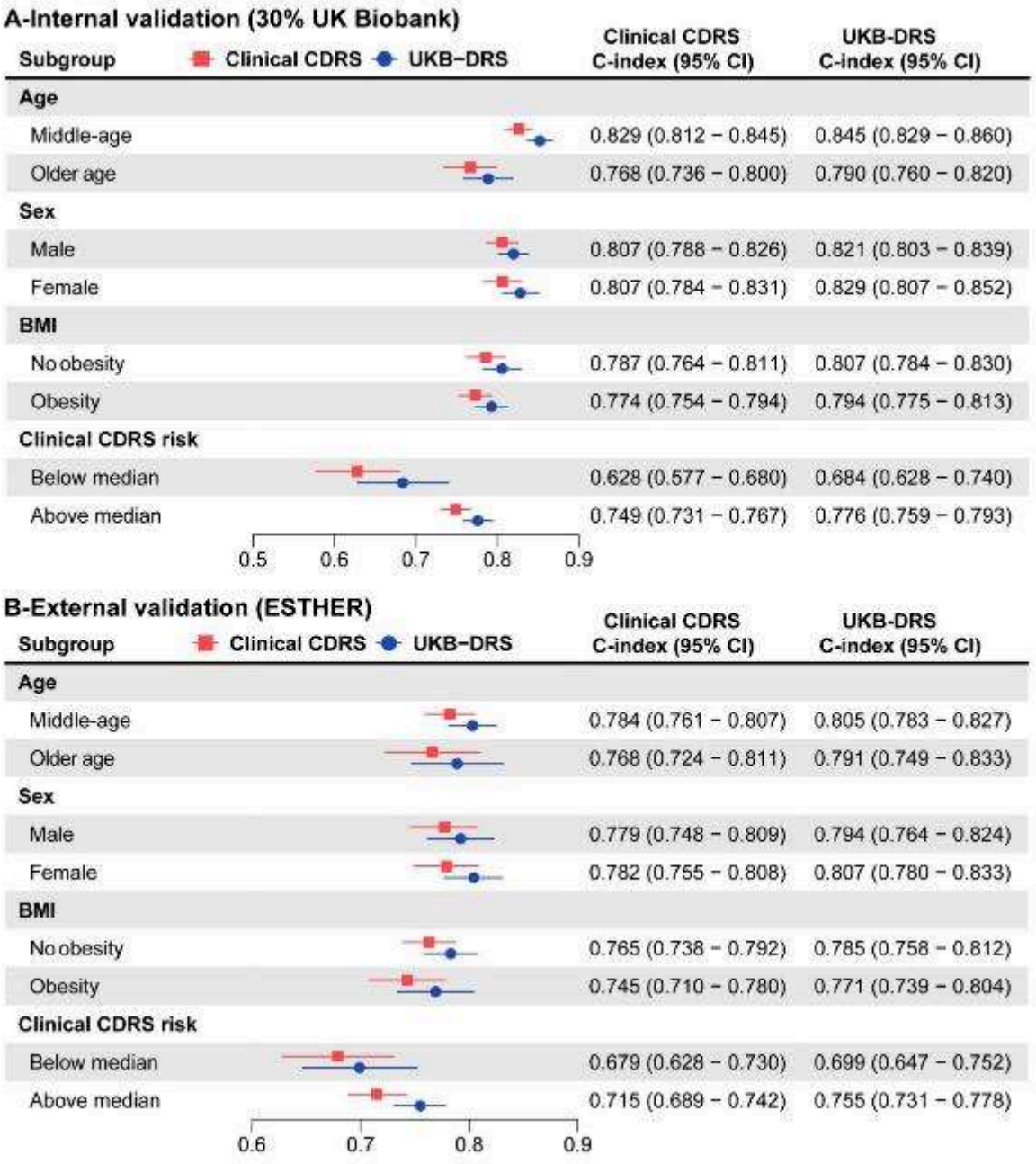
**Appendix 32. Comparison of the predictive performance of the clinical CDRS plus glucose and the UKB-DRS for 10-year type 2 diabetes risk prediction**

| **Metrics** | |
|---|---|
| **Training set (70% of UK Biobank)** | |
| **Sample size / incident type 2 diabetes** | **N=60,362 / 2,477** |
| C-index (clinical CDRS[a] + glucose ) | 0.835 (0.826, 0.843) |
| C-index (UKB-DRS[b]) | 0.847 (0.838, 0.855) |
| Δ C-index; *P*-value | +0.012; *P*<0.001 |
| **Test set (30% of UK Biobank)** | |
| **Sample size / incident type 2 diabetes** | **N=25,870 / 1,059** |
| C-index (clinical CDRS[a] + glucose ) | 0.820 (0.805, 0.834) |
| C-index (UKB-DRS[b]) | 0.834 (0.820, 0.847) |
| Δ C-index; *P*-value | +0.014; *P*<0.001 |
| Continuous NRI (%); *P*-value | 26.3 (21.2, 31.4); *P*<0.001 |
| IDI; *P*-value | 0.0015 (-0.0019, 0.0048); *P*=0.386 |
| **External validation set (ESTHER)** | |
| **Sample size / incident type 2 diabetes** | **N=4,383 / 495** |
| C-index (clinical CDRS[a] + glucose ) | 0.773 (0.755, 0.791) |
| C-index (UKB-DRS[b]) | 0.798 (0.779, 0.817) |
| Δ C-index; *P*-value | +0.025; *P*<0.001 |
| Continuous NRI (%); *P*-value | 13.2 (7.1, 19.2); *P*<0.001 |
| IDI; *P*-value | 0.0013 (0.0004, 0.0022); *P*<0.001 |

**Abbreviations:** CDRS, Cambridge Diabetes Risk Score; IDI, integrated discrimination index; NRI, net reclassification index, UKB-DRS (UK Biobank Diabetes Risk Score)

[a] The Cambridge Diabetes Risk Score (CDRS) consists of the following diabetes risk factors: (age, sex, BMI, smoking status, family history of diabetes, prescription of anti-hypertensive medication and steroids).

[b] The UK Biobank Diabetes Risk Score (UKB-DRS) is based on the traditional diabetes risk factors of the clinical CDRS (see above) and the following 11 metabolites: glucose, citrate, lactate, pyruvate, 3-hydroxybutyrate, acetate, glutamine, tyrosine, IDL-CE-pct, LA-pct, and M-LDL-TG-pct.




**Appendix 33. Subgroup analyses comparing the C-index of the UKB-DRS and the clinical CDRS in internal validation (30% UK Biobank, N=25,870) and external validation (ESTHER, N=4,383)**

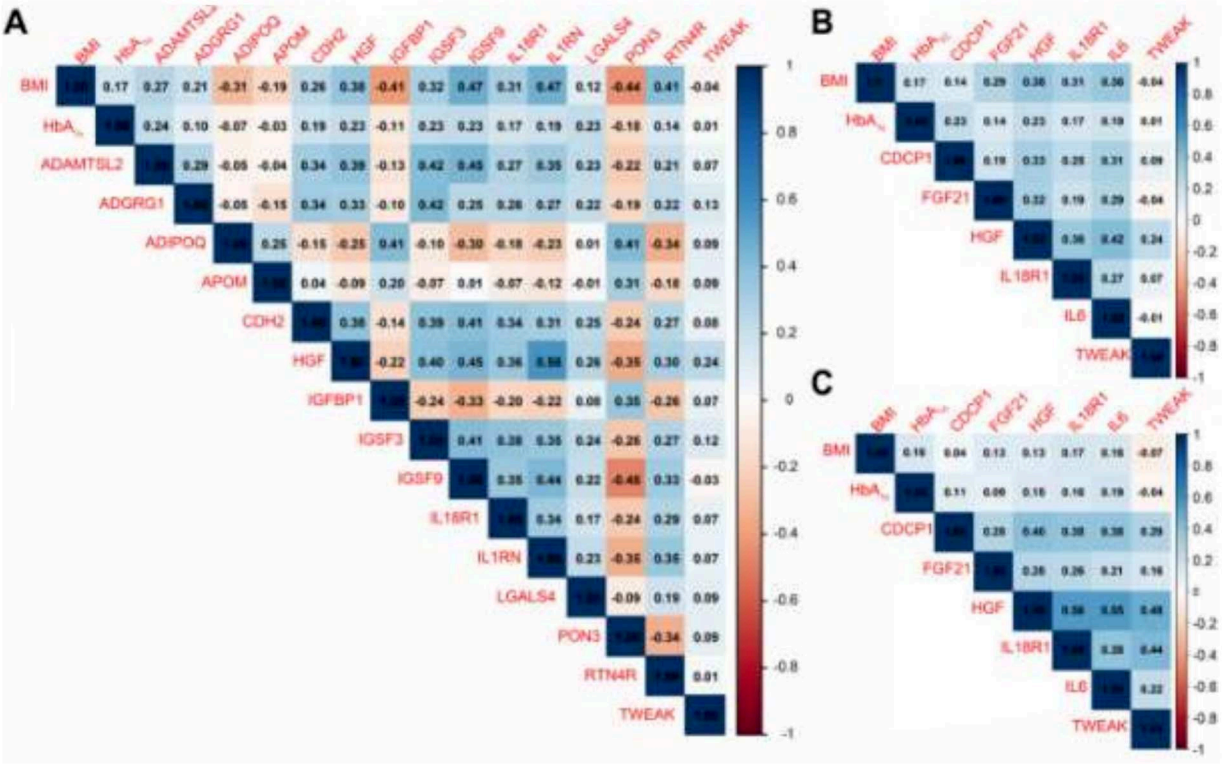
**Notes:** Middle-aged is defined as <65 years and older age as ≥65 years. Obesity is defined as a BMI ≥30 kg/m².
**Abbreviations:** CDRS, Cambridge Diabetes Risk Score; UKB-DRS, UK Biobank Diabetes Risk Score; CI, confidence interval.

**Appendix 34. NRI and IDI if each of the 11 selected metabolite is added one by one to the clinical CDRS in the test set (30% of UK Biobank, N=25,870) and external validation set (ESTHER, N=4,383)**

| Metabolites | Test set (N=25,870) | | External validation set (N=4,383) | |
|---|---|---|---|---|
| | NRI (95% CI) | IDI (95% CI)) | NRI (95% CI)) | IDI (95% CI)) |
| 3-Hydroxybutyrate | 2.5% (-2.4%, 7.5%) | **0.000 (0.000, 0.001)** | 14.2% (-2.4%, 27.0%) | **0.000 (0.000, 0.000)** |
| Acetate | **3.9% (1.0%, 5.8%)** | **0.000 (0.000, 0.001)** | 4.6% (-3.2%, 10.1%) | **0.000 (0.000, 0.000)** |
| Citrate | **7.1% (2.0%, 12.2%)** | **0.002 (0.001, 0.002)** | 2.4% (-5.7%, 6.5%) | 0.000 (-0.001, 0.001) |
| Glutamine | 3.1% (-2.2%, 8.3%) | **0.000 (0.000, 0.000)** | 1.0% (-6.3%, 4.0%) | 0.001 (-0.001, 0.002) |
| Glucose | **22.5% (17.3%, 27.6%)** | **0.009 (0.005, 0.013)** | 6.6% (-0.8%, 13.9%) | 0.001 (-0.006, 0.008) |
| IDL-CE-pct | **27.5% (22.4%, 32.7%)** | **0.005 (0.002, 0.007)** | **18.7% (11.6%, 25.7%)** | -0.002 (-0.004, 0.001) |
| LA-pct | **24.2% (19.2%, 29.2%)** | 0.001 (-0.002, 0.003) | **8.8% (2.3%, 15.2%)** | 0.003 (-0.001, 0.007) |
| Lactate | 4.1% (-1.2%, 9.3%) | 0.005 (-0.006, 0.016) | 6.5% (-1.1%, 14.1%) | -0.001 (-0.004, 0.001) |
| M-LDL-TG-pct | **28.4% (23.2%,33.5%)** | **0.006 (0.003, 0.009)** | **11.9% (5.0%, 18.8%)** | **0.001 (0.000, 0.001)** |
| Pyruvate | **8.8% (3.5%, 14.0%)** | **0.001 (0.000, 0.001)** | **5.1% (0.2%, 13.8%)** | **0.001 (0.000, 0.002)** |
| Tyrosine | **14.1% (8.9%, 19.3%)** | **0.001 (0.000, 0.002)** | **13.6% (6.0%, 23.2%)** | **0.000 (0.000, 0.000)** |

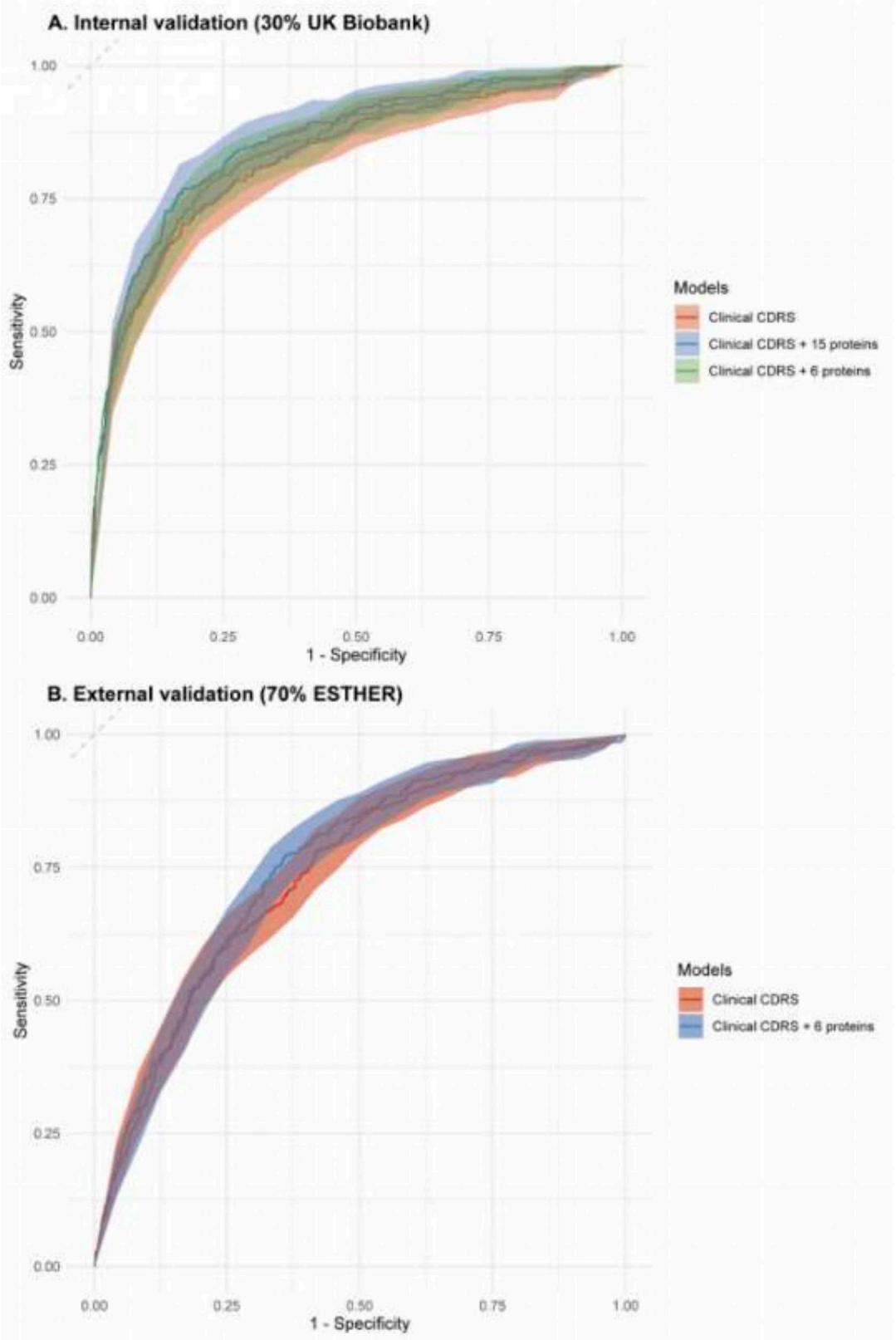
**Abbreviations:** CI, confidence interval; IDL-CE-pct, cholesteryl esters to total lipids in IDL percentage; LA-pct, linoleic acid to total fatty acids percentage; M-LDL-TG-pct, triglycerides to total lipids in medium LDL percentage; SD, standard deviation.

