



Metabolomics and Coronary Heart Disease Risk in UK Biobank

Clinical Trial Service Unit & Epidemiological Studies Unit (CTSU)
Nuffield Department of Population Health

Danyao Jin, BMed MSc

Green Templeton College, Oxford

Thesis submitted for the degree of Doctor of Philosophy
Michaelmas Term 2023

Abstract

Background: Metabolomics is the comprehensive study of numerous metabolites from multiple molecular pathways using high-throughput methods such as nuclear magnetic resonance (NMR). Although the importance of some metabolites to coronary heart disease (CHD) risk is well established, such as low-density lipoprotein cholesterol (LDL-C), the relevance of many hundreds of other metabolites to CHD risk remains unclear.

Objective: To assess the associations of NMR-derived plasma metabolites with incident CHD, and to evaluate the additional benefit of such metabolites to the prediction of 10-year cardiovascular risk.

Methods: UK Biobank is a prospective study of 0.5 million adults aged 40-69 recruited in 2006-2010; in 2012-2013, a subset of the participants were resurveyed. NMR metabolites are currently available in a random subset of 118k participants. After excluding participants with prior cardiovascular disease and those taking statins, Cox regression models (adjusted for major potential confounders) were used to assess the association of each metabolite with CHD risk. Hazard ratios were corrected for regression dilution using the correlation of baseline and resurvey metabolic measures. The additional predictive value of metabolic profiling to an established risk score (QRISK) was evaluated with both Cox regression and machine-learning approaches.

Results: After exclusions, 86,499 participants remained. There were 4,196 (4.9%) incident cases of CHD during a mean follow-up time of 11.5 years. CHD risk was positively associated with the concentrations of very-low-density (VLDL), intermediate-density (IDL), and low-density lipoproteins (LDL), and inversely associated with high-density lipoproteins (HDL). Hazard ratios (HRs) per standard deviation (SD) were 1.20 (95% CI 1.16-1.23), 1.15 (1.12-1.19), 1.17 (1.14-1.21) and 0.90 (0.87-0.91), respectively. There was little evidence that particle size of each of these lipoproteins was related to CHD risk. Given lipoprotein particle concentrations, lipid composition (including cholesterol) was not strongly related to CHD risk, except for triglyceride in LDL particles (HR 1.17 [1.11-1.23]). Independently of other fatty acids, CHD risk was positively associated with saturated fatty acids and monounsaturated fatty acids, inversely associated with omega-3 polyunsaturated fatty acids (PUFA), but

there was no strong evidence of an association with omega-6 PUFA: HRs per SD higher fatty acid were 1.12 (1.07-1.17), 1.13 (1.08-1.18), 0.93 (0.90-0.97), and 1.02 (0.98-1.07), respectively. Independently of cholesterol and triglycerides, several non-lipid metabolites, including omega-3 fatty acids, glycine, histidine, albumin and glycoprotein acetyls, remained significantly associated with CHD risk. However, compared with QRISK, adding selected metabolites did not significantly improve measures of discrimination, reclassification or calibration.

Conclusion: Of the lipoprotein characterization assessed, CHD risk was most strongly related to particle concentrations, suggesting that the major mechanism by which plasma lipids are related to CHD risk is through the concentration of lipoprotein particles that transport them. Independent of lipids, omega-3 PUFA and several low-molecular-weight metabolites remained significantly associated with CHD risk. However, there was no evidence that metabolic profiling improved cardiovascular risk prediction in UK primary care. Future research should further explore causality and mechanism pathways, and should include the assessment of the associations and predictive value in more diverse populations and wider-range of metabolites.

Acknowledgements

First of all, I would like to express my deepest gratitude and appreciation to my supervisors Prof Sarah Lewington, Dr Ben Lacey, Dr Eirini Trichia, and Dr Nazrul Islam, for their unwavering support, guidance, and invaluable mentorship throughout the incredible journey of completing my PhD thesis. Their expertise, insightful feedback, and dedication have been instrumental in shaping the direction of my research and fostering my academic growth.

I am also profoundly thankful to the members of my thesis committee, Dr Fiona Bragg, Prof Ruth Travis, and Prof Derrick Bennett. Their constructive critiques, thoughtful suggestions, and expertise that have significantly enriched the content and scope of this thesis. Meanwhile, I would like to express my deep appreciation to my college advisor, Prof Apostolos Tsiachristas, for his guidance and encouragement in the past three years.

My sincere appreciation extends to Prof Robert Clark and Prof David Preiss for their valuable insight when I was doing the fatty acids part and lipoproteins part of the thesis. I am also grateful to Jennifer Carter, Paul Sherliker, and Diego Aguilar Ramirez for their statistical support during the thesis.

I want to thank my family who have stood by me with understanding and love, and express my apologies for not being able to go home in the past three years. I also want to thank my friends for their unwavering encouragement and companionship. Your constant belief in me has been a wellspring of motivation to support me get through the hard times.

Last but not least, I would like to acknowledge the participants of UK Biobank and the supporting staff for their contribution to the resource, and appreciate the financial support from my college, the department, and Medical Science Division of Oxford.

In closing, the completion of this thesis would not have been possible without the collective support and encouragement of all those mentioned above and many others who have contributed in various ways. Thank you for being a part of this remarkable journey.

Statement of contribution to work

I drafted the text and produced all tables and figures presented in the current thesis. All analysis was conducted by myself, and I produced the pre-specified analysis plans for Chapters 4 to 9 with advice from my supervisors Ben Lacey, Sarah Lewington, Eirini Trichia, and Nazrul Islam. I used SAS programs developed by Paul Sherliker and Alison Offer for some analyses, modifying them where required to adapt to UK Biobank data. To produce some of the figures presented in Chapters 5 and 8, I used R packages developed by Diego Aguilar Ramirez. Study selection for the systematic review (Chapter 2) was conducted independently by Ayodipupo Oguntade and me.

The research presented in Chapter 6, 7 and 9 has been published or under review (List of publications and presentations), for which I was first author. I drafted the text, and Ben Lacey, Sarah Lewington, Eirini Trichia, Nazrul Islam and Jelena Besevic provided advice on the analysis plan and the manuscript.

This thesis used the UK Biobank resource (application number 31461). This work uses data provided by patients and collected by the NHS as part of their care and support.

Table of Contents

Abstract.....	i
Acknowledgements	iii
Statement of contribution to work	iv
List of publications and presentations	viii
List of abbreviations	ix
List of tables	x
List of figures.....	xi
CHAPTER 1: Introduction.....	1
Chapter summary	1
1.1 Coronary heart disease	2
1.2 Clinical biomarkers and risk scores	3
1.3 Metabolomics and NMR spectroscopy	5
1.4 Research gaps	7
1.5 Research objectives	9
CHAPTER 2: Systematic review	10
Chapter summary	10
2.1 Background	11
2.2 Methods	11
2.3 Results	13
2.4 Discussion	16
2.5 Conclusion.....	20
Chapter 2 Tables and Figures	22
CHAPTER 3: Study design and participants	28
Chapter summary	28
3.1 Overall study design	29
3.2 Blood sample processing and chemistry-measured markers	29
3.3 Measurement of NMR-derived metabolites	30
3.4 Definition of end point	31
3.5 Processing of other data.....	31
3.6 Participants exclusions	32
3.7 Baseline population characteristics	33
Chapter 3 Tables and Figures	35
CHAPTER 4: Baseline description of metabolites	45
Chapter summary	45

4.1 Background	46
4.2 Methods	47
4.3 Distribution of metabolites	48
4.4 Cross-correlations between metabolites	50
4.5 Correlations with chemistry biomarkers	51
4.6 Discussion	52
4.7 Conclusion.....	54
Chapter 4 Tables and Figures	55
CHAPTER 5: Within-person variability of metabolites	66
Chapter summary	66
5.1 Background	67
5.2 Methods	68
5.3 Baseline characteristics of resurvey	69
5.4 Estimation of within-person variability	70
5.5 Discussion	72
5.6 Conclusion.....	74
Chapter 5 Tables and Figures	75
CHAPTER 6: Prospective association of lipids and lipoproteins with CHD risk	83
Chapter summary	83
6.1 Background	84
6.2 Methods	85
6.3 Association of lipoprotein and apolipoprotein concentrations	87
6.4 Association of lipoprotein subclasses and particle sizes	88
6.5 Association of lipoprotein particle compositions	88
6.6 Association of lipids	89
6.7 Sensitivity analyses	90
6.8 Discussion	91
6.9 Conclusion.....	97
Chapter 6 Tables and Figures	98
CHAPTER 7: Prospective association of fatty acids with CHD risk.....	112
Chapter summary	112
7.1 Background	113
7.2 Methods	114
7.3 Association of fatty acids	115
7.4 Association of fatty acids independent of lipids.....	117

7.5 Sensitivity analyses	117
7.6 Discussion	118
7.7 Conclusion.....	123
Chapter 7 Tables and Figures	124
CHAPTER 8: Prospective associations of non-lipid and non-fatty acids metabolic biomarkers with CHD risk.....	129
Chapter summary	129
8.1 Background	130
8.2 Methods	131
8.3 Associations of other metabolic biomarkers	132
8.4 Associations of low-molecular-weight metabolites independent of lipids	133
8.5 Sensitivity analyses	133
8.6 Discussion	134
8.7 Conclusion.....	139
Chapter 8 Tables and Figures	140
CHAPTER 9: Predictive value of metabolic profiling in cardiovascular risk....	143
Chapter summary	143
9.1 Background	144
9.2 Methods	145
9.3 Prediction performance of the established risk score.....	147
9.4 Metabolites selected by different algorithms	148
9.5 Prediction performance after adding metabolites	148
9.6 Sensitivity analyses	149
9.7 Discussion	150
9.8 Conclusion.....	153
Chapter 9 Tables and Figures	154
CHAPTER 10: Overall conclusions.....	159
Chapter summary	159
10.1 Main findings	160
10.2 Strengths and limitations	163
10.3 Clinical and etiologic implications.....	164
10.4 Future directions	165
10.5 Conclusion	167
References	168
Appendix.....	185

List of publications and presentations

Publications

Jin D, Trichia E, Lewington S, Lacey B. Predictive value of circulating metabolites in assessing atherosclerotic cardiovascular disease risk. *J Epidemiol Community Health*. 2023;jech-2023-220801. (abstract in *Atherosclerosis* 379, S64-S66)

Jin D, Trichia E, Islam N, Lewington S, Lacey B. Independent associations of lipoprotein characteristics by NMR with incident coronary heart disease: a study of 90,000 individuals in UK Biobank. *J Am Heart Assoc*. 2023;0:e029552.

Jin D, Trichia E, Islam N, Lewington S, Lacey B. Associations of circulating fatty acids with incident coronary heart disease: a prospective study of 89,242 individuals in UK Biobank. 2023. *BMC Cardiovasc Disord* 23, 365.

Trichia E, Malden D, **Jin D**, et al. Independent relevance of adiposity measures to coronary heart disease risk among 0.5 million adults in the UK Biobank. (Accepted by *Int J Epidemiol*)

Oguntade AS, Islam N, Malouf R, Taylor H, **Jin D**, Lewington S, Lacey B. Body Composition and Risk of Incident Heart Failure in 1 Million Adults: A Systematic Review and Dose-Response Meta-Analysis of Prospective Cohort Studies. *J Am Heart Assoc*. 2023;e029062. (abstract in *Open Heart* 8 (1), e001632)

Tang X, Lu K, **Jin D**, et al. Incidence and Survival of Aortic Dissection in Urban China: Results from the National Insurance Claims. *Lancet Reg Health West Pac*. 2021;17:100280.

Oguntade AS, **Jin D**, Islam N, et al. Body composition and risk of heart failure: a systematic review and meta-analysis. *Open Heart*. 2021;8(1):e001632.

Jin D, Rich-Edwards JW, Chen C, et al. Gestational Diabetes Mellitus: Predictive Value of Fetal Growth Measurements by Ultrasonography at 22-24 Weeks. *Nutrients*. 2020;12(12):3645.

Presentations

Invited speaker, HDR UK Cambridge seminar series Metabolomics and coronary heart disease in UK Biobank	09/2023, Cambridge
Presentation, 91st European Atherosclerosis Society Congress Predictive value of high-throughput metabolites in assessing atherosclerotic cardiovascular disease risk	05/2023, Manheim
Poster, British Atherosclerosis Society Annual Meeting 2022 Independent associations of lipoprotein characteristics on risk of coronary heart disease	08/2022, Oxford
UK Biobank Scientific Conference 2021 Independent associations of lipoprotein characteristics with incident coronary heart disease	11/2021, Online

List of abbreviations

ApoA-1	ApolipoproteinA-1
ApoB	ApolipoproteinB
ASCVD	atherosclerotic cardiovascular disease
BCAA	branched-chain amino acids
BMI	body-mass index
CHD	coronary heart disease
CI	confidence interval
DHA	docosahexaenoic acid
EPA	eicosapentaenoic acid
FDR	false discovery rate
GlycA	glycoprotein acetyls
HDL	high-density lipoprotein
HR	hazard ratio
IDI	integrated discrimination improvement
IDL	intermediate-density lipoprotein
LDL	low-density lipoprotein
IQR	interquartile range
LR	likelihood ratio
MUFA	monounsaturated fatty acids
NMR	nuclear magnetic resonance
NRI	net reclassification improvement
PUFA	polyunsaturated fatty acids
RDR	regression dilution ratio
SBP	systolic blood pressure
SFA	saturated fatty acids
SD	standard deviation
UK	United Kingdom
VLDL	very-low-density lipoprotein

List of tables

Tables	Title	Page
Table 2.1	Summary of articles included in systematic review	24-27
Table 3.1	NMR-measured metabolites in main analyses	37-38
Table 3.2	ICD-10 and operation code of outcomes	39-40
Table 3.3	Baseline characteristics of study population by sex	42-44
Table 4.1	Baseline distribution of NMR-derived metabolites	56-59
Table 4.2	Correlations of baseline lipoproteins characteristics and lipids	62
Table 4.3	Cross-correlations of baseline fatty acids concentration	63
Table 4.4	Correlations of baseline fatty acids concentration and lipids	64
Table 4.5	Comparison of metabolites by clinical chemistry and NMR spectroscopy	65
Table 5.1	Baseline characteristics of participants with repeat assessment	76
Table 5.2	Within-person variability of metabolites, by sex	78-81
Table 5.3	Within-person variability of metabolites, by measurement method	82
Table 6.1	Baseline characteristics by incident coronary heart disease	99
Table 6.2	Lipoprotein and apolipoprotein concentrations vs coronary heart disease risk	100
Table 6.3	Lipoprotein particle concentration vs coronary heart disease risk, by particle subclass	104
Table 6.4	Lipoprotein particle size vs coronary heart disease risk	105
Table 6.5	Lipoprotein particle composition vs coronary heart disease risk	107
Table 6.6	Lipids and apolipoproteins vs coronary heart disease risk, by clinical chemistry and NMR spectroscopy	109
Table 6.7	Lipids vs coronary heart disease risk	110
Table 7.1	Association of fatty acids concentration with coronary heart disease	125
Table 8.1	Low-molecular-weight metabolites vs coronary heart disease risk	141
Table 9.1	Characteristics of baseline QRISK factors by 10-year incident ASCVD	155
Table 9.2	Associations of clinical metabolites independent from QRISK3	156
Table 9.3	Prediction performance of 10-year ASCVD risk w/o metabolites	157

List of figures

Figures	Title	Page
Figure 2.1	Flowchart of article selection and inclusion in literature review	23
Figure 3.1	Recruitment pattern of UK Biobank	36
Figure 3.2	Flowchart of exclusion criteria for study population in prospective analyses	41
Figure 4.1	Cross-correlations of NMR-derived metabolites at baseline	60
Figure 4.2	Cross-correlations of baseline lipoproteins characteristics	61
Figure 5.1	Regression dilution ratios for metabolites	77
Figure 6.1	Lipoprotein particle concentration vs coronary heart disease risk Regression	101
Figure 6.2	LDL particle and apolipoprotein B concentration vs coronary heart disease risk	102
Figure 6.3	VLDL particle and apolipoprotein B concentration vs coronary heart disease risk	103
Figure 6.4	Lipoprotein size vs coronary heart disease risk, independently of lipoprotein particle concentration	106
Figure 6.5	Lipoprotein composition of triglycerides vs coronary heart disease risk, independently of lipoprotein particle concentration	108
Figure 6.6	Total triglycerides and triglycerides in LDL vs coronary heart disease risk, independently of lipoprotein particle concentration	111
Figure 7.1	Risk of coronary heart disease by usual fatty acids concentration, with progressively adjustment for other fatty acids	126
Figure 7.2	Fatty acids concentration vs coronary heart disease risk, with adjustment for other fatty acids	127
Figure 7.3	Risk of coronary heart disease by usual fatty acids concentration, with further adjustment for lipids	128
Figure 8.1	Metabolite profiling vs coronary heart disease risk	142
Figure 9.1	Calibration of risk prediction models for 10-year CVD risk	158

CHAPTER 1: Introduction

Chapter summary

Coronary heart disease (CHD), caused by inadequate perfusion of the myocardium, is a leading cause of premature death and disability worldwide. Metabolomics is the large-scale study of small molecules, commonly known as metabolites, which are the substrates and intermediate products of the life-sustaining cellular regulatory processes. The comprehensive analysis of metabolites in a biological specimen has the potential to provide insights into multiple physiological processes and pathways underlying atherosclerosis, and thereby identify novel risk factors and improve risk prediction. High-throughput spectroscopy allows for quantifying a wider range of metabolites in the circulation and more detailed characterisation of some measures than standard clinical chemistry. However, the clinical value of such measures has not yet been thoroughly investigated using large-scale epidemiological studies. This chapter describes the background of metabolomics and cardiovascular disease, summaries the research gaps, and introduces the objectives of the thesis on investigating the prospective associations of NMR-derived metabolic biomarkers with risk of CHD in the UK Biobank.

1.1 Coronary heart disease

Epidemiology

CHD, also called ischaemic heart disease or coronary artery disease, is the major form of cardiovascular disease and the leading cause of premature mortality worldwide.¹ It is also the greatest cause of disability-adjusted life-years in both the 50-74 year and 75-years-and-older age groups.² The European Heart Network has estimated that CHD and related cardiovascular disease (CVD) events represents the highest component of healthcare costs in European countries and costs the EU economy more than €200 billion a year.³ Although CHD mortality in the UK has decreased in the past 40 years, it is still the single biggest cause of death in the UK (15% of total deaths in 2019), and the hospital admissions for all CVD increased by over 46,000 between 2010/2011 and 2013/2014.^{4,5}

Pathophysiology

CHD is caused by inadequate perfusion of the myocardium, when the formation of atherosclerotic plaque in the arteries obstructs the blood flow through the arteries, and decreases the supply of oxygen and nutrients to the heart muscle.^{6,7} One of the main underlying pathological processes that leads to CHD and stroke is known as atherosclerosis, an inflammatory process involved with a series of highly specific cellular and molecular responses affecting the medium- and large-sized blood vessels throughout the cardiovascular system.^{6,8} Endothelial dysfunction was suggested to be the first step in the atherosclerosis. Possible causes of endothelial dysfunction include elevated and modified levels of low-density lipoprotein (LDL), free radicals (caused by smoking, hypertension, and diabetes mellitus), genetic alterations, elevated

plasma homocysteine concentrations, infectious microorganisms, and combinations of these or other factors.⁸ Recent evidence also suggested coronary microvascular dysfunction and vasospasm may also contribute to the multifaceted pathophysiology of CHD.⁹

Risk factors

The World Health Organization has categorised several modifiable risk factors of CVD, and highlighted their impacts to the CVD burden worldwide. These factors include 1) metabolic risks, including high systolic blood pressure (SBP), high LDL cholesterol, high body-mass index (BMI), high fasting plasma glucose and kidney dysfunction; 2) behavioural risks, including dietary risks, tobacco smoking, alcohol use and low physical activity; and 3) environmental risks, including air pollution and non-optimal temperature.¹⁰ The European guideline issued by European Society of Cardiology lists five main causal and modifiable risk factors of atherosclerotic cardiovascular disease (ASCVD), including blood apolipoproteinB (ApoB)-containing lipoproteins, high BP, cigarette smoking, diabetes mellitus and adiposity.¹¹ Both the European guideline and US guideline (issued by American Heart Association /American College of Cardiology) have provided instructions on the management of these risk factors for primary and secondary prevention.¹¹⁻¹³

1.2 Clinical biomarkers and risk scores

Clinical biomarkers

Many biomarkers related to CHD pathology have been suggested to improve risk stratification including lipids and lipoproteins, C-reactive protein, fibrinogen, natriuretic peptides and high-sensitivity cardiac troponin.^{11,14} Some of them may be causal, whereas others may reflect underlying mechanisms or indicate

early cardiac damage.¹¹ However, except for LDL cholesterol and other ApoB-containing lipoproteins, which has been demonstrated to be critical in the development of ASCVD beyond any doubt,¹¹ most of the proposed association of other biomarkers may be inflated by publication bias or limited in additional predictive value of reclassification.¹⁴⁻¹⁷ Most of these biomarkers still need further evidence to prove their predictive value. Recent studies developed renewed interest in lipoprotein(a),^{18,19} and some other evidence provided insight on cardiac biomarkers;^{20,21} however, studies have not comprehensively assessed the improvement in risk discrimination on adding new biomarkers to established risk scores, and more evidence is required to assess reliably their usefulness in clinical practice.

Risk scores

Several risk scores, derived from different populations, have been developed to estimate individual cardiovascular risk in primary care. The Framingham Risk Score was the first widely used risk score for CHD developed separately for men and women based on the individual's age, total and high-density lipoprotein (HDL) cholesterol levels, cigarette smoking, SBP and hypertension treatment.^{22,23} However, since the score was developed during the peak incidence of CVD in the US over 40 years ago and has been shown to overestimate substantially the risk in contemporary European populations,²⁴ Systematic COronary Risk Evaluation (SCORE) was later developed for estimating 10-year risk of fatal CVD in Europe,²⁵ and SCORE2 was an updated algorithm for sex-specific risk prediction of fatal and non-fatal CVD, scoring by age, smoking status, SBP, and total and HDL cholesterol.²⁶ Pooled Cohort Equations is another score for 10-year ASCVD risk derived from a US

population, which improves the risk estimation among Black adults.^{27,28} Other risk scores have also been developed for specific areas and populations, such as the PROCAM score for acute coronary event risk based on a European population and the QRISK score for CVD risk that is widely used in UK primary care.²⁹⁻³¹ Globorisk has been developed for fatal and total CVD risk and can be recalibrated for application in different countries with routinely available information.³² WHO also developed CVD risk charts in 21 global regions to help adapt risk prediction in low-income and middle-income countries.³³

1.3 Metabolomics and NMR spectroscopy

Metabolites and metabolomics

Metabolites are the substrates and intermediate products of the life-sustaining cellular regulatory processes in organisms that drive essential cellular functions, such as energy production and storage, signal transduction and apoptosis.³⁴ Metabolomics, which is the comprehensive study of metabolic profiling (numerous metabolites from multiple molecular pathways), lies at the heart of omics profiling technologies that reflects the information expressed by genome, transcriptome and proteome.³⁵ As the omics discipline closest to the phenotype, metabolomics can detect the subtle alterations in biological pathways and provide insight into the mechanisms that underlie various physiological conditions and disease pathogenesis.³⁴ The application of metabolomics in clinical practice also allows for the discovery of novel biomarkers to improve the prognosis and diagnosis of disease.

High-throughput spectroscopy and nuclear magnetic resonance

The recent application of metabolomics in large-scale population-based studies has been enabled in part by the development of the new high-throughput

measurement technologies, namely spectroscopy, which have made it feasible to analyse hundreds or thousands of metabolites from biological fluids or biopsies from larger sample sizes than before. Mass spectrometry and NMR spectroscopy are the two primary high-throughput techniques to detect metabolites from a single serum or plasma sample, and perform metabolic profiling beyond the scope of the conventional chemistry methods³⁶. Mass spectrometry has the strength of high sensitivity that enables quantification of very low molecular concentrations; when coupled with chromatography, mass spectrometry has the capability of detecting thousands of metabolites.³⁷ NMR differs from mass spectrometry by its virtual absence of batch effects, high reproducibility, quick analysis, and comparatively low cost.³⁸ The choice of profiling platform for epidemiological studies is usually a trade-off between larger metabolite coverage (mass spectrometry) and larger sample size (NMR).³⁹

The NMR platform by Nightingale Health Ltd provides quantitative molecular data on 14 lipoprotein subclasses and their lipid compositions, apolipoprotein A-I (ApoA-1) and ApoB, multiple lipids and fatty acids measures, as well as numerous low-molecular-weight metabolites, including amino acids, glycolysis-related measures and ketone bodies. This platform has been used in numerous large-scale cohorts worldwide, including FINRISK cohorts, UK Biobank, China Kadoorie Biobank, and Mexico City Prospective Study, and was highly replicable in metabolic measures across cohorts.⁴⁰ In the UK Biobank, NMR metabolites are currently available in the baseline plasma samples of 118k participants (a random subset of the initial cohort), including 249 metabolic measures (168 in absolute levels and 81 ratio measures).

1.4 Research gaps

As applications of high-throughput spectroscopy in large-scale population-based studies has been increasing in recent years, various biological questions can readily be addressed to fill the research gaps. In terms of cardiovascular research, although some previous studies have investigated the associations of several metabolites with cardiovascular risk, none of them included metabolites data from a single cohort of more than 50k participants, and the evidence was especially limited for low-molecular-weight metabolites. Meanwhile, the high heterogeneity in measurement assay and biological samples among previous studies made it difficult to generalise robust evidence of the association for each metabolite.⁴¹ In addition, none of these studies have considered the within-person variability of the metabolic biomarkers over time. Therefore, more large-scale studies with uniform settings and repeated measurements are still necessary for the future association analyses (**Chapter 2**).

Although the associations of plasma lipids with CHD risk are well described,⁴²⁻⁴⁴ more recent evidence suggests these associations may be driven more directly by the total circulating concentration of ApoB-containing lipoprotein particles.^{42,45} As traditional lipids measures are dependent on both the concentration of lipoproteins and their composition (size and lipid content), the atherogenic potential of lipoproteins may well be dependent on both these characteristics, but most previous studies have not described lipid composition for each of the different lipoprotein classes. Therefore, the evidence on the independent association of each lipoprotein characteristic may provide more insight on how lipids and lipoproteins contribute to the atherosclerosis.

Circulating levels of fatty acids are the main constitutional component of circulating lipid classes and have been shown to modulate lipid metabolism.⁴⁶⁻

⁴⁸ However, the associations and the underlying biological mechanisms between circulating fatty acids and CHD risk, especially the associations independently from the effects from lipids and lipoproteins, have not been well described. Meanwhile, there has been a debate on the protective effect of omega-3 PUFA supplement to cardiovascular risk. The inconsistent results from clinical trials have raised the hypothesis that the formulation of omega-3 PUFA supplementation (Eicosapentaenoic acid [EPA] plus docosahexaenoic acid [DHA] versus pure EPA) may be the key aspect of the effects to cardiovascular event.^{49,50} Circulating levels of EPA and DHA are strongly influenced by dietary intake, and it is meaningful to explore the associations with circulating levels, which remain unclear; the ongoing RESPECT-EPA trial will provide further evidence.⁵¹

High-throughput metabolic biomarkers are thought to have great potential to improve risk stratification in clinical care. However, current evidence on the predictive value of metabolites has only focused on a limited number of biomarkers with linear associations with CVD, which may not reflect the complex pathophysiology of atherosclerosis.^{52,53} In addition, current evidence has not compared the predictive performance of adding metabolites to any established cardiovascular risk scores, so the additional benefit of metabolic profiling in clinical practice needs to be further confirmed.