**Bold print:** Statistically significant (P<0.05).




**Appendix 35. Correlation matrix showing Spearman correlation coefficients between BMI, $HbA_{1c}$, and the selected proteins**

Panel A, 30% of UK Biobank (N=6,339) and 15 proteins selected from the OLINK Explore; Panel B, 30% of UK Biobank (N=6,339) and 6 proteins selected from OLINK Target Inflammation panel; Panel C, 70% of ESTHER (N=3,118) and 6 proteins selected from OLINK Target Inflammation panel.


A. Internal validation (30% UK Biobank)
Sensitivity
1 - Specificity
0.00
0.25
0.50
0.75
1.00
Models
Clinical CDRS
Clinical CDRS + 15 proteins
Clinical CDRS + 6 proteins
B. External validation (70% ESTHER)
Sensitivity
1 - Specificity
Models
Clinical CDRS
Clinical CDRS + 6 proteins

**Appendix 36. ROC curves of the clinical CDRS with and without proteomics data for 10-year type 2 diabetes risk prediction in the internal validation (30% of UK Biobank, N=6,339) and external validation (70% of ESTHER, N=3,118)**

**Appendix 37. ß-coefficients of the variables of the 15 proteins extended clinical CDRS for 10-year prediction of type 2 diabetes**

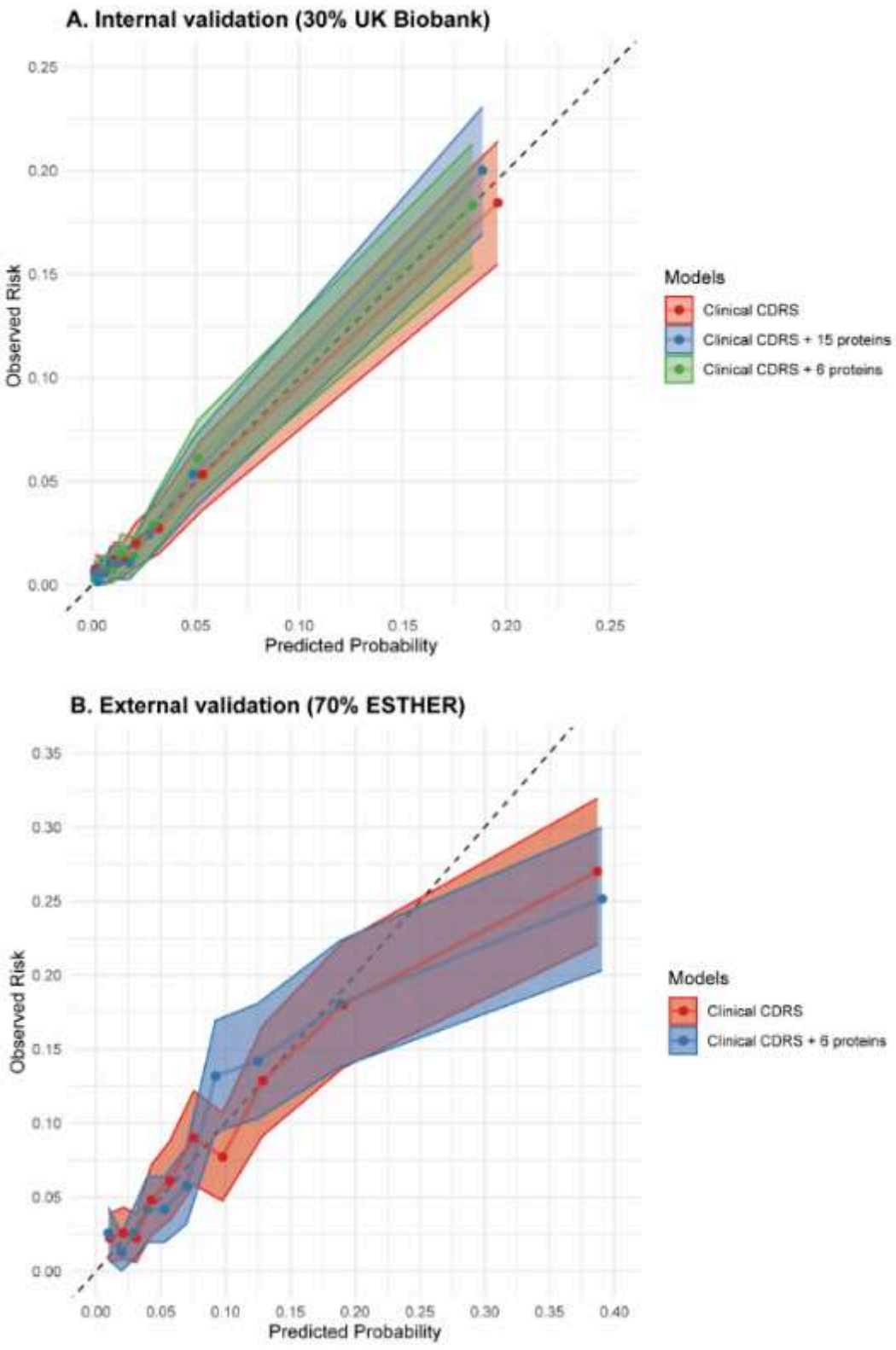
| Risk factor (units) | ß-coefficient |
|---|---|
| **Clinical CDRS variables** | |
| Female | -0.2074 |
| Prescribed anti-hypertension medication | -0.0369 |
| Prescribed steroids | 0.9681 |
| Age (per years) | 0.0121 |
| Body mass index <25 kg/m$^2$ | Ref |
| Body mass index 25-27.49 kg/m$^2$ | -0.1543 |
| Body mass index 27.5-29.99 kg/m$^2$ | -0.0799 |
| Body mass index ≥ 30 kg/m$^2$ | 0.1483 |
| No first degree relative with diabetes | Ref |
| Parent or sibling with diabetes | 0.3918 |
| Parent and sibling with diabetes | 0.6208 |
| None-smoker | Ref |
| Ex-smoker | 0.1475 |
| Current smoker | -0.0661 |
| $HbA_{1c}$ (per 1%) | 2.5019 |
| **Added proteins (per 1 NPX)** | |
| ADAMTSL2 | -0.1586 |
| ADGRG1 | 0.0051 |
| Adiponectin | -0.1281 |
| Apolipoprotein M | -0.4961 |
| CDH2 | 0.5629 |
| HGF | 0.3801 |
| IGFBP1 | -0.1312 |
| IGSF3 | -0.0182 |
| IGSF9 | 0.1307 |
| IL18R1 | 0.3150 |
| IL1RN | 0.0342 |
| LGALS4 | 0.0721 |
| PON3 | -0.2585 |
| RTN4R | 0.2623 |
| TWEAK | -0.4359 |

**Abbreviations:** ADAMTSL2, ADAMTS-like protein 2; ADGRG1, Adhesion G-protein coupled receptor G1; CDH2, Cadherin-2; CDRS, Cambridge Diabetes Risk Score; HGF, Hepatocyte growth factor; IGFBP1, Insulin-like growth factor-binding protein 1; IGSF, Immunoglobulin superfamily member; IL1RN, Interleukin-1 receptor antagonist protein; IL18R1, Interleukin-18 receptor 1; PON3, Serum paraoxonase/lactonase 3; Ref, reference; RTN4R, Reticulon-4 receptor; TWEAK, TNF-related weak inducer of apoptosis.

**Appendix 38. ß-coefficients of the variables of the 6 proteins extended clinical CDRS for 10-year prediction of type 2 diabetes in UK Biobank and ESTHER study**

| Risk factor (units) | UK Biobank ß-coefficient | ESTHER ß-coefficient |
|---|---|---|
| **Clinical CDRS variables** | | |
| Female | -0.5146 | 0.0686 |
| Prescribed anti-hypertension medication | -0.0445 | 0.3604 |
| Prescribed steroids | 0.7184 | -0.1315 |
| Age (per years) | 0.0149 | -0.0153 |
| Body mass index <25 kg/m$^2$ | Ref | Ref |
| Body mass index 25-27.49 kg/m$^2$ | 0.0194 | 0.4507 |
| Body mass index 27.5-29.99 kg/m$^2$ | 0.2339 | 0.9075 |
| Body mass index ≥ 30 kg/m$^2$ | 0.6311 | 1.0740 |
| No first degree relative with diabetes | Ref | Ref |
| Parent or sibling with diabetes | 0.4606 | 0.1697 |
| Parent and sibling with diabetes | 0.7630 | 0.0463 |
| None-smoker | Ref | Ref |
| Ex-smoker | 0.1152 | -0.1039 |
| Current smoker | -0.2600 | 0.2561 |
| $HbA_{1c}$ (per 1%) | 2.7175 | 1.8360 |
| **Added proteins (per 1 NPX)** | | |
| CDCP1 | -0.0012 | 0.1443 |
| FGF21 | 0.1110 | 0.1326 |
| HGF | 0.6310 | -0.1236 |
| IL18R1 | 0.5284 | 0.5146 |
| IL6 | 0.0474 | -0.0209 |
| TWEAK | -0.5795 | -0.5885 |

**Abbreviations:** CDCP1, CUB domain-containing protein 1; CDRS, Cambridge Diabetes Risk Score; FGF21, Fibroblast growth factor 21; HGF, Hepatocyte growth factor; IL18R1, Interleukin-18 receptor 1; IL6, Interleukin-6; Ref, reference; TWEAK, TNF-related weak inducer of apoptosis.




**Appendix 39. Calibration curves of the clinical CDRS with and without proteomics data for 10-year type 2 diabetes risk prediction in the internal validation (30% of UK Biobank, N=6,339) and external validation (70% of ESTHER, N=3,118)**

**Appendix 40. Comparison of 6 selected proteins levels by fasting status in the ESTHER study**

| Proteins (NPX) | ESTHER (N=4,454) | | |
|---|---|---|---|
| | Fasting (N=3,879) | Non-fasting (N=425) | |
| | Mean (95%CI) | Mean (95%CI) | ANOVA *p*-value |
| CDCP1 | 3.68 (3.61, 3.76) | 3.69 (3.67, 3.72) | 0.808 |
| FGF21 | 5.39 (5.26, 5.53) | 5.80 (5.76, 5.84) | <0.001 |
| HGF | 9.78 (9.71, 9.84) | 9.75 (9.73, 9.77) | 0.402 |
| IL18R1 | 9.04 (8.89, 9.09) | 9.04 (9.02, 9.05) | 0.975 |
| IL6 | 2.92 (2.75, 3.11) | 2.68 (2.63, 2.73) | 0.013 |
| TWEAK | 9.20 (9.13, 9.25) | 9.25 (9.23, 9.27) | 0.065 |

**Abbreviations:** CDCP1, CUB domain-containing protein 1; CI, confidence interval; FGF21, Fibroblast growth factor 21; HGF, Hepatocyte growth factor; IL18R1, Interleukin-18 receptor 1; IL6, Interleukin-6; NPX, Normalised Protein eXpression; TWEAK, TNF-related weak inducer of apoptosis.

**Appendix 41. Comparison of the predictive performance of the clinical Cambridge Diabetes Risk Score with and without proteomic biomarkers for 10-year type 2 diabetes risk prediction in UK Biobank without non-random selected participants**

| Metrics | |
|---|---|
| **Derivation set (70% of UK Biobank)** | |
| **Sample size / incident type 2 diabetes** | **N=13,470 / 460** |
| C-index (Clinical CDRS[a]) | 0.855 (0.837, 0.874) |
| C-index (+15 proteins derived from OLINK Explore[b]); *P*-value[c] | **0.887 (0.871, 0.904); *P*<0.001** |
| C-index (+6 proteins derived from OLINK Target Inflammation panel[d]); *P*-value[c] | **0.878 (0.861, 0.895); *P*<0.001** |
| **Internal validation set (30% of UK Biobank)** | |
| **Sample size / incident type 2 diabetes** | **N=5,773 / 194** |
| C-index (Clinical CDRS[a]) | 0.847 (0.814, 0.879) |
| C-index (+15 proteins derived from OLINK Explore[b]); *P*-value[c] | **0.869 (0.841, 0.897); *P*<0.001** |
| Continuous NRI (%); *P*-value[c] | **21.7 (7.0, 36.4); *P*<0.001** |
| IDI; *P*-value[c] | -0.003 (-0.011, 0.005); *P*=0.478 |
| C-index (+6 proteins derived from OLINK Target Inflammation panel[d]); *P*-value[c] | **0.858 (0.828, 0.888); *P*<0.001** |
| Continuous NRI (%) ; *P*-value[c] | **25.5 (10.7, 40.3); *P*<0.001** |
| IDI; *P*-value[c] | -0.000 (-0.008, 0.007); *P*=0.946 |

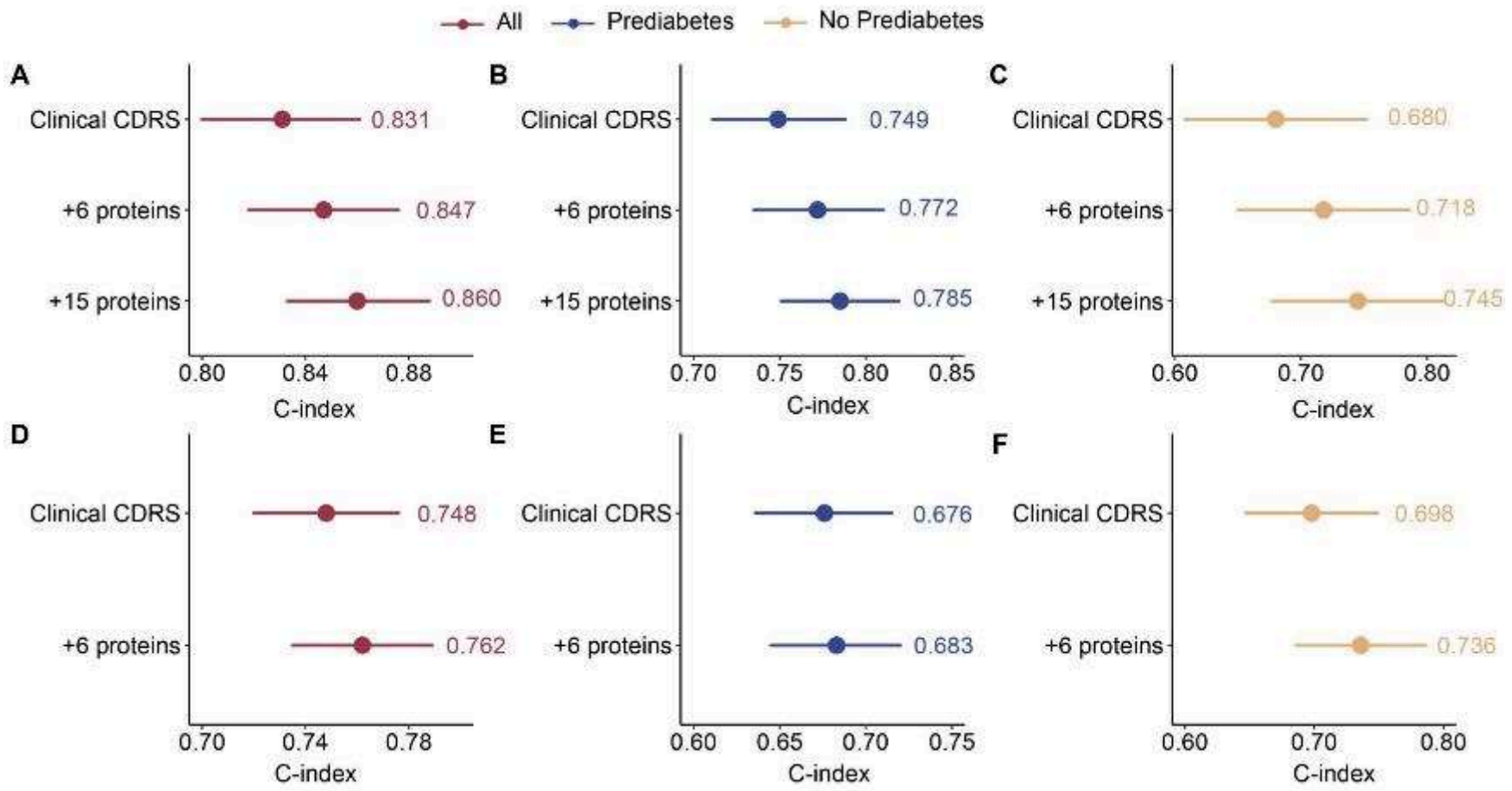
**Abbreviations:** CDRS, Cambridge Diabetes Risk Score; IDI, integrated discrimination improvement; NRI, net reclassification improvement

[a] The clinical Cambridge Diabetes Risk Score (CDRS) consists of the following diabetes risk factors: (age, sex, BMI, smoking status, family history of diabetes, prescription of anti-hypertensive medication and steroids, and $HbA_{1c}$).

[b] The clinical CDRS (see above) and the following 15 proteins: ADAMTSL2, ADGRG1, Adiponectin, Apolipoprotein M, CDH2, HGF, IGFBP1, IGSF3, IGSF9, IL18R1, IL1RN, LGALS4, PON3, RTN4R and TWEAK (see legend of Figure 2 for protein abbreviations).

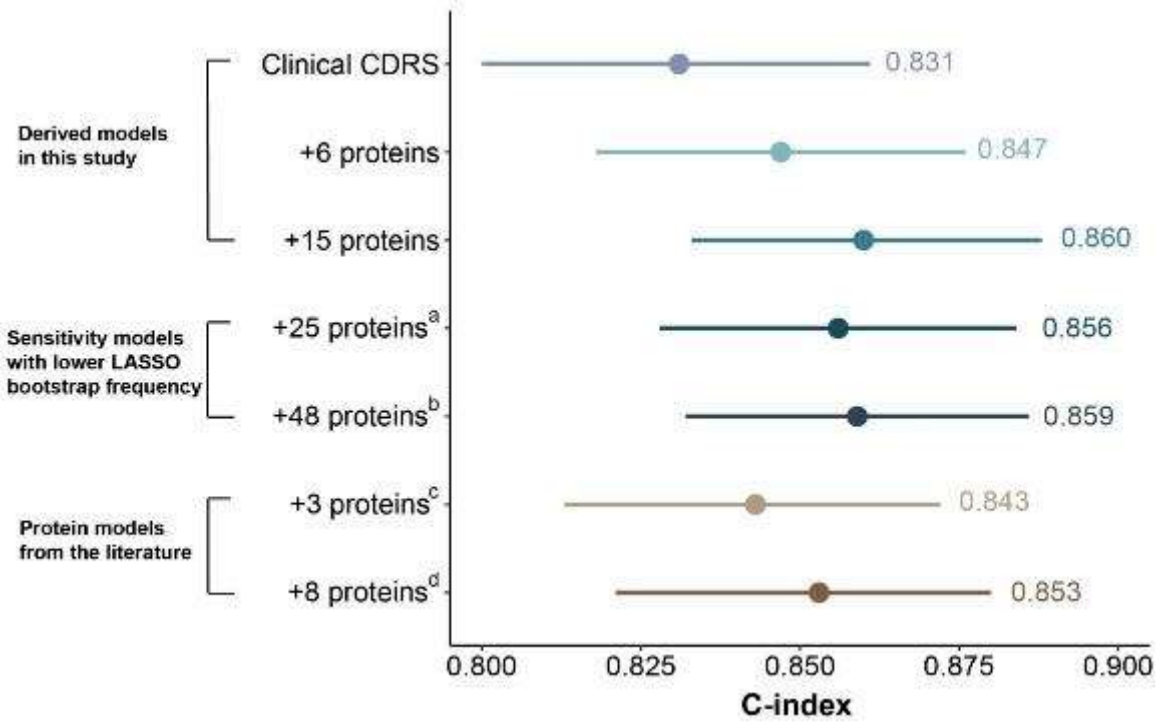
[c] *P*-value for comparison with the clinical CDRS.

[d] The clinical CDRS (see above) and the following 6 proteins: CDCP1, FGF21, HGF, IL18R1, IL6, and TWEAK (see legend of Figure 2 for proteins abbreviations).




**Appendix 42. Comparison of the incremental improvements in the C-index by proteomic biomarkers added to the clinical CDRS in the total population, subjects with and without pre-diabetes status in the UKB internal validation and external validation with the ESTHER study**

(A) UKB total population (n=6,339), (B) UKB participants with pre-diabetes (n=994), (C) UKB participants without pre-diabetes (n=5,576), (D) ESTHER total population (n=3,118), (E) ESTHER participants with pre-diabetes (n=1,102), and (F) ESTHER participants without pre-diabetes (n=2,016).




**Appendix 43. Comparison of the C-index of the 6- and 15-protein-extended models derived in this study with different protein selection approaches with LASSO bootstrapping and suggested protein panels from the literature in the internal validation set (30% of UK Biobank, N=6,339)**

[a] Clinical CDRS + 25 proteins derived from a 90% LASSO-bootstrap selection frequency.

[b] Clinical CDRS + 48 proteins derived from an 80% LASSO-bootstrap selection frequency.

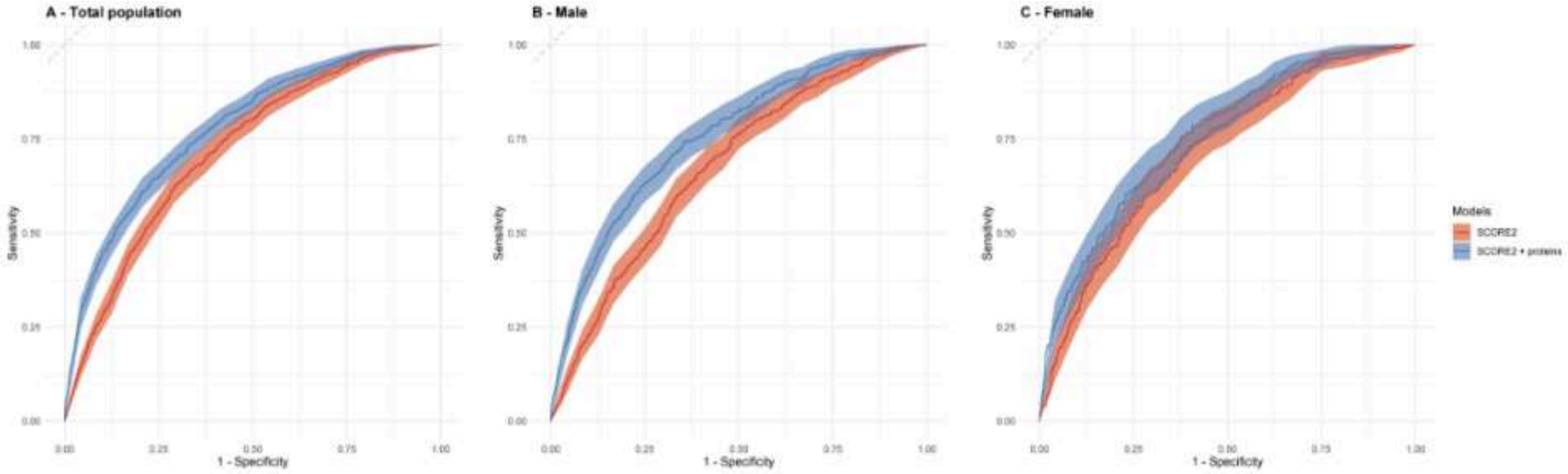
[c] Clinical CDRS + 3 proteins identified in the Fenland study for identification of impaired glucose tolerance: reticulon-4 receptor (RTN4R), carboxypeptidase M (CBPM), and growth hormone receptor (GHR).

[d] Clinical CDRS + 8 out of 10 proteins identified in the EPIC Norfolk study for type 2 diabetes risk prediction. Two proteins were excluded due to unavailability in the OLINK Explore platform: SLIK3 (SLIT and NTRK (Neurotrophic Receptor Tyrosine Kinase) like family member 3) and HTRA1 (high temperature requirement protease A serine peptidase 1).

**Appendix 44. ß-coefficients of the variables of the SCORE2 model extended by 18 proteins for 10-year prediction of MACE**

| Risk factor (units) | Coefficient | |
|---|---|---|
| | **Male** | **Female** |
| **SCORE2 variables** | | |
| Age (per 5 years) | 0.1882 | 0.2833 |
| Current smoking | 0.1601 | 0.4543 |
| Systolic blood pressure (per 20mmHg) | 0.1157 | 0.2152 |
| Total cholesterol (per 1 mmol/L) | 0.0772 | 0.0027 |
| HDL cholesterol (per 0.5 mmol/L) | -0.1286 | -0.0466 |
| Smoking interaction with age | -0.0378 | -0.0678 |
| SBP interaction with age | -0.0839 | -0.0464 |
| Total cholesterol interaction with age | -0.0351 | -0.0242 |
| HDL interaction with age | -0.0213 | 0.0326 |
| **Additional proteins (per 1 SD)** | | |
| ADAMTS13 | -0.5045 | - |
| ADGRG2 | - | -0.5115 |
| BCAN | -0.4317 | -0.3065 |
| CCL7 | 0.1692 | - |
| CRYBB2 | - | 0.1065 |
| CXCL17 | 0.1976 | - |
| EDA2R | - | 0.0039 |
| EGFR | - | -0.5132 |
| GDF15 | 0.2204 | 0.3188 |
| HAVCR1 | - | 0.1832 |
| IL6 | 0.0826 | - |
| LRRN1 | -0.1956 | - |
| MMP12 | 0.2937 | 0.2286 |
| NEFL | 0.2009 | - |
| NTproBNP | 0.1906 | 0.1801 |
| PRSS8 | 0.2465 | - |
| SEPTIN8 | - | 0.3420 |
| WFDC2 | -0.0738 | 0.2344 |

**Abbreviations:** ADAMTS13, A disintegrin and metalloproteinase with thrombospondin motifs 13; ADGRG2, Adhesion G-protein coupled receptor G2; BCAN, Brevican core protein; CCL7, C-C motif chemokine 7; CRYBB2, Beta-crystallin B2; CXCL17, C-X-C motif chemokine 17; EDA2R, Tumor necrosis factor receptor superfamily member 27; EGFR, Epidermal growth factor receptor; GDF15, Growth/differentiation factor 15; HAVCR1, Hepatitis A virus cellular receptor 1; IL6, Interleukin-6; LRRN1, Leucine-rich repeat neuronal protein 1; MMP12, Macrophage metalloelastase; NEFL, Neurofilament light polypeptide; NTproBNP, N-terminal prohormone of brain natriuretic peptide; PRSS8, Prostasin; SD, standard deviation; SEPTIN8, Septin-8; WFDC2, WAP four-disulfide core domain protein.




**Appendix 45. Receiver operating characteristic (ROC) curves for the SCORE2 model with and without proteomic biomarkers in the validation cohort (30% of UK Biobank, N=14,216)**

The ROC curves for the SCORE2 model (red line) and the SCORE2 model augmented with proteomic biomarkers (blue line) for predicting 10-year major adverse cardiovascular events (MACE). The results are shown for the total population (**Panel A**), males (**Panel B**), and females (**Panel C**).

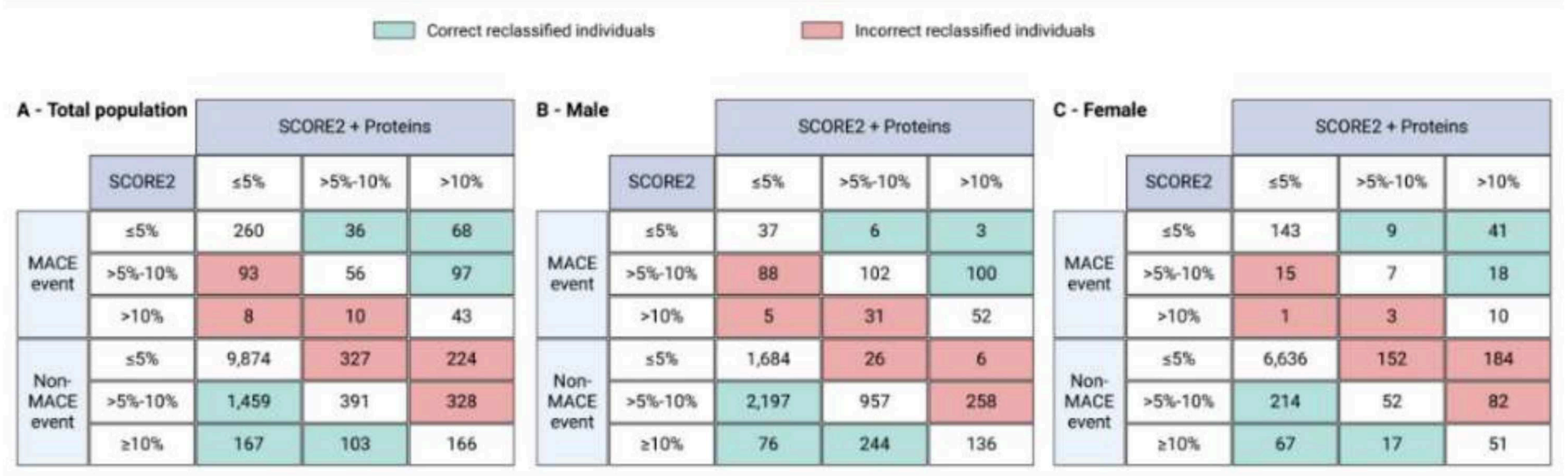
**Appendix 46. Metrics of the predictive performance of the (protein extended) SCORE2 model for 10-year MACE risk including only randomly selected participants**

| Metrics | Male * | Female † | Overall |
|---|---|---|---|
| **Derivation set (70% of UK Biobank)** | | | |
| **Total sample size (N=29,585) / MACE case number (N=1,164)** | | | |
| C-Statistics (SCORE2) | 0.676 (0.658,0.694) | 0.726 (0.704,0.749) | 0.727 (0.713,0.740) |
| C-Statistics (+Proteins) | 0.753 (0.735,0.770) | 0.774 (0.753, 0.794) | 0.780 (0.767,0.793) |
| *P*-values for C-Statistics comparisons | **<0.001** | **<0.001** | **<0.001** |
| **Validation set (30% of UK Biobank)** | | | |
| **Total sample size (N=12,683) / MACE case number (N=515)** | | | |
| C-Statistics (SCORE2) | 0.677 (0.649, 0.705) | 0.713 (0.678, 0.769) | 0.724 (0.703, 0.745) |
| C-Statistics (+Proteins) | 0.756 (0.729,0.783) | 0.747 (0.712,0.787) | 0.772 (0.752,0.792) |
| *P*-values for C-Statistics comparisons | **<0.001** | **<0.001** | **<0.001** |
| NRI total (%) | **22.1 (15.3, 29.2)** | **16.2 (8.7, 22.6)** | **20.2 (14.7, 24.7)** |
| NRI events (%) | 14.2 (-0.1, 22.6) | **19.5 (12.0, 26.7)** | **22.1 (16.5, 26.7)** |
| NRI non-events (%) | **7.0 (1.4, 21.7)** | **-3.3 (-4.2, -2.6)** | **-2.0 (-2.8, -1.0)** |
| IDI | **0.032 (0.022, 0.048)** | **0.015 (0.006, 0.028)** | **0.025 (0.018, 0.034)** |

**Abbreviations:** IDI, integrated discrimination index; MACE, major cardiovascular events; NRI, net reclassification index.

* Proteins included in the combined model for males were: ADAMTS13, BCAN, CCL7, CXCL17, GDF15, IL6, LRRN1, MMP12, NEFL, NTproBNP, PRSS8, WFDC2.

† Proteins included in the combined model for females were: ADGRG2, BCAN, CRYBB2, EDA2R, EGFR, GDF15, HAVCR1, MMP12, NTproBNP, SEPTIN8, WFDC2.