1.5 Research objectives

The primary aims of this thesis are to assess the prospective associations of NMR-derived plasma metabolites to CHD risk in the UK Biobank, a prospective cohort study of approximately 0.5 million adults in the United Kingdom aged 40-69 years and to evaluate whether such metabolites improve prediction of future cardiovascular risk. The secondary aims include:

- To conduct a systematic review of the published literature on the associations between metabolites and risk of CHD;
- To assess the within-person temporal variation in NMR measures, and their reliability by comparing the levels of biomarkers measured using both NMR spectroscopy and clinical chemistry;
- To examine whether the associations of lipoprotein concentration with CHD risk differ by lipoprotein class, and whether size and lipid composition are independently associated with CHD risk;
- To quantify the associations of circulating fatty acids with CHD risk, independently of each other and major lipid fractions;
- To investigate the associations of low-molecular-weight metabolites with CHD risk, independent of major lipid fractions lipids;
- To evaluate whether adding high-throughput circulating metabolites to an established risk score improves the prediction of 10-year cardiovascular risk

CHAPTER 2: Systematic review

Chapter summary

High-throughput techniques for the measurement of circulating metabolites have been increasingly used in recent population-based studies. Since this review identified heterogeneity in the associations of metabolites with different subtypes of CVD, a systematic review was conducted focussing on risk of CHD to identify the latest evidence on the associations of circulating metabolites measured by high-throughput techniques with incident CHD. After exclusions, 33 articles were selected that were published prior to April 2023: nine publications were related to metabolic profiling, 16 of them solely focused on the association of lipoprotein-related metabolites, and another eight articles focused on a specific type of metabolite other than lipoproteins. The review demonstrated the potential of several metabolites as clinical biomarkers for risk stratification and early identification of CHD. However, due to the heterogeneity in measurement assay and biological samples, it is difficult to pool the results and generate robust evidence from current studies.

2.1 Background

Metabolomics is one of the new omics technologies that are thought to have great potential to better understand the mechanism of atherosclerosis, and improve risk stratification and the early identification of CVD.⁵⁴⁻⁵⁶ High-throughput technologies make it possible to measure hundreds or thousands of metabolites from biological fluids or biopsies, and therefore push the development of metabolomics research from limited enzymatic reactions or single pathways to comprehensive metabolic networks⁵⁷.

High-throughput technology, including NMR spectroscopy and mass spectrometry, has been used in population-based studies in the past few decades, and has been increasingly used in large-scale cohorts in the past ten years.³⁹ However, there have been no recent comprehensive systematic reviews on the associations of metabolites and metabolic profiling measured by high-throughput technology with risk of CHD. Therefore, this chapter aims to assess and summarise the existing observational literature on metabolites measured by high-throughput methods, including both comprehensive metabolic profiling and any specific type of metabolites, and CHD risk through a systematic literature review.

2.2 Methods

Data sources and search strategy

The systematic review was conducted following *the Cochrane Handbook of systematic review*.⁵⁹ The protocol of the review was registered in the PROSPERO registry (CRD42021225785) following the PRISMA-P (Preferred Reporting Items for systematic review and Meta-Analysis Protocols)

statement.^{60,61} We conducted the systematic search from MEDLINE and EMBASE database from inception through 1st April 2023.

Study selection

Eligible studies included those with: prospective cohort study design; adult participants; analyses on blood metabolites measured using NMR or other high-throughput technology; exposures included a wide range of metabolites or any specific type of metabolites; and CHD as the outcome. Exclusion criteria included: cohorts of highly selected groups (such as pregnant women or those with a given disease), solely genetic instrumental variants of metabolites; or studies focused solely on the value of metabolites in risk prediction models (i.e. where the association with specific metabolites, adjusted for potential confounders, could not be extracted). The detailed search strategy and search terms are shown in **S3.1**. Titles and abstracts were extracted independently by two study authors (D.J. and A.O.).

Data extraction and quality assessment

For each included study, we extracted information including the publication details (authors, year, journal), study design and follow-up years, population characteristics (country/region, ethnicity, age, sex, sample size, other inclusion/exclusion criteria), exposures and their measurement (detected metabolite profiling, biospecimen type, metabolomics measuring technique), outcomes (CHD or CHD subtypes, number of cases), analysis approach (targeted [identifies a limited number of known metabolites] or untargeted [measures all detectable signals in a given biospecimen capturing both known and unknown metabolites]),^{62,63} statistical model type and adjusted covariates (capturing from the final model with complete adjustments), multiple testing and

regression dilution bias correction (whereby use of measurements made at a single time point systematically underestimate the relationship with long-term 'usual' values),⁶⁴ results and findings (selected metabolites and their effect size). Newcastle-Ottawa Quality Assessment Form (quality assessment for meta-analysis of cohort study) was used for the quality assessment.

2.3 Results

The process of the literature search and study selection is described in **Figure 2.1**. After removing duplicates, 11,088 records were identified from two databases for abstract screening. After final exclusions, 33 articles were selected: 16 of them solely focused on the association of lipoprotein-related metabolites and CHD risk⁶⁵⁻⁸⁰, eight articles focused on a specific type of metabolite other than lipoproteins,⁸¹⁻⁸⁸ and nine were related to metabolic profiling^{52,58,61,89-94}.

The characteristics of the included literature were summarised in **Table 2.1**. All the articles had overall sample size larger than 1,000. The largest study had 30,208 participants and 5,766 CHD cases (focusing on LDL triglycerides),⁸⁰ and the largest study focusing on metabolic profiling was a prediction modelling study with 7,256 participants in the discovery set, and 6,185 participants in the validation set.⁵² Except for one Chinese cohort (China Kadoorie Biobank),⁵⁸ one African American cohort (Jackson Heart Study),⁹³ and five studies from two multi-ethnic cohorts (Multi-Ethnic Study of Atherosclerosis, the Atherosclerosis Risk In Communities study) in the US,^{72,76,79,88,91} other included studies were conducted in populations of European descent. All lipoprotein-related studies were measured by NMR spectroscopy, while the studies focusing on other

metabolites or on metabolic profiling were measured by either NMR or mass spectrometry.

Among the lipoprotein-related literature, the studies before 2018 mainly focused on the associations of lipoprotein concentrations and particle size with CHD risk, while the studies after 2018 started to focus on triglycerides in lipoproteins.^{70,71,80} All the literature on other types of metabolites, including fatty acids, amino acids, ketone bodies, and gut bacteria-related metabolites, were published after 2014.

All nine articles on metabolic profiling were published after 2014, and five of them validated their results in a new dataset^{52,61,90,92,93}. All included studies had more than 5 years of follow-up in at least one of the discovery or validation datasets. Four studies on metabolic profiling used fasting blood samples in at least one of the datasets^{52,89,90,92}, and five studies used mass spectrometry^{61,90,91,93,94}. Two studies used an untargeted approach to identify both unknown and known compounds^{90,91}, and others used targeted metabolic profiling, with the numbers of candidate metabolites ranging from 58⁹³ to 371⁶¹.

For the statistical methods, the associations of metabolites with CHD risk were estimated using Cox proportional hazards model or logistic regression, but none of the included studies dealt with regression dilution bias. All the studies on metabolic profiling identified the metabolites significantly associated with CHD risk and corrected for multiple testing using false discovery rate or the Bonferroni method. Two of them further applied stepwise selection or the Least Absolute Shrinkage and Selection Operator (LASSO) algorithm to select novel metabolites^{91,92}. Six studies examined the predictive value of metabolites on

estimating cardiovascular risk, but none of them added metabolites to an established cardiovascular risk score.^{58,61,89,91,92,94}

A number of studies have assessed the associations of NMR-derived measures of lipoprotein concentrations with CHD risk.^{65,73,74} These studies have consistently described positive associations with very-low-density lipoprotein (VLDL) and LDL particle concentrations, and inverse associations with HDL particle concentrations. Subclasses of lipoproteins can be categorized by either particle size or density. Previous studies on the association of size-determined lipoprotein subclasses with CHD risk focused mainly on LDL and HDL^{58,65,74} and did not find differences in the strength of associations across LDL subclasses.⁵⁸ However, previous evidence on HDL subclasses has been less consistent, with some describing heterogeneity and others not.^{58,65,74} Triglycerides composition has been assessed in three recent studies: one focused on triglycerides in LDL,⁸⁰ one on triglycerides in HDL and HDL subclasses,⁷¹ and the other comparing the associations of triglycerides composition across different lipoproteins.⁷⁰ The strength of associations were found to vary by different lipoproteins and by different subclasses of HDL, and the associations of triglycerides in HDL showed inconsistent results.^{70,71}

In terms of other metabolites, circulating fatty acids have been associated with CHD risk in several studies, but heterogeneity still existed among each subtype of fatty acids.^{52,58,82} In the category of cholines, positive associations with CHD risk have been observed for phosphatidylcholines (PCs)^{61,89,94}, and inverse associations for lysoPCs^{90,92,93}, but the evidence of sphingomyelins remained inconsistent^{89,90,93,94}. Although many previous studies have observed amino

acids as significant biomarkers of CHD risk, including histidine, glutamine, alanine, and branched-chain amino acids, the results were not consistent across studies.^{58,61,91,92,94}

Three of six studies that evaluated the predictive value of metabolites observed a significant improvement in risk prediction^{61,91,92}. In one of the three studies, adding arginine and two LysoPCs to the conventional cardiovascular risk factors has been observed to increase the C-statistics of myocardial infarction risk prediction from 0.70 to 0.78 ($p=0.001$) in the discovery set and from 0.70 to 0.76 ($p=0.02$) in one of the validation sets⁹², but not in the other validation set using the non-fasting blood samples. The changes of C-statistics were 0.016 (95% CI [0.008, 0.024]) and 0.03 ($p=0.03$) in two other studies showing significant improvement in risk prediction, and one of them added principal components rather than the novel metabolites to the established risk factors⁹⁵. Another study that added DHA, monounsaturated fatty acids (MUFA) to total fatty acids ratio, omega-6 fatty acids and phenylalanine failed to enhance the C-statistics, although it observed the improvement in risk classification among participants in 5% to 10% risk range.⁵²

2.4 Discussion

This systematic review identified 33 studies (published before April 2023) assessing the prospective associations between metabolites measured by high-throughput methods and CHD risk. Half of the included literature solely focused on lipoprotein-related metabolites, and another nine studies assessed the whole metabolic profiling, among which the largest one included 13k participants. Two types of high-throughput spectroscopy, NMR and mass

spectrometry, and two types of blood sample, serum and plasma, were used for metabolites measurement across the literature. There was inconsistent evidence on the associations of several metabolites with CHD risk, and the additional predictive values of these metabolites, especially compared with an established risk score, remained unclear.

A previous systematic review identified 12 papers published before 2016 on comprehensive metabolic profiling and incident CVD, but showed lack of consistency in the results of included studies because of the heterogeneity in measurement assay and biological samples.⁴¹ Additionally, CVD is a complex disease, and previous evidence has proved the heterogeneity between associations of metabolites with CHD and those with stroke. It is unclear whether the metabolites that have been identified to be associated with CVD risk remain relevant for CHD risk. Therefore, the systematic review in this session included the studies of both metabolic profiling and specific types of metabolites to capture more studies, and limited the outcome to CHD.

Among the studies on metabolic profiling, two of the included articles applied untargeted assays. The difference between targeted and untargeted profiling is that targeted profiling identifies a limited number (tens to hundreds) of known metabolites, while the untargeted approach, applied by mass spectrometry, measures all detectable signals in a given biospecimen capturing both known and unknown metabolites^{62,63}. Although untargeted assays provide an opportunity to discover unknown metabolites, it is sometimes difficult to separate the signal peaks from background interferences, and the large detection range may lead to serious multiple testing problem^{62,96,97}.

NMR and mass spectrometry are the two primary techniques to detect and identify numerous metabolites across multiple biological pathways³⁶. Mass spectrometry has the strength of high sensitivity that can quantify very low molecular concentrations; when it is coupled with chromatography, mass spectrometry has the capability of detecting thousands of metabolites³⁷. However, mass spectrometry coupled with chromatography has the problem of batch effect due to the interaction with the instrument, and it is more time-consuming. NMR has been increasingly applied in recent large population-based studies because of its quick analysis, low cost, high reproducibility (consistency), and free of batch effects³⁸. Meanwhile, NMR provides detailed characterizations of lipoproteins, and was therefore applied in all included studies specifically assessing the associations of lipoproteins. However, NMR has the limitations of lower sensitivity and narrower detection range compared with mass spectrometry³⁷. In summary, mass spectrometry and NMR are complementary, and it is critical to choose the more suitable technique according to the study objective and the nature of the sample.

Some evidence has suggested that plasma and serum should generate similar results in clinical and biological studies when applying the same blood preparation procedure, which is consistent with one of the included studies using plasma samples in the discovery set and serum samples in validation. Additionally, plasma samples have better reproducibility, while serum samples provide more sensitive results in biomarker detection.⁹⁸ This conclusion explained the reason that plasma samples were widely used in the studies focusing on specific types of metabolites, while serum samples were more commonly used for the studies on detection of novel metabolites from metabolic

profiling. On the other hand, most of studies in this review did not use fasting samples, and did not adjust for fasting time in the association analyses, which may affect the measurement accuracy of some metabolites that were strongly related to dietary intake, such as fatty acids and amino acids.

Clinical Perspectives

This review of 33 prospective studies has demonstrated the potential of several metabolites, including phosphatidylcholines, fatty acids and amino acids, as clinical biomarkers for risk stratification and early identification of cardiovascular disease. However, the additional prediction value of adding metabolites on top of traditional risk factors varied among studies, and there was no evidence on the predictive value compared with any established cardiovascular risk score.

Strengths and limitations

This systematic review comprehensively summarised the observational studies on both specific metabolites and metabolic profiling to CHD risk, and updated the most recent evidence after the wide applications of NMR in population-based cohorts. Compared with the previous review, this review included more studies with larger sample size and validation set, and had more accurate endpoint of CHD to avoid the interference of various effects in different CVD subtypes.

However, because of heterogeneity in the measurement assay and biological samples, and no report of effect size for the non-significant metabolites, it is still difficult to pool the results from different studies. Although some studies showed little difference in results between serum and plasma samples^{90,98}, other procedures in sample collection and preparation, such as clotting time and

clotting temperature can also affect the accuracy of the measurements⁹⁹. Therefore, more convincing evidence from large cohorts and different ethnic groups, as well as the pooled results from studies in unified candidate profiling and sample collection, are the key points for results generalisation. Additionally, statistical analysis needs to be refined in future studies, including adjustment of fasting time if using non-fasting samples, and taking regression dilution bias into consideration.

The application of a genetic approach could be one of the future directions of metabolomics research. Genome-wide studies have the advantage of more standardized techniques and shared databases, and can better deal with unmeasured confounding, which could possibly solve some of the problems in current observational studies⁴¹. One of the included studies applied GWAS and Mendelian randomisation to identify the unreported metabolite-genetic locus associations.⁹⁰ Several recent studies have integrated genomic and metabolic profiling to explore the reaction networks¹⁰⁰⁻¹⁰². The understanding of the genomic role and corresponding pathway through metabolites to disease will facilitate the understanding of atherosclerosis mechanisms and the exploration of environmental effects to the disease.

2.5 Conclusion

This review summarized current evidence on circulating metabolites and incident CHD, and demonstrated the potential of several metabolites as clinical biomarkers for risk stratification and early identification of cardiovascular disease. However, it is still difficult to pool the results and generate robust evidence from current studies, largely due to the heterogeneity in measurement

assays and biological samples. This review suggests future studies increase the sample size or include multiple cohorts as validation, explore the evidence from diverse ethnic groups, consider regression dilution bias in association analyses, and evaluate the additional prediction value of metabolites compared with any established risk score. The application of genetic approaches in future metabolomics studies will allow exploration of the casual pathways and reaction networks.

Chapter 2 Tables and Figures

Tables	Title
Table 2.1	Summary of articles included in systematic review

Figures	Title
Figure 2.1	Flowchart of article selection and inclusion in literature review

Figure 2.1 Flowchart of article selection and inclusion in literature review

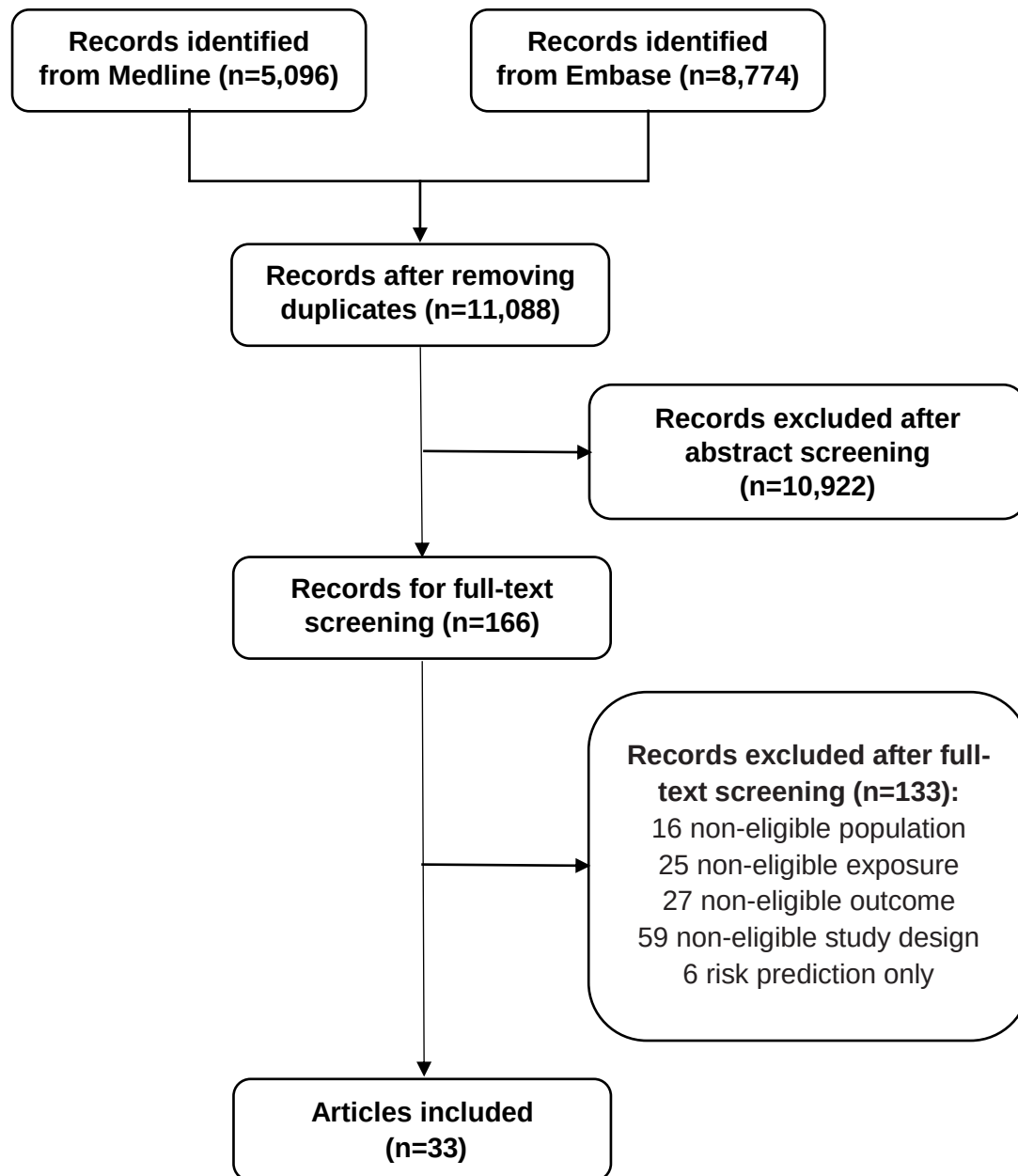


Table 2.1 Summary of articles included in systematic review

A. Articles on lipoproteins only

First author, year	Study name/ country	Study design	Follow-up years	Sample size (cases)	Target metabolites	Assay	Biological sample	Results
Balling, 2023	Copenhagen General Population Study /Denmark	cohort	9.2 yrs	30208 (5766)	TG in LDL	NMR	plasma	TG in LDL associated with increased risk of CHD, even given ApoB
Joshi, 2020	BRHS, the Whitehall II study, SABRE/UK	cohort	> 7 yrs	11560 (1031)	TG in lipoprotein subclasses	NMR	serum	Total TG and most TG sub-fractions (with the exception of TG in HDL subclass) were associated with increased risk of CHD. TG within VLDL subclass had strongest association with CHD
Kettunen, 2019	FINRISK & DILGOM, SABRE/Finland, UK	cohort	8 yrs	14180 (616)	TG in lipoprotein subclasses	NMR	serum	TG compositions of all HDL subclass were positively associated with CHD, independent of concentrations of chol and TG. Adjusting for LDL-C had only very minor effects on the associations of both HDL-related TG concentrations and TG compositions of HDL
Pichler, 2018	the Hortega Study/Spain	cohort	12.4 yrs	1162 (159)	LDL_D subclasses	NMR	plasma	LDL_P and subclasses were associated with incident CVD. A change from large to medium and small LDL particles was associated with an increased CHD risk
Chandra, 2015	DHS/US	cohort	9.3 yrs	1977 (226)	HDL_C*, HDL_P	NMR	plasma	HDL-C, adjusted for risk factors and HDL-P, was inversely associated with incident CHD in both ethnicity; HDL-P, adjusted for risk factors and HDL-C, was inversely associated with incident CHD with no interaction by ethnicity
Steffen, 2015	MESA/US	cohort	8.5 yrs	6814 (272)	LDL_P, HDL_P	NMR	plasma	LDL-P and LDL-P/HDL-P ratio were positively associated with CHD events, and remained significant after adjustment for TC/HDL-C.

* metabolites mark with * was measured by chemistry method instead of NMR; DILGOM=Genetic determinants of Obesity and Metabolic syndrome; SABRE=the Southall And Brent Revisited Study; BRHS=the British Regional Heart Study; DHS=The Dallas Heart Study; WHS= Women's Health Study; HPS=Heart Protection Study; MESA=Multi-Ethnic Study of Atherosclerosis; EPIC=the European Prospective Investigation into Cancer and Nutrition; CHS=Cardiovascular Health Study. sdLDL-C=small dense low-density lipoprotein cholesterol content; VLDL=very-low-density lipoproteins; IDL=Intermediate-density lipoproteins; LDL=low-density lipoproteins; HDL=High-density lipoproteins; D=diameter; C=cholesterol; TG=triglycerides; P=particle concentration; Apo=Apolipoprotein

A. Articles on lipoproteins only (continued)

First author, year	Study name/ country	Study design	Follow-up years	Sample size (cases)	Target metabolites	Assay	Biological sample	Results
Mora, 2014	WHS/US	cohort	17.2 yrs	27533 (1070)	LDL-C*, LDL-P, NHDL-C*	NMR	plasma	For women with discordant LDL-related measures, coronary risk may be under or overestimated when relying on LDL-C alone
Tsai, 2014	MESA/US	cohort	8.5 yrs	8769(479)	sdLDL-C	NMR	plasma	NMR-sdLDL-C was not associated with significant risk of CHD
Akinkuolie, 2014	WHS/US	cohort	17 yrs	26332 (969)	HDL_P subclasses	NMR	plasma	Adjusted for correlations of HDL subclasses with each other (and metabolic and LP variables), associations were attenuated (still inverse) for L, M, and S_HDL-P, and became null for XL and XS_HDL-P
Parish, 2012	HPS/UK	cohort	5.3 yrs	19988 (1797)	LDL_D, LDL_P subclasses	NMR	plasma	CHD was equally strongly associated with the cholesterol- and particle-based total LDL measures; given LDL-P, the association became weaker for HDL-C than for HDL-P
Mackey, 2012	MESA/US	cohort	6 yrs	5598 (227)	HDL_C*, HDL_P	NMR	plasma	Adjusted for each other and LDL-P, HDL-C was no longer associated with CHD, but HDL-P independently associated with CHD
Arsenault, 2010	EPIC-Norfolk study	nested case-control	6 yrs	2940 (1025)	LDL_D, HDL_D	NMR	plasma	The association between HDL size and CAD risk was abolished on adjustment for apoB and TG levels
Harchaoui, 2009	EPIC-Norfolk study	nested case-control	6 yrs	2223 (822)	HDL_P, HDL_D	NMR	plasma	HDL-P and HDL size are independently (and inversely) associated with CHD. Association between HDL size and CHD was abolished when adjusting for ApoB
Steeg, 2008	EPIC-Norfolk study	nested case-control	6 yrs	2349 (858)	HDL_C*, HDL_D	NMR	plasma	HDL-C and HDL particle size, at very high values, holding apoA-1 and apoB constant, positively associated with CAD
Harchaoui, 2007	EPIC-Norfolk study	nested case-control	6 yrs	2888 (1003)	LDL_P, LDL_D	NMR	plasma	LDL-P and non-HDL-C positively and closely related to CAD. Stronger association than LDL-C. LDL size association was abolished after adjusting for LDL-P
Kuller, 2002	CHS/US	nested case-control	>3 yrs	683 (434)	LDL_P, LDL_D, HDL_P, HDL_D	NMR	plasma	After adjusting for TG and HDL_C, LDL_P (but not LDL size) remained significantly related to myocardial infarction and angina

* metabolites mark with * was measured by chemistry method instead of NMR; DILGOM=Genetic determinants of Obesity and Metabolic syndrome; SABRE=the Southall And Brent Revisited Study; BRHS=the British Regional Heart Study; DHS=The Dallas Heart Study; WHS= Women's Health Study; HPS=Heart Protection Study; MESA=Multi-Ethnic Study of Atherosclerosis; EPIC=the European Prospective Investigation into Cancer and Nutrition; CHS=Cardiovascular Health Study. sdLDL-C=small dense low-density lipoprotein cholesterol content; VLDL=very-low-density lipoproteins; IDL=Intermediate-density lipoproteins; LDL=low-density lipoproteins; HDL=High-density lipoproteins; D=diameter; C=cholesterol; TG=triglycerides; P=particle concentration.

B. Articles on specific types of metabolites other than lipoproteins

First author, year	Study name/ country	Study design	Follow-up years	Sample size (cases)	Target metabolites	Assay	Biological sample	Results
Satos, 2023	PREDIMED trial	case-cohort	4.8 yrs	791 (231)	Citric acid cycle metabolites	LC-MS	Plasma (fasting)	no metabolites remained significant after correction for multiple comparisons, although nominal significant associations with MI were found for 2-hydroxyglutarate, fumarate and malate
Shemesh, 2023	MESA/US	cohort	13.6 yrs	6796 (336)	Ketone bodies	NMR	plasma	Total ketone bodies were associated with higher hard CVD risk, but not with risk of MI
Borges, 2020	6 studies of UCLEB consortium/UK	5 cohorts and 1 nested case-control (individual meta)	10-25 yrs	16126 (3022)	fatty acids	NMR	serum	Docosahexaenoic acid, but not linoleic acid, was related to lower CHD risk; monounsaturated fatty acids were associated with higher CHD risk; Saturated fatty acids were not related to increased CHD risk
Heianza, 2020	NHS/US	nested case-control	10 yrs	760 (380)	TMAO	LC-MS	plasma	10-year increases in TMAO were significantly associated with an increased risk of CHD
Ottosson, 2020	MDC study, Malmo Preventive Project)/Swedish	case-cohort and cohort	7.2 and 19.0 yrs	3361 (423) and 880 (384)	gut bacteria-related metabolites	LC-MS	plasma	Phenylacetylglutamine (PAG) was associated with an increased risk of CAD
Ottosson, 2019	MDC study, Malmo Preventive Project)/Swedish	case-cohort and cohort	7.2 and 19.0 yrs	1049 (384) and 3799(423)	DMGV	LC-MS	plasma	Baseline levels of DMGV associated with incident coronary artery disease
Tobias, 2018	WHS/US	cohort	18.6 yrs	27041 (551)	BCAA	NMR	plasma	Total BCAAs were associated with MI ,the same positive association for Isoleucine, Leucine, Valine
Akinkuolie, 2014	WHS/US	cohort	17.2 yrs	27491 (1089)	GlycA	NMR	plasma	Baseline GlycA was associated with incident CHD, consistent with a possible role for protein glycans in inflammation and CVD

UCLEB= UCL-LSHTM-Edinburgh-Bristol (UCLEB) consortium; WHS= Women's Health Study; NHS=Nurses' Health Study; MDC study=Malmo Diet and Cancer Study; MESA=Multi-Ethnic Study of Atherosclerosis. BCAA= Branched-Chain Amino Acids; TMAO=trimethylamine N-oxide; DMGV= Dimethylguanidino Valerate .

C. Articles on numerous metabolites (metabolic profiling)

First author, year	Study name/ country	Study design	Follow-up years	Sample size (cases)	Assay	Biological sample	Results (metabolites found to be associated with CHD risk)
Cruz, 2022	discovery: JHS/US, validation: WHI	discovery: cohort, validation: nested case-control	discovery: 11.7 yrs, validation: ~5 yrs	discovery: 2346 (190); validation: 1588 (793)	LC-MS	plasma (fasting)	homoarginine, histidine, methionine, uridine, 2 amino acids, 2 lipids, 3 nucleosides, 1 endocannabinoid, and 1 monocarboxylic acid anion
Cavus, 2019	BiomarCaRE/6 European cohorts	case-cohort	9.2 yrs	10741 (2166)	LC-MS	serum	2 acyl-alkyl-PC, 3 diacyl-5 PC
Wang, 2019	ARIC study/US	cohort	30 yrs	3598 (633)	LC-MS & GC-MS	plasma	8 amino acids, 1 carbohydrate, 3 lipids, 2 nucleotide, 1 peptide, 4 Xenobiotics
Floegel, 2018	EPIC/2 European cohorts	case-cohort	7.8 and 7.3 yrs	2,984 (564)	NMR	serum (fasting)	1 acylalkyl-PC, 2 diacyl-PC
Holmes, 2018	CKB/China	nested case-control	7-10 yrs	4,662 (912)	NMR	plasma (non-fasting)	lipoproteins, ApoB, size of HDL, cholesterol and TG in most lipoprotein subclasses, glutamine, glucose, bOHBut, Gp, most fatty acids
Paynter, 2018	discovery: WHI-OS; replication: WHI-HT/US	nested case-control	discovery: 5.8 yrs; replication: 3.7 yrs	discovery: 944 (472); validation: 627 (312)	LC-MS	plasma	2 hydroxy-PC, 1 CMP, 3 HETE, 1 glutamate, 1 glutamine
Ward-Caviness, 2017	discovery: KORA S4 /German; replication: KORA S2/German, AGES and REFINE/Iceland	case-cohort	8-10 yrs	discovery: 1342 (67); validation: 661(112) & 254 (87)	NMR	serum (non-fasting for KORA S2, fasting for others)	1 arginine, 2 lysoPC
Würtz, 2015	discovery: FINRISK; replication: SABRE, BWHHS/UK	cohort	discovery: 15 yrs; replication: 12-23 yrs	discovery: 7256 (800); validation: 2622 & 3563 (941 in total)	NMR	serum (fasting for replication cohorts)	phenylalanine, monounsaturated and polyunsaturated fatty acids
Ganna, 2014	discovery: ULSAM, validation: TwinGene /Sweden	discovery: cohort, validation: case-cohort	discovery: 10 yrs; validation: 3.9 yrs	discovery: 1028 (131), validation: 1670 (282)	LC-MS	discovery: plasma, validation: serum(fasting)	2 lysoPC, 1 sphingomyelin, 1 monoglyceride

* Except Holmes's paper, other literature did not included lipids and lipoproteins as their candidate metabolites. JHS, Jackson Heart Study; WHI-OS, Women's Health Initiative Observational Study; WHI-HT, the Women Health Initiative Hormone Therapy Trials; BiomarCaRE, the Biomarkers for Cardiovascular Risk Assessment in Europe; ARIC, the Atherosclerosis Risk In Communities; EPIC, the European Prospective Investigation into Cancer and Nutrition; CKB, the China Kadoorie Biobank; KORA, The Cooperative Health Research in the Region of Augsburg; AGES and REFINE, The Age, Gene/Environment Susceptibility -Reykjavik and the Risk Evaluation For Infarct Estimates; SABRE, the Southall and Brent Revisited study; BWHHS, the British Women's Heart and Health Study; ULSAM, The Uppsala Longitudinal Study of Adult Men. LC-MS, liquid chromatography–mass spectrometry; GC-MS, gas chromatography–mass spectrometry; NMR, nuclear magnetic resonance spectroscopy. PC, phosphatidylcholines; SM=sphingomyelin; bOHBut=3-Hydroxybutyrate; Gp=Glycoprotein acetyls; HETE, hydroxyeicosatetraenoic acid; CMP, cytidine monophosphate.

CHAPTER 3: Study design and participants

Chapter summary

UK Biobank is a large-scale prospective cohort of 502,485 adults aged 40-69 years, recruited from the general population of the United Kingdom from 2006 to 2010. NMR spectroscopy was used for high-throughput metabolic profiling of the baseline plasma samples of 117,980 participants (a random subset of the initial cohort). Metabolic profiling included 249 metabolites, including particle concentration, particle size and lipid contents of lipoproteins, as well as apolipoproteins, total lipids, fatty acids, and low-molecular-weight metabolites. After exclusions, a total of 86,499 individuals with NMR-metabolomics profiling and without CHD or statin use at baseline were used in the main analysis.

3.1 Overall study design

UK Biobank is a prospective cohort study of 502,485 adults in the United Kingdom aged 40-69 years. Participants were recruited from 2006 to 2010, through National Health Service (NHS) registers, and invited to one of 20 assessment centres around the UK (**Figure 3.1**). Information about sociodemographic, lifestyle and health-related characteristics was collected by computer-based questionnaires, and clinical measurements including anthropometrics and blood pressure were taken by trained nurses. Blood samples were collected for long-term storage. All participants provided consent for follow-up through linkage to their health-related records. A subset of 20,346 participants, who lived within a 35 km radius of the assessment centre at Stockport, received a resurvey approximately 5 years after baseline (2012-2013), comprising the full baseline assessment. The study protocol, baseline assessment and population characteristics are described in greater depth elsewhere.^{103,104}

Ethics approval for the UK Biobank study was given by the National Health Service North West Multicentre Research Ethics Committee. All participants provided informed written consent to take part in the study. The UK Biobank application number for this research is 31461.

3.2 Blood sample processing and chemistry-measured markers

During baseline assessment, approximately 45 ml of blood sample was collected from each participant and stored in the appropriate aliquots for long-term storage. Serum samples were collected in Silica clot accelerator (serum separator tubes), and plasma samples were collected in Ethylenediaminetetraacetic acid (EDTA) tube. All samples were processed with an average time of 24 ± 2.5 hours between venepuncture and sample storage.¹⁰⁵

Overall, 29 chemistry-measured markers from serum samples were selected for assay in all 500,000 participants, including cholesterol levels and markers of inflammation, liver function, kidney function and insulin sensitivity. Further details of the measurements and the assay procedures are provided online by the UK Biobank.¹⁰⁵ A comparison between chemistry-measured markers and NMR-derived markers is included in Chapter 4.

3.3 Measurement of NMR-derived metabolites

NMR spectroscopy (Nightingale Health, Finland) was used for high-throughput metabolic profiling of the baseline plasma samples of 117,980 participants (a random subset of the initial cohort), including 249 metabolic measures (168 in absolute levels and 81 ratio measures) are quantified per EDTA plasma sample.^{39,106} Samples were generally drawn non-fasting. Overall 5,306 participants (a random set of 20,346 resurveyed participants) underwent repeat measurement of blood sample by NMR spectroscopy.

Data on the following metabolites were used in the main analyses: concentration of lipoproteins; concentration of 13 lipoprotein subclasses (as defined by particle size); the average diameters of lipoprotein particles; concentration of seven lipids (including cholesterol esters, free cholesterol, phospholipids and triglycerides) from each lipoprotein and its subclass; and the average lipid composition of each lipoprotein and its subclass; concentration of two key apolipoproteins (apolipoproteinA-1 and apolipoprotein B) and their ratio; concentration of 11 lipids or lipid fractions; eight fatty acids; four cholines; nine amino acids; four glycolysis-related metabolites; four ketone bodies; two fluid balance-related metabolites; and one inflammation-related metabolite. The analysis for prospective associations (Chapter 6, 7, 8) focus on these aforementioned metabolites, and 39 of them all measured with comparable validity to

the clinical chemistry (called clinical-validated metabolites) are included as candidate metabolites in prediction analysis (Chapter 9). The concentrations and ratios of lipid composition in 14 lipoprotein subclasses (totalling 157 measures) were not included in analyses - reliable assessment of the major lipoprotein characteristics was prioritised over analyses of their subclasses, especially given the limited power to assess associations with the many subclasses after accounting for multiple testing. A full list of the 92 biomarkers in the prospective association analyses is provided in **Table 3.1**.

3.4 Definition of end point

Incident CHD (outcome for Chapters 6, 7, 8) was defined as the first-ever myocardial infarction, unstable angina pectoris, coronary-related death, or relevant procedure. Incident events were identified from Hospital Episode Statistics (HES) and the Office for National Statistics cause of death data using codes of the 10th edition of the International Classification of Disease (ICD-10), and coronary-related procedures (coronary artery bypass surgery or percutaneous transluminal angioplasty stent placement) were identified from HES by the Office of Population Censuses and Surveys (OPCS) Classification of Interventions and Procedures. The events prior to recruitment were identified through HES or baseline self-report. A detailed definition and code list of the outcomes are included in **Table 3.2** (the codes marked with * were used to identify incident CHD, and all the codes in the table were used to exclude baseline CVD).

3.5 Processing of other data

In prospective analyses (Chapters 6, 7, 8), selected sociodemographic and lifestyle characteristics at baseline assessment were categorised into groups: age (5-year age groups); sex (men, women); educational attainment (University education, AS/A-level,

GSE/GCSE/O-level, other qualification); UK region (ten regions); ethnicity (White, other); smoking status (current regular, current occasional, former, never); and alcohol consumption (current regular, current occasional, former, never). Continuous covariates, including Townsend Deprivation Index,¹⁰⁷ body mass index, waist circumference, waist-hip ratio, systolic blood pressure, and fasting time, were categorised into seven equal groups. Baseline medical history, including baseline diabetes, hypertension treatment, and regular steroid tablets use, identified from self-report and HES, were used as binary variables. Self-reported baseline dietary and supplement intake were also categorised into groups: whole grains (0 serving/day, <1 serving/day, 1-<3 servings/day, ≥3 servings/day), fresh fruit (<1 serving/day, 1-<2 servings/day, 2-<3 servings/day, ≥3 servings/day), vegetables (<1 serving/day, 1-<2 servings/day, 2-<3 servings/day, ≥3 servings/day), cheese (<1 time/week, 1 time/week, 2-<5 times/week, ≥5 times/week), red meat (0 time/week, 1 time/week, ≥1 times/week), oily fish (0 time/week, 1 time/week, ≥1 times/week), dairy products (yes, never), supplements (vitamin supplement, mineral supplement, fish oil, glucosamine, multiple supplement, no supplement).

3.6 Participants exclusions

Of the 117,980 participants who attended the baseline assessment and had NMR-metabolomics profiling, 17,855 participants were further excluded because they had withdrawn consent from the UK Biobank at the time of analysis; had NMR data quality control issues (duplicate samples, technical errors or missing data for all metabolic biomarkers); had missing or outlying values in the clinical-validated metabolites and the key variables; or had prior cardiovascular disease at baseline (identified by HES records or baseline self-report). Those who reported taking lipid-lowering medication at baseline (13,381 participants) were also excluded because of the effect modification

of lipid-lowering medication to the associations. Participants with any NMR-derived metabolite outside the range of 4 standard deviations (SD) and out of 0.003% of either side of the distribution were defined as outliers (**Table 3.1**). The flowchart of exclusion criteria for the study population is presented in **Figure 3.2**. Participants with both baseline and repeat NMR-assessment were defined as the resurvey population, and analyses of resurvey data followed the same exclusion criteria as the study population at baseline (more details provided in Chapter 5).

3.7 Baseline population characteristics

The study population consisted of a total of 86,499 individuals without a history of CVD who had complete information on NMR-metabolomics profiling at baseline. The baseline characteristics of the study population showed no difference with the whole UK Biobank participants (**S3.1**). **Table 3.3A** describes the socio-demographic characteristics of participants included in the main analyses. The mean baseline age of the full study population was 55.3 (SD 8.0) years and 57% were women. In total, 95% of the study population were white and 40% were educated to college or higher. The proportion of current smokers and current alcohol drinkers were 10% and 93%, respectively. On average, men had higher education compared with women, and a larger proportion of men consumed alcohol regularly and were current or former smokers.

The overall mean BMI was 27 kg/m², and the mean blood pressure was 137 mmHg. About 2% of the study population had a history of diabetes, and 12% were on blood pressure-lowering medication. In terms of clinical biomarkers, the median fasting time was 3.0 hours. The mean clinical measured total cholesterol, HDL cholesterol, and total triglycerides were 5.9, 1.5, and 1.7 mmol/L, respectively. Men had higher levels of creatinine (64 umol/L in men; 81 umol/L in women) and GGT (43 vs. 29 U/L,

respectively) compared with women. HbA1c level was similar between sexes (35 mmol/mol). Detailed characteristics are summarised in **Table 3.3B**.

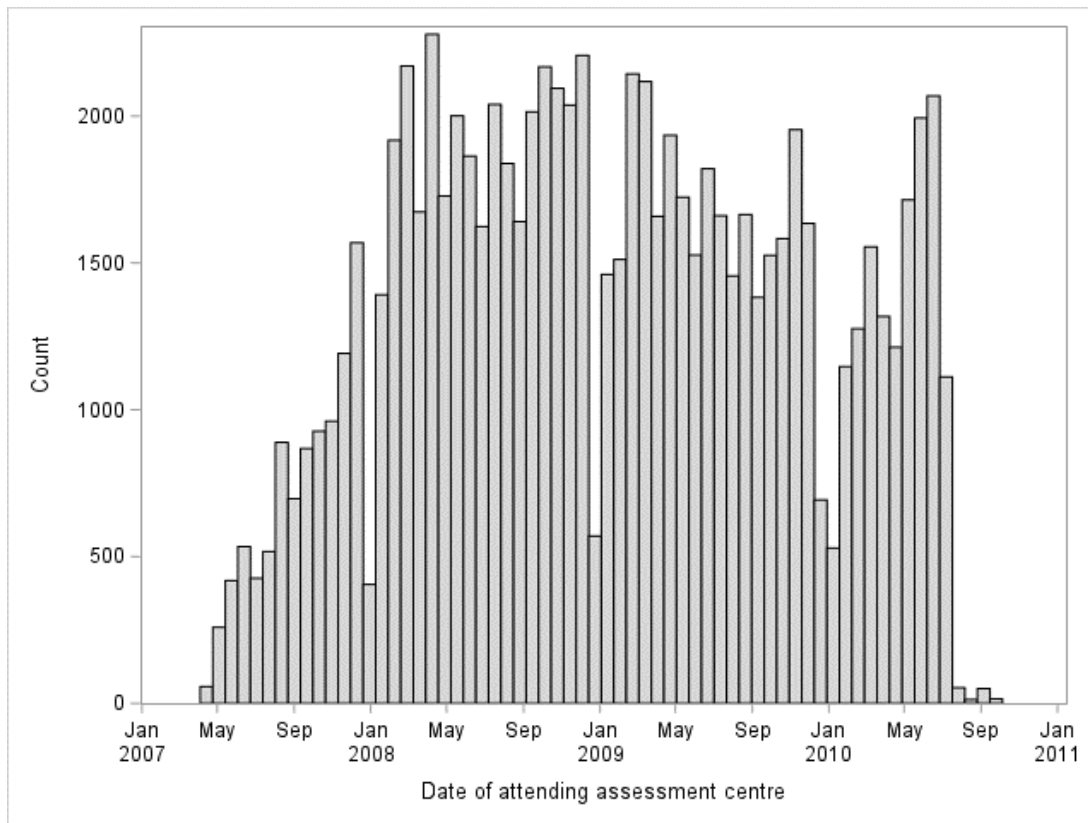
Almost two-thirds (65%) of participants took supplements regularly, among whom 30% took fish oil regularly. The percentages eating at least three servings/day of fresh fruit and vegetables were 36% and 30%, respectively. About 10% of participants had vegetables intake less than one serves per day or were vegetarian. Over half of participants (55%) reported eating oily fish more than once per week while 89% reported eating red meat at least weekly. About half of the participants (47%) eating whole grains one or two servings per day, and 45% of the participants eating cheese for 2-4 times per week. About 2% of participants did not report to consume dairy products. (**Table 3.3C**)

Chapter 3 Tables and Figures

Tables	Title
Table 3.1	NMR-measured metabolites in main analyses
Table 3.2	ICD-10 and operation code of outcomes
Table 3.3	Baseline characteristics of study population by sex

Figures	Title
Figure 3.1	Recruitment pattern of UK Biobank
Figure 3.2	Flowchart of exclusion criteria for study population in prospective analyses

Figure 3.1 Recruitment pattern of UK Biobank



Recruitment pattern represents the monthly recruitment numbers of those participants included in the main analyses i.e. without missing/outlying values or a previous diagnosis of cardiovascular disease at baseline assessment.

Table 3.1 NMR-measured metabolites in main analyses

Biomarker name*	Biomarker abbreviation	Quantification method*	N Missing	N Zero	N Outlier
Total lipids, mmol/L					
Total cholesterol*	Total_C	Derived	0	0	4
VLDL cholesterol*	VLDL_C	Derived	0	0	4
IDL cholesterol	IDL_C	Direct	0	0	4
LDL cholesterol*	LDL_C	Derived	0	0	4
HDL cholesterol*	HDL_C	Derived	0	0	4
Total cholesterol minus HDL-C	non_HDL_C	Derived	0	0	4
Remnant cholesterol	Remnant_C	Derived	0	0	4
Total esterified cholesterol	Total_CE	Derived	0	0	4
Total free cholesterol	Total_FC	Derived	0	0	4
Total phospholipids	Total_PL	Derived	0	0	4
Total triglycerides*	Total_TG	Derived	0	0	4
Lipoprotein particle concentration, mmol/L					
Chylomicrons&extremely large VLDL	XXL_VLDL_P	Direct	0	0	4
Very large VLDL	XL_VLDL_P	Direct	0	0	4
Large VLDL	L_VLDL_P	Direct	0	0	4
Medium VLDL	M_VLDL_P	Direct	0	0	4
Small VLDL	S_VLDL_P	Direct	0	0	4
Very small VLDL	XS_VLDL_P	Direct	0	0	4
Total VLDL	VLDL_P	Derived	0	0	4
IDL	IDL_P	Direct	0	0	4
Large LDL	L_LDL_P	Direct	0	0	4
Medium LDL	M_LDL_P	Direct	0	0	4
Small LDL	S_LDL_P	Direct	0	0	4
Total LDL	LDL_P	Derived	0	0	4
Very large HDL	XL_HDL_P	Direct	0	0	4
Large HDL	L_HDL_P	Direct	0	0	4
Medium HDL	M_HDL_P	Direct	0	0	6
Small HDL	S_HDL_P	Direct	0	0	8
Total HDL	HDL_P	Derived	0	0	8
Mean lipoprotein particle size, nm					
VLDL	VLDL_D	Direct	0	0	22
LDL	LDL_D	Direct	0	0	4
HDL	HDL_D	Direct	0	0	24
Lipoprotein particle composition, average molecules per particle					
Esterified cholesterol per VLDL	VLDL_CE	Derived	0	0	4
Free cholesterol per VLDL	VLDL_FC	Derived	0	0	4
Phospholipids per VLDL	VLDL_PL	Derived	0	0	8
Triglycerides per VLDL	VLDL_TG	Derived	0	0	4
Esterified cholesterol per IDL	IDL_CE	Derived	0	0	8
Free cholesterol per IDL	IDL_FC	Derived	0	0	7
Phospholipids per IDL	IDL_PL	Derived	0	0	7
Triglycerides per IDL	IDL_TG	Derived	0	0	4
Esterified cholesterol per LDL	LDL_CE	Derived	0	1	8
Free cholesterol per LDL	LDL_FC	Derived	0	0	8
Phospholipids per LDL	LDL_PL	Derived	0	0	8
Triglycerides per LDL	LDL_TG	Derived	0	0	4
Esterified cholesterol per HDL	HDL_CE	Derived	0	0	6
Free cholesterol per HDL	HDL_FC	Derived	0	0	4
Triglycerides per HDL	HDL_TG	Derived	0	0	4
Phospholipids per HDL	HDL_PL	Derived	0	0	4
Apolipoproteins, g/L					
Apolipoprotein A-I*	ApoA-1	Direct	0	0	6
Apolipoprotein B*	ApoB	Direct	0	0	4
Apolipoprotein B to A-1 ratio*	ApoB_ApoA-1	Derived	0	0	4

Biomarker name	Biomarker abbreviation	Quantification method*	N Missing	N Zero	N Outlier
Fatty acids (concentration), mmol/L					
Polyunsaturated fatty acids*	PUFA	Direct	82	0	4
Monounsaturated fatty acids*	MUFA	Direct	82	0	4
Saturated fatty acids*	SFA	Direct	82	0	4
Docosahexaenoic acid*	DHA	Direct	82	0	4
Linoleic acid*	LA	Direct	82	0	4
Omega-3 fatty acids*	FAw3	Direct	82	0	4
Omega-6 fatty acids*	FAw6	Direct	82	0	4
Total fatty acids*	TotFA	Direct	82	0	16
Fatty acids ratio, %					
Polyunsaturated fatty acids to total*	PUFA_FA	Derived	82	0	5
Monounsaturated fatty acids to total*	MUFA_FA	Derived	82	0	4
Saturated fatty acids to total*	SFA_FA	Derived	82	0	4
Docosahexaenoic acid to total*	DHA_FA	Derived	82	0	4
Linoleic acid to total*	LA_FA	Derived	82	0	4
Omega-3 fatty acids to total*	FAw3_FA	Derived	82	0	4
Omega-6 fatty acids to total*	FAw6_FA	Derived	82	0	5
Polyunsaturated to monounsaturated fatty acids*	PUFA_MUFA	Derived	82	0	4
Omega-6 to omega-3 fatty acids*	FAw6_FAw3	Derived	82	0	4
Cholines, mmol/L					
Total cholines	TotCho	Direct	82	0	4
Phosphatidylcholine	PC	Direct	82	0	4
Sphingomyelins	SM	Direct	82	0	4
Phosphoglycerides	Phosphoglyc	Direct	82	0	4
Amino acids, mmol/L					
Alanine*	Ala	Direct	4	0	4
Glutamine	Gln	Direct	350	0	8
Glycine*	Gly	Direct	106	676	4
Histidine*	His	Direct	186	0	5
Isoleucine*	Ile	Direct	3	0	4
Leucine*	Leu	Direct	4	0	4
Valine*	Val	Direct	31	0	4
Branched-chain amino acids*	BCAA	Derived	32	0	4
Phenylalanine*	Phe	Direct	55	0	4
Tyrosine*	Tyr	Direct	168	0	5
Glycolysis related metabolites, mmol/L					
Lactate*	Lac	Direct	280	0	4
Citrate	Cit	Direct	16	0	4
Glucose*	Glc	Direct	210	0	4
Pyruvate	Pyruvate	Direct	351	0	4
Ketone bodies, mmol/L					
Acetate	Ace	Direct	36	523	4
Aceto acetate	AcAce	Direct	3	1235	4
Acetone	Acetone	Direct	3	0	4
Beta-hydroxybutyrate	bOHBut	Direct	1521	465	4
Fluid balance					
Albumin*, g/L	Alb	Direct	16	0	18
Creatinine*, mmol/L	Crea_nmr	Direct	5252	0	4
Inflammation, mmol/L					
Glycoprotein acetyls*	Gp	Direct	0	0	5

*The clinical-validated metabolites used for risk prediction. *Direct means the measure was quantified directly from the NMR platform; Derived means the measure was generated from direct measures.

Table 3.2 ICD-10 and operation codes of outcomes

ICD/OPCS category	Disease category	Code definition
I20	Angina pectoris	I20.0* Unstable angina I20.1 Angina pectoris with documented spasm I20.8 Other forms of angina pectoris I20.9 Angina pectoris, unspecified angina
I21	Acute myocardial infarction	I21.0* Acute transmural myocardial infarction of anterior wall I21.1* Acute transmural myocardial infarction of inferior wall I21.2* Acute transmural myocardial infarction of other sites I21.3* Acute transmural myocardial infarction of unspecified site I21.4* Acute subendocardial myocardial infarction I21.9* Acute myocardial infarction, unspecified
I22	Subsequent myocardial infarction	I22.0* Subsequent myocardial infarction of anterior wall I22.1* Subsequent myocardial infarction of inferior wall I22.8* Subsequent myocardial infarction of other sites I22.9* Subsequent myocardial infarction of unspecified site
I23	Certain current complications following acute myocardial infarction	I23.0* Haemopericardium as current complication following acute myocardial infarction; I23.1* Atrial septal defect as current complication following acute myocardial infarction; I23.2* Ventricular septal defect as current complication following acute myocardial infarction; I23.3* Rupture of cardiac wall without haemopericardium as current complication following acute myocardial infarction; I23.4* Rupture of chordae tendineae as current complication following acute myocardial infarction I23.5* Rupture of papillary muscle as current complication following acute myocardial infarction I23.6* Thrombosis of atrium, auricular appendage, and ventricle as current complications following acute myocardial infarction; I23.8* Other current complications following acute myocardial infarction
I24	Other acute ischaemic heart diseases	I24.0 Coronary thrombosis not resulting in myocardial infarction I24.1* Dressler's syndrome I24.8* Other forms of acute ischaemic heart disease I24.9* Acute ischaemic heart disease, unspecified (excl. ischaemic heart disease (chronic) NOS)
I25	Chronic ischaemic heart disease	I25.0 Atherosclerotic cardiovascular disease, so described I25.1* Atherosclerotic heart disease I25.2* Old myocardial infarction I25.3 Aneurysm of heart I25.4 Coronary artery aneurysm I25.5* Ischaemic cardiomyopathy I25.6* Silent myocardial ischaemia I25.8* Other forms of chronic ischaemic heart disease - Any condition in I21-I22 and I24.- specified as chronic I25.9* Chronic ischaemic heart disease, unspecified -Ischaemic heart disease (chronic) NOS
I63	Cerebral infarction	I63.0 Cerebral infarction due to thrombosis of precerebral arteries I63.1 Cerebral infarction due to embolism of precerebral arteries I63.2 Cerebral infarction due to unspecified occlusion or stenosis of precerebral arteries I63.3 Cerebral infarction due to thrombosis of cerebral arteries I63.4 Cerebral infarction due to embolism of cerebral arteries I63.5 Cerebral infarction due to unspecified occlusion or stenosis of cerebral arteries I63.6 Cerebral infarction due to cerebral venous thrombosis, nonpyogenic I63.8 Other cerebral infarction I63.9 Cerebral infarction, unspecified
I64	Stroke	Stroke, not specified as haemorrhage or infarction

ICD/OPCS category	Disease category	Code definition
G45	Transient cerebral ischaemic attacks	G45.0 Vertebro-basilar artery syndrome G45.1 Carotid artery syndrome (hemispheric) G45.2 Multiple and bilateral precerebral artery syndromes G45.3 Amaurosis fugax G45.4 Transient global amnesia G45.8 Other transient cerebral ischaemic attacks and related syndromes G45.9 Transient cerebral ischaemic attack, unspecified
410	Acute myocardial infarction	410.1 Acute myocardial infarction of other anterior wall, episode of care unspecified 410.2 Acute myocardial infarction of inferolateral wall, initial episode of care 410.3 Acute myocardial infarction of inferoposterior wall, episode of care unspecified 410.4 Acute myocardial infarction of other inferior wall, initial episode of care 410.6 True posterior wall infarction, initial episode of care 410.9 Acute myocardial infarction
411	Other acute and subacute forms of ischaemic heart disease	411.1 Intermediate coronary syndrome 411.8 Other acute and subacute forms of ischemic heart disease 411.9 Other acute and subacute forms of ischaemic heart disease
412	Old myocardial infarction	412.9 Old myocardial infarction
413	Angina pectoris	413.0 Angina decubitus 413.1 Prinzmetal angina 413.9 Angina pectoris
414	Other forms of chronic ischaemic heart disease	414.0 Coronary atherosclerosis 414.1 Aneurysm of heart 414.8 Other specified forms of chronic ischaemic heart disease 414.9 Chronic ischaemic heart disease, unspecified
434	Occlusion of cerebral arteries	434.9 Occlusion of cerebral arteries, unspecified
436	Acute but ill-defined cerebrovascular disease	436.9 Acute but ill-defined cerebrovascular disease
K40*		Saphenous vein graft replacement of coronary artery
K41*		Other autograft replacement of coronary artery
K42*		Allograft replacement of coronary artery
K43*		Prosthetic replacement of coronary artery
K44*		Other replacement of coronary artery
K45*		Connection of thoracic artery to coronary artery
K46*		Other bypass of coronary artery
K47*		Repair of coronary artery
K49*		Transluminal balloon angioplasty of coronary artery
K50*		Other therapeutic transluminal operations on coronary artery
K75*		Percutaneous transluminal balloon angioplasty and insertion of stent into coronary artery

All listed codes were used to identify cardiovascular disease at baseline, and the bold codes marked with * were used to identify incident CHD (ICD-9 codes were not used to identify incident disease in this dataset because they were not used in clinical practice after 2015).