**A - Total population**

| | SCORE2 | SCORE2 + Proteins ≤5% | SCORE2 + Proteins >5%-10% | SCORE2 + Proteins >10% |
|---|---|---|---|---|
| MACE event | ≤5% | 260 | 36 | 68 |
| | >5%-10% | 93 | 56 | 97 |
| | >10% | 8 | 10 | 43 |
| Non-MACE event | ≤5% | 9,874 | 327 | 224 |
| | >5%-10% | 1,459 | 391 | 328 |
| | ≥10% | 167 | 103 | 166 |

**B - Male**

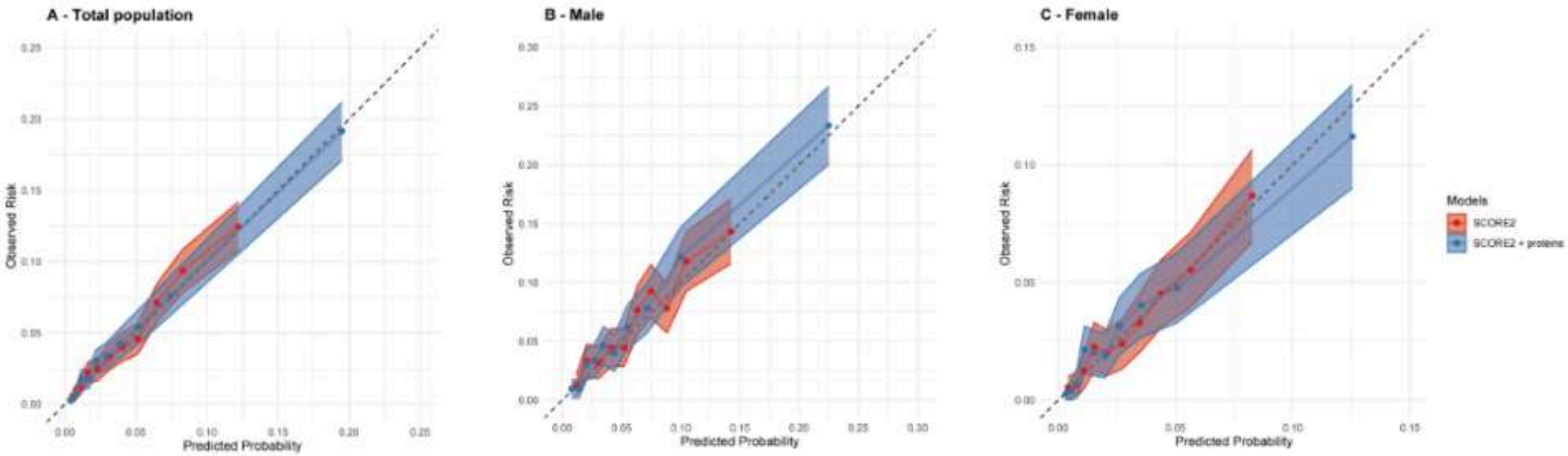
| | SCORE2 | SCORE2 + Proteins ≤5% | SCORE2 + Proteins >5%-10% | SCORE2 + Proteins >10% |
|---|---|---|---|---|
| MACE event | ≤5% | 37 | 6 | 3 |
| | >5%-10% | 88 | 102 | 100 |
| | >10% | 5 | 31 | 52 |
| Non-MACE event | ≤5% | 1,684 | 26 | 6 |
| | >5%-10% | 2,197 | 957 | 258 |
| | ≥10% | 76 | 244 | 136 |

**C - Female**

| | SCORE2 | SCORE2 + Proteins ≤5% | SCORE2 + Proteins >5%-10% | SCORE2 + Proteins >10% |
|---|---|---|---|---|
| MACE event | ≤5% | 143 | 9 | 41 |
| | >5%-10% | 15 | 7 | 18 |
| | >10% | 1 | 3 | 10 |
| Non-MACE event | ≤5% | 6,636 | 152 | 184 |
| | >5%-10% | 214 | 52 | 82 |
| | ≥10% | 67 | 17 | 51 |

**Appendix 47. Reclassification of the SCORE2 model with and without proteomic biomarkers in the validation set (30% of UK Biobank, N=14,216)**

The reclassification of individuals into cardiovascular risk categories (≤5%, >5%-10%, >10%) for major adverse cardiovascular events (MACE) using the SCORE2 model alone and with the addition of proteomic biomarkers. The results are shown for the total population (**Panel A**), males (**Panel B**), and females (**Panel C**).




**Appendix 48. Calibration curves of the SCORE2 model with and without proteomics data for 10-year MACE risk prediction in the validation set (30% of UK Biobank, N=14,216)**

**Appendix 49. Comparison of the standardized circulating concentration of the 7 selected proteins by fasting status in the ESTHER study**

| Proteins (NPX) | Fasting (N=5,413) | Non-fasting (N=648) | ANOVA *p*-value |
|---|---|---|---|
| | Mean (95%CI) | Mean (95%CI) | |
| CDCP1 | 0.060 (0.038, 0.082) | 0.126 (0.059, 0.193) | 0.055 |
| CCL7 | 1.037 (0.962, 1.112) | 1.356 (1.117, 1.559) | 0.007 |
| CXCL9 | 0.062 (0.039, 0.086) | 0.048 (-0.017, 0.114) | 0.709 |
| HGF | -0.040 (-0.057, -0.023) | 0.031 (-0.017, 0.080) | 0.007 |
| IL6 | 0.457 (0.408, 0.506) | 0.649 ( 0.499, 0.799) | 0.012 |
| TNFRSF9 | -0.004 (-0.019, 0.011) | -0.047 ( -0.092, -0.003) | 0.068 |
| TNFRSF11B | -0.029 (-0.043, -0.015) | -0.048 (-0.089, -0.008) | 0.381 |

**Abbreviations:** CDCP1, CUB domain-containing protein 1; CCL7, C-C motif chemokine 7; CXCL9, C-X-C motif chemokine 9; HGF, Hepatocyte growth factor; IL6, Interleukin-6; NPX, Normalized Protein eXpresstion;TNFRSF9, Tumor necrosis factor receptor superfamily member 9; TNFRSF11B, Tumor necrosis factor receptor superfamily member 11B.

**Appendix 50. ß-coefficients of the variables of the SCORE2 model extended by 7 proteins for 10-year prediction of MACE in the derivation set of the UK Biobank**

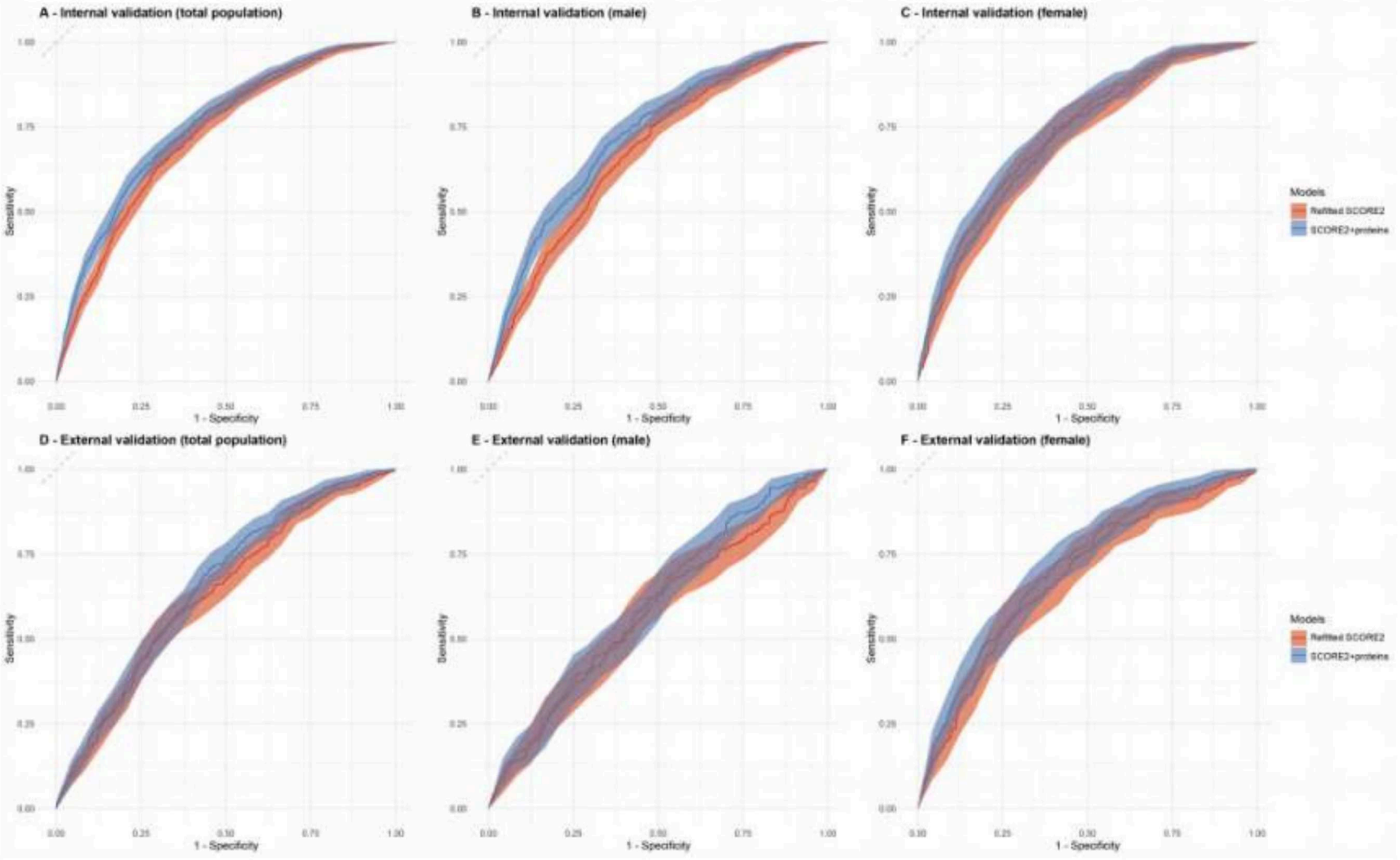
| Risk factor (units) | ß coefficients | |
|---|---|---|
| | Male | Female |
| **SCORE2 variables** | | |
| Age (per 5 years) | 0.3741 | 0.4126 |
| Current smoking | 0.6407 | 0.9844 |
| Systolic blood pressure (per 20mmHg) | 0.1117 | 0.2526 |
| Total cholesterol (per 1 mmol/L) | -0.0008 | -0.0502 |
| HDL cholesterol (per 0.5 mmol/L) | -0.0722 | -0.1139 |
| Smoking interaction with age | -0.0485 | -0.0578 |
| SBP interaction with age | -0.0820 | -0.0399 |
| Total cholesterol interaction with age | -0.0452 | -0.0286 |
| HDL interaction with age | -0.0434 | 0.0379 |
| **Additional proteins (per 1 SD)** | | |
| CDCP1 | 0.2192 | 0.1276 |
| CCL7 | 0.2023 | - |
| CXCL9 | - | 0.1639 |
| HGF | 0.3362 | - |
| IL6 | 0.1542 | - |
| TNFRSF9 | - | 0.2343 |
| TNFRSF11B | 0.2067 | 0.2531 |

**Abbreviations:** CDCP1, CUB domain-containing protein 1; CCL7, C-C motif chemokine 7; CXCL9, C-X-C motif chemokine 9; HDL, High density lipoprotein; HGF, Hepatocyte growth factor; IL6, Interleukin-6; NPX, Normalized Protein eXpresstion; SD, standard deviation; TNFRSF9, Tumor necrosis factor receptor superfamily member 9; TNFRSF11B, Tumor necrosis factor receptor superfamily member 11B.

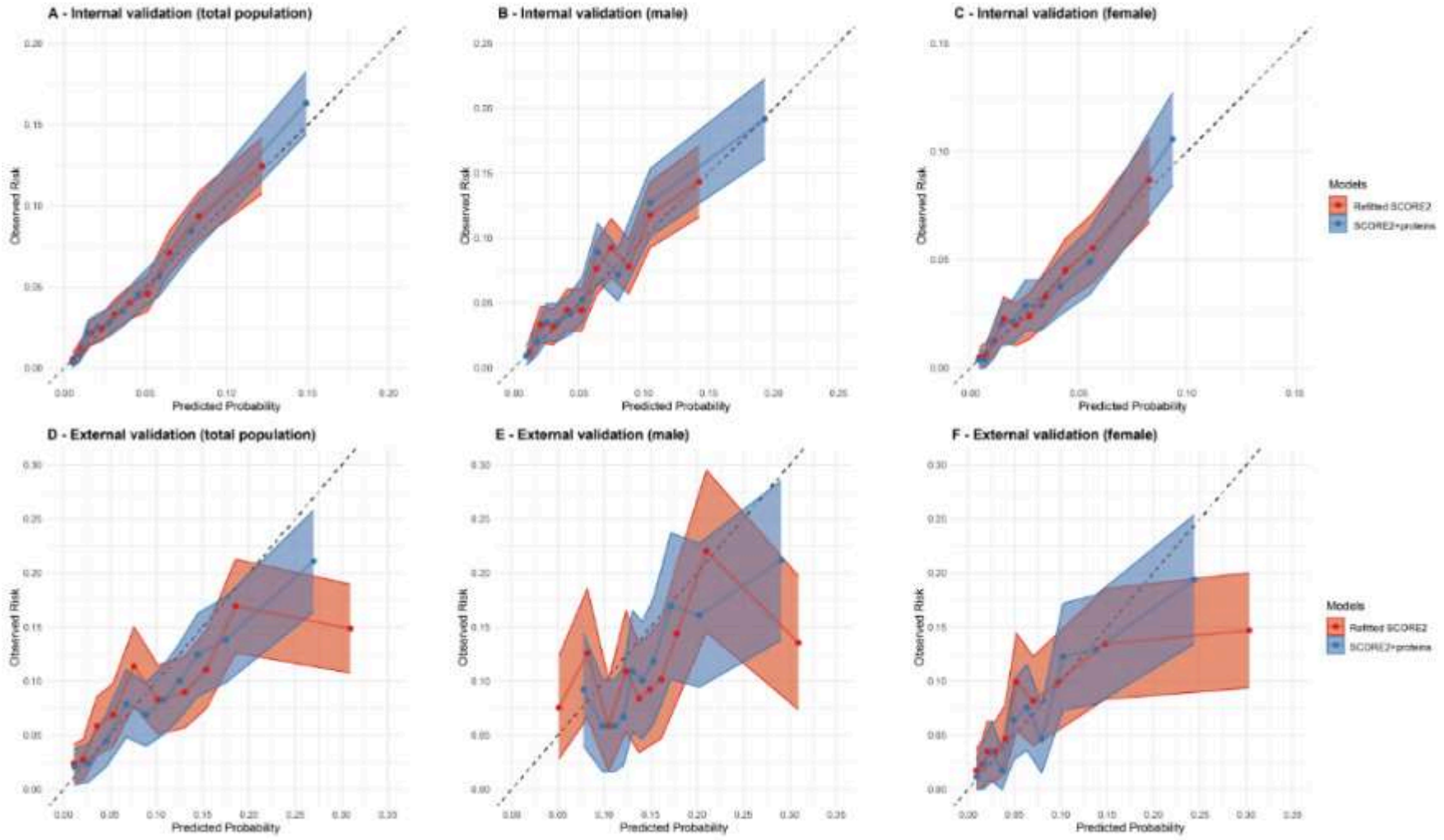
**Appendix 51. ß-coefficients of the variables of the SCORE2 model extended by 7 proteins for 10-year prediction of MACE in the calibration set of the ESTHER study**

| Risk factor (units) | ß coefficients | |
|---|---|---|
| | Male | Female |
| **SCORE2 variables** | | |
| Age (per 5 years) | 0.2364 | 0.5314 |
| Current smoking | 0.7539 | 0.9767 |
| Systolic blood pressure (per 20mmHg) | 0.1431 | 0.2696 |
| Total cholesterol (per 1 mmol/L) | 0.1087 | 0.0164 |
| HDL cholesterol (per 0.5 mmol/L) | -0.1201 | -0.1987 |
| Smoking interaction with age | -0.1960 | -0.0489 |
| SBP interaction with age | 0.0459 | -0.0255 |
| Total cholesterol interaction with age | -0.0587 | -0.0259 |
| HDL interaction with age | 0.0380 | -0.0237 |
| **Additional proteins (per 1 SD)** | | |
| CDCP1 | 0.0565 | 0.2127 |
| CCL7 | -0.0029 | - |
| CXCL9 | - | 0.0838 |
| HGF | -0.1677 | - |
| IL6 | 0.0311 | - |
| TNFRSF9 | - | 0.0202 |
| TNFRSF11B | -0.1314 | -0.3630 |

**Abbreviations:** CDCP1, CUB domain-containing protein 1; CCL7, C-C motif chemokine 7; CXCL9, C-X-C motif chemokine 9; HDL, High density lipoprotein; HGF, Hepatocyte growth factor; IL6, Interleukin-6; NPX, Normalized Protein eXpresstion; SD, standard deviation; TNFRSF9, Tumor necrosis factor receptor superfamily member 9; TNFRSF11B, Tumor necrosis factor receptor superfamily member 10B.