Figure 3.2 Flowchart of exclusion criteria for study population in prospective analysis

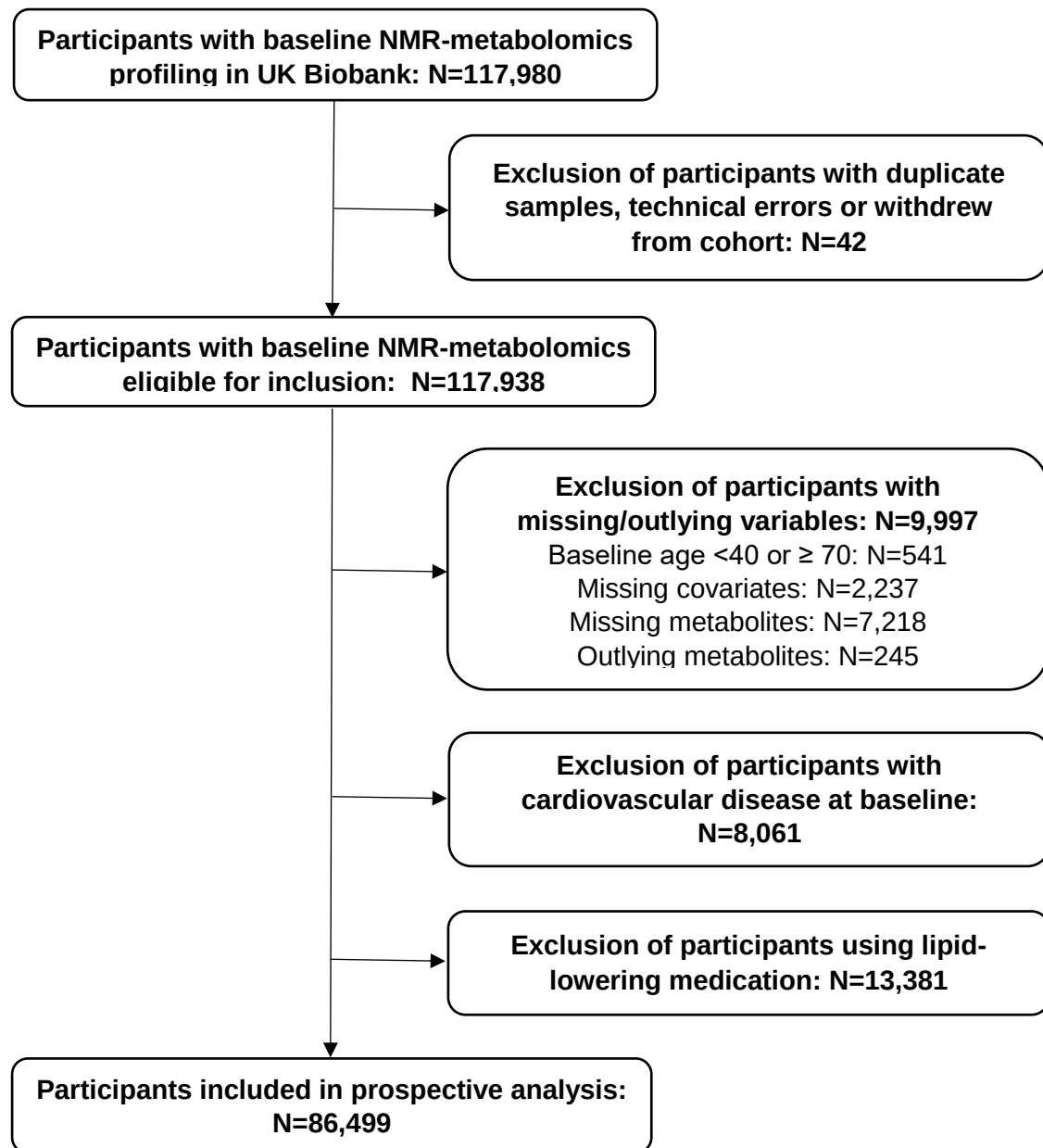


Table 3.3 Baseline characteristics of study population by sex**A. Social-demographic and lifestyle characteristics**

	Women	Men	All
No. of participants	49,489	37,010	86,499
White, %	95.2	95.1	95.2
Baseline age groups, %			
40-44	11.7	13.3	12.4
45-49	15.1	16.7	15.4
50-54	17.3	16.3	16.9
55-59	19.1	18.1	18.7
60-64	22.5	21.7	22.2
65-69	14.2	14.9	14.5
Mean, years	55.4 (7.9)	55.2 (8.2)	55.3 (8.0)
Education, %			
Higher education	39.2	42.3	40.5
AS/A-level	26.9	29.3	27.9
CSE/GCSE/O-Level	19.2	14.1	17.0
No qualification	14.6	14.4	14.5
Assessment region, %			
London	13.3	12.3	12.9
Wales	3.9	3.9	3.9
Northwest	16.8	17.3	17.0
Northeast	10.6	10.7	10.7
Yorkshire & Humber	16.4	16.2	16.3
West Midlands	7.7	9.2	8.3
East Midlands	6.7	6.5	6.6
Southeast	8.7	8.8	8.8
Southwest	8.5	8.2	8.4
Scotland	7.4	6.8	7.1
Townsend Deprivation Index*	-1.6 (2.9)	-1.5 (2.9)	-1.6 (2.9)
Frequency of alcohol intake, %†			
Never	5.0	2.5	3.9
Previous	3.3	3.0	3.2
Current occasional	27.1	18.9	22.3
Current regular	64.6	78.6	70.6
Frequency of smoking, %‡			
Never	60.5	52.9	57.2
Previous	30.8	35.0	32.6
Current occasional	2.2	3.5	2.7
Current regular	6.6	8.6	7.5

Continuous variables are presented as mean (standard deviation), and categorical variables are presented as column percentages. *Area-level measure of material deprivation; †Regular alcohol intake means drinks alcohol once or more than week; ‡Regular smoking means smoking on most or all days.

B. Physiological factors and medical history

	Women	Men	All
No. of participants	49,489	37,010	86,499
Anthropometry			
Body mass index, kg/m ²	26.7 (4.9)	27.4 (4.0)	27.0 (4.6)
Waist Circumference, cm	83.5 (11.9)	95.5 (10.8)	88.6 (12.9)
Waist to hip ratio	0.8 (0.1)	0.9 (0.1)	0.9 (0.1)
Systolic blood pressure, mmHg	134.0 (18.9)	140.3 (17.3)	136.7 (18.5)
Clinical biochemistry Biomarkers			
Total cholesterol, mmol/l (missing=3,837)	6.0 (1.0)	5.8 (1.0)	5.9 (1.1)
LDL cholesterol, mmol/l (missing=3,958)	3.7 (0.8)	3.7 (0.8)	3.7 (0.8)
HDL cholesterol, mmol/l (missing=10,783)	1.6 (0.4)	1.3 (0.3)	1.5 (0.4)
Total triglycerides, mmol/l (missing=3,896)	1.5 (0.8)	1.9 (1.1)	1.7 (1.0)
Apolipoprotein A-I, g/l (missing=11,217)	1.6 (0.3)	1.4 (0.2)	1.5 (0.3)
Apolipoprotein B, g/l (missing=4,240)	1.0 (0.2)	1.1 (0.2)	1.1 (0.2)
Creatinine, umol/L (missing=3,885)	64.1 (11.0)	80.9 (13.4)	71.3 (14.7)
Albumin, g/L (missing=10,756)	44.9 (2.6)	45.6 (2.6)	45.2 (2.6)
C-reactive protein, mg/L (missing=3,988)	2.6 (4.2)	2.4 (4.3)	2.5 (4.2)
ALP, U/L (missing=3,831)	83.6 (26.3)	81.4 (23.7)	82.7 (25.3)
GGT, U/L (missing=3,860)	28.5 (29.4)	42.8 (41.0)	34.6 (35.6)
HbA1c, mmol/mol (missing=3,937)	35.0 (4.6)	35.0 (5.3)	35.0 (4.9)
Fasting time, h	3.0 (2.0)	3.0 (2.0)	3.0 (2.0)
Prevalence of diagnosed, treatment and medications, %			
Baseline diabetes	1.7	2.5	2.0
Hypertension treatment	12.3	12.6	12.4
Atypical antipsychotic medication	0.2	0.3	0.2
Regular steroid tablets	0.8	0.8	0.8

Continuous variables are presented as mean (standard deviation) (except for the skewed variable fasting time, presented as median [quartile range]), and categorical variables are presented as column percentages. ALP= Alkaline phosphatase; GGT= Gamma glutamyltransferase; HbA1c=Glycated haemoglobin.

C. Supplements and dietary habits

	Women	Men	All
No. of participants	49,489	37,010	86,499
Supplements, %			
Vitamin supplement (missing=316)	35.9	26.4	31.8
Mineral supplement (missing=147)	26.4	18.6	23.1
Fish oil (missing=147)	31.9	28.5	30.5
Glucosamine (missing=147)	8.4	5.4	7.1
No supplement (missing=204)	39.5	28.8	34.9
Fresh fruit (missing=0)			
< 1 serving/day*	6.8	13.0	9.4
1-<2 servings/day	22.9	31.1	26.4
2-<3 servings/day	29.5	26.6	28.3
≥ 3 servings/day	40.8	29.3	35.9
Vegetables (missing=0)			
<1 servings/day†	8.0	13.8	10.5
1-<2 servings/day	23.9	28.8	26.0
2-<3 servings/day	34.8	31.1	33.2
≥ 3 servings/day	33.3	26.3	30.3
Red meat (missing=268)			
0 times/week	9.0	4.9	7.3
1 times/week	4.1	2.7	3.5
> 1 times/week	86.7	92.1	89.0
Oily fish (missing=467)			
0 times/week	10.7	11.7	11.1
1 times/week	32.4	35.9	33.9
> 1 times/week	56.5	51.8	54.5
Whole grains (missing=218)			
0 servings/day*	16.0	24.5	19.7
<1 servings/day	19.1	16.9	18.2
1-<3 servings/day	53.0	38.6	46.8
≥ 3 servings/day	11.7	19.8	15.1
Cheese (missing=1880)			
<1 times/week	21.1	15.6	18.8
1 times/week	21.1	19.6	20.4
2<5 times/week	42.5	47.8	44.8
≥ 5 times/week	13.1	15.0	13.9
No dairy products, %	2.2	1.9	2.0

Categorical variables are presented as column percentages. *One prune, or one dried apricot, or 10 raisins as one serving; †Two heaped tablespoons of vegetables of one serving; *One slice of bread, or one bowl of cereal as one serving.

CHAPTER 4: Baseline description of metabolites

Chapter summary

NMR spectroscopy was used for high-throughput metabolic profiling of the baseline plasma samples of the study population, and 92 metabolic biomarkers were included in the main analyses. This chapter gives an overall baseline description of these metabolites, before subsequent chapters examine their prospective associations and predictive value for CHD. The distributions of most metabolites were approximately normal, except for a few such as ketone bodies, phenylalanine, and creatinine, which were skewed to the right. The average concentrations of most metabolites showed differences (in either direction) by sex, and most were getting gradually higher or showing no difference by age. The cross-correlations of all the selected metabolites were estimated using Spearman partial correlation coefficients: high correlations were observed between the concentrations of lipids, lipoproteins, apolipoproteins, and some fatty acids, and lower correlations between low-molecular-weight metabolites with other metabolites. Overall, NMR-measured metabolites had high consistency with the same metabolites measured using clinical chemistry method.

4.1 Background

Describing the distribution of the exposure provides an overall understanding of the circulating metabolites. Previous studies standardised the biomarkers to facilitate comparisons between metabolites with different distributions and measured from different platforms.^{52,58} Some studies additionally log-transformed the metabolite before standardisation to improve normality of the metabolites⁵². However, applying log-transform will decrease the interpretability of the associations (e.g., interpreting as ‘change of risk with every standard deviation higher concentration of the metabolite’ is easier to understand than ‘change of risk with every standard deviation higher concentration of the log-metabolite’). Moreover, because of large variations in the distributions of metabolites, log-transformation may not improve normality for all the metabolites. For the above reasons, it is necessary to check the distribution of every metabolite, and find the most appropriate way to transform each biomarker before the association analyses.

NMR-derived metabolic profiling includes a large proportion of lipids and lipoproteins-related metabolites, and these biomarkers in the same categories may have high cross-correlations. On the other hand, because of the complex network of metabolites, some metabolites can be generated from, or be the composition of, other metabolites (e.g. fatty acids are the main constitutional component of lipids; the dynamic transport of cholesterol and triglycerides in lipoproteins lead to the transfer from VLDL to intermediate-density lipoproteins [IDL] to LDL). However, when examining metabolites as exposures, their high correlations may lead to collinearity problems and reduce the interpretability of the models. Therefore, it is necessary to check cross-correlations of the

metabolites before further analyses, in order to better inform model adjustments and interpretations of effect sizes.

Biochemistry markers have been widely used in clinical practice for disease prediction and diagnosis. Compared with traditional clinical chemistry, NMR spectroscopy enables measurement of more low-molecular-weight metabolites, and has the strength of separating lipoprotein classes and providing detailed measures of characteristics.³⁹ Conversely, some biomarkers were more accurately measured by NMR spectroscopy. For example, routinely chemistry-measured LDL-C includes the cholesterol carried in IDL and some VLDL particles, while high-throughput techniques, such as NMR, can achieve size-specific quantification of lipoproteins and accurately quantify the cholesterol level in LDL particles.¹⁰⁸ However, to better interpret the results from NMR-derived metabolites, it is necessary to compare the consistency between the biomarkers measured from both chemistry and NMR spectroscopy. The FINRISK study reported high consistency between the routine lipids measured by clinical chemistry assays and NMR spectroscopy.¹⁰⁹ Except for lipids biomarkers, UK Biobank also included biomarkers of apolipoproteins and fluid balance that were measured by both chemistry assays and NMR spectroscopy. It should be noted, however, that the NMR-based metabolites profiling platform in the UK Biobank used EDTA plasma samples from aliquot 3, which was reported to be more affected by dilution compared with aliquot 1 that used for chemistry measurements.¹⁰⁶

4.2 Methods

The mean, standard deviation, median and interquartile range (IQR) were calculated for each metabolite, and the distribution of each metabolite was

described using histograms. The histograms of the metabolites after log-transformation were also plotted as comparison. Means of the biomarker by each age and sex group are also presented. Because of the skewed distributions of some metabolites, the cross-correlations of all the metabolites were estimated using Spearman partial correlation, adjusting for age and sex, and were plotted in heat maps.

The mean, standard deviation, median and interquartile range were calculated for the lipids, apolipoproteins, and creatinine biomarkers that were measured by both clinical chemistry and NMR spectroscopy. The cross-correlation of each pair of chemistry-measured and NMR-measured metabolite was estimated using Spearman partial correlation coefficients, adjusted for age and sex.

4.3 Distribution of metabolites

The means, standard deviations, medians, and IQR of the selected NMR metabolites are given in **Table 4.1**. In terms of lipoproteins, VLDL had the lowest concentration (mean $1.5 [SD\ 0.4] \times 10^{-4}$ mmol/L), while the concentration was about eight times higher for LDL ($1.2 [0.3] \times 10^{-3}$ mmol/L), and 100 times higher for HDL ($1.5 [0.2] \times 10^{-2}$ mmol/L). The concentration of ApoA-1 and ApoB were 1.4 (0.2) and 0.9 (0.2) g/L, respectively. The particle size of lipoproteins decreased from VLDL (38.5 [1.2] nm) to HDL (9.7 [0.2] nm). The average molecules of lipid per lipoprotein particle decreased as the particle became smaller. Triglycerides had largest proportion in each VLDL particle ($5.9 [0.2] \times 10^3$ molecules/particle) compared with other lipid contents, and esterified cholesterol was highest in IDL and LDL ($2.0 [0.2]$ and $1.1 [0.1] \times 10^3$ molecules/particle, respectively). Esterified cholesterol and phospholipids were higher (68 [9] and 102 [9] molecules/particle) in HDL. The total concentration of

cholesterol, triglycerides, fatty acids and cholines were 4.7 (0.9), 1.3 (0.6), 11.9 (2.3) and 2.6 (0.4) mmol/L, respectively. Amino acids and ketone bodies had much lower concentrations, both of which were lower than 1.0 mmol/L.

The histogram of each metabolite is shown in **S4.1**. The distributions of most metabolites were approximately normal, except for a few such as ketone bodies, phenylalanine, and creatinine, which were skewed to the right. However, for other metabolites, log-transformation did not make too much of a difference compared with the original distribution, and for some metabolites, such as, glycine, average esterified cholesterol and free cholesterol per LDL particle, the skewness shifted to the left after log-transformation (**S4.2**).

S4.3 shows the means of biomarkers by age and sex. The average concentrations of most metabolites showed differences by sex. For VLDL particle concentrations, ApoB to ApoA-1 ratio, VLDL particle size, average triglycerides levels in each lipoprotein particles, saturated fatty acids (SFA) and MUFA to total fatty acids, most amino acids (except for glycine and tyrosine), lactate and creatinine, the average concentrations/ratios were consistently higher in men than women across all age groups. While the average levels in all age groups were higher in women than men for HDL particle concentrations, ApoA-1, LDL and HDL particle size, cholesterol composition in IDL, LDL, and HDL, all cholines, polyunsaturated fatty acids (PUFA) to total fatty acids, glycine, citrate, and pyruvate. The average concentrations of ketone bodies and glucose were similar between men and women. Meanwhile, most of biomarkers were getting gradually higher or showing no difference by age, and only a few biomarkers, such as polyunsaturated fatty acid to total fatty acids ratio in women, histidine, and albumin in both men and women, were getting gradually

lower with age. The changing patterns by age were different between sexes (interactions by sex) for some of the biomarkers, including particle concentration of VLDL, IDL and LDL, concentration of ApoB, non-HDL cholesterol, total triglycerides and total fatty acids. The level of these metabolites were higher in older women, while they showed no difference or were in slightly lower level in elder men.

4.4 Cross-correlations between metabolites

Figure 4.1 describes the Spearman partial correlations between each pair of the metabolites. The cross-correlations were not high (Spearman partial correlation coefficient, $\rho < 0.5$) between low-molecular-weight metabolites and other classes of metabolites, and between low-molecular-weight metabolites in different subclasses.

Very high correlations ($\rho > 0.9$) were observed between Apolipoproteins and lipoprotein concentrations: the Spearman's correlation coefficient was 0.99 between ApoB and LDL concentration, and 0.96 between ApoA-1 and HDL concentration. ApoB concentration was relatively weaker but still highly correlated with IDL and VLDL (0.90 and 0.87, respectively). Very high correlations were also observed among total cholesterol and LDL cholesterol (0.93), and among composition of free cholesterol and phospholipids in VLDL (0.92) (**Figure 4.2**).

Table 4.2 describes the correlations between lipoproteins and apolipoproteins with lipid-related biomarkers. Baseline level of non-HDL-C had very high correlations with concentrations of LDL particle and ApoB (both 0.98), and a relative weaker correlation with VLDL particle (0.82). HDL-C level was most

highly correlated with ApoA-1 (0.93), and was also correlated with HDL particle (0.85) and HDL diameter (0.80). Except for a moderate correlation between average levels of esterified cholesterol in LDL and HDL-C, the partial correlation coefficients of other lipoprotein compositions with non-HDL-C and HDL-C were lower than 0.7. Total triglycerides was highly and positively correlated with both VLDL particle concentration (0.85) and VLDL particle size (0.86), as well as average levels of phospholipids and triglycerides in VLDL (0.87 and 0.81, respectively), and the average level of triglycerides in HDL (0.84). Average levels of esterified cholesterol in HDL was negatively moderately associated with total triglycerides (-0.71).

In terms of fatty acids, the concentration of SFA was highly correlated with the concentration of MUFA (0.93). The concentration of omega-6 PUFA was moderately correlated with the concentration of SFA (0.75) and MUFA (0.73), while omega-3 PUFA had lower correlations with the other types of fatty acid (0.38 to 0.47) (**Table 4.3**). When comparing the correlations between fatty acids and lipids, both SFA and MUFA had high correlations with total triglycerides (0.87 and 0.93, respectively), and omega-6 PUFA had a high correlation with non-HDL-C (0.82). The correlations between omega-3 PUFA and all lipid biomarkers were low (0.16 to 0.40) (**Table 4.4**).

4.5 Correlations with chemistry biomarkers

The means, standard deviations of the lipids, apolipoproteins, and fluid balance biomarkers that were measured by both clinical chemistry and NMR spectroscopy, as well as their Spearman partial correlation coefficients are shown in **Table 4.5**. In terms of lipids and apolipoproteins, there was high

consistency between each pair of chemistry-measured and NMR-measured metabolites, with partial correlations in the range of 0.84 to 0.91. For fluid balance biomarkers, consistency was lower but still remained for creatinine (0.79). The lowest correlation chemistry-measured and NMR-measured biomarker was for albumin (0.56).

4.6 Discussion

This chapter provides an overall description of baseline NMR-derived metabolites, and compares them with the same metabolites measured by clinical chemistry. The distribution of most of metabolites was approximately normal, indicating no need to perform log-transformations prior to analysis. Only a few metabolites, including ketone bodies, phenylalanine, and creatinine were right-skewed, and became approximately normal after log-transformation. Therefore, the associations using log-transformed metabolites will be examined in the sensitivity analyses. Conversely, the changing patterns by age and sex for most metabolites indicated that Cox proportional-hazards models of the association analyses should be stratified by sex and age-at-risk.

LDL and ApoB concentrations had extremely high correlations with LDL-C and non-HDL-C, and ApoA-1 was highly correlated with HDL-C. However, the correlations of routinely measured lipids with VLDL particle concentration, as well as with particle size of each lipoprotein were relatively lower, and the correlations with lipoprotein compositions per particle were also moderate. Such correlation patterns indicated that lipoprotein characteristics may provide additional information that has been overlooked when only measuring traditional lipid biomarkers, and therefore highlighted the value of examining the associations of lipoproteins and fatty acids to cardiovascular risk independent

of lipids. Moreover, the low correlations between low-molecular-weight metabolites and lipids also indicated the potential predictive values from these metabolites that could not be captured when using the clinical chemistry method.

High consistency was observed between lipids, lipoproteins and creatinine biomarkers measured using both NMR and clinical chemistry methods. The relatively low correlation between chemistry-measured and NMR-measured albumin was probably due to two reasons. Firstly, NMR-derived metabolic biomarkers were measured using EDTA plasma samples, while clinical biochemistry markers were measured from serum samples. Previous research has found that the stability of albumin were different in serum and plasma samples under different storing conditions,¹¹⁰ and the results in Chapter 5 also showed much greater within-person variability of plasma albumin by NMR than serum albumin by the chemistry method; secondly, technical issues were another potential reason. Blood sample aliquot 3 in the UK Biobank, which was used for NMR spectroscopy measurement, was reported to be more affected by dilution compared with aliquot 1 that was used for chemistry measurements. According to the assay quality report from the UK Biobank, assays with a naturally wide biological range are typically far less importantly affected from an epidemiological perspective by a given dilution, but albumin had a narrow biological range, which was the only circulating metabolites mentioned in this report, and hence more strongly affected by the dilution compared with other biomarkers.¹⁰⁵ Furthermore, a recent paper evaluated the technical variation of NMR-derived metabolic biomarker in the UK Biobank, and found that the source of technical variation of albumin was not due to the NMR quantification pipeline,

but rather arose during the UK Biobank sample plating process.¹¹¹ A recent study based on China Kadoorie Biobank showed high correlations for plasma creatinine ($\rho=0.98$) and plasma albumin ($\rho=0.80$) between measured by clinical chemistry and NMR spectroscopy, also indicating that the low correlation for albumin in the study was probably due to the difference in sample type or technical variation in the UK Biobank rather than the measurement technique.⁵⁸

4.7 Conclusion

The distribution of most of metabolites was approximately normal. Lipids had extremely high correlations with most lipoproteins and apolipoproteins concentrations, high but relatively lower correlations with VLDL concentration and other characteristics of lipoproteins, moderate correlations with most fatty acids, and relatively lower correlations with low-molecular-weight metabolites. Overall, NMR-measured metabolites had high consistency with the same metabolites measured using clinical chemistry.

Chapter 4 Tables and Figures

Tables	Title
Table 4.1	Baseline distribution of NMR-derived metabolites
Table 4.2	Correlations of baseline lipoproteins characteristics and lipids
Table 4.3	Cross-correlations of baseline fatty acids concentration
Table 4.4	Correlations of baseline fatty acids concentration and lipids
Table 4.5	Comparison of metabolites measured by clinical chemistry and NMR spectroscopy

Figures	Title
Figure 4.1	Cross-correlations of NMR-derived metabolites at baseline
Figure 4.2	Cross-correlations of baseline lipoproteins characteristics

Table 4.1 Baseline distribution of NMR-derived metabolites**A. Concentration of lipoproteins and apolipoproteins**

	Mean	SD	Median	IQR
Lipoproteins (*10⁻⁵), mmol/L				
VLDL	14.7	4.4	14.3	5.9
Chylomicrons & extremely large	0.2	0.1	0.1	0.2
Very large	0.3	0.2	0.3	0.3
Large	1.0	0.5	0.9	0.8
Medium	3.6	1.2	3.5	1.6
Small	4.0	1.3	3.8	1.7
Very small	5.7	1.3	5.5	1.7
IDL	32.3	7.1	31.7	9.5
LDL	123.1	26.8	120.8	35.1
Large	75.5	16.6	73.9	21.9
Medium	30.2	7.3	29.7	9.6
Small	17.4	3.6	17.1	4.7
HDL	1511.7	234.6	1495.3	304.3
Very large	24.3	9.3	22.1	10.8
Large	145.1	76.3	128.6	101.2
Medium	383.6	90.7	376.2	120.3
Small	958.7	126.1	952.3	161.5
Apolipoproteins				
Apolipoprotein A-I, g/L	1.4	0.2	1.4	0.3
Apolipoprotein B, g/L	0.9	0.2	0.9	0.2
Apolipoprotein B to A-1 ratio	0.6	0.2	0.6	0.2

VLDL=very-low-density lipoproteins; IDL=intermediate-density lipoproteins; LDL=low-density lipoproteins; HDL=high-density lipoproteins.

B. Size and composition of lipoproteins

	Mean	SD	Median	IQR
Lipoprotein diameter, nm*				
VLDL	38.5	1.2	38.5	1.8
LDL	23.9	0.1	23.9	0.1
HDL	9.7	0.2	9.6	0.3
Lipoprotein composition (*10²), molecules/particle[†]				
VLDL				
Esterified cholesterol	30.3	2.5	30.6	3.2
Free cholesterol	19.9	1.6	20.0	2.2
Phospholipids	31.9	3.4	31.9	4.7
Triglycerides	58.8	1.8	56.2	2.6
IDL				
Esterified cholesterol	20.0	1.8	20.0	2.6
Free cholesterol	7.1	0.7	7.1	1.0
Phospholipids	9.5	0.8	9.4	1.2
Triglycerides	3.1	0.8	3.0	0.9
LDL				
Esterified cholesterol	10.6	0.7	10.5	0.8
Free cholesterol	3.9	0.4	3.9	0.6
Phospholipids	5.1	0.3	5.1	0.4
Triglycerides	1.2	0.3	1.1	0.3
HDL				
Esterified cholesterol	0.7	0.1	0.7	0.1
Free cholesterol	0.2	<0.1	0.2	<0.1
Phospholipids	1.0	0.1	1.0	0.1
Triglycerides	0.1	<0.1	0.1	<0.1

*Diameter of IDL particles not given in the UK Biobank. [†]Lipoprotein composition derived by dividing the lipidic contents by the particle concentration. VLDL=very-low-density lipoproteins; IDL=intermediate-density lipoproteins; LDL=low-density lipoproteins; HDL=high-density lipoproteins.

C. Total Lipids and fatty acids

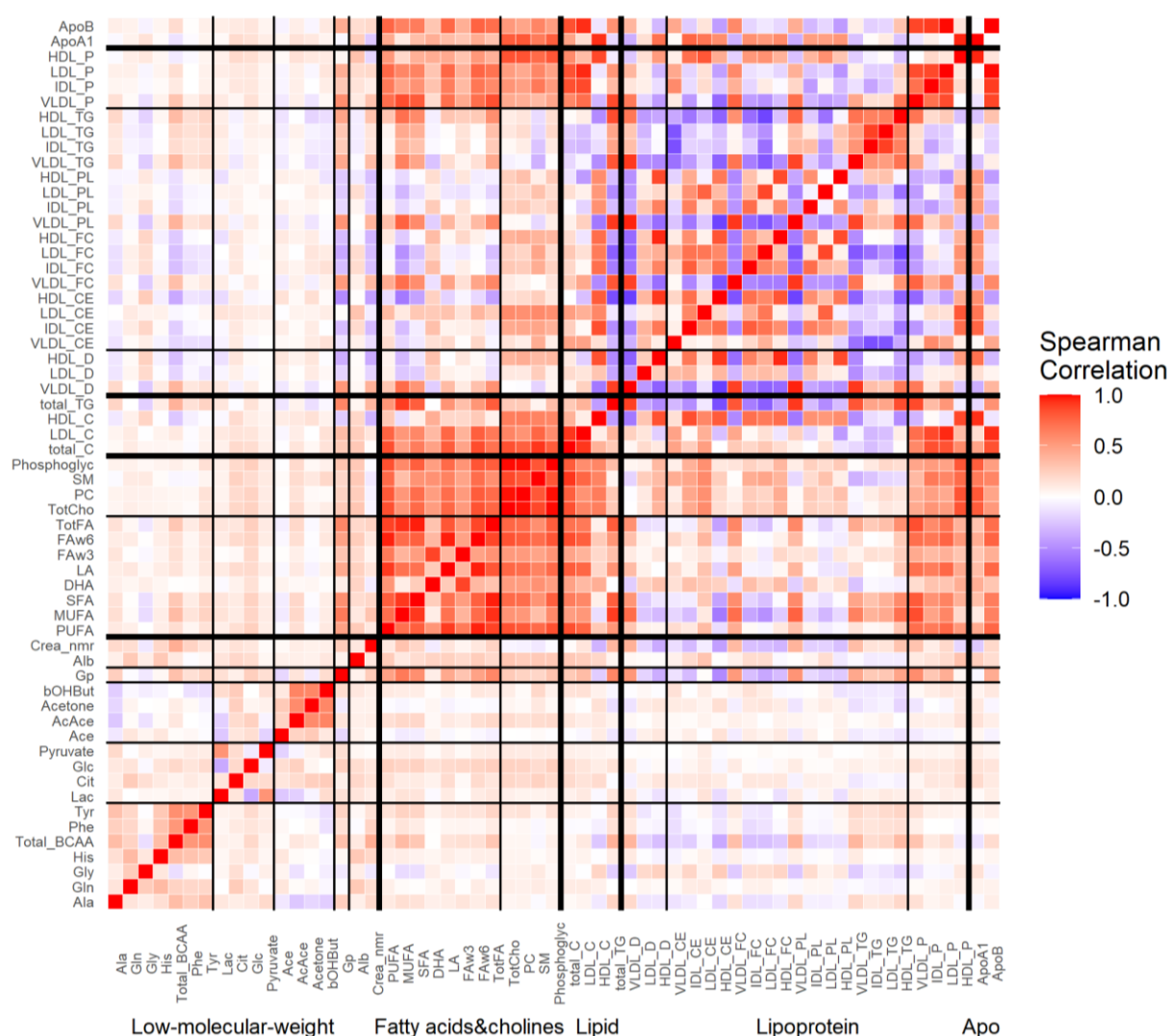
	Mean	SD	Median	IQR
Total plasma lipids, mmol/L				
Total Cholesterol	4.7	0.9	4.7	1.1
VLDL Cholesterol	0.7	0.2	0.7	0.3
IDL Cholesterol	0.8	0.2	0.9	0.3
LDL Cholesterol	1.8	0.4	1.7	0.5
HDL Cholesterol	1.3	0.3	1.3	0.4
Non-HDL Cholesterol*	3.4	0.8	3.3	1.0
Remnant Cholesterol	1.6	0.4	1.6	0.5
Total Esterified Cholesterol	3.4	0.6	3.4	0.8
Total Free Cholesterol	1.3	0.2	1.3	0.3
Total Phospholipids	2.9	0.4	2.9	0.6
Total Triglycerides	1.3	0.6	1.2	0.7
Fatty acids concentration, mmol/L				
Polyunsaturated fatty acids	5.1	0.8	5.0	1.0
Monounsaturated fatty acids	2.8	0.8	2.7	1.0
Saturated fatty acids	4.1	0.9	3.9	1.1
Docosahexaenoic acid	0.2	0.1	0.2	0.1
Linoleic acid	3.5	0.7	3.5	0.8
Omega-3 fatty acids	0.5	0.2	0.5	0.3
Omega-6 fatty acids	4.5	0.7	4.5	0.8
Total fatty acids	11.9	2.3	11.7	3.0
Fatty acids ratio to total*				
Polyunsaturated fatty acids	42.7	3.6	43.2	4.6
Monounsaturated fatty acids	23.3	2.6	23.0	3.4
Saturated fatty acids	34.0	1.9	33.9	2.3
Docosahexaenoic acid	2.0	0.7	1.9	0.8
Linoleic acid	29.6	3.2	29.8	4.1
Omega-3 fatty acids	4.3	1.5	4.1	1.8
Omega-6 fatty acids	38.4	3.5	38.9	4.4
Other fatty acids ratio				
Omega-6 to Omega-3 fatty acids	10.1	4.4	9.3	4.4
Poly- to mono-unsaturated fatty acids	1.9	0.3	1.9	0.5

*Total cholesterol minus HDL cholesterol; *Ratio of concentration of given type of fatty acid to concentration of total fatty acid.

D. Other low-molecular-weight metabolites

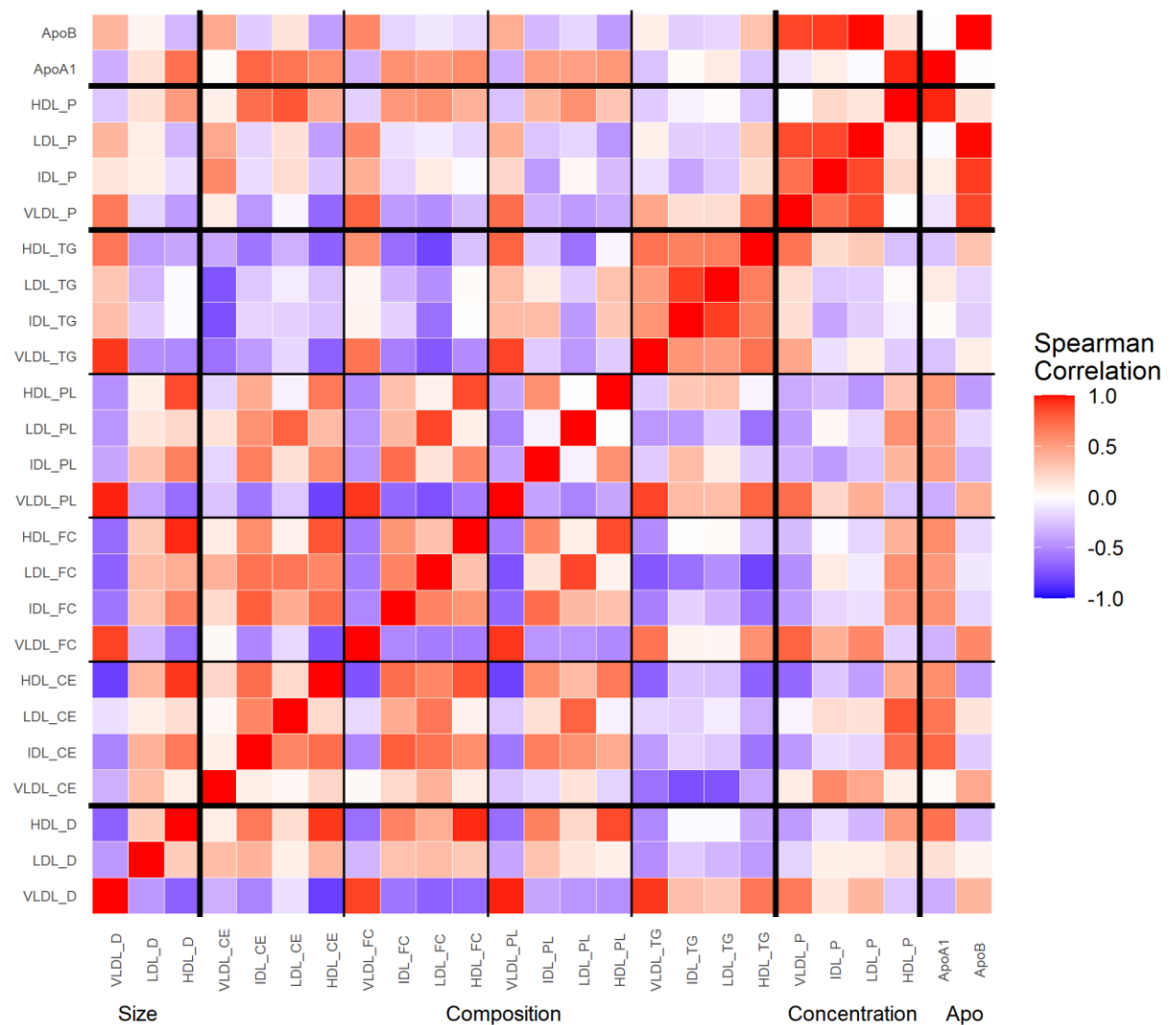
	Mean	SD	Median	IQR
Cholines concentration, mmol/L				
Phosphatidylcholine	2.11	0.36	2.09	0.47
Sphingomyelin	0.45	0.07	0.45	0.09
Phosphoglycerides	2.28	0.38	2.26	0.50
Total cholines	2.57	0.39	2.55	0.51
Amino acids, mmol/L				
Alanine	0.29	0.07	0.29	0.10
Glutamine	0.53	0.08	0.53	0.10
Glycine	0.17	0.06	0.15	0.07
Histidine	0.06	0.01	0.06	0.01
Isoleucine	0.05	0.02	0.05	0.02
Leucine	0.10	0.03	0.10	0.03
Valine	0.20	0.04	0.20	0.05
Branched-chain amino acids	0.35	0.08	0.34	0.10
Phenylalanine	0.04	0.01	0.04	0.01
Tyrosine	0.06	0.01	0.06	0.02
Glycolysis related metabolites, mmol/L				
Lactate	3.78	1.07	3.66	1.36
Citrate	0.08	0.01	0.06	0.01
Glucose	3.46	0.86	3.40	0.83
Pyruvate	0.08	0.03	0.07	0.03
Ketone bodies, mmol/L				
Acetate	0.015	0.011	0.014	0.007
Aceto acetate	0.012	0.012	0.009	0.009
Acetone	0.014	0.005	0.013	0.004
Beta-hydroxybutyrate	0.057	0.058	0.040	0.038
Fluid balance				
Albumin, g/L	39.98	3.24	38.94	3.96
Creatinine, mmol/L	0.06	0.01	0.06	0.02
Inflammation, mmol/L				
Glycoprotein A	0.79	0.11	0.78	0.15

Figure 4.1 Cross-correlations of baseline NMR-derived metabolites



Partial Spearman correlations (adjusted for age and sex) among lipoproteins, apolipoproteins. Lipids, fatty acids, and other low-molecular-weight metabolites (amino acids, glycolysis-related metabolites, and ketone bodies), Abbreviations: Ala=Alanine; Gln=Glutamine; Gly=Glycine; His=Histidine; Ile=Isoleucine; Leu=Leucine; Phe=Phenylalanine; Tyr=Tyrosine; Val=Valine; Cit=Citrate; Glc=Glucose; Lac=Lactate; Pyruvate=Pyruvate; Ace=Acetate; AcAce=Acetoacetate; Acetone=Acetone; bOHBut=3-Hydroxybutyrate; Gp=Glycoprotein acetyls; Alb=Albumin; Crea_nmr=Creatinine; DHA=Docosahexaenoic acid; LA=Linoleic acid; MUFA= Monounsaturated fatty acids; FAW3=Omega-3 fatty acids; FAW6=Omega-6 fatty acids; PUFA=Polyunsaturated fatty acids; SFA=Saturated fatty acids; TotFA=Total fatty acids; TotCho=Total choline; PC=Phosphatidylcholines; Phosphoglyc=Phosphoglycerides; SM=Sphingomyelins; VLDL=very-low-density lipoproteins; IDL=Intermediate-density lipoproteins; LDL=low-density lipoproteins; HDL=High-density lipoproteins; D=diameter; C=cholesterol; CE=cholesterol esters; FC=free cholesterol; PL=phospholipids; TG=triglycerides; P=particle concentration; Apo=Apolipoprotein

Figure 4.2 Cross-correlations of baseline lipoproteins characteristics



Spearman partial correlations (adjusted for age and sex) among lipoproteins (including particle size, concentration, and composition) and apolipoprotein concentration. VLDL=very-low-density lipoproteins; IDL=Intermediate-density lipoproteins; LDL=low-density lipoproteins; HDL=High-density lipoproteins; P=particle concentration; D=particle diameter; CE=cholesterol esters; FC=free cholesterol; PL=phospholipids; TG=triglycerides; Apo=Apolipoprotein.

Table 4.2 Correlations of baseline lipoproteins characteristics and lipids

Lipoproteins	non-HDL-C	HDL-C	Total triglycerides
Lipoprotein concentration			
VLDL	0.82	-0.30	0.85
LDL	0.98	-0.12	0.52
HDL	0.31	0.85	0.01
Apolipoprotein concentration			
Apolipoprotein A-I	0.14	0.93	-0.06
Apolipoprotein B	0.98	-0.12	0.54
Lipoprotein diameter			
VLDL	0.32	-0.51	0.86
LDL	0.10	0.24	-0.36
HDL	-0.26	0.80	-0.42
Lipoprotein composition			
VLDL			
Esterified cholesterol	0.48	0.06	-0.36
Free cholesterol	0.56	-0.47	0.73
Phospholipids	0.35	-0.53	0.87
Triglycerides	-0.01	-0.40	0.81
LDL			
Esterified cholesterol	0.30	0.64	-0.07
Free cholesterol	0.06	0.66	-0.67
Phospholipids	-0.02	0.56	-0.47
Triglycerides	-0.23	-0.06	0.53
HDL			
Esterified cholesterol	-0.36	0.76	-0.71
Free cholesterol	-0.14	0.68	-0.31
Phospholipids	-0.44	0.53	-0.17
Triglycerides	0.19	-0.45	0.84

Spearman partial correlations between lipoprotein characteristics and lipid-related biomarkers, adjusted for age and sex. VLDL= very-low-density lipoproteins; LDL=Low-density lipoproteins; HDL=High-density lipoproteins; C=Cholesterol.

Table 4.3 Cross-correlations of baseline fatty acids concentration

Fatty acids	Total FA	SFA	MUFA	Omega-6 PUFA	LA	Omega-3 PUFA	DHA
Total fatty acids	1	0.97	0.95	0.86	0.81	0.53	0.28
SFA		1	0.93	0.75	0.68	0.47	0.21
MUFA			1	0.73	0.69	0.38	0.07
Omega-6 PUFA				1	0.98	0.43	0.30
Linoleic acids					1	0.33	0.22
Omega-3 PUFA						1	0.91
DHA							1

Spearman partial correlations of fatty acids concentration, adjusted for age and sex. FA=fatty acids; SFA=Saturated fatty acids; MUFA=Monounsaturated fatty acids; PUFA=Polyunsaturated fatty acids; DHA= Docosahexaenoic acid.

Table 4.4 Correlations of baseline fatty acids concentration and lipids

Fatty acids	non-HDL-C	HDL-C	Total triglycerides
SFA	0.64	0.09	0.87
MUFA	0.57	-0.11	0.93
PUFA			
Omega-6 PUFA	0.82	0.26	0.59
Linoleic acids	0.79	0.20	0.58
Omega-3 PUFA	0.40	0.16	0.37
DHA	0.34	0.34	0.03

Spearman partial correlations between fatty acids concentration and lipid-related biomarkers, adjusted for age and sex. SFA=Saturated fatty acids; MUFA=Monounsaturated fatty acids; PUFA=Polyunsaturated fatty acids; DHA= Docosahexaenoic acid; LDL=Low-density lipoproteins; HDL=High-density lipoproteins; C=Cholesterol.

Table 4.5 Comparison of metabolites measured by clinical chemistry and NMR spectroscopy

	N	Mean (SD)		Correlation Coefficient [†]
		Clinical chemistry*	NMR spectroscopy	
Lipids (mmol/L)				
Total cholesterol	82,662	5.9 (1.1)	4.7 (0.9)	0.86
LDL cholesterol	82,541	3.7 (0.8)	1.8 (0.4)	0.87
HDL cholesterol	75,716	1.5 (0.4)	1.3 (0.3)	0.91
Triglycerides	82,603	1.7 (1.0)	1.3 (0.6)	0.87
Apolipoproteins (g/L)				
Apolipoprotein A-1	75,282	1.5 (0.3)	1.4 (0.2)	0.84
Apolipoprotein B	82,259	1.1 (0.2)	0.9 (0.2)	0.90
Fluid balance				
Creatinine, μmol/L	82,614	71.3 (14.7)	64.9 (12.5)	0.79
Albumin, g/L	75,743	45.2 (2.6)	39.0 (3.2)	0.56

*NMR-derived metabolic biomarkers were measured using EDTA plasma samples from aliquot 3, while clinical biochemistry markers were measured from serum samples primarily from aliquot 1; [†]Correlations are Spearman partial correlation coefficients, adjusted for age and sex; [§]The biomarker of LDL-cholesterol measured by NMR spectroscopy was 'size specific' which only for the cholesterol carried in LDL particles, while for LDL-cholesterol measured by clinical chemistry assay, the measurement also include cholesterol in IDL and small VLDL particles. LDL=low-density lipoproteins; HDL=High-density lipoproteins.

CHAPTER 5: Within-person variability of metabolites

Chapter summary

Random fluctuations in exposure, caused by measurement error or short-term biological variability (collectively named as within-person variability), have been proved to underestimate the prospective epidemiological associations. The amount by which the true association is underestimated by using the baseline exposure value is called regression dilution ratio (RDR). There is limited evidence from large-scale prospective studies on the within-person variability of NMR-derived metabolites. This chapter assesses the within-person variability using ~1,000 participants in the UK Biobank with both baseline and repeat NMR assessment. The RDRs of most metabolites were in the range of 0.5 to 0.7, while some of the low-molecular-weight metabolites had higher within-person variability with RDRs under 0.4. Within-person variability should be corrected to avoid underestimation when evaluating the prospective associations of metabolites with CHD risk, but because of the instability of some measurements, a cautious interpretation of the prospective association of the low-molecular-weight metabolites is warranted.

5.1 Background

Within-person variability, a reflection of random fluctuations in exposure caused by measurement error, or short-term biological variability, has been proved to underestimate the prospective epidemiological associations.¹¹²⁻¹¹⁴ Although the within-person variability of cholesterol has been measured in several large-scale studies,¹¹³⁻¹¹⁵ there is still no evidence on the within-person variability of metabolites measured by high-throughput methods, and no evidence on whether the variability of the same biomarker is similar among different measurement methods. All previous studies on the associations of NMR-derived biomarkers with cardiovascular risk used metabolites levels measured on a single occasion at the start of the study only. Without taking within-person variability into account, the estimated associations in previous evidence may have resulted in systematic underestimation due to random fluctuations from measurement error or biological variability.¹¹⁶

The effect of variability in exposure on disease associations is called regression dilution bias, and the amount by which the true association is underestimated by using baseline values is called regression dilution ratio (RDR), representing the ratio of the uncorrected association to that of the real association.⁶⁴ RDRs can be derived by relating baseline measurements to replicate measurements from a reasonably representative sample of study participants after an interval.⁶⁴ The repeat measurements in a subset of participants in the UK Biobank enabled the evaluation of RDR of each metabolite and the estimation of the prospective associations after the correction of regression dilution bias.

There are a number of methods to estimate RDR based on different situations and assumptions. The aim of this section is to choose the most suitable method and evaluate the within-person variability of each metabolite.

5.2 Methods

MacMahon-Peto, Rosner's regression and self-correlation are the three methods that are used to estimate RDR based on different situations and assumptions:

- The MacMahon-Peto method is a non-parametric method that calculates RDR as the ratio of differences between repeat exposure and baseline measurements across equal groups defined by baseline risk factor distributions.¹¹⁶ The method divides the data into equal groups based on the baseline measurement and assumes the mean value of the resurvey measurement in each baseline quintile group is an unbiased estimate of the mean value in these individuals. The absolute differences in mean values between the top and bottom fifths are calculated at both time points, and their ratio is used as the estimate of the RDR (RDR_{MP}). The limitation is that this method does not use all of the available data.
- Rosner's regression is a parametric method that calculates RDR by regressing the resurvey measurement on the baseline measurement.¹¹⁷ This method assumes a linear relationship between baseline and resurvey measurements, and asks for additional assumptions when applying linear regression, including independence, homoscedasticity and normality of residuals. Potential confounders, assuming measured without error or variability, can be adjusted for in Rosner's regression.

- Self-correlation method is another parametric method that calculates RDR by the Pearson or Spearman correlation coefficient between baseline and resurvey measurements, and assumes constant variance between the two measurements.¹¹⁸ Partial correlations can be used to allow for adjustment for potential confounders.

The RDRs were estimated by all three methods as comparison. Age group and sex were adjusted for in Rosner's regression and self-correlation method. The assumptions of each method were tested, and one method was chosen for the correction of regression dilution bias in subsequent analyses. Fasting time was further adjusted for in sensitivity analyses. The RDRs of metabolites measured by the clinical chemistry method were also calculated and compared with the RDRs of the same biomarkers measured by NMR.

5.3 Baseline characteristics of resurvey

Among the study population, a subset of 20,346 participants took part in a resurvey after baseline, and among these participants, a subset of 1,042 had repeat NMR assessment. After excluding the missing and outlying metabolites in the resurvey using the same criteria as baseline (Chapter 3), 953 participants with both baseline and repeat NMR assessment were included for the analyses of within-person variability. The interval between baseline and resurvey was 4.2 years on average. Except for higher percentage of college or university education (54% in resurvey versus 41% in whole study population) and higher percentage of never smokers (64% in resurvey versus 57% in whole study population), the levels of the other baseline characteristics were similar between the study population and the resurvey subset. The median (IQR) of

fasting time was 3.0 (2.0) hours at baseline assessment and 4.0 (1.0) hours at repeat assessment. Detailed baseline characteristics of participants with repeat NMR assessment at resurvey is provided in **Table 5.1**. The mean and the variance of each metabolite in resurvey were similar to the baseline level among the participants with repeat assessment (**S5.1**).

5.4 Estimation of within-person variability

There was very little difference in RDRs estimated by the three methods, and the difference in estimates when using Rosner's regression and using self-correlation were all less than 0.1 across all the metabolites. For most metabolites, the RDRs measured by MacMahon-Peto method were slightly larger compared with using the two other methods, but most of the differences were still under 0.1, with only two exceptions for sphingomyelin and very large HDL particle concentration (the differences in RDRs using the MacMahon-Peto method and the self-correlation method were more than 0.1 for these two biomarkers) (**S5.2**). For the assumptions of Rosner's regression, the normality of residual was slightly violated for ketone bodies. Self-correlation method was therefore chosen for the correction of regression dilution bias in subsequent analyses, due to its strengths of fewer assumptions and use of all available data, and partial spearman correlation was applied to account for the skewed distribution of some metabolites.

Figure 5.1 and **Table 5.2** show the within-person variability of each metabolite using self-correlation method and adjusting for age group and sex. The RDRs of most lipids, lipoproteins and apolipoproteins were in the range of 0.5 to 0.7, with only exception of LDL particle diameter and average molecule numbers of

esterified cholesterol in LDL that were lower than 0.5. Except for the ratio of saturated fatty acids to total fatty acids, the RDRs of other fatty acids concentrations and ratios were in the range of 0.5 to 0.6. The within-person variability of cholines was similar to fatty acids, with RDRs around 0.55. However, except for glycine, creatinine and glycoprotein A, with RDRs also larger than 0.50, the within-person variability of other low-molecular-weight metabolites (including amino acids, ketone bodies, fluid balance, and glycolysis related metabolites) were much greater, with RDRs lower than 0.4. Further adjusting for fasting time did not materially change the estimates of RDRs in all types of metabolites (**S5.3**).

Most metabolites showed no difference in their within-person variability by sex, and the differences in RDRs were less than 0.1. However, the within-person variability was greater in men than women for some of the lipid-related biomarkers (their confidence intervals did not overlap), including lipoprotein concentration of IDL, LDL and very large HDL, concentration of ApoB, ratio of ApoB to ApoA-1, as well as some lipids composition biomarkers of HDL (**Table 5.2**). When comparing by measurement method, the within-person variability of most metabolites measured by NMR spectroscopy was slightly larger than the same types of metabolites measured by the clinical chemistry, but the confidence intervals of most lipids and apolipoprotein biomarkers still overlapped. The RDRs were in the range of 0.5 to 0.8 for lipids and apolipoprotein biomarkers by both NMR and clinical chemistry method. However, the variations of within-person variability by measurement method were greater among fluid balance biomarkers, with RDRs 0.66 (0.61, 0.70) by

the clinical chemistry and 0.54 (0.49, 0.59) by NMR for creatinine, and 0.44 (0.38, 0.50) and 0.26 (0.19, 0.33) respectively for albumin (**Table 5.3**).

5.5 Discussion

This chapter describes the within-person variability of NMR-measured metabolites in a subset of about one thousand participants with repeat metabolites measures. The results showed the variations of RDRs among different types of metabolites, and highlighted the importance of taking within-person variability into account when assessing the prospective associations of metabolites in epidemiological research.

Our study estimated the within-person variability of metabolites in a subset of participants with repeat metabolites measures. Although the resurveyed population was not entirely representative of the study population, the baseline distribution of each metabolite in the subset was similar to the whole study population. The RDRs of total cholesterol and HDL cholesterol by clinical chemistry method in the study were very close to the RDRs of the same biomarkers in previous studies by the same measurement method¹¹³⁻¹¹⁵, which suggests the generalisability of the results based on this subset.

There were very little material difference in RDRs estimated by MacMahon-Peto, Rosner's regression and self-correlation methods. The self-correlation method was chosen for the correction of regression dilution bias in subsequent analyses of the study, due to its strengths of fewer assumptions and use of all available data. According to the results, all NMR-derived metabolites showed some degree of within-person variability, and this was substantially greater for most of the low-molecular-weight metabolites, demonstrating the necessity of

consideration of within-person variability in the prospective associations of metabolites.

The within-person variability of NMR-derived lipids, lipoproteins, fatty acids, and cholines were similar according to the results, and the RDRs of lipids measured by NMR was slightly lower but still similar to the RDRs by the clinical chemistry method. In terms of amino acids, ketone bodies, fluid balance, and glycolysis related metabolites, the within-person variability was much greater, but no evidence from other cohorts can be found on the RDRs of low-molecular-weight metabolites. However, when comparing the RDRs of creatinine and albumin, the only two low-molecular-weight metabolites available from both NMR and the clinical chemistry method in the UK Biobank, the RDRs by NMR spectroscopy were lower than the corresponding RDRs by the clinical chemistry, and the within-person variability of albumin, regardless of measurement method, was larger than all other lipid measures. This result suggests that the large within-person variability of most low-molecular-weight metabolites may be due to multiple reasons of both instability of the metabolite itself and the measurement error from NMR spectroscopy.