**Appendix 52. Receiver operating characteristic (ROC) curves of the SCORE2 model with and without proteomics data for 10-year MACE risk prediction in the internal validation (30% of UK Biobank, N=14,216) and external validation (70% of ESTHER, N=4,397)**




**Appendix 53. Calibration curves of the SCORE2 model with and without proteomics data for 10-year MACE risk prediction in the internal validation (30% of UK Biobank, N=14,216) and external validation (70% of ESTHER, N=4,397)**

**Appendix 54. Metrics of the predictive performance of the SCORE2 model for 10-year MACE risk without and with extension by proteins in random selected participants**

| Metrics | Male [a] | Female [b] | Overall |
|---|---|---|---|
| **Derivation set (70% of UK Biobank)** | | | |
| **Total sample size (N=29,585) / MACE case number (N=1,164)** | | | |
| C-Statistics (SCORE2) | 0.673 (0.655,0.691) | 0.724 (0.702,0.747) | 0.725 (0.711,0.738) |
| C-Statistics (Refitted SCORE2) | 0.676 (0.658,0.694) | 0.726 (0.704,0.749) | 0.727 (0.713,0.740) |
| C-Statistics (+Proteins) | 0.715 (0.698,0.733) | 0.742 (0.720, 0.764) | 0.750 (0.737,0.763) |
| *P*-values, C-Statistics comparisons[c] | **<0.001** | **0.001** | **<0.001** |
| **Internal validation set (30% of UK Biobank)** | | | |
| **Total sample size (N=12,683) / MACE case number (N=515)** | | | |
| C-Statistics (SCORE2) | 0.675 (0.647, 0.703) | 0.710 (0.675, 0.766) | 0.722 (0.701, 0.743) |
| C-Statistics (Refitted SCORE2) | 0.677 (0.649, 0.705) | 0.713 (0.678, 0.769) | 0.724 (0.703, 0.745) |
| C-Statistics (+Proteins) | 0.725 (0.698,0.752) | 0.730 (0.696,0.764) | 0.751 (0.730,0.771) |
| *P*-values, C-Statistics comparisons[c] | **<0.001** | **0.027** | **<0.001** |
| NRI categorical total (%)[d] | **14.8 (2.6, 22.2)** | **14.9 (6.0, 22.8)** | **8.9 (2.6, 15.6)** |
| NRI categorical events (%)[d] | -0.9 (-7.8, 7.3) | **16.9 (7.6, 24.4)** | 6.2 (-0.8, 13.3) |
| NRI categorical non-events (%)[d] | 15.7 (-1.6, 22.0) | **-2.0 (-2.9, -1.3)** | **2.7 (1.2, 4.4)** |
| IDI | **0.016 (0.006, 0.028)** | **0.006 (0.000, 0.015)** | **0.014 (0.007, 0.021)** |
| **External validation set (70% of ESTHER)** | | | |
| **Total sample size (N=4,209) / MACE case number (N=387)** | | | |
| C-Statistics (SCORE2) | 0.622 (0.578, 0.665) | 0.728 (0.690, 0.767) | 0.689 (0.662, 0.716) |
| C-Statistics (Refitted SCORE2) | 0.616 (0.570, 0.661) | 0.726 (0.687, 0.766) | 0.686 (0.658, 0.714) |
| C-Statistics (+Proteins) | 0.645 (0.599,0.691) | 0.751 (0.712,0.790) | 0.709 (0.681,0.737) |
| *P*-values, C-Statistics comparisons[c] | **0.044** | **0.009** | **<0.001** |
| NRI categorical total (%)[e] | 3.0 (-2.2, 9.1) | **28.6 (0.6, 44.4)** | **11.2 (1.5, 36.1)** |
| NRI categorical events (%)[e] | **6.3 (0.1, 11.4)** | -14.7 (-21.1, 6.6) | 4.9 (-11.6, 10.3) |

| | | | |
|---|---|---|---|
| NRI categorical non-events (%)[e] | **-3.3 (-6.4, -1.3)** | 43.3 (-1.6, 57.9) | 6.3 (-3.1, 43.9) |
| IDI | 0.003 (-0.007, 0.013) | **0.015 (0.004, 0.029)** | 0.008 (-0.002, 0.016) |

**Abbreviations:** IDI, integrated discrimination improvement; NRI, net reclassification improvement.

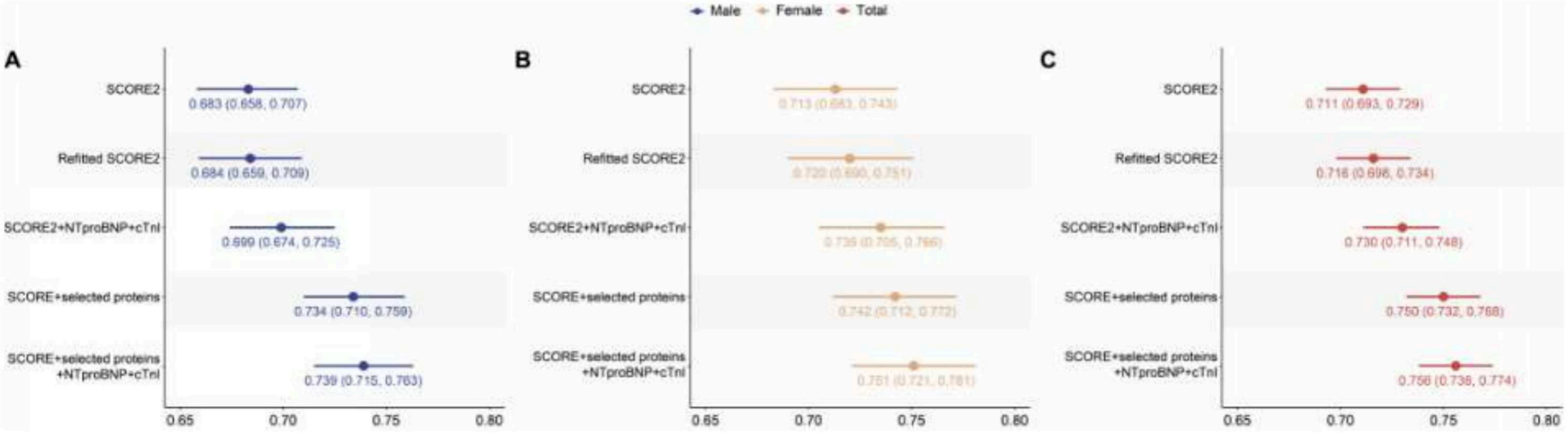
[a] Proteins that were included for male were CCL7, CDCP1, HGF, IL6, and TNFRSF11B.

[b] Proteins that were included for female were CDCP1, CXCL9, TNFRSF9, and TNFRSF11B.

[c] P-values compare the refitted SCORE2 model with the SCORE2 model extended by proteins.

[d] NRI calculated with the pre-specified 10-year MACE risk categories of 0–5%, >5–10%, and >10%.

[e] NRI calculated with the pre-specified 10-year MACE risk categories of 0–7.5%, >7.5–15%, and >15%.




**Appendix 55. Comparison of C-index for five risk prediction models based on SCORE2 and its extensions in the internal validation (30% of UK Biobank, N=14,216)**

C-statistics for five models are shown: (1) the original SCORE2 model; (2) the refitted SCORE2 model; (3) SCORE2 plus cardiac-specific biomarkers (NT-proBNP and cTnI); (4) SCORE2 plus the seven selected inflammation-related proteins; and (5) SCORE2 plus both cardiac-specific and inflammation-related biomarkers. The results are shown for males (**Panel A**), females (**Panel B**), and the total population (**Panel C**).

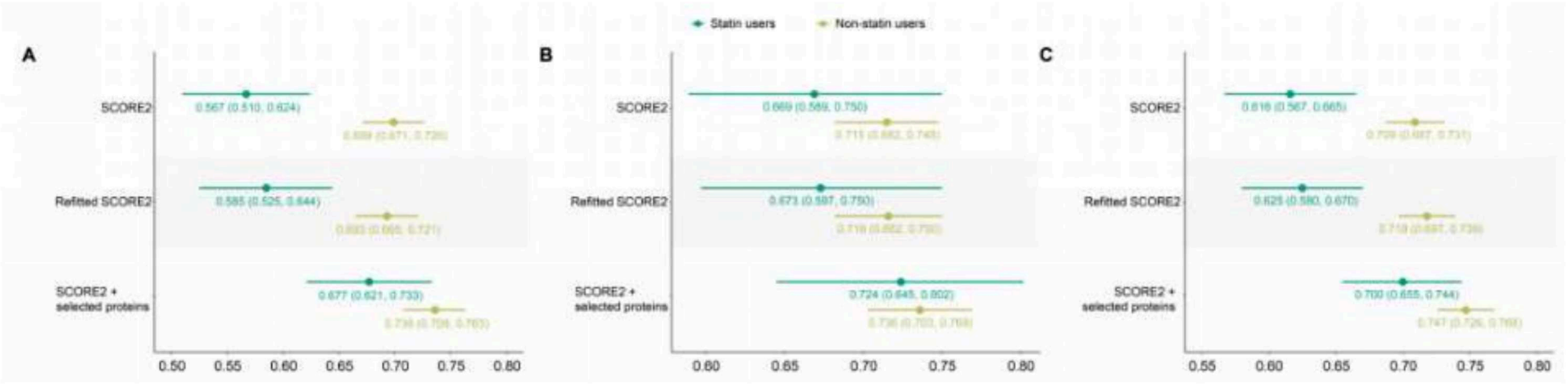
**Abbreviations:** cTnI, cardiac troponin I; NTproBNP, N-terminal pro-brain natriuretic peptide.

**Appendix 56. Comparison of adjusted standardized circulating protein concentrations by statin use in the UK Biobank**

| Proteins (NPX) | Statin users (N=5,296) | Non-statin users (N=42,086) | β (95% CI)[a], *p*-value |
|---|---|---|---|
| | Mean (95%CI) | Mean (95%CI) | |
| CDCP1 | 0.113 (0.096, 0.130) | 0.093 (0.084, 0.102) | 0.020 (0.003, 0.037), 0.020 |
| CCL7 | 0.154 (0.134, 0.173) | 0.076 (0.066, 0.086) | 0.077 (0.058, 0.097), <0.001 |
| CXCL9 | 0.081 (0.059, 0.103) | 0.033 (0.022, 0.044) | 0.048 (0.026, 0.069), <0.001 |
| HGF | 0.135 (0.123, 0.147) | 0.072 (0.066, 0.078) | 0.063 (0.051, 0.074), <0.001 |
| IL6 | 0.203 (0.178, 0.229) | 0.159 (0.147, 0.172) | 0.044 (0.019, 0.069), <0.001 |
| TNFRSF9 | 0.060 (0.047, 0.073) | 0.050 (0.043, 0.057) | 0.010 (-0.003, 0.023), 0.148 |
| TNFRSF11B | 0.166 (0.152, 0.179) | 0.109 (0.102, 0.116) | 0.010 (0.000, 0.019), 0.044 |

[a] Linear regression estimate adjusted for age, sex, systolic blood pressure, total cholesterol, HDL cholesterol, and smoking status (SCORE2 variables).

**Abbreviations:** CDCP1, CUB domain-containing protein 1; CCL7, C-C motif chemokine 7; CXCL9, C-X-C motif chemokine 9; HGF, Hepatocyte growth factor; IL6, Interleukin-6; NPX, Normalized Protein eXpresstion; TNFRSF9, Tumor necrosis factor receptor superfamily member 9; TNFRSF11B, Tumor necrosis factor receptor superfamily member 11B.




**Appendix 57. Performance of the SCORE2 model and its extension with inflammation-related proteins stratified by statin use in the internal validation (30% of UK Biobank, N=14,216)**

C-statistics (with 95% confidence intervals) for three models: original SCORE2, refitted SCORE2, and the SCORE2 model extended with selected inflammation-related proteins are shown stratified by statin use. The results are presented for males (**Panel A**), females (**Panel B**), and the total population (**Panel C**) in the UK Biobank validation set.

**Appendix 58. ß-coefficients of the variables of the SCORE2-Diabetes model extended by 15 proteins for 10-year prediction of MACE**

| Risk factor (units) | Coefficient | |
|---|---|---|
| | Male | Female |
| **SCORE2-Diabetes variables** | | |
| Age (per 5 years) | -0.0897 | 0.6155 |
| Current smoking | -0.1200 | 0.0700 |
| SBP (per 20 mmHg) | -0.0720 | 0.0787 |
| Total cholesterol (per 1 mmol/L) | 0.2928 | 0.1959 |
| HDL cholesterol (per 0.5 mmol/L) | -0.0420 | -0.2396 |
| Smoking interaction with age | -0.0378 | -0.0678 |
| SBP interaction with age | 0.1047 | -0.1130 |
| Total cholesterol interaction with age | -0.0626 | -0.1599 |
| HDL interaction with age | 0.0193 | 0.2325 |
| Diabetes age at diagnosis (per 5 years) | 0.0113 | -0.0997 |
| $HbA_{1c}$ (per 9.34 mmol/mol) | 0.1009 | -0.0455 |
| Ln eGFR (per 0.15 ml/min/1.73m$^2$) | 0.2262 | 0.1099 |
| Ln $eGFR^2$ (quadratic term) | 0.0101 | 0.0137 |
| $HbA_{1c}$ interaction with age | 0.0192 | -0.1586 |
| Ln eGFR interaction with age | 0.0197 | 0.0691 |
| **Additional proteins (per 1 SD)** | | |
| ACTA2 | 0.3303 | - |
| CHGA | 0.1079 | - |
| CRNN | -0.2856 | - |
| CTRC | - | -0.7099 |
| EFHD1 | - | 0.5865 |
| ENPP5 | -0.3111 | - |
| GAST | 0.0965 | - |
| HS3ST3B1 | - | 0.4250 |
| IL22 | - | 0.4918 |
| NT-proBNP | 0.2030 | - |
| POSTN | 0.3851 | - |
| SMOC1 | - | 0.4752 |
| TNC | - | 0.6270 |
| TRIM21 | -0.3095 | - |
| WFDC2 | 0.8349 | 0.4915 |

**Abbreviations:** ACTA2, Actin Alpha 2; CHGA, Chromogranin-A; CRNN, Cornulin; CTRC, Chymotrypsin-C; EFHD1, EF-hand domain-containing protein D1; eGFR, estimated glomerular filtration rate; ENPP5, Ectonucleotide pyrophosphatase/phosphodiesterase family member 5; GAST, Gastrin; $HbA_{1c}$, glycated hemoglobin; HDL, High-density lipoprotein; HS3ST3B1, Heparan sulfate glucosamine 3-O-sulfotransferase 3B1; IL22, Interleukin-22; NT-proBNP, N-terminal prohormone of brain natriuretic peptide; POSTN, Periostin; SBP,

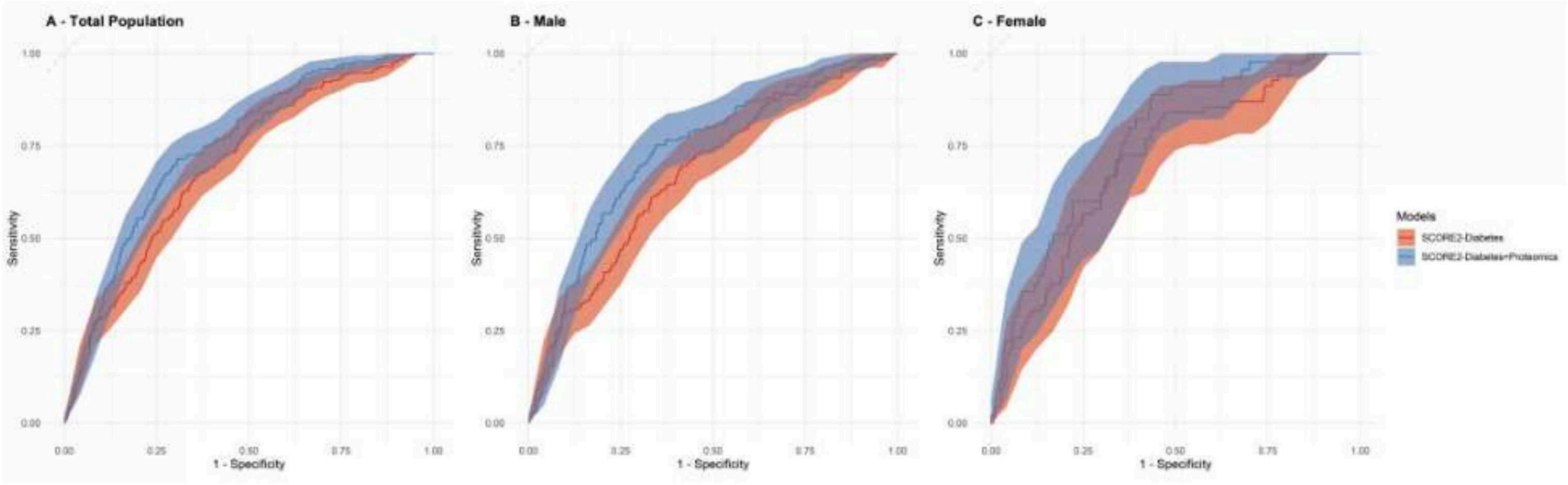
systolic blood pressure; SD, standard deviation; SMOC1, SPARC-related modular calcium-binding protein 1; TNC, Tenascin-C; TRIM21; E3 ubiquitin-protein ligase TRIM21; WFDC2, WAP four-disulfide core domain protein.

**Appendix 59. ß-coefficients of the variables of the SCORE2-Diabetes model extended by selected multi-omics biomarkers for 10-year prediction of MACE**

| Risk factor (units) | Coefficient | |
|---|---|---|
| | **Male** | **Female** |
| **SCORE2-Diabetes variables** | | |
| Age (per 5 years) | -0.6712 | 0.9063 |
| Current smoking | -0.3326 | -0.0181 |
| SBP (per 20 mmHg) | -0.2602 | 0.0900 |
| Total cholesterol (per 1 mmol/L) | 0.2097 | 0.2351 |
| HDL cholesterol (per 0.5 mmol/L) | 0.0064 | -0.5594 |
| Smoking interaction with age | -0.0378 | -0.0678 |
| SBP interaction with age | 0.2976 | -0.0918 |
| Total cholesterol interaction with age | -0.2730 | -0.0340 |
| HDL interaction with age | -0.0264 | 0.2162 |
| Diabetes age at diagnosis (per 5 years) | 0.0051 | -0.1638 |
| $HbA_{1c}$ (per 9.34 mmol/mol) | 0.0470 | 0.0147 |
| Ln eGFR (per 0.15 ml/min/1.73m$^2$) | -0.2007 | -0.1060 |
| Ln eGFR$^2$ (quadratic term) | -0.0115 | 0.0291 |
| $HbA_{1c}$ interaction with age | -0.0093 | -0.2434 |
| Ln eGFR interaction with age | 0.0085 | 0.0364 |
| **Additional PRS (per 1 SD)** | | |
| CVD-PRS | 0.0673 | 0.0403 |
| **Additional metabolites (per 1 SD)** | | |
| Acetate | -0.1210 | - |
| Albumin | 0.3158 | 0.1281 |
| Creatinine | -0.3306 | -0.3978 |
| GlycA | - | -0.0636 |
| Omega-3-pct | -0.1954 | - |
| Lactate | - | 0.1075 |
| VLDL-size | -0.0901 | - |
| **Additional proteins (per 1 SD)** | | |
| ACTA2 | 0.1979 | - |
| CHGA | 0.0220 | - |
| CRNN | -0.1997 | - |
| CTRC | - | -0.8490 |
| EFHD1 | - | 0.1114 |
| ENPP5 | -0.3899 | - |
| GAST | 0.2080 | - |
| HS3ST3B1 | - | 0.5037 |
| IL22 | - | 0.5579 |

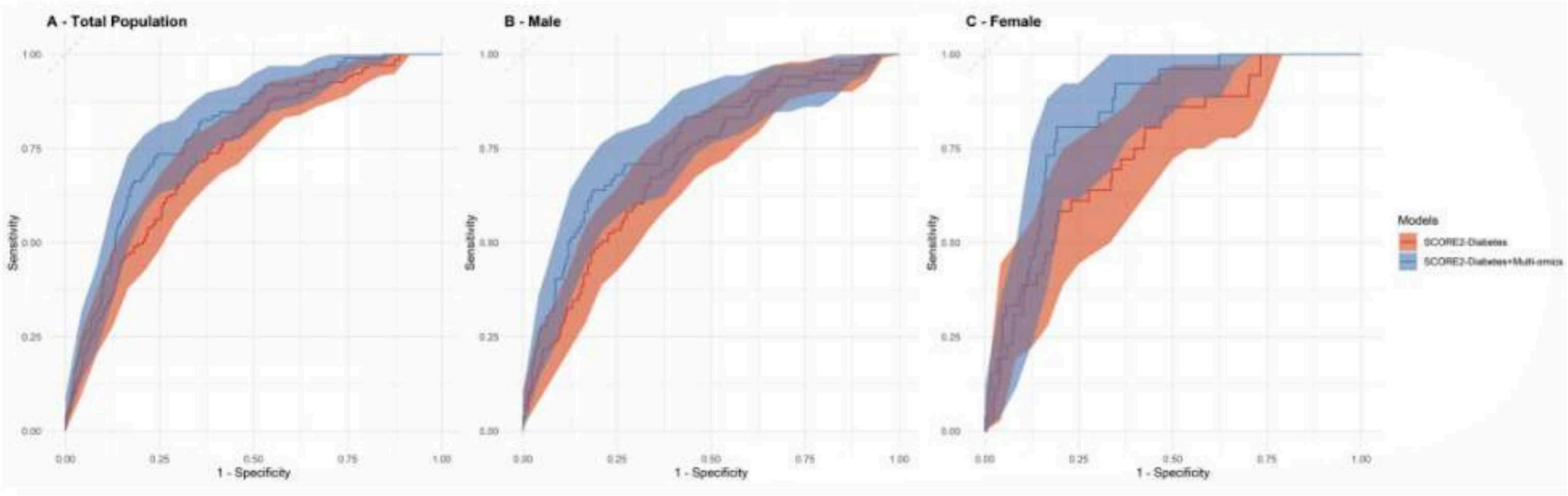
| | | |
|---|---|---|
| NT-proBNP | 0.1940 | - |
| POSTN | 0.4796 | - |
| SMOC1 | - | 0.2765 |
| TNC | - | 0.6748 |
| TRIM21 | -0.4930 | - |
| WFDC2 | 0.8517 | 0.5199 |

**Abbreviations:** ACTA2, Actin Alpha 2; CHGA, Chromogranin-A; CRNN, Cornulin; CTRC, Chymotrypsin-C; CVD, Cardiovascular Disease; EFHD1, EF-hand domain-containing protein D1; eGFR, estimated glomerular filtration rate; ENPP5, Ectonucleotide pyrophosphatase/phosphodiesterase family member 5; GAST, Gastrin; $HbA_{1c}$, glycated hemoglobin; HDL, High-density lipoprotein; HS3ST3B1, Heparan sulfate glucosamine 3-O-sulfotransferase 3B1; IL22, Interleukin-22; NT-proBNP, N-terminal prohormone of brain natriuretic peptide; Omega-3-pct, The Omega-3 fatty acids percentage of total fatty acids; PRS, polygenic risk score; SBP, Systolic blood pressure; SD, standard deviation; SMOC1, SPARC-related modular calcium-binding protein 1; TNC, Tenascin-C; TRIM21; E3 ubiquitin-protein ligase TRIM21; VLDL-size, Average diameter for very-low-density lipoprotein particles; WFDC2, WAP four-disulfide core domain protein.



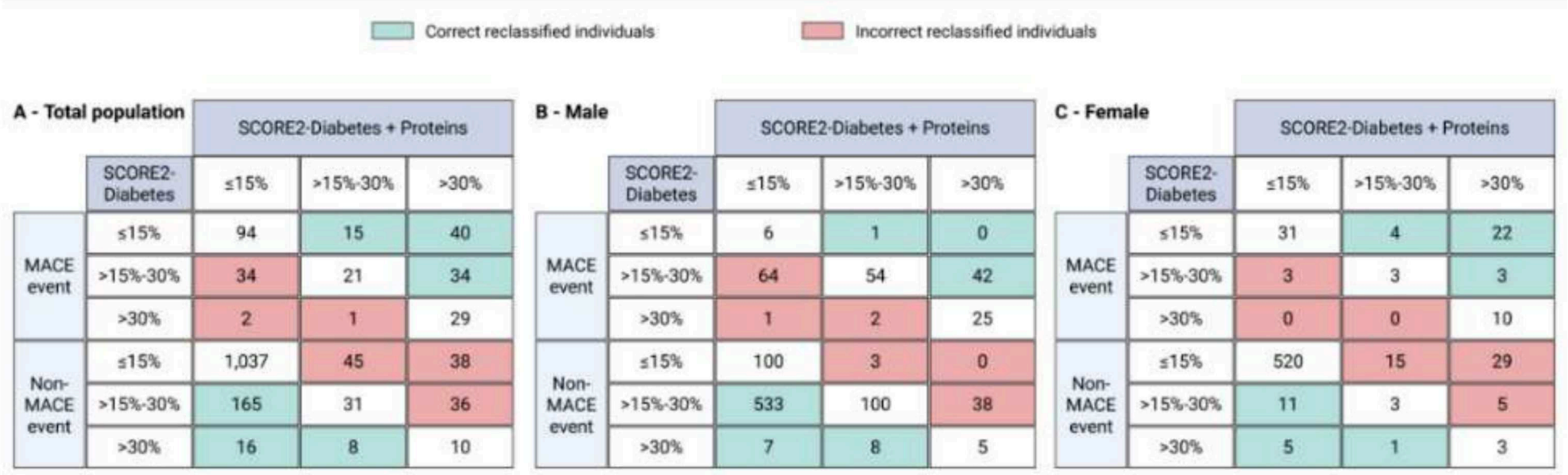

**Appendix 60. ROC curves comparing the discrimination of the SCORE2-Diabetes model with and without proteomics extendsion for 10-year MACE risk (N=1,751)**

Panels show receiver operating characteristic (ROC) curves for the SCORE2-Diabetes model (red) and the model extended with sex-specific proteomic signatures (blue) in the total population (**Panel A**), male participants (**Panel B**), and female participants (**Panel C**). Shaded areas represent 95% confidence intervals.




**Appendix 61. ROC curves comparing the discrimination of the SCORE2-Diabetes model with and without multi-omics extendsion for 10-year MACE risk (N=990)**

Panels display receiver operating characteristic (ROC) curves for the SCORE2-Diabetes model (red) and the model extended with multi-omics biomarkers (blue) in the total population (**Panel A**), male participants (**Panel B**), and female participants (**Panel C**). The multi-omics model refers to the integration of proteomics, metabolomics, and the CVD-PRS.




**A - Total population**

| | SCORE2-Diabetes | SCORE2-Diabetes + Proteins ≤15% | >15%-30% | >30% |
|---|---|---|---|---|
| MACE event | ≤15% | 94 | 15 | 40 |
| | >15%-30% | 34 | 21 | 34 |
| | >30% | 2 | 1 | 29 |
| Non-MACE event | ≤15% | 1,037 | 45 | 38 |
| | >15%-30% | 165 | 31 | 36 |
| | >30% | 16 | 8 | 10 |

**B - Male**

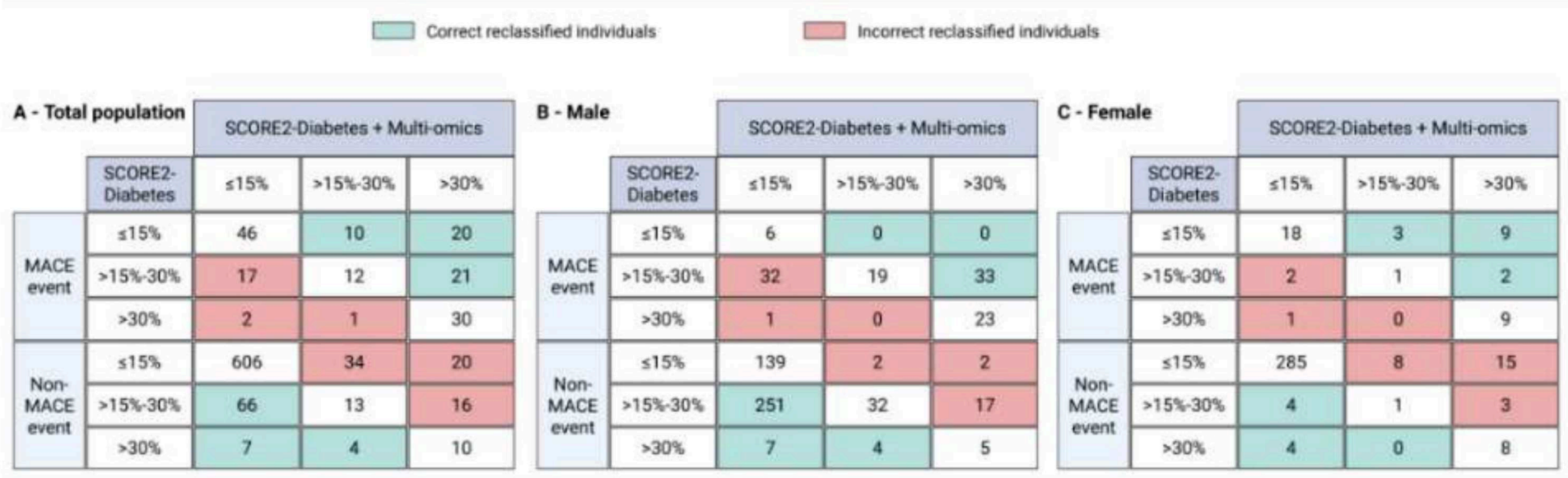
| | SCORE2-Diabetes | SCORE2-Diabetes + Proteins ≤15% | >15%-30% | >30% |
|---|---|---|---|---|
| MACE event | ≤15% | 6 | 1 | 0 |
| | >15%-30% | 64 | 54 | 42 |
| | >30% | 1 | 2 | 25 |
| Non-MACE event | ≤15% | 100 | 3 | 0 |
| | >15%-30% | 533 | 100 | 38 |
| | >30% | 7 | 8 | 5 |

**C - Female**

| | SCORE2-Diabetes | SCORE2-Diabetes + Proteins ≤15% | >15%-30% | >30% |
|---|---|---|---|---|
| MACE event | ≤15% | 31 | 4 | 22 |
| | >15%-30% | 3 | 3 | 3 |
| | >30% | 0 | 0 | 10 |
| Non-MACE event | ≤15% | 520 | 15 | 29 |
| | >15%-30% | 11 | 3 | 5 |
| | >30% | 5 | 1 | 3 |

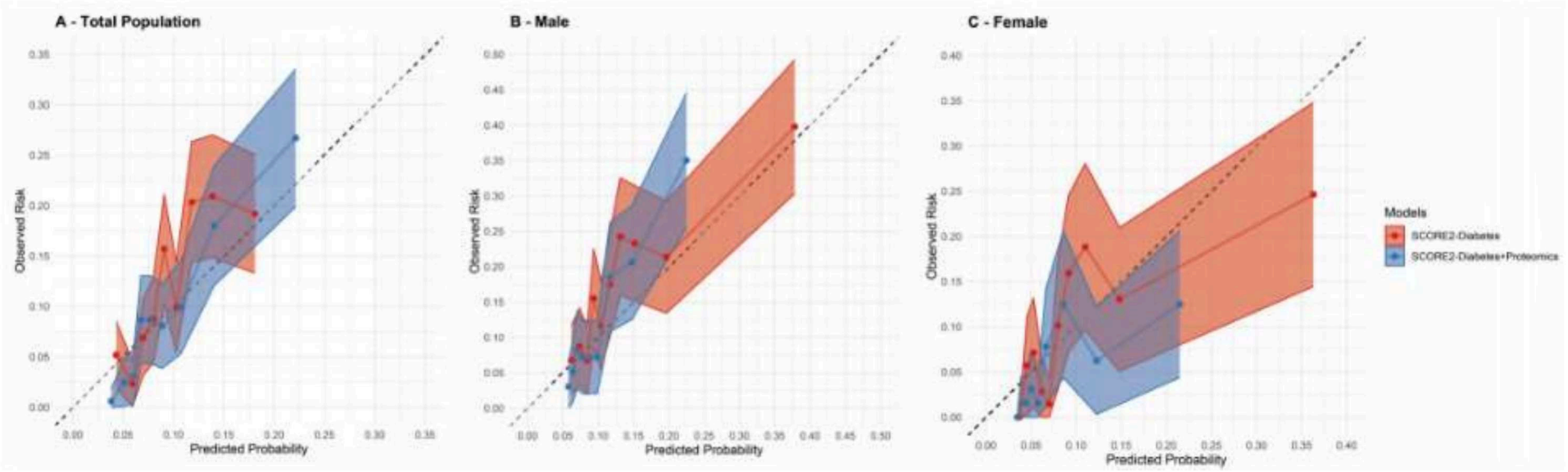
**Appendix 62. Reclassification of cardiovascular risk by adding proteomics to the SCORE2-Diabetes model (N=1,751)**

The reclassification of individuals into cardiovascular risk categories (≤15%, >15%-30%, >30%) for major adverse cardiovascular events (MACE) using the SCORE2-Diabetes model alone and with the addition of proteomic biomarkers. The results are shown for the total population (**Panel A**), males (**Panel B**) and females (**Panel C**).