Fasting time is another possible reason for the within-person variability. Participants in the UK Biobank were not asked to fast before attending the screening clinic, but were asked about the time since their last meal. The fasting time is skewed at both baseline and resurvey, with median of 3 and 4 hours, respectively. Because of the very limited number of fasting participants (4%), we were not able to compare the RDRs by fasting state. However, we have adjusted for fasting time in the sensitivity analyses and did not find changes in the RDRs of any metabolite. The spearman correlations between fasting time

and the concentrations of the major measurements in the study were all less than 0.3. A recent study also showed that applying postprandial metabolites measurements using NMR spectroscopy is robust in biomarker research¹¹⁹.

5.6 Conclusion

This chapter investigates the within-person variability of NMR-measured metabolites. Similar within-person variability was observed among NMR-derived lipids, lipoproteins, apolipoproteins, fatty acids and cholines, while much greater variability was observed among most of the small-molecular-weight metabolites, except for glycine, creatinine and glycoprotein A. These findings suggest the importance of considering within-person variability when assessing and interpreting the prospective associations of metabolites in the following chapters.

Chapter 5 Tables and Figures

Tables	Title
Table 5.1	Baseline characteristics of participants with repeat assessment
Table 5.2	Within-person variability of metabolites, by sex
Table 5.3	Within-person variability of metabolites, by measurement method

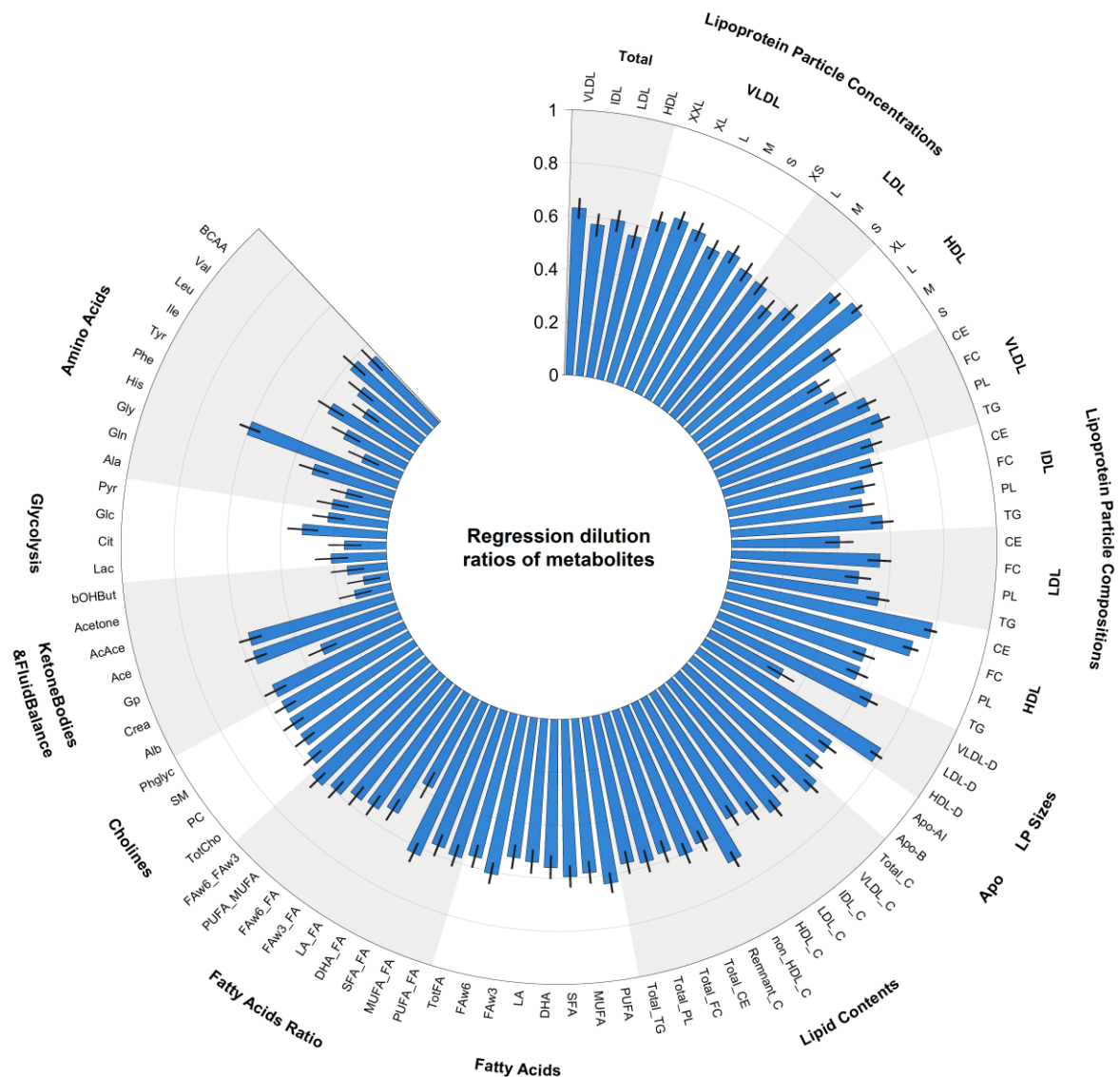
Figures	Title
Figure 5.1	Regression dilution ratios for metabolites

Table 5.1 Baseline characteristics of participants with repeat assessment

	Women	Men	All
No. of participants	516	437	953
White, %	97.0	97.4	97.2
Baseline age, years	55.7 (7.3)	56.7 (7.8)	56.2 (7.5)
Education, %			
Higher education	51.4	57.1	53.9
AS/A-level	25.7	26.8	26.2
CSE/GCSE/O-Level	14.8	8.8	12.1
No qualification	8.2	7.3	7.8
Assessment region, %			
Northwest	63.2	64.4	63.7
Yorkshire & Humber	31.1	27.9	29.7
West Midlands	5.6	7.7	6.5
Southeast	0.2	0.0	0.1
Townsend Deprivation Index*	-2.0 (2.6)	-2.0 (2.8)	-2.0 (2.7)
Frequency of alcohol intake, %†			
Never	2.4	2.8	2.6
Previous	2.4	1.9	2.2
Current occasional	26.9	13.5	20.9
Current regular	68.2	81.8	74.3
Frequency of smoking, %‡			
Never	68.9	58.4	64.2
Previous	27.8	35.2	31.1
Current occasional	0.7	1.5	1.1
Current regular	2.6	4.9	3.7
Anthropometry			
Body mass index, kg/m ²	26.4 (4.8)	26.9 (3.5)	27.0 (4.6)
Waist Circumference, cm	82.4 (11.3)	93.8 (10.3)	87.5 (12.3)
Waist to hip ratio	0.8 (0.1)	0.9 (0.1)	0.9 (0.1)
Systolic blood pressure, mmHg	135.1 (18.1)	140.4 (17.1)	137.5 (17.8)
Baseline diabetes, %	0.9	2.8	1.7
Baseline fasting time, h	3.0 (2.0)	3.0 (2.0)	3.0 (2.0)
Resurvey fasting time, h	4.0 (1.0)	4.0 (1.0)	4.0 (1.0)

Continuous variables are presented as mean (standard deviation) (except for skew variable fasting time, presented as median [quartile range]), and categorical variables are presented as column percentages. *Area-level measure of material deprivation; †Regular alcohol intake means drinks a certain amount of alcohol no less than once a week; ‡Regular smoking means smoking on most or all days.

Figure 5.1 Within-person variability of metabolites



Regression dilution ratios of metabolites by self-correlation method adjusted for age groups and sex. Ala=Alanine; Gln=Glutamine; Gly=Glycine; His=Histidine; Ile=Isoleucine; Leu=Leucine; Phe=Phenylalanine; Tyr=Tyrosine; Val=Valine; Cit=Citrate; Glc=Glucose; Lac=Lactate; Pyruvate=Pyruvate; Ace=Acetate; AcAce=Acetoacetate; Acetone=Acetone; bOHBut=3-Hydroxybutyrate; Gp=Glycoprotein acetyls; Alb=Albumin; Crea_nmr=Creatinine; DHA=Docosahexaenoic acid; LA=Linoleic acid; MUFA=Monounsaturated fatty acids; Faw3=Omega-3 fatty acids; Faw6=Omega-6 fatty acids; PUFA=Polyunsaturated fatty acids; SFA=Saturated fatty acids; TotFA=Total fatty acids; TotCho=Total cholines; PC=Phosphatidylcholines; Phosphoglyc=Phosphoglycerides; SM=Sphingomyelins; VLDL=very-low-density lipoproteins; IDL=Intermediate-density lipoproteins; LDL=low-density lipoproteins; HDL=High-density lipoproteins; D=diameter; C=cholesterol; CE=cholesterol esters; FC=free cholesterol; PL=phospholipids; TG=triglycerides; P=particle concentration; Apo=Apolipoprotein.

Table 5.2 Within-person variability of metabolites, by sex

A. Concentration of lipoproteins and apolipoproteins

	Women	Men	All
Lipoproteins, mmol/L			
VLDL	0.65 (0.59, 0.70)	0.59 (0.52, 0.65)	0.63 (0.59, 0.67)
Chylomicrons & extremely large	0.63 (0.57, 0.68)	0.65 (0.59, 0.70)	0.64 (0.60, 0.67)
Very large	0.66 (0.61, 0.70)	0.67 (0.61, 0.72)	0.67 (0.63, 0.70)
Large	0.64 (0.59, 0.69)	0.65 (0.59, 0.70)	0.65 (0.61, 0.69)
Medium	0.64 (0.59, 0.69)	0.56 (0.49, 0.62)	0.61 (0.57, 0.65)
Small	0.65 (0.59, 0.69)	0.61 (0.55, 0.66)	0.64 (0.60, 0.67)
Very small	0.64 (0.59, 0.69)	0.55 (0.48, 0.61)	0.61 (0.56, 0.64)
IDL	0.63 (0.58, 0.68)	0.49 (0.41, 0.56)	0.58 (0.53, 0.62)
LDL	0.64 (0.59, 0.69)	0.53 (0.46, 0.59)	0.60 (0.56, 0.64)
Large	0.65 (0.59, 0.69)	0.52 (0.44, 0.58)	0.60 (0.55, 0.64)
Medium	0.58 (0.51, 0.63)	0.47 (0.39, 0.54)	0.54 (0.49, 0.58)
Small	0.63 (0.58, 0.68)	0.52 (0.45, 0.59)	0.59 (0.55, 0.63)
HDL	0.56 (0.49, 0.61)	0.55 (0.48, 0.61)	0.56 (0.51, 0.60)
Very large	0.80 (0.77, 0.83)	0.69 (0.63, 0.73)	0.75 (0.72, 0.78)
Large	0.80 (0.77, 0.83)	0.76 (0.72, 0.80)	0.79 (0.76, 0.81)
Medium	0.62 (0.56, 0.67)	0.57 (0.51, 0.63)	0.61 (0.57, 0.65)
Small	0.50 (0.43, 0.56)	0.49 (0.41, 0.56)	0.50 (0.45, 0.55)
Apolipoproteins			
Apolipoprotein A-I, g/L	0.62 (0.57, 0.67)	0.58 (0.52, 0.64)	0.61 (0.57, 0.65)
Apolipoprotein B, g/L	0.65 (0.60, 0.70)	0.54 (0.47, 0.60)	0.61 (0.57, 0.65)
Apolipoprotein B to A-1 ratio	0.71 (0.66, 0.75)	0.57 (0.50, 0.63)	0.66 (0.62, 0.70)

VLDL=very-low-density lipoproteins; IDL=intermediate-density lipoproteins; LDL=low-density lipoproteins; HDL=high-density lipoproteins.

B. Size and composition of lipoproteins

	Women	Men	All
Lipoprotein diameter, nm*			
VLDL	0.63 (0.58, 0.68)	0.67 (0.61, 0.72)	0.66 (0.62, 0.69)
LDL	0.30 (0.22, 0.38)	0.32 (0.24, 0.41)	0.32 (0.26, 0.37)
HDL	0.80 (0.77, 0.83)	0.74 (0.70, 0.78)	0.78 (0.75, 0.80)
Lipoprotein composition, molecules/particle[†]			
VLDL			
Esterified cholesterol	0.57 (0.51, 0.62)	0.50 (0.43, 0.57)	0.54 (0.49, 0.58)
Free cholesterol	0.64 (0.59, 0.69)	0.59 (0.52, 0.65)	0.63 (0.59, 0.67)
Phospholipids	0.65 (0.59, 0.70)	0.65 (0.59, 0.70)	0.65 (0.62, 0.69)
Triglycerides	0.57 (0.51, 0.63)	0.62 (0.56, 0.67)	0.59 (0.55, 0.63)
IDL			
Esterified cholesterol	0.59 (0.53, 0.64)	0.52 (0.45, 0.59)	0.57 (0.53, 0.61)
Free cholesterol	0.55 (0.48, 0.61)	0.47 (0.39, 0.54)	0.52 (0.47, 0.56)
Phospholipids	0.53 (0.47, 0.59)	0.44 (0.36, 0.51)	0.50 (0.45, 0.55)
Triglycerides	0.56 (0.50, 0.62)	0.59 (0.52, 0.65)	0.57 (0.53, 0.62)
LDL			
Esterified cholesterol	0.41 (0.34, 0.48)	0.40 (0.31, 0.47)	0.41 (0.36, 0.46)
Free cholesterol	0.56 (0.49, 0.61)	0.58 (0.51, 0.64)	0.56 (0.52, 0.60)
Phospholipids	0.49 (0.43, 0.56)	0.47 (0.39, 0.54)	0.49 (0.44, 0.53)
Triglycerides	0.57 (0.51, 0.62)	0.58 (0.51, 0.64)	0.57 (0.52, 0.61)
HDL			
Esterified cholesterol	0.80 (0.76, 0.83)	0.78 (0.75, 0.82)	0.79 (0.76, 0.81)
Free cholesterol	0.78 (0.74, 0.81)	0.66 (0.60, 0.71)	0.73 (0.70, 0.76)
Phospholipids	0.62 (0.56, 0.67)	0.49 (0.42, 0.56)	0.57 (0.53, 0.61)
Triglycerides	0.54 (0.48, 0.60)	0.61 (0.54, 0.66)	0.57 (0.53, 0.61)

*Diameter of IDL particles not given in the UK Biobank. †Lipoprotein composition derived by dividing the lipidic contents by the particle concentration. VLDL=very-low-density lipoproteins; IDL=intermediate-density lipoproteins; LDL=low-density lipoproteins; HDL=high-density lipoproteins.

C. Total Lipids and fatty acids

	Women	Men	All
Total plasma lipids, mmol/L			
Total Cholesterol	0.58 (0.52, 0.64)	0.52 (0.45, 0.59)	0.57 (0.53, 0.61)
VLDL Cholesterol	0.66 (0.61, 0.71)	0.57 (0.50, 0.63)	0.63 (0.59, 0.67)
IDL Cholesterol	0.61 (0.56, 0.67)	0.52 (0.45, 0.59)	0.59 (0.54, 0.63)
LDL Cholesterol	0.58 (0.52, 0.63)	0.50 (0.43, 0.57)	0.56 (0.51, 0.60)
HDL Cholesterol	0.72 (0.68, 0.76)	0.66 (0.61, 0.71)	0.70 (0.67, 0.73)
Non-HDL Cholesterol*	0.62 (0.56, 0.67)	0.52 (0.45, 0.59)	0.59 (0.54, 0.63)
Remnant Cholesterol	0.64 (0.58, 0.69)	0.53 (0.46, 0.60)	0.60 (0.56, 0.64)
Total Esterified Cholesterol	0.57 (0.51, 0.63)	0.52 (0.45, 0.59)	0.57 (0.52, 0.61)
Total Free Cholesterol	0.60 (0.55, 0.66)	0.52 (0.45, 0.59)	0.58 (0.54, 0.62)
Total Phospholipids	0.55 (0.49, 0.61)	0.55 (0.48, 0.61)	0.57 (0.52, 0.61)
Total Triglycerides	0.61 (0.55, 0.66)	0.66 (0.60, 0.71)	0.63 (0.59, 0.67)
Fatty acids concentration, mmol/L			
Polyunsaturated fatty acids	0.55 (0.48, 0.61)	0.57 (0.50, 0.63)	0.59 (0.54, 0.63)
Monounsaturated fatty acids	0.57 (0.51, 0.63)	0.60 (0.53, 0.65)	0.60 (0.55, 0.64)
Saturated fatty acids	0.54 (0.47, 0.60)	0.55 (0.48, 0.61)	0.56 (0.51, 0.60)
Docosahexaenoic acid	0.55 (0.49, 0.61)	0.53 (0.46, 0.60)	0.54 (0.50, 0.59)
Linoleic acid	0.53 (0.47, 0.59)	0.50 (0.42, 0.56)	0.53 (0.49, 0.58)
Omega-3 fatty acids	0.59 (0.53, 0.64)	0.61 (0.55, 0.67)	0.61 (0.57, 0.65)
Omega-6 fatty acids	0.55 (0.48, 0.60)	0.54 (0.47, 0.60)	0.56 (0.52, 0.60)
Total fatty acids	0.55 (0.49, 0.61)	0.57 (0.50, 0.63)	0.58 (0.54, 0.62)
Fatty acids ratio to total*			
Polyunsaturated fatty acids	0.56 (0.49, 0.61)	0.57 (0.50, 0.63)	0.57 (0.52, 0.61)
Monounsaturated fatty acids	0.60 (0.54, 0.65)	0.65 (0.59, 0.70)	0.63 (0.59, 0.67)
Saturated fatty acids	0.39 (0.31, 0.46)	0.37 (0.28, 0.44)	0.38 (0.32, 0.43)
Docosahexaenoic acid	0.54 (0.48, 0.60)	0.49 (0.42, 0.56)	0.52 (0.48, 0.57)
Linoleic acid	0.57 (0.51, 0.63)	0.53 (0.46, 0.60)	0.55 (0.51, 0.60)
Omega-3 fatty acids	0.57 (0.50, 0.62)	0.56 (0.49, 0.62)	0.56 (0.52, 0.60)
Omega-6 fatty acids	0.56 (0.50, 0.62)	0.62 (0.55, 0.67)	0.60 (0.55, 0.64)
Other fatty acids ratio			
Omega-6 to Omega-3 fatty acids	0.57 (0.51, 0.62)	0.58 (0.51, 0.64)	0.57 (0.53, 0.62)
Poly- to mono-unsaturated fatty acids	0.60 (0.54, 0.65)	0.63 (0.57, 0.69)	0.62 (0.58, 0.66)

*Total cholesterol minus HDL cholesterol; *Ratio of concentration of given type of fatty acid to concentration of total fatty acid.

D. Other low-molecular-weight metabolites

	Women	Men	All
Cholines concentration, mmol/L			
Phosphatidylcholine	0.54 (0.47, 0.60)	0.55 (0.48, 0.61)	0.56 (0.51, 0.60)
Sphingomyelin	0.55 (0.49, 0.61)	0.52 (0.44, 0.58)	0.55 (0.50, 0.59)
Phosphoglycerides	0.52 (0.46, 0.58)	0.54 (0.47, 0.60)	0.55 (0.50, 0.59)
Total cholines	0.53 (0.47, 0.59)	0.54 (0.47, 0.60)	0.55 (0.51, 0.60)
Amino acids, mmol/L			
Alanine	0.19 (0.10, 0.27)	0.16 (0.07, 0.25)	0.17 (0.11, 0.23)
Glutamine	0.33 (0.26, 0.41)	0.30 (0.21, 0.38)	0.32 (0.26, 0.38)
Glycine	0.69 (0.64, 0.73)	0.49 (0.42, 0.56)	0.60 (0.56, 0.64)
Histidine	0.17 (0.08, 0.25)	0.15 (0.06, 0.24)	0.16 (0.10, 0.22)
Isoleucine	0.22 (0.13, 0.30)	0.25 (0.16, 0.34)	0.24 (0.17, 0.30)
Leucine	0.28 (0.20, 0.36)	0.32 (0.23, 0.40)	0.30 (0.24, 0.36)
Valine	0.35 (0.27, 0.42)	0.42 (0.34, 0.50)	0.38 (0.33, 0.44)
Branched-chain amino acids	0.32 (0.24, 0.40)	0.37 (0.29, 0.45)	0.35 (0.29, 0.40)
Phenylalanine	0.24 (0.16, 0.32)	0.28 (0.19, 0.37)	0.26 (0.20, 0.32)
Tyrosine	0.30 (0.22, 0.38)	0.42 (0.34, 0.49)	0.36 (0.30, 0.41)
Glycolysis related metabolites, mmol/L			
Lactate	0.13 (0.04, 0.21)	0.18 (0.09, 0.27)	0.16 (0.10, 0.22)
Citrate	0.32 (0.24, 0.40)	0.30 (0.22, 0.39)	0.32 (0.26, 0.38)
Glucose	0.21 (0.13, 0.29)	0.25 (0.16, 0.33)	0.23 (0.16, 0.29)
Pyruvate	0.17 (0.09, 0.26)	0.27 (0.18, 0.35)	0.22 (0.16, 0.28)
Ketone bodies, mmol/L			
Acetate	0.15 (0.06, 0.23)	0.13 (0.03, 0.22)	0.14 (0.08, 0.21)
Aceto acetate	0.10 (0.02, 0.19)	0.09 (0.00, 0.19)	0.09 (0.02, 0.15)
Acetone	0.12 (0.04, 0.21)	0.18 (0.09, 0.27)	0.15 (0.09, 0.22)
Beta-hydroxybutyrate	0.22 (0.13, 0.30)	0.20 (0.10, 0.29)	0.21 (0.15, 0.27)
Fluid balance			
Albumin, g/L	0.35 (0.27, 0.42)	0.27 (0.18, 0.36)	0.32 (0.26, 0.38)
Creatinine, mmol/L	0.59 (0.53, 0.65)	0.52 (0.44, 0.58)	0.57 (0.52, 0.61)
Inflammation, mmol/L			
Glycoprotein A	0.52 (0.46, 0.58)	0.59 (0.53, 0.65)	0.57 (0.52, 0.61)

Table 5.3: Within-person variability of metabolites, by measurement method

Trait	Regression dilution ratio	
	Clinical Chemistry*	NMR spectroscopy
Lipids		
Total cholesterol	0.61 (0.57,0.66)	0.55 (0.50,0.60)
LDL-cholesterol [§]	0.62 (0.57,0.67)	0.55 (0.50,0.60)
HDL-cholesterol	0.79 (0.76,0.82)	0.71 (0.67,0.75)
Triglycerides	0.57 (0.51,0.62)	0.62 (0.57,0.67)
Apolipoproteins		
ApoA-1	0.70 (0.66,0.73)	0.61 (0.56,0.65)
ApoB	0.65 (0.61,0.69)	0.60 (0.55,0.65)
Fluid balance		
Creatinine, $\mu\text{mol/L}$	0.66 (0.61,0.70)	0.54 (0.49,0.59)
Albumin, g/L	0.44 (0.38,0.50)	0.26 (0.19,0.33)

Regression dilution ratios among 696 participants with both baseline and repeat measures by two measurement methods, estimated by self-correlation method, adjusted for age groups and sex; *NMR-based metabolic biomarkers were measured using EDTA plasma samples from aliquot 3, while clinical biochemistry markers were measured from serum samples primarily from aliquot 1; §The biomarker of LDL-cholesterol measured by NMR spectroscopy was 'size specific' which only for the cholesterol carried in LDL particles, while for LDL-cholesterol measured by clinical chemistry assay, the measurement also include cholesterol in IDL and small VLDL particles. LDL=low-density lipoproteins; HDL=High-density lipoproteins; ApoB=Apolipoprotein B.

CHAPTER 6: Prospective association of lipids and lipoproteins with CHD risk

Chapter summary

Associations of CHD with plasma lipids are well-described, but the associations with characteristics of lipoproteins (which transport lipids) remains unclear. This chapter quantifies the independent associations of NMR-defined lipoprotein characteristics with CHD risk. CHD risk was most strongly related to particle concentrations of VLDL, IDL, LDL and HDL, and direct measurement of lipoprotein concentrations may be of greater value than their indirect measurement by ApoB, which was largely determined by LDL concentration alone. Independently of lipoprotein particle concentrations, there was strong evidence of a positive association with mean triglycerides molecules per LDL particle, but little evidence of associations with total triglycerides or other lipid and lipoprotein fractions.

6.1 Background

Lipoproteins are complex particles that transport lipids in the circulation. Although the associations of plasma lipids with CHD risk are well described,⁴²⁻⁴⁴ more recent evidence suggests these associations may be driven more directly by the total circulating concentration of ApoB-containing lipoprotein particles.^{42,45} Although total ApoB levels strongly predict CHD risk, ApoB-containing particles differ substantially by class. VLDL, IDL and LDL vary between and within class by size and lipid composition.³⁹ Despite considerable research in this area, the independent relevance of these characteristics to CHD risk remains unclear.

Previous studies comparing the strength of associations between lipids and ApoB-containing lipoproteins relied on total ApoB levels as an indirect measure of the total concentration of all ApoB-containing particles (each VLDL, IDL and LDL particle has a single ApoB molecule).^{42,45} However, VLDL and IDL are present in the plasma at much lower concentrations than LDL, and thus independent variations in VLDL, IDL and LDL particle concentration may be obscured by measurements of total ApoB concentration alone. Furthermore, as traditional lipids measures are dependent on both the concentration of lipoproteins and their composition (size and lipid content), the atherogenic potential of lipoprotein may well be dependent on both these characteristics, but most previous studies have not described lipid composition for each of the different lipoprotein classes.

NMR spectroscopy is a high-throughput technology for metabolic profiling, with the strength of measuring blood samples in much larger sample sizes and providing details on many characteristics of major classes and subclasses of

lipoproteins.³⁹ The high consistency between NMR biomarkers with the same biomarkers measured using traditional clinical chemistry methods has been previously shown.^{95,109} The aims of this section are to quantify reliably the associations of CHD risk with concentration of each of the major lipoprotein particle classes, and to examine the effects of lipoprotein size and lipid composition independently of particle concentration.

6.2 Methods

The study followed the same criteria of study population selection, outcome ascertainment and processing of metabolites that was described in Chapter 3. Cox proportional-hazards models, stratified by sex and age, and adjusted for ethnicity, education, region, Townsend Deprivation Index, smoking, alcohol intake, and BMI were used to estimate hazard ratios (HR) for the associations of lipoprotein characteristics, as well as ApoA-1 and ApoB concentrations, with incident CHD. To examine the independent effect of these characteristics on CHD risk, analyses of the size and composition of lipoproteins were further adjusted for lipoprotein concentrations of VLDL, LDL and HDL (IDL concentration was not included for further adjustment because of its high correlation with LDL). The associations of lipids and apolipoproteins measured using either clinical chemistry or NMR method, are given for comparison.

Hazard ratios and their 95% confidence intervals (CIs) were corrected for regression dilution bias (i.e. categorizing people by their baseline concentrations and estimating the long-term average concentration in each category using the correlation between resurvey and baseline measurements), and are therefore described associations of usual lipoprotein levels with CHD risk.¹²⁰ The values of the regression dilution ratios are described in Chapter 5.

'Usual' levels in the plot are estimated from the mean value at resurvey within each baseline-defined group. This provides an unbiased estimate of the long-term average level in each baseline-defined group. CIs in the plots were calculated using the variance of the log risk, which appropriately attributes variance to all groups, including the reference.¹²¹

The change in the log-likelihood (LR) χ^2 between models with and without the lipoprotein parameter is a measure of extent of variance explained by the lipoprotein parameter in addition to the other variables in the model. This statistic provides not only a significance test for the improvement in fit from including the main lipoprotein term, but also a quantitative measure of the extent to which the lipoprotein term improves risk prediction in different models (e.g. with and without adjustment for confounders or other lipoprotein characteristics).¹¹²

Sensitivity analysis of the main associations were conducted by excluding CHD events in the first two years of follow-up, and by further adjusting for other potential confounders, including waist circumference, systolic blood pressure, diabetes and chronic kidney disease at baseline, fasting time, and spectrometer. Discordance analysis (used to deal with the challenge of discriminating among highly correlated markers) was also used to judge which biomarker is more responsible for risk among highly correlated metabolites by comparing the hazard ratios among the participants who are discordant in two highly correlated biomarkers.^{122,123}

6.3 Association of lipoprotein and apolipoprotein concentrations

During a mean follow-up of 11.5 years, 4,196 participants had an incident CHD event. Compared to participants that did not have an incident CHD event, those with incident CHD were on average slightly older, more likely to be men and current smokers, have lower education, slightly higher BMI and systolic blood pressure (**Table 6.1**).