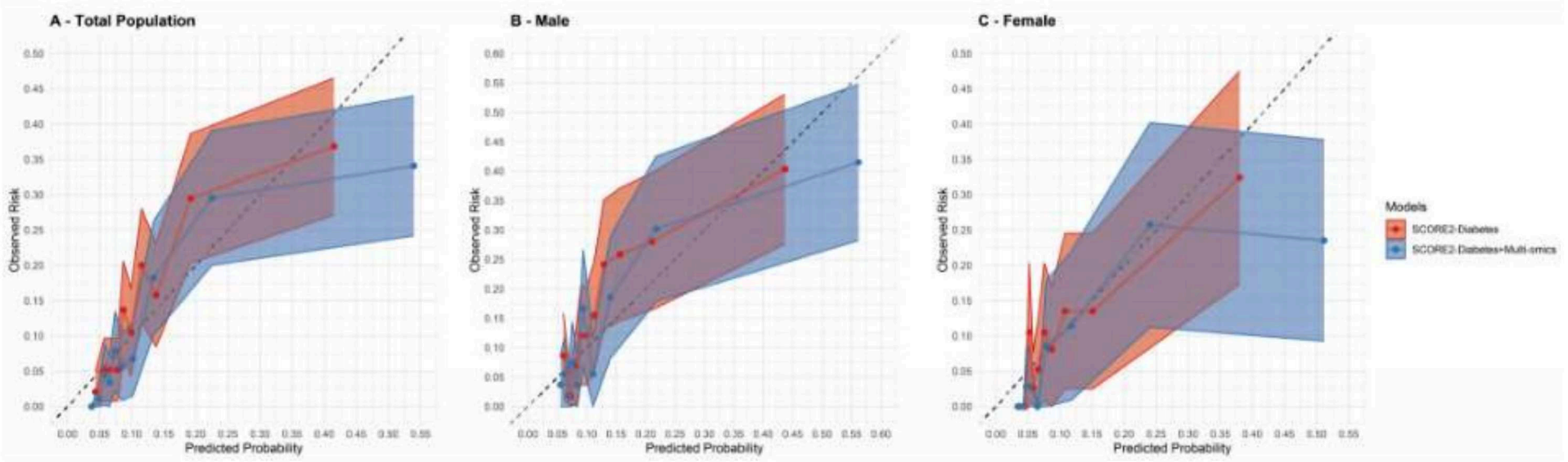

**Appendix 63. Reclassification of cardiovascular risk by adding multi-omics to the SCORE2-Diabetes model (N=990)**

The reclassification of individuals into cardiovascular risk categories (≤15%, >15%-30%, >30%) for major adverse cardiovascular events (MACE) using the SCORE2-Diabetes model alone and with the addition of multi-omic biomarkers. The results are shown for the total population (**Panel A**), males (**Panel B**) and females (**Panel C**). The multi-omics model refers to the integration of proteomics, metabolomics, and the CVD-PRS.




**Appendix 64. Calibration curves of the SCORE2-Diabetes model with and without proteomics extension for 10-year MACE risk prediction (N=1,751)**

Calibration of the SCORE2-Diabetes model (red) and the extended model incorporating proteomic biomarkers (blue) is shown in the total population (Panel A), male participants (Panel B), and female participants (Panel C).




**Appendix 65. Calibration curves of the SCORE2-Diabetes model with and without multi-omics extension for 10-year MACE risk prediction (N=990)**

Calibration of the SCORE2-Diabetes model (red) and the extended model incorporating multi-omic biomarkers (blue) is shown in the total population (Panel A), male participants (Panel B), and female participants (Panel C). The multi-omics model refers to the integration of proteomics, metabolomics, and the CVD-PRS.

**Appendix 66. ß-coefficients of the variables of the SCORE2 model extended by multi-omics for 10-year prediction of MACE**

| Risk factor (units) | Coefficient | |
|---|---|---|
| | **Male** | **Female** |
| **SCORE2 variables** | | |
| Age (per 5 years) | 0.1248 | 0.3475 |
| Current smoking | -0.0638 | 0.5276 |
| Systolic blood pressure (per 20mmHg) | 0.0710 | 0.2138 |
| Total cholesterol (per 1 mmol/L) | 0.0728 | 0.0370 |
| HDL cholesterol (per 0.5 mmol/L) | -0.0004 | -0.1207 |
| Smoking interaction with age | -0.0223 | -0.1793 |
| SBP interaction with age | -0.1158 | -0.0273 |
| Total cholesterol interaction with age | -0.0116 | -0.0449 |
| HDL interaction with age | 0.0099 | -0.0325 |
| **Additional PRS (per 1 SD)** | | |
| CVD-PRS | 0.2114 | 0.1476 |
| **Additional metabolites (per 1 SD)** | | |
| Acetate | -0.0247 | - |
| Acetoacetate | - | 0.0017 |
| Albumin | -0.0093 | -0.0656 |
| GlycA | 0.0304 | 0.0180 |
| IDL-CE-pct | -0.0130 | -0.0916 |
| LA-pct | - | 0.0041 |
| L-LDL-TG-pct | - | 0.0061 |
| L-VLDL-TG-pct | -0.0522 | - |
| Omega-3-pct | -0.0466 | - |
| S-HDL-CE | -0.0436 | - |
| Val | - | -0.0276 |
| VLDL-size | -0.0387 | - |
| XL-HDL-FC | 0.0640 | - |
| **Additional proteins (per 1 SD)** | | |
| ADAMTS13 | -0.1388 | - |
| ADGRG2 | - | -0.1080 |
| BCAN | -0.1774 | -0.0816 |
| CCL7 | 0.1626 | - |
| CRYBB2 | - | 0.0921 |
| CXCL17 | 0.0742 | - |
| EDA2R | - | 0.0325 |
| EGFR | - | -0.0822 |
| GDF15 | 0.1704 | 0.1990 |

| Risk factor (units) | Coefficient | |
|---|---|---|
| | **Male** | **Female** |
| HAVCR1 | - | 0.0648 |
| IL6 | -0.0039 | - |
| LRRN1 | -0.0892 | - |
| MMP12 | 0.2227 | 0.0713 |
| NEFL | 0.1276 | - |
| NTproBNP | 0.2052 | 0.1898 |
| PRSS8 | 0.1012 | - |
| SEPTIN8 | - | 0.1523 |
| WFDC2 | 0.0435 | 0.2055 |

**Abbreviations:** ADAMTS13, A disintegrin and metalloproteinase with thrombospondin motifs 13; BCAN, Brevican core protein; CCL7, C-C motif chemokine 7; CRYBB2, Beta-crystallin B2; CXCL17, C-X-C motif chemokine 17; EDA2R, Tumor necrosis factor receptor superfamily member 27; EGFR, Epidermal growth factor receptor; GDF15, Growth/differentiation factor 15; GlycA, Glycoprotein acetyls; IDL-CE-pct, The cholesteryl esters percentage of total lipids in intermediate-density lipoprotein; IL6, Interleukin-6; LA-pct, The linoleic acid percentage of total fatty acids; L-LDL-TG-pct, Triglycerides to total lipids in large low-density lipoprotein percentage; L-VLDL-TG-pct, The triglycerides percentage of total lipids in large very-low-density lipoprotein; LRRN1, Leucine-rich repeat neuronal protein 1; MMP12, Macrophage metalloelastase; NEFL, Neurofilament light polypeptide; NTproBNP, N-terminal prohormone of brain natriuretic peptide; Omega-3-pct, The Omega-3 fatty acids percentage of total fatty acids; PRS, polygenic risk score; PRSS8, Prostasin; S-HDL-CE, Cholesteryl esters in small high-density lipoprotein; SD, standard deviation; SEPTIN8, Septin-8; Val, Valine; VLDL-size, Average diameter for very-low-density lipoprotein particles; WFDC2, WAP four-disulfide core domain protein; XL-HDL-FC, Free cholesterol in very large high-density lipoprotein.

**Appendix 67. ß-coefficients of the variables of the optimal multi-omics combination models for 10-year prediction of MACE**

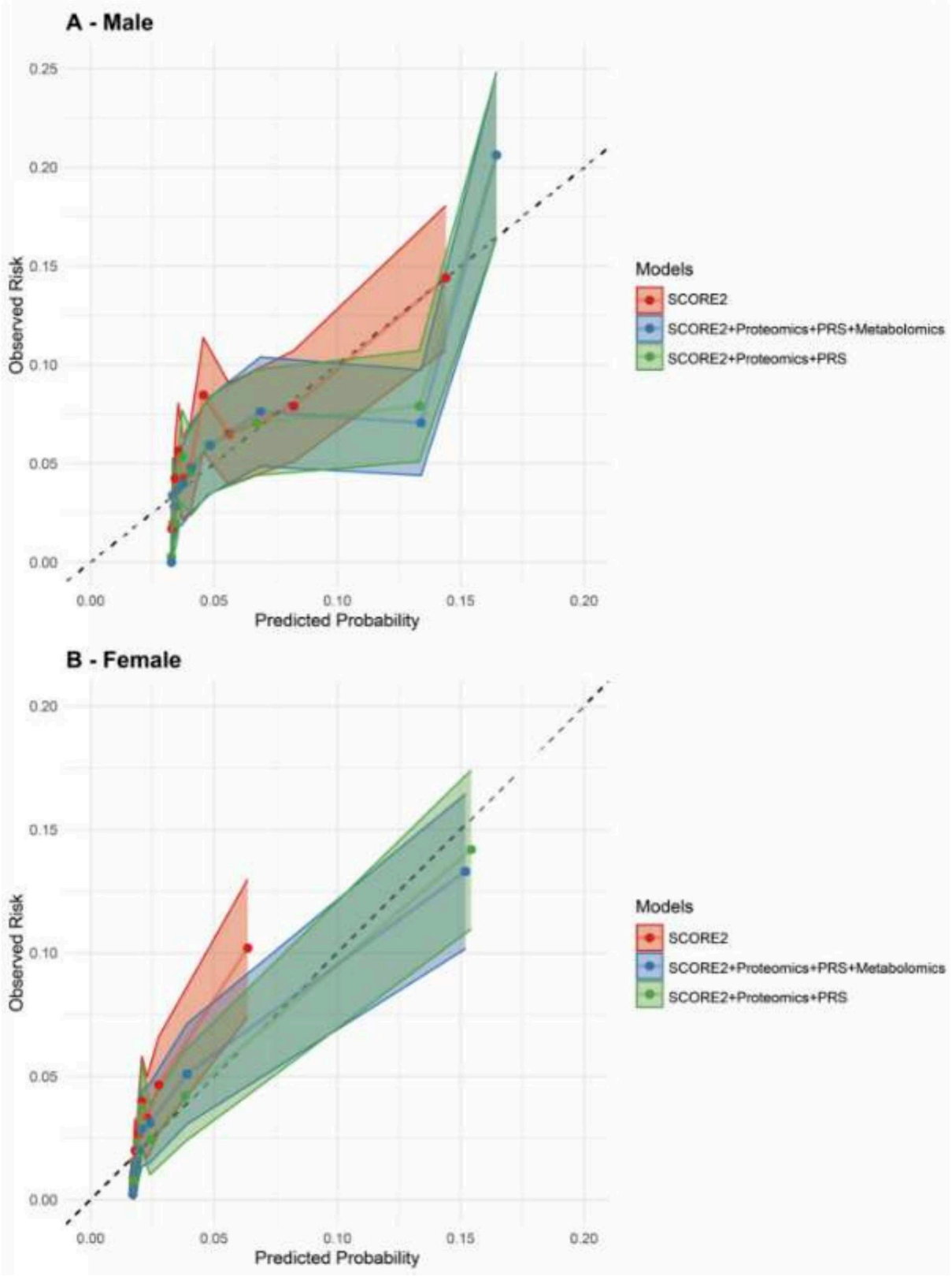
| Risk factor (units) | ß coefficients | |
|---|---|---|
| | **Male** | **Female** |
| **SCORE2 variables** | | |
| Age (per 5 years) | 0.1115 | 0.3393 |
| Current smoking | -0.0499 | 0.5188 |
| Systolic blood pressure (per 20mmHg) | 0.0677 | 0.2107 |
| Total cholesterol (per 1 mmol/L) | 0.0773 | 0.0119 |
| HDL cholesterol (per 0.5 mmol/L) | 0.0008 | -0.1737 |
| Smoking interaction with age | -0.0212 | -0.1711 |
| SBP interaction with age | -0.0697 | -0.0255 |
| Total cholesterol interaction with age | -0.0133 | -0.0447 |
| HDL interaction with age | 0.0186 | -0.0330 |
| **Additional PRS (per 1 SD)** | | |
| CVD-PRS | 0.2107 | 0.1478 |
| **Additional proteins (per 1 SD)** | | |
| ADAMTS13 | -0.1468 | - |
| ADGRG2 | - | -0.1169 |
| BCAN | -0.1749 | -0.0881 |
| CCL7 | 0.1733 | - |
| CRYBB2 | - | 0.0901 |
| CXCL17 | 0.0745 | - |
| EDA2R | - | 0.0345 |
| EGFR | - | -0.0941 |
| GDF15 | 0.1903 | 0.2186 |
| HAVCR1 | - | 0.0808 |
| IL6 | 0.0090 | - |
| LRRN1 | -0.0945 | - |
| MMP12 | 0.2263 | 0.0720 |
| NEFL | 0.1278 | - |
| NTproBNP | 0.2139 | 0.1931 |
| PRSS8 | 0.0639 | - |
| SEPTIN8 | - | 0.1662 |
| WFDC2 | 0.0671 | 0.2129 |

**Abbreviations:** ADAMTS13, A disintegrin and metalloproteinase with thrombospondin motifs 13; BCAN, Brevican core protein; CCL7, C-C motif chemokine 7; CRYBB2, Beta-crystallin B2; CXCL17, C-X-C motif chemokine 17; EDA2R, Tumor necrosis factor receptor superfamily member 27; EGFR, Epidermal growth factor receptor; GDF15, Growth/differentiation factor 15; IL6, Interleukin-6; LRRN1, Leucine-rich repeat neuronal protein 1; MMP12, Macrophage metalloelastase; NEFL, Neurofilament light polypeptide; NTproBNP, N-terminal prohormone of brain natriuretic peptide; PRS, polygenic risk score; PRSS8, Prostasin; SD, standard deviation; SEPTIN8, Septin-8; WFDC2, WAP four-disulfide core domain protein.