CHD risk was positively associated with usual VLDL, IDL and LDL particle concentrations, and inversely associated with HDL concentration (**Figure 6.1**). After adjusting for confounders and correcting for regression dilution, hazard ratios per usual SD higher level of VLDL, IDL, LDL and HDL particle concentration were 1.20 (95% CI 1.16-1.23), 1.15 (1.12-1.19), 1.17 (1.14-1.21) and 0.90 (0.87-0.91), respectively. With the same adjustment, the confidence interval of the HR of ApoB concentration (1.18 [1.14-1.22]) overlapped the HR of LDL particle concentration, and the overlap was also between the HR of ApoA-1 (0.88 [0.85-0.91]) and HDL concentration (**Table 6.2**).

Total ApoB concentration (also measured by NMR) was highly correlated with LDL particle concentration ($r=0.99$, **S6.1**), and the strength of the association between total ApoB level and CHD risk was the same as LDL particle concentration (**Figure 6.2**). ApoB concentration was still highly but relatively more weakly correlated with VLDL ($r=0.87$). Given total ApoB concentration, concentrations of VLDL particles remained positively associated with CHD risk,

while the association of total ApoB level given VLDL concentration attenuated to null (**Figure 6.3**).

6.4 Association of lipoprotein subclasses and particle sizes

The strength of the associations of lipoprotein concentration and CHD risk varied across subclass of VLDL, with progressively stronger associations for smaller subclasses: HR was 1.20 (95% CI 1.16-1.23) for the very small subclass of VLDL particles, but 1.09 (1.06-1.13) for the extremely large subclass of VLDL (**Table 6.3**). By contrast, the strength of the associations of lipoprotein concentration and CHD risk were similar across LDL subclasses. For HDL, the strength of associations slightly varied by subclass, but there was no evidence of progressively increasing or decreasing trend from very large to small HDL particles. The findings on lipoprotein subclass were consistent with analyses of lipoprotein particle diameter and CHD risk: there was no evidence of associations with either usual LDL diameter (1.04 [95% CI 0.99-1.07]) or usual HDL diameter (0.97 [0.92-1.02]), after adjusting for lipoprotein concentrations, but an inverse association with usual VLDL diameter (0.92 [0.87-0.97]) (**Table 6.4, Figure 6.4**).

6.5 Association of lipoprotein particle compositions

Table 6.5 shows the associations of lipoprotein composition and CHD risk. In analyses of cholesterol level in each type of lipoprotein particle, mean free cholesterol molecules per VLDL was inversely associated with CHD risk (0.93 [95% CI 0.88-0.98]) after adjustment for VLDL, LDL and HDL particle concentrations, but the proportional reduction in the LR χ^2 statistic (from 119 to 8, a reduction of 93%) with progressive adjustments indicated the strong

likelihood for residual confounding in this association. There was no evidence of associations with mean molecules level of cholesterol or cholesteryl esters per VLDL particle, neither with any type of cholesterol per IDL, LDL or HDL after adjustment for particle concentrations. The composition of phospholipids showed no evidence of an association except in LDL (1.07 [1.01-1.13]); however, the proportional reduction in the LR χ^2 statistic (from 85 to 6, a reduction of 93%) also indicated residual confounding in this association.

For triglyceride levels in each type of lipoprotein particle, independently of lipoprotein concentration, mean triglycerides molecules per VLDL were inversely associated with CHD risk (0.93 [0.89, 0.98]), while mean triglycerides molecules per IDL, LDL and HDL showed positive associations (1.10 [1.05, 1.15], 1.17 [1.11-1.23] and 1.17 [1.07, 1.28], respectively, **Table 6.5, Figure 6.5**). Given the proportional reduction in LR χ^2 statistic, residual confounding may still exist in HDL (from 171 to 12, a reduction of 93%); however, there was strong evidence of a positive association with mean triglyceride molecules in LDL, showing a 17% increase of CHD risk per 1-SD-higher level.

6.6 Association of lipids

Chapter 4 showed high consistency among lipid and apolipoprotein measurements using traditional clinical chemistry and those using NMR (**Table 4.5**). And among participants with both measurements, the associations to CHD risk were similar (overlaps of CIs) between the clinical and NMR measurements (**Table 6.6**).

When assessing the NMR-derived lipids measures, similar to the results of the derived cholesterol composition measures, further adjustment for lipoprotein

particle concentrations attenuated all the cholesterol-related associations to null, but the confidence intervals were wider due to the collinearity with lipid content measures (e.g. LDL cholesterol represents sum of cholesterol in all LDL particles, which has high correlation with LDL concentration) (**Table 6.7**). Independent of lipoprotein particle concentrations, the association of total triglycerides with CHD risk also attenuated to null (0.95 [95% CI 0.87, 1.04]) (**Figure 6.6**).

6.7 Sensitivity analyses

In sensitivity analyses, further adjustment for other potential confounders or exclusion of the first two years of follow-up did not materially change the main associations, and neither did mutual adjustment for other particle concentrations for the association of lipoprotein concentrations (**S6.2**).

In analyses on the improvement of model fit when adding different measures, the addition of the mean triglyceride molecules per LDL particle to the models using either cholesterol measures or lipoprotein particle concentrations improved the LR χ^2 statistic (by about 30 in both cases), whereas there was little evidence of improvement on addition of total triglycerides (**S6.3**). Similarly in discordance analyses, among participants with absolute difference of total triglycerides and mean triglyceride in LDL particles above the 20th percentile, total triglycerides were not associated with CHD after accounting for HDL-C and non-HDL-C, or particles concentrations of VLDL, LDL and HDL, while triglycerides in LDL remained to be a significant marker of CHD risk (**S6.4**).

6.8 Discussion

In this large-scale prospective study, CHD risk was positively associated with VLDL, IDL and LDL particle concentrations, and inversely associated with HDL particle concentration. There was no evidence that the strength of these associations varied by sizes of LDL and HDL particles, but larger subclasses of VLDL were found to be less strongly related to CHD risk. Independently of lipoprotein particle concentrations, there was strong evidence of a positive association between mean triglyceride molecules per LDL particle and CHD risk, but little evidence of associations with total triglyceride. Of the lipoprotein characteristics assessed, CHD risk was most strongly related to lipoprotein particle concentration, and direct measurement of lipoprotein concentrations may be of greater value than their indirect measurement by ApoB, which was largely determined by LDL concentration alone.

A number of previous studies have also assessed the associations of NMR-derived measures of lipoprotein concentrations with CHD risk.^{65,124,125} These studies have consistently described positive associations with VLDL and LDL particle concentrations, and inverse associations with HDL particle concentrations, as in the present report. However, the associations in these studies are somewhat weaker than in the present report, likely reflecting associations with baseline measures rather than long-term average ('usual') levels of lipoprotein concentration. For example, the placebo arm of a large randomized trial of simvastatin versus placebo showed adjusted hazard ratios per 1-SD-higher level were 1.11 (95% CI 1.05-1.17; SD=0.34 $\mu\text{mol/L}$) for LDL particles concentration and 0.88 (0.83, 0.92; SD=5.18 $\mu\text{mol/L}$) for HDL particle concentration.¹²⁴

Subclasses of lipoproteins can be categorized by either particle size or density. Previous studies on the association of size-determined lipoprotein subclasses with CHD risk focused mainly on LDL and HDL.^{65,95,124} Consistent with the present report, these studies did not find differences in the strength of associations across LDL subclasses.⁹⁵ However, the evidence for the associations of CHD risk with HDL subclasses has been less consistent, with some describing heterogeneity and others not.^{65,95,124} The reason for this is unclear but may reflect different sets of adjustments; studies that have further adjusted for LDL concentration tend to find little evidence of heterogeneity among HDL subclasses.^{65,124} Similarly, although HDL size was significantly associated with CHD risk in some studies,⁹⁵ there was little evidence of associations of LDL and HDL particle size after accounting for particle concentrations.^{68,124} One possible explanation is that although smaller particles are easier to penetrate into the artery wall, the ratio of surface cholesterol to phospholipid in the particle also decreases significantly as the size becomes smaller, leading to decrease in binding affinity, and therefore making them have equal possibility to get trapped within the arterial intima.^{126,127} While some small variations in atherogenicity observed among lipoproteins in different size are more likely due to the variations in concentrations. On the other hand, small dense LDL, another subclass of LDL categorized by density, has been found to be more atherogenic than other lipoproteins.^{126,128} This is probably because small dense LDL is mostly derived from triglyceride-rich LDL, following a different hydrolysis mechanism in hypertriglyceridemic status, and its more atherogenic property is probably due to their susceptibility to chemical modification that increases their atherogenicity.^{126,129}

We also examined the effect on CHD risk with variation in lipid composition of lipoproteins, independently of lipoprotein concentration. Standard measures of plasma lipids are the sum of the lipid carried by lipoprotein particles and therefore strongly correlated with particle concentration, which limits the potential to assess their independent effects. In contrast to previous studies, we divided the lipid concentration by the corresponding lipoprotein particle concentration to examine the associations with average lipid composition (molecules per particle) for each type of lipoprotein, which is not as strongly correlated with lipoprotein concentration. Associations with lipid composition showed substantial attenuation on adjustment for potential confounders and particle concentrations, with the exception of mean triglycerides molecules per LDL particle. Although large observational studies have demonstrated a strong positive association of LDL-C and remnant lipoprotein particle (composed primarily of VLDL and IDL) cholesterol with cardiovascular disease,¹³⁰⁻¹³² in the present report there was little association of mean molecules of cholesterol (either esterified cholesterol or free cholesterol or adding both) in any type of lipoprotein, independently of particle concentrations.

The consistency between biomarkers measured by NMR and by clinical chemistry has been shown in three large cohorts, including UK Biobank.^{58,109} Before the application of NMR, a number of previous studies have indirectly assessed the effect of lipoprotein particle composition on CHD risk by comparing the strengths of associations of ApoB-containing particles concentration and those with standard measures of plasma lipids.^{42,73,122,123,133} These studies have found similar associations of total ApoB concentration, LDL-C and non-HDL-C to CHD risk. However, few have directly assessed the

relevance of particle composition independently of particle concentration. A recent cohort study also found that the association with myocardial infarction was best captured by the number of ApoB-containing lipoproteins, independent of standard measures of lipid content.⁴⁵ Furthermore, two Mendelian randomisation studies have supported the causality of the relation between ApoB and CHD risk, but not with plasma cholesterol independently of total ApoB concentration.^{134,135} It has been suggested that cholesterol enters the arterial wall within ApoB-containing particles, and as such the number of ApoB-containing particles within the lumen of an artery may be the primary determinant of atherosclerosis.¹³⁶

In this study, total ApoB concentration was strongly correlated with LDL concentration, but less strongly correlated with VLDL. Total ApoB concentration may only capture the variation in LDL concentration, but fail to adequately capture the variations in other ApoB-containing lipoproteins with much lower concentrations. The strong association of VLDL concentration given ApoB, indicates that VLDL concentration is relevant to CHD risk over and above that of total ApoB concentration alone, and highlights the importance of using direct measures of lipoprotein concentration to assess the independent effects of other lipoprotein characteristics.

Accumulating epidemiologic and genetic evidence has supported raised triglycerides as an additional cause of cardiovascular disease,^{43,44,137} while uncertainty still existed on whether triglycerides provides additional information on vascular risk independently of cholesterol or ApoB.^{42,130} The results showed little evidence for an association with mean triglyceride molecules levels in

VLDL particle, but positive associations with triglycerides in LDL. This finding is consistent with a recent genetic study showing that triglycerides in LDL are associated with increased CHD risk independently of cholesterol.¹³⁸ One hypothesis is that excess triglycerides in LDL, generated in hypertriglyceridemic states, will finally convert to small, dense LDL which was found to be closely associated with increased CVD risk.^{128,129,138} Given lipoprotein concentrations, the results found that the addition of single measurement of LDL triglyceride to the model improved the LR χ^2 statistic, whereas little evidence of improvement was shown on addition of total triglycerides. Discordance analysis also showed that after accounting for HDL-C and non-HDL-C (as measured by NMR) or lipoprotein concentrations, LDL triglyceride remained significantly associated with CHD risk, while the association of total triglycerides was largely accounted for by the concentration of VLDL particles. These differences may partly explain the previous inconsistency in the evidence on triglycerides, which mostly used total triglycerides concentration as the exposure. Therefore, the findings indicated the value of disaggregating the standard measures of total triglycerides by type of lipoprotein particle. Further work is necessary to assess causality, but if causal, triglycerides in LDL might become a valuable additional biomarker of CHD risk.

Clinical Perspectives

NMR spectroscopy provides a wealth of information about lipoprotein particles, including their concentration, size and content of lipid. Separate measurement of VLDL and LDL particle concentrations, by NMR, may be of greater value than the single measurement by ApoB, which was largely determined by LDL concentration alone. Meanwhile, the findings also indicated that triglycerides in

LDL might become a valuable additional biomarker of CHD risk. The practicalities of performing this new method in a typical clinical setting, however, (including logistics and cost-effectiveness) remain to be explored.

Strengths and limitations

This study has a number of key strengths. This is one of few large-scale studies to assess robustly the independent associations of lipoprotein particle concentration, size and composition. Although multiple metabolite measurements allowing for assessment of within-person variation were not available for all participants, the study conducted a resurvey to allow correction for regression dilution, enabling estimates of associations with long-term average levels of lipoprotein characteristics. Furthermore, linkage to NHS electronic health records and national death registries limited loss to follow-up and allowed reliable ascertainment of CHD events. However, it is not possible to exclude the potential for residual confounding (because of confounders not measured or measured with random noise) or reverse causality in observational studies. Lack of measurement of the lipid density of lipoprotein limited a deeper exploration of relevant mechanisms, which needs to be considered when additional data on metabolomics are released by UK Biobank. Finally, although UK Biobank is not representative of the UK population, it has been shown that associations with health outcomes can be generalized. Future analyses should further assess the effect modification by other metabolic factors and also the causality of these associations, including using Mendelian randomisation approaches.

6.9 Conclusion

This large-scale prospective study utilizes NMR to quantify the independent associations of lipoprotein particle concentration, size and lipid composition with CHD risk. CHD risk was most strongly related to VLDL, LDL and HDL particle concentrations, but there was some additional effect from mean triglyceride molecules per LDL particle. The results also suggest that separate measurement of VLDL and LDL particles by NMR method may be superior to measurement of total ApoB levels by the same method, which was found to be highly correlated with LDL particle concentration alone.

Chapter 6 Tables and Figures

Tables	Title
Table 6.1	Baseline characteristics by incident coronary heart disease
Table 6.2	Lipoprotein and apolipoprotein concentrations vs coronary heart disease risk
Table 6.3	Lipoprotein particle concentration vs coronary heart disease risk, by particle subclass
Table 6.4	Lipoprotein particle size vs coronary heart disease risk
Table 6.5	Lipoprotein particle composition vs coronary heart disease risk
Table 6.6	Lipids and apolipoproteins vs coronary heart disease risk, by clinical chemistry and NMR spectroscopy
Table 6.7	Lipids vs coronary heart disease risk

Figures	Title
Figure 6.1	Lipoprotein particle concentration vs coronary heart disease risk Regression
Figure 6.2	LDL particle and apolipoprotein B concentration vs coronary heart disease risk
Figure 6.3	VLDL particle and apolipoprotein B concentration vs coronary heart disease risk
Figure 6.4	Lipoprotein size vs coronary heart disease risk, independently of lipoprotein particle concentration
Figure 6.5	Lipoprotein composition of triglycerides vs coronary heart disease risk, independently of lipoprotein particle concentration
Figure 6.6	Total triglycerides and triglycerides in LDL vs coronary heart disease risk, independently of lipoprotein particle concentration

Table 6.1 Baseline characteristics by incident coronary heart disease

	Incident coronary heart disease		All
	No	Yes	
No. of participants	82,303	4,196	86,499
Sociodemographic factors			
Baseline age, years	55.1 (8.0)	59.4 (7.1)	55.2 (8.0)
Men, %	41.7	63.8	42.8
White, %	94.8	95.5	94.9
University education, %	41.0	31.2	40.5
Townsend Deprivation Index*	-1.4 (3.0)	-1.2 (3.1)	-1.4 (3.0)
Lifestyle factors			
Current smoker, %	9.9	15.7	10.2
Current regular alcohol drinker, %	70.8	67.9	70.6
Anthropometry			
Body Mass Index, kg/m ²	26.9 (4.6)	27.9 (4.6)	27.0 (4.6)
Waist circumference, cm	88.4 (12.8)	93.7 (12.5)	88.7 (12.9)
Waist to hip ratio	0.86 (0.09)	0.90 (0.08)	0.86 (0.09)
Lipids and apolipoproteins measured by clinical chemistry			
LDL cholesterol, mmol/l	3.7 (0.8)	3.9 (0.8)	3.7 (0.8)
HDL cholesterol, mmol/l	1.5 (0.4)	1.3 (0.3)	1.5 (0.4)
Total triglycerides, mmol/l	1.7 (1.0)	2.0 (1.1)	1.7 (1.0)
Apolipoprotein A-I, g/l	1.5 (0.3)	1.5 (0.3)	1.5 (0.3)
Apolipoprotein B, g/l	1.1 (0.2)	1.1 (0.2)	1.1 (0.2)
Blood pressure and diabetes			
Systolic blood pressure, mmHg	136.3 (18.4)	144.6 (18.8)	136.7 (18.5)
Baseline diabetes, %	1.9	4.3	2.0
Fasting time, h	3.0 (2.0)	3.0 (1.0)	3.0 (2.0)

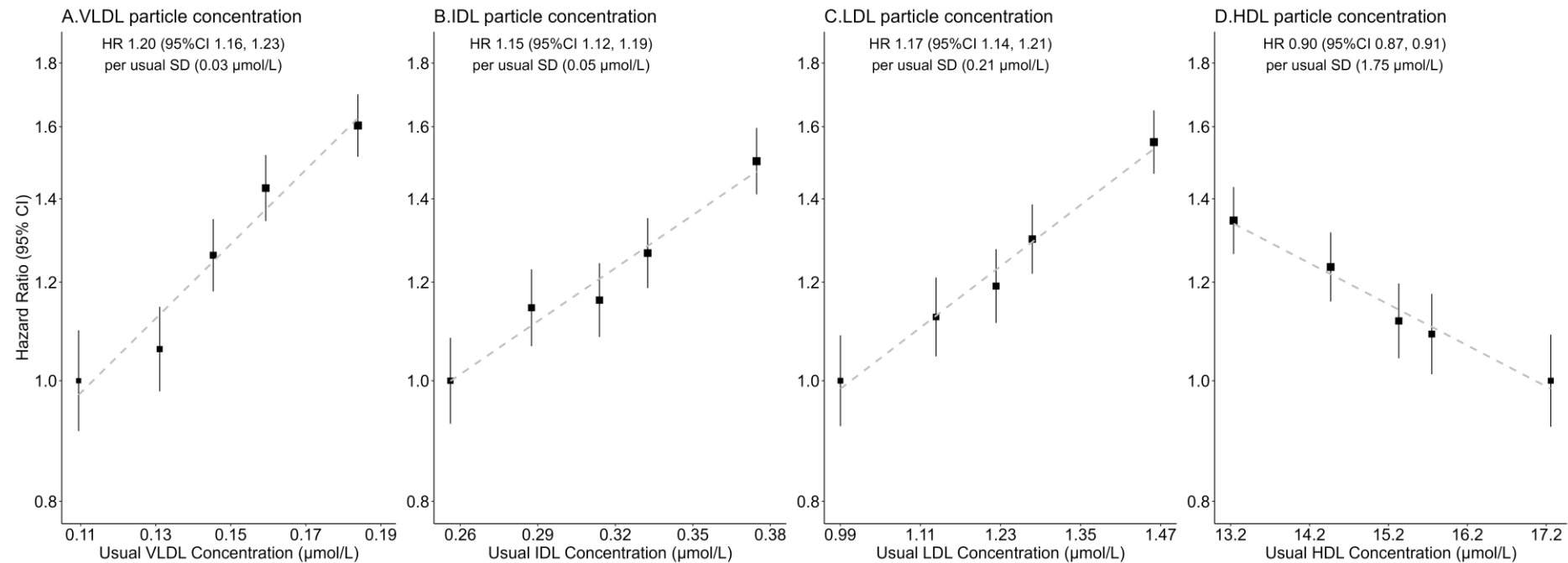
Baseline characteristics of those with and those without incident coronary heart disease during follow up. Analyses exclude those with missing or outlying values for metabolites or key covariates, and those with prior coronary heart disease or taking statins at baseline. Continuous variables are presented as mean (standard deviation) (except for skew variable fasting time, presented as median [quartile range]), and categorical variables are presented as column percentages. *Area-level measure of material deprivation. LDL=low-density lipoproteins; HDL=high-density lipoproteins.

Table 6.2 Lipoprotein and apolipoprotein concentrations vs coronary heart disease risk

	Concentration ($\mu\text{mol/L}$)		A. Adjusted for age and sex only		B. Adjusted for age, sex and other confounders	
	Baseline mean	Usual SD	HR (95% CI)	LR χ^2	HR (95% CI)	LR χ^2
Lipoproteins						
VLDL	0.15	0.03	1.25 (1.22,1.29)	216.3	1.20 (1.16,1.23)	125.7
IDL	0.32	0.05	1.17 (1.13,1.21)	98.2	1.15 (1.12,1.19)	78.6
LDL	1.23	0.21	1.20 (1.16,1.24)	137.8	1.17 (1.14,1.21)	104.1
HDL	15.12	1.75	0.85 (0.82,0.88)	85.9	0.90 (0.87,0.91)	48.2
Apolipoproteins						
ApoA-1	1.45	0.19	0.82 (0.79,0.85)	124.8	0.88 (0.85,0.91)	48.2
ApoB	0.87	0.15	1.21 (1.17,1.25)	148.6	1.18 (1.14,1.22)	109.5

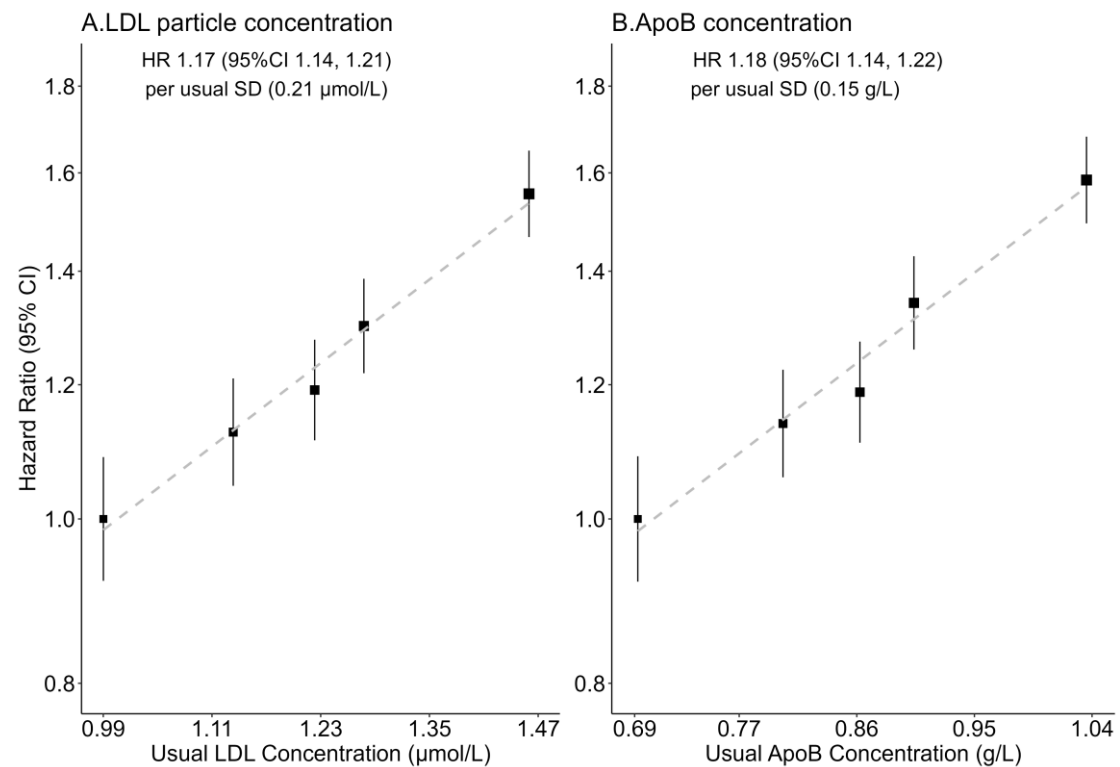
Hazard ratios (HR) per usual SD higher level of particle concentration. HRs calculated by Cox proportional-hazards models with: (A) stratification by age and sex; (B) stratification by age and sex, and adjusted for ethnicity, education, region, Townsend Deprivation Index, smoking, alcohol, and body-mass index. Likelihood ratio (LR) χ^2 improvement with the addition of the exposure variable to the model with stated adjustments. 'Usual SD' was obtained from the baseline SD through multiplication by square root of the regression dilution ratio, representing an unbiased estimate of the long-term average level in each baseline-defined group. VLDL=very-low-density lipoproteins; IDL=intermediate-density lipoproteins; LDL=low-density lipoproteins; HDL=high-density lipoproteins.

Figure 6.1 Lipoprotein particle concentration vs coronary heart disease risk



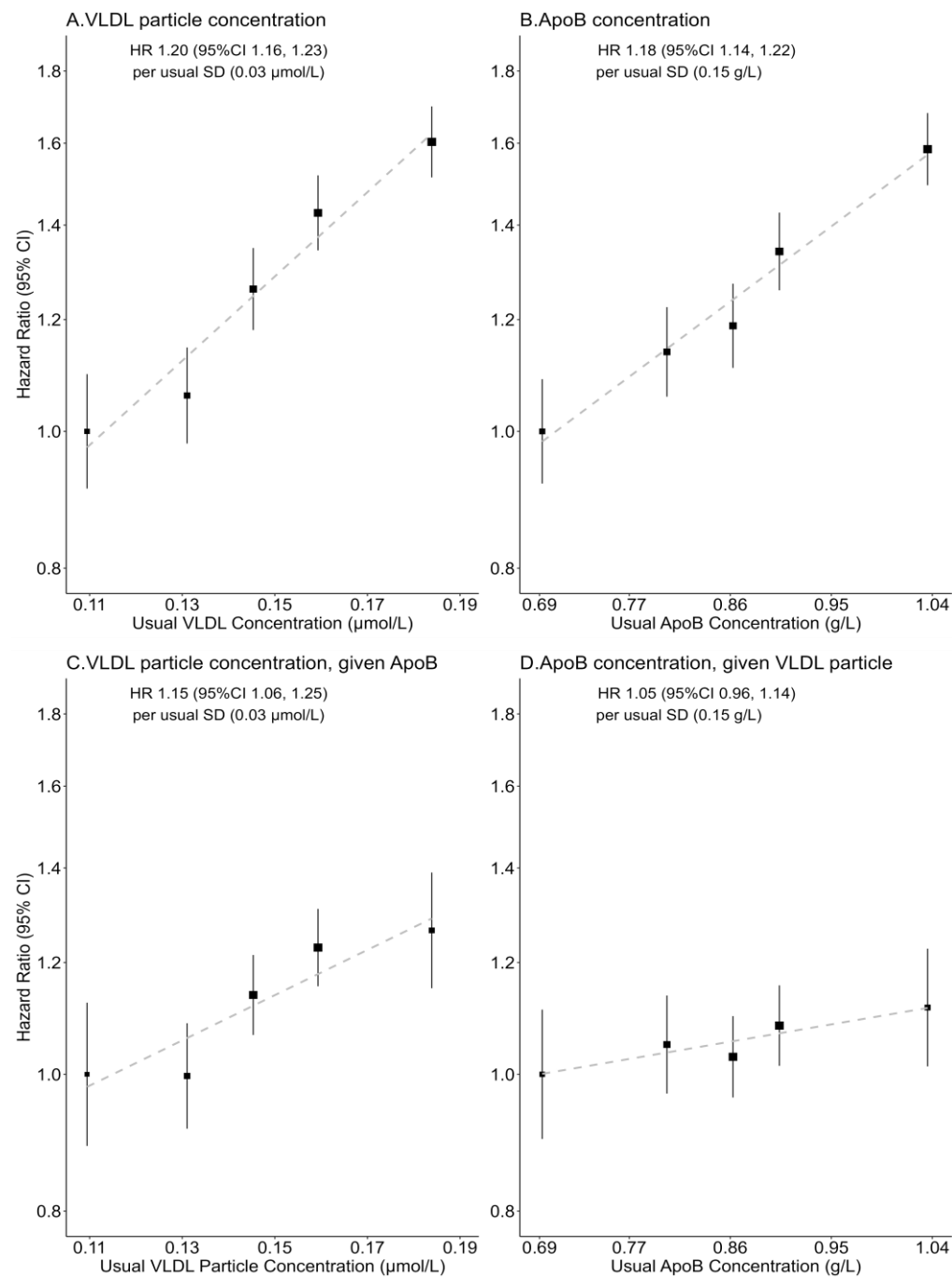
Hazard ratios (HR) per usual SD higher level of particle concentration (stratified by quintiles). HRs calculated by Cox proportional-hazards models, stratified by age and sex, and adjusted for ethnicity, education, region, Townsend Deprivation Index, smoking, alcohol, and body-mass index. 'Usual' levels are estimated from the mean value at resurvey within each baseline defined group, providing an unbiased estimate of the long-term average level in each baseline-defined group. 'Usual SD' was obtained from the baseline SD through multiplication by square root of the regression dilution ratio. Area of the square is inversely proportional to the variance of the category-specific log risk. VLDL=very-low-density lipoproteins; IDL=intermediate-density lipoproteins; LDL=low-density lipoproteins; HDL=high-density lipoproteins.