**Appendix 68. Reclassification table of the SCORE2 model with Proteomics + PRS + Metabolomics or optimal multi-omics combination in the validation set (30% of UK Biobank, N=7,214)**

The reclassification of individuals into cardiovascular risk categories (≤7.5%, >7.5%-15%, >15%) for major adverse cardiovascular events (MACE) using the SCORE2 model alone and with the addition of proteomic biomarkers. The results are shown for the males (**Panel A**) and females (**Panel B**).



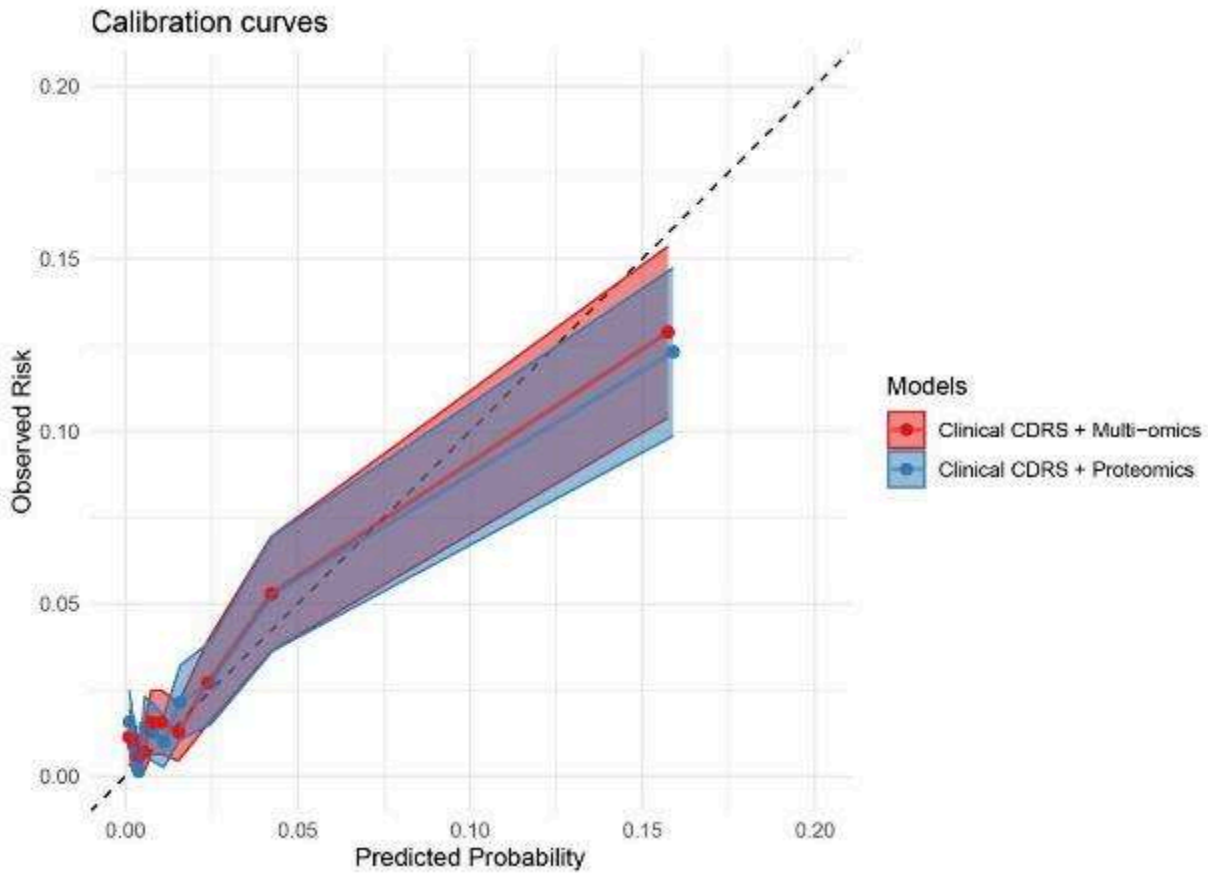

**Appendix 69. Calibration curves of the SCORE2 model with and without Proteomics + PRS + Metabolomics or optimal multi-omics combination extended for 10-year MACE risk prediction in the validation set (30% of UK Biobank, N=7,214)**

**Appendix 70. ß-coefficients of the different extensions of the clinical CDRS model with multi-omics data for 10-year prediction of T2D**

| **Risk factor (units)** | **ß-coefficient** | |
|---|---|---|
| **Clinical CDRS variables** | **Clinical CDRS + Proteomics** | **Clinical CDRS + Multi-omics** |
| Female | -0.3040 | -0.2743 |
| Prescribed anti-hypertension medication | 0.0334 | 0.0168 |
| Prescribed steroids | 0.4104 | 0.4647 |
| Age (per years) | 0.0077 | 0.0077 |
| Body mass index <25 kg/m$^2$ | Ref | Ref |
| Body mass index 25-27.49 kg/m$^2$ | -0.0833 | -0.1195 |
| Body mass index 27.5-29.99 kg/m$^2$ | 0.2421 | 0.1798 |
| Body mass index ≥ 30 kg/m$^2$ | 0.3493 | 0.3178 |
| No first degree relative with diabetes | Ref | Ref |
| Parent or sibling with diabetes | 0.2846 | 0.2655 |
| Parent and sibling with diabetes | 0.5445 | 0.4719 |
| None-smoker | Ref | Ref |
| Ex-smoker | 0.1552 | 0.1504 |
| Current smoker | -0.0466 | -0.0350 |
| $HbA_{1c}$ (per %) | 2.6941 | 2.4053 |
| **Added proteins (per 1 SD)** | | |
| ADAMTSL2 | 0.0262 | 0.0360 |
| ADGRG1 | 0.0316 | 0.0297 |
| ADIPOQ | 0.0642 | 0.0905 |
| APOM | -0.0930 | -0.0876 |
| CDH2 | 0.1098 | 0.0817 |
| HGF | 0.1062 | 0.1669 |
| IGFBP1 | -0.1341 | -0.1308 |
| IGSF3 | -0.0043 | -0.0206 |
| IGSF9 | 0.1384 | 0.0878 |
| IL18R1 | 0.1452 | 0.1194 |
| IL1RN | 0.0389 | 0.0532 |
| LGALS4 | 0.1180 | 0.0993 |
| PON3 | -0.1297 | -0.1084 |
| RTN4R | 0.1406 | 0.1595 |
| TNFSF12 | -0.0615 | -0.0672 |
| **Additional metabolites (per 1 SD)** | | |
| 3-hydroxybutyrate | - | -0.0144 |
| Acetate | - | 0.0186 |
| Citrate | - | 0.0300 |

| | | |
|---|---|---|
| Gln | - | 0.0479 |
| Glucose | - | 0.2071 |
| IDL-CE-pct | - | -0.0088 |
| LA-pct | - | -0.0751 |
| Lactate | - | 0.0602 |
| M-LDL-TG-pct | - | 0.0549 |
| Pyruvate | - | -0.1157 |
| Tyr | - | -0.0469 |
| **Additional PRS (per 1 SD)** | | |
| T2D-PRS | - | 0.2988 |

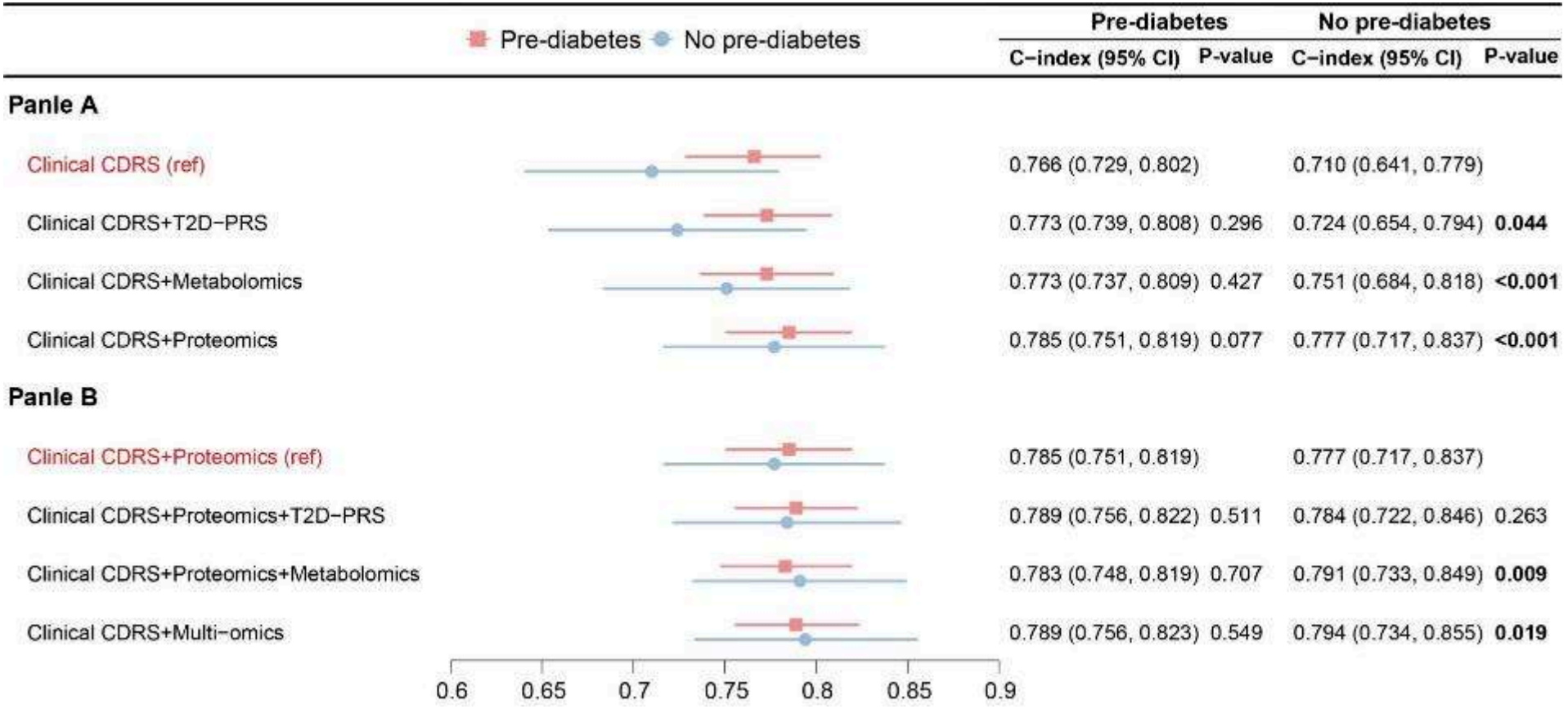
**Abbreviations:** ADAMTSL2, A disintegrin and metalloproteinase with thrombospondin repeats–like 2; ADGRG1, Adhesion G-protein coupled receptor G1; ADIPOQ, Adiponectin; APOM, Apolipoprotein M; CDH2, Cadherin-2; CDRS, Cambridge Diabetes Risk Score; HGF, Hepatocyte growth factor; IDL-CE-pct, cholesteryl esters to total lipids in IDL percentage; IGFBP1, Insulin-like growth factor-binding protein 1; IGSF, Immunoglobulin superfamily member; IL1RN, Interleukin-1 receptor antagonist; IL18R1, Interleukin-18 receptor 1; LA-pct, linoleic acid to total fatty acids percentage; M-LDL-TG-pct, triglycerides to total lipids in medium LDL percentage; PON3, paraoxonase/lactonase 3; PRS, polygenic risk score; Ref, reference; RTN4R, Reticulon-4 receptor; SD, standard deviation; T2D, type 2 diabetes; TNFSF12, TNF-related weak inducer of apoptosis.




**Appendix 71. Calibration curves of the different extensions of the clinical CDRS model with multi-omics data for 10-year T2D risk prediction in the test set (30% of UK Biobank, N=6,998)**

The multi-omics model refers to the integration of proteomics, metabolomics, and the T2D-PRS.

**Abbreviations:** CDRS, Cambridge Diabetes Risk Score.




**Appendix 72. Subgroup analysis of predictive performance by prediabetes status (30% of UK Biobank, N=6,998)**

The number of individuals and cases was 1,009 (135 cases, 874 non-cases) in the group with prediabetes and 5,989 (66 cases, 5,923 non-cases) in the group without prediabetes.

Panels A and B show the C-index and 95% confidence intervals for the clinical CDRS model and its multi-omics extensions in two subgroups: individuals without prediabetes ($HbA_{1c}$ < 5.7% [< 39 mmol/mol]; Panel A) and with prediabetes ($HbA_{1c}$ ≥ 5.7% [≥ 39 mmol/mol]; Panel B), based on ADA criteria. In Panel A, the clinical CDRS was used as the reference model; in Panel B, Clinical CDRS + Proteomics served as the reference.

All P values represent comparisons of the C-index with the respective reference model. Values in bold indicate $P < 0.05$.

The multi-omics model refers to the integration of proteomics, metabolomics, and the T2D-PRS.

**Abbreviations:** CDRS, Cambridge Diabetes Risk Score; T2D-PRS, polygenic risk score for type 2 diabetes.

# Curriculum Vitae

| **Personal Details** | |
|---|---|
| Name | Ruijie Xie |
| Date of birth | 13.01.1995 |
| Place of birth | Hunan, China |
| Nationality | Chinese |
| **Education** | |
| 04.2023 – | PhD candidate in Medicine<br>Ruprecht Karl University of Heidelberg, Heidelberg |
| 08.2019 – 06.2022 | Master of Medicine in Clinical Medicine<br>University of South China |
| 09.2013 – 06.2018 | Bachelor of Medicine in Clinical Medicine<br>Xiangnan University |
| **Research/Working Experience** | |
| 04.2023 – | PhD candidate<br>German Cancer Research Center (DKFZ), Heidelberg, Germany |
| 08.2019 – 06.2022 | Resident Physician<br>The Affiliated Nanhua Hospital, University of South China, Hengyang, China |
| 05.2017 – 04.2018 | Internship<br>The Affiliated Nanhua Hospital, University of South China, Hengyang, China |

# Acknowledgements

I wish to express my sincere and profound gratitude to all who have played a pivotal role in my academic journey and the successful completion of this dissertation.

First and foremost, my deepest appreciation goes to my supervisor, Prof. Dr. Ben Schöttker. His outstanding guidance, unwavering dedication, and enduring patience were instrumental in helping me overcome the numerous challenges encountered throughout my doctoral studies. His rigorous approach to research, combined with his patient encouragement and tailored mentorship, has profoundly influenced and shaped me as a researcher. The achievements I have made to date would not have been attainable without the valuable opportunities and dedicated guidance he provided.

I am also deeply grateful to Prof. Dr. Hermann Brenner for generously providing the ESTHER study data, without which this work would not have been possible. I also thank him for his valuable critique of my manuscripts. Furthermore, I extend my sincere appreciation to my Thesis Advisory Committee members, Prof. Christian Herder and Dr. Silvia Calderazzo, for their invaluable insights, guidance, and encouragement, which were crucial to my research.

My thanks are also extended to the departmental staff, Ms. Iris Lingen and Dr. Utz Benscheid, for their kind care and assistance. I would also like to thank my colleagues and friends, including Dr. Sha Sha, Dr. Li-Ju Chen, Qiming Yin, Xianzhe Li, Youqing Wang, Dr. Joshua Stevenson-Hoare, Dr. Kira Trares, and Tomislav Vlaski, for their collaboration and friendship.

Last but by no means least, I wish to express my heartfelt gratitude to my family. To my parents and my wife, I offer my deepest thanks for your unwavering support, understanding, and patience throughout my PhD journey and my life over the past three years.

## Eidesstattliche Versicherung

1. Bei der eingereichten Dissertation zu dem Thema

   *„ Improvement of prediction models for cardiovascular disease and type 2 diabetes mellitus with multi-omics"*

   handelt es sich um meine eigenständig erbrachte Leistung.

2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht.

3. Die Arbeit oder Teile davon habe ich bislang nicht an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt.

4. Die Richtigkeit der vorstehenden Erklärungen bestätige ich.

5. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt. Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

Ort und Datum Unterschrift