Figure 6.2 LDL particle and apolipoprotein B concentration vs coronary heart disease risk



Hazard ratios (HR) per usual SD higher level of particle concentration (all biomarkers in the figure were measured by NMR). HRs are calculated by Cox proportional-hazards models, stratified by age and sex, and adjusted for ethnicity, education, region, Townsend Deprivation Index, smoking, alcohol, and body-mass index. 'Usual' levels are estimated from the mean value at resurvey within each baseline defined group, providing an unbiased estimate of the long-term average level in each baseline-defined group. 'Usual SD' was obtained from the baseline SD through multiplication by square root of the regression dilution ratio. Area of the square is inversely proportional to the variance of the category-specific log risk. ApoB=Apolipoprotein B; LDL=Low-density lipoproteins.

Figure 6.3 VLDL particle and apolipoprotein B concentration vs coronary heart disease risk



Hazard ratios (HR) per usual SD higher level of particle concentration (all biomarkers in the figure were measured by NMR). HRs estimated by the stratified Cox proportional-hazards model with: (A,B) stratification by age, sex, and further adjusted for ethnicity, education, region, Townsend Deprivation Index, smoking, alcohol, body mass index; (C) model A with further adjustment for ApoB concentration; (D) model B with further adjustment for VLDL concentration. 'Usual' levels are estimated from the mean value at resurvey within each baseline defined group, providing an unbiased estimate of the long-term average level in each baseline-defined group. 'Usual SD' was obtained from the baseline SD through multiplication by square root of the regression dilution ratio. Area of the square is inversely proportional to the variance of the category-specific log risk. ApoB=Apolipoprotein B; VLDL=Very low-density lipoproteins.

Table 6.3 Lipoprotein particle concentration vs coronary heart disease risk, by particle subclass

Lipoprotein subclass	Particle concentration (μmol/L)		A. Adjusted for age and sex only		B. Adjusted for age, sex and other confounders	
	Baseline mean	Usual SD	HR (95% CI)	LR χ^2	HR (95% CI)	LR χ^2
VLDL						
Extremely large	0.002	0.001	1.17 (1.14,1.20)	118.4	1.09 (1.06,1.13)	34.9
Very large	0.003	0.002	1.22 (1.18,1.25)	173.6	1.14 (1.11,1.18)	68.6
Large	0.010	0.004	1.22 (1.19,1.26)	178.8	1.15 (1.12,1.19)	76.8
Medium	0.036	0.009	1.23 (1.19,1.26)	170.5	1.18 (1.14,1.21)	104.0
Small	0.040	0.010	1.25 (1.21,1.29)	212.6	1.19 (1.15,1.22)	118.3
Very small	0.057	0.010	1.24 (1.20,1.28)	187.3	1.20 (1.16,1.23)	125.7
IDL	0.323	0.054	1.17 (1.13, 1.21)	98.2	1.15 (1.12, 1.19)	78.6
LDL						
Large	0.755	0.128	1.19 (1.16,1.23)	128.6	1.17 (1.14,1.21)	103.4
Medium	0.302	0.054	1.18 (1.15,1.22)	116.2	1.15 (1.12,1.19)	79.4
Small	0.174	0.028	1.22 (1.18,1.26)	164.1	1.18 (1.15,1.22)	112.7
HDL						
Very large	0.243	0.081	0.84 (0.81,0.88)	80.2	0.94 (0.90,0.98)	10.0
Large	1.451	0.677	0.75 (0.72,0.78)	212.3	0.84 (0.80,0.87)	69.6
Medium	3.836	0.707	0.81 (0.78,0.84)	145.0	0.86 (0.83,0.89)	63.8
Small	9.587	0.893	1.00 (0.96,1.03)	0.1	0.99 (0.96,1.02)	0.5

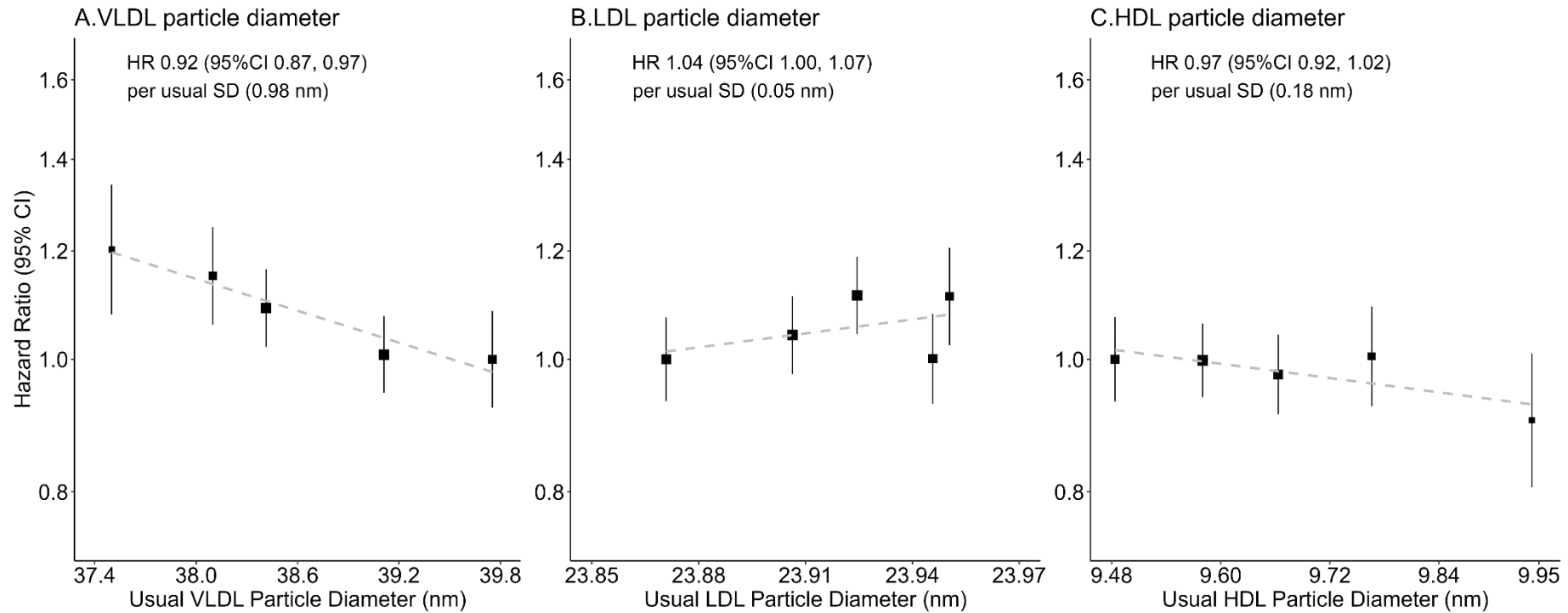
Hazard ratios (HR) per usual SD higher level of particle concentration. HRs calculated by Cox proportional-hazards models with: (A) stratification by age and sex; (B) stratification by age and sex, and adjusted for ethnicity, education, region, Townsend Deprivation Index, smoking, alcohol, and body-mass index. Likelihood ratio (LR) χ^2 improvement with the addition of the exposure variable to the model with stated adjustments. 'Usual SD' was obtained from the baseline SD through multiplication by square root of the regression dilution ratio, representing an unbiased estimate of the long-term average level in each baseline-defined group. VLDL=very-low-density lipoproteins; IDL=intermediate-density lipoproteins; LDL=low-density lipoproteins; HDL=high-density lipoproteins.

Table 6.4 Lipoprotein particle size vs coronary heart disease risk

Lipoprotein type	Particle diameter (nm)		A. Adjusted for age and sex only		B. Adjusted for age, sex and other confounders		C. Further adjusted for lipoprotein particle concentration	
	Baseline mean	Usual SD	HR (95% CI)	LR χ^2	HR (95% CI)	LR χ^2	HR (95% CI)	LR χ^2
VLDL	38.5	1.0	1.17 (1.13,1.21)	90.4	1.08 (1.04,1.12)	19.2	0.92 (0.87,0.97)	9.9
LDL	23.9	<0.1	0.98 (0.95,1.01)	1.4	1.02 (0.99,1.05)	2.1	1.04 (0.99,1.07)	3.7
HDL	9.7	0.2	0.77 (0.74,0.80)	198.8	0.85 (0.81,0.88)	65.1	0.97 (0.92,1.02)	1.7

Hazard ratios (HR) per usual SD higher level of average particle diameter. HRs calculated by Cox proportional-hazards models with: (A) stratification by age and sex; (B) stratification by age and sex, and adjusted for ethnicity, education, region, Townsend Deprivation Index, smoking, alcohol, and body-mass index; (C) model B with further adjusted for concentrations of VLDL, LDL and HDL particles. Likelihood ratio (LR) χ^2 improvement with the addition of the exposure variable to the model with stated adjustments. 'Usual SD' was obtained from the baseline SD through multiplication by square root of the regression dilution ratio, representing an unbiased estimate of the long-term average level in each baseline-defined group. Particle diameter of IDL is unavailable in Nightingale NMR platform. VLDL=very-low-density lipoproteins; LDL=low-density lipoproteins; HDL=high-density lipoproteins.

Figure 6.4 Lipoprotein size vs coronary heart disease risk, independently of lipoprotein particle concentration



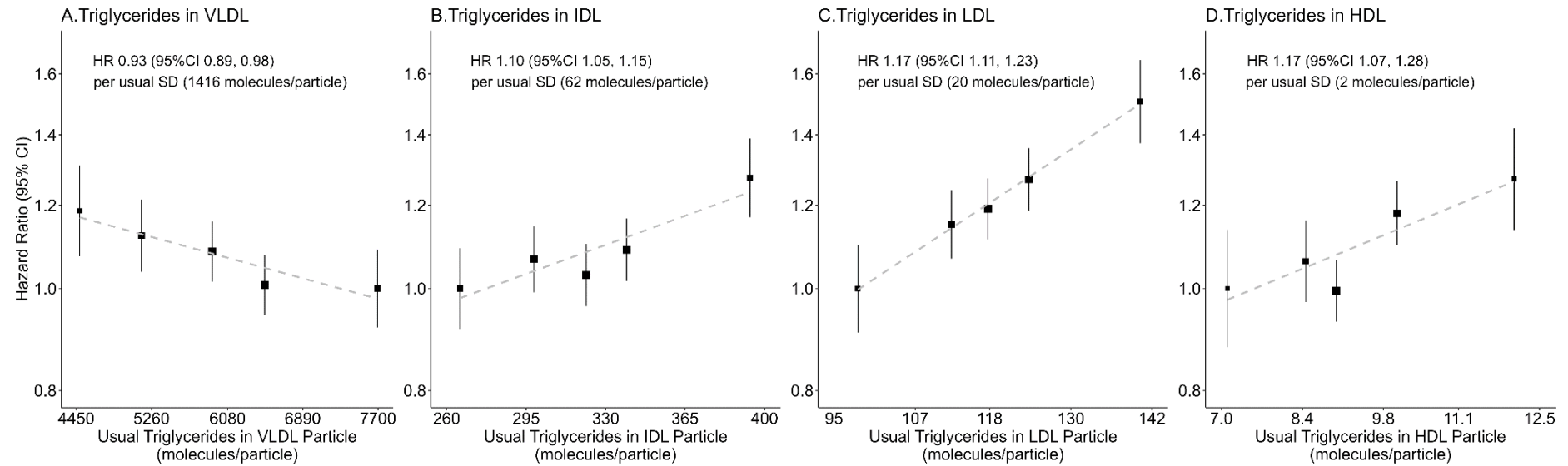
Hazard ratios (HR) per usual SD higher level of particle diameter (stratified by quintiles). HRs are estimated by the stratified Cox proportional-hazards model by age, sex, and further adjusted for ethnicity, education, region, Townsend Deprivation Index, smoking, alcohol, body mass index, concentrations of VLDL, LDL and HDL particles. 'Usual' levels are estimated from the mean value at resurvey within each baseline defined group, providing an unbiased estimate of the long-term average level in each baseline-defined group. 'Usual SD' was obtained from the baseline SD through multiplication by square root of the regression dilution ratio. Area of the square is inversely proportional to the variance of the category-specific log risk. VLDL=very-low-density lipoproteins; LDL=Low-density lipoproteins; HDL=High-density lipoproteins.

Table 6.5 Lipoprotein particle composition vs coronary heart disease risk

Lipoprotein type	Particle composition (molecules/particle)		A. Adjusted for age and sex only		B. Adjusted for age, sex and other confounders		C. Further adjusted for lipoprotein particle concentration	
	Baseline mean	Usual SD	HR (95% CI)	LR χ^2	HR (95% CI)	LR χ^2	HR (95% CI)	LR χ^2
VLDL								
Cholesterol	5023.1	230.1	1.09 (1.06,1.13)	29.5	1.09 (1.05,1.13)	25.1	0.97 (0.92,1.02)	1.6
Cholesteryl esters	3034.1	181.1	1.00 (0.97,1.03)	<0.1	1.04 (1.01,1.07)	5.5	1.00 (0.96,1.05)	<0.1
Free cholesterol	1988.9	129.8	1.21 (1.17,1.26)	118.5	1.13 (1.09,1.17)	40.9	0.93 (0.88,0.98)	8.3
Phospholipids	3188.5	272.2	1.22 (1.18,1.26)	137.1	1.12 (1.08,1.16)	38.9	0.95 (0.90,1.01)	3.1
Triglycerides	5875.6	1415.8	1.10 (1.07,1.13)	37.6	1.02 (0.99,1.06)	1.8	0.93 (0.89,0.98)	7.2
IDL								
Cholesterol	2717.6	183.2	0.79 (0.77,0.82)	193.9	0.87 (0.84,0.90)	66.1	0.97 (0.91,1.03)	1.2
Cholesteryl esters	2002.7	138.0	0.79 (0.77,0.82)	195.5	0.86 (0.84,0.89)	69.8	0.96 (0.90,1.02)	1.5
Free cholesterol	714.9	48.5	0.83 (0.81,0.86)	130.0	0.90 (0.87,0.93)	35.7	0.99 (0.94,1.04)	0.2
Phospholipids	945.8	56.6	0.89 (0.86,0.91)	54.8	0.95 (0.91,0.98)	10.9	1.01 (0.98,1.05)	0.6
Triglycerides	312.9	61.9	1.10 (1.07,1.12)	41.6	1.05 (1.02,1.08)	10.4	1.10 (1.05,1.15)	14.8
LDL								
Cholesterol	1446.0	69.2	0.86 (0.83,0.89)	89.3	0.91 (0.88,0.94)	32.3	1.05 (0.97,1.12)	1.5
Cholesteryl esters	1055.3	43.7	0.89 (0.86,0.92)	52.0	0.93 (0.90,0.96)	21.1	1.04 (0.97,1.10)	1.2
Free cholesterol	390.7	32.2	0.84 (0.82,0.87)	123.9	0.90 (0.87,0.93)	39.7	1.04 (0.96,1.13)	1.1
Phospholipids	506.3	21.7	0.86 (0.84,0.89)	85.4	0.90 (0.87,0.93)	45.8	1.07 (1.01,1.13)	6.1
Triglycerides	117.8	20.4	1.10 (1.07,1.13)	46.3	1.05 (1.02,1.08)	9.3	1.17 (1.11,1.23)	33.9
HDL								
Cholesterol	87.0	9.9	0.75 (0.72,0.78)	242.5	0.83 (0.79,0.86)	85.9	0.96 (0.91,1.01)	2.1
Cholesteryl esters	67.6	7.8	0.74 (0.71,0.76)	283.4	0.81 (0.78,0.84)	111.1	0.93 (0.88,0.99)	4.9
Free cholesterol	19.4	2.2	0.85 (0.82,0.89)	72.9	0.94 (0.90,0.98)	9.8	1.01 (0.97,1.05)	0.2
Phospholipids	102.0	6.9	0.87 (0.84,0.90)	68.4	0.91 (0.88,0.95)	25.6	1.02 (0.97,1.07)	0.7
Triglycerides	9.5	2.4	1.21 (1.18,1.25)	170.8	1.14 (1.10,1.17)	70.3	1.17 (1.07,1.28)	12.1

Hazard ratios (HR) per usual SD higher level of average molecules of the lipid per lipoprotein particle. HRs calculated by Cox proportional-hazards models with: (A) stratification by age and sex; (B) stratification by age and sex, and adjusted for ethnicity, education, region, Townsend Deprivation Index, smoking, alcohol, and body-mass index; (C) model B with further adjusted for concentrations of VLDL, LDL and HDL particles (no further adjusted for IDL concentration because its high correlation with LDL concentration). Likelihood ratio (LR) χ^2 improvement with the addition of the exposure variable to the model with stated adjustments. 'Usual SD' was obtained from the baseline SD through multiplication by square root of the regression dilution ratio, representing an unbiased estimate of the long-term average level in each baseline-defined group. VLDL=very-low-density lipoproteins; LDL=low-density lipoproteins; HDL=high-density lipoproteins.

Figure 6.5 Lipoprotein composition of triglycerides vs coronary heart disease risk, independently of lipoprotein particle concentration



Hazard ratios (HR) per usual SD higher level of particle composition (stratified by quintiles). HRs are estimated by the stratified Cox proportional-hazards model by age, sex, and further adjusted for ethnicity, education, region, Townsend Deprivation Index, smoking, alcohol, body mass index, concentration of VLDL, LDL and HDL particles. 'Usual' levels are estimated from the mean value at resurvey within each baseline defined group, providing an unbiased estimate of the long-term average level in each baseline-defined group. 'Usual SD' was obtained from the baseline SD through multiplication by square root of the regression dilution ratio. Area of the square is inversely proportional to the variance of the category-specific log risk. VLDL=very-low-density lipoproteins; IDL=intermediate-density lipoproteins; LDL=Low-density lipoproteins; HDL=high-density lipoproteins.

Table 6.6 Lipids and apolipoproteins vs coronary heart disease risk, by clinical chemistry and NMR spectroscopy

Lipids/ Apolipoproteins	Clinical chemistry				NMR spectroscopy			
	Baseline mean	Usual SD	HR (95% CI)	LR χ^2	Baseline mean	Usual SD	HR (95% CI)	LR χ^2
Lipids (mmol/L)								
Total Cholesterol	5.90	0.82	1.14 (1.11,1.18)	59.6	4.72	0.64	1.09 (1.05,1.13)	23.9
Non-HDL Cholesterol	4.42	0.80	1.21 (1.17,1.25)	116.5	3.40	0.59	1.14 (1.11,1.18)	62.2
LDL Cholesterol ^s	3.72	0.62	1.19 (1.15,1.23)	101.6	1.78	0.30	1.12 (1.09,1.16)	46.6
HDL Cholesterol	1.47	0.33	0.81 (0.78,0.85)	87.6	1.32	0.26	0.84 (0.80,0.87)	64.9
Triglycerides	1.69	0.71	1.10 (1.07,1.14)	36.6	1.28	0.44	1.11 (1.08,1.15)	42.6
Apolipoproteins (g/L)								
ApoA-1	1.55	0.22	0.85 (0.82,0.89)	60.4	1.44	0.18	0.88 (0.85,0.92)	38.4
ApoB	1.07	0.19	1.22 (1.18,1.26)	135.5	0.87	0.15	1.17 (1.13,1.21)	85.2

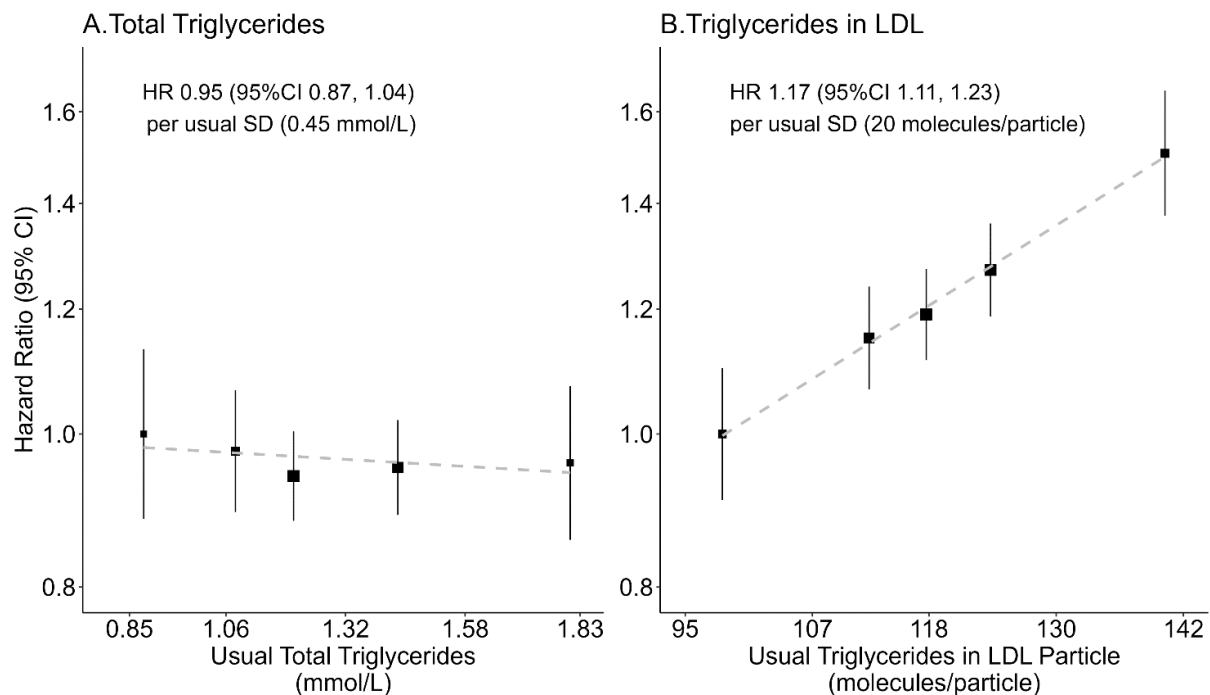
Hazard ratios (HR) per usual SD higher level of concentration among 74,741 participants with both NMR and clinical chemistry measures. HRs calculated by Cox proportional-hazards models, stratified by age and sex, and adjusted for ethnicity, education, region, Townsend Deprivation Index, smoking, alcohol, and body-mass index. Likelihood ratio (LR) χ^2 improvement with the addition of the exposure variable to the model with stated adjustments. 'Usual SD' was obtained from the baseline SD through multiplication by square root of the regression dilution ratio, representing an unbiased estimate of the long-term average level in each baseline-defined group. ^sThe biomarker of LDL-cholesterol measured by NMR spectroscopy was 'size specific' which is only for the cholesterol carried in LDL particles, while for LDL-cholesterol measured by clinical chemistry assay, the measurement also includes cholesterol in IDL and small VLDL particles. LDL=low-density lipoproteins; HDL=high-density lipoproteins; ApoA-1=Apolipoprotein A-1; ApoB=Apolipoprotein B

Table 6.7 Lipids vs coronary heart disease risk

Lipids (mmol/L)	Concentration		A. Adjusted for age and sex only		B. Adjusted for age, sex and other confounders		C. Further adjusted for lipoprotein particle concentration	
	Baseline mean	Usual SD	HR (95% CI)	LR χ^2	HR (95% CI)	LR χ^2	HR (95% CI)	LR χ^2
Total cholesterol	4.73	0.65	1.07 (1.04,1.10)	17.7	1.09 (1.06,1.13)	30.2	1.03 (0.85,1.24)	0.1
Remnant cholesterol	1.62	0.30	1.19 (1.16,1.23)	124.1	1.17 (1.13,1.20)	94.1	1.07 (0.94,1.22)	1.0
Non-HDL-cholesterol	3.40	0.60	1.17 (1.13,1.20)	94.9	1.15 (1.12,1.29)	78.6	1.22 (0.97,1.53)	3.0
LDL-cholesterol	1.78	0.31	1.13 (1.10,1.17)	64.7	1.13 (1.10,1.16)	60.0	1.23 (0.97,1.55)	3.0
HDL-cholesterol	1.33	0.27	0.76 (0.73,0.79)	214.5	0.83 (0.79,0.87)	80.1	0.93 (0.84,1.03)	2.2
Triglycerides	1.28	0.45	1.20 (1.16,1.23)	151.4	1.13 (1.09,1.16)	60.2	0.95 (0.87,1.04)	1.4

Hazard ratios (HR) per usual SD higher level of concentration (all biomarkers in this tables were measured by NMR). HRs calculated by Cox proportional-hazards models with: (A) stratification by age and sex; (B) stratification by age and sex, and adjusted for ethnicity, education, region, Townsend Deprivation Index, smoking, alcohol, and body-mass index; (C) model B with further adjusted for concentrations of VLDL, LDL and HDL particles. Likelihood ratio (LR) χ^2 improvement with the addition of the exposure variable to the model with stated adjustments. 'Usual SD' was obtained from the baseline SD through multiplication by square root of the regression dilution ratio, representing an unbiased estimate of the long-term average level in each baseline-defined group. LDL=low-density lipoproteins; HDL=high-density lipoproteins.

Figure 6.6 Total triglycerides and triglycerides in LDL vs coronary heart disease risk, independently of lipoprotein particle concentration



Hazard ratios (HR) per usual SD higher level of triglycerides (stratified by quintiles). Total triglyceride is triglyceride concentration in all lipoprotein particles, and triglyceride in LDL is average triglyceride molecules per LDL particle. HRs calculated by Cox proportional-hazards models stratified by age, sex, and adjusted for ethnicity, education, region, Townsend Deprivation Index, smoking, alcohol, body mass index, and concentration of VLDL, LDL and HDL particles. 'Usual' levels are estimated from the mean value at resurvey within each baseline defined group, providing an unbiased estimate of the long-term average level in each baseline-defined group. 'Usual SD' was obtained from the baseline SD through multiplication by square root of the regression dilution ratio. Area of the square is inversely proportional to the variance of the category-specific log risk. LDL=Low-density lipoproteins.

CHAPTER 7: Prospective association of fatty acids with CHD risk

Chapter summary

The role of fatty acids in CHD remains uncertain. There is little evidence from large-scale epidemiological studies on the relevance of circulating fatty acids levels to CHD risk. This chapter examines the independent associations of the major circulating types of fatty acids with CHD risk. CHD risk was positively associated with circulating SFA and MUFA, inversely related to omega-3 PUFA, and unrelated to omega-6 PUFA. After correcting for regression dilution, the strength of the associations were similar for SFA and for MUFA. The study also indicates that although the associations of circulating SFA and MUFA were mostly attributed to their relationship with triglycerides levels, the inverse association with omega-3 PUFA is unlikely to be mediated by major lipid fractions.

7.1 Background

Dietary guidelines commonly recommend reducing total fat intake, and replacing saturated fat with polyunsaturated fat to lower CVD risk.^{46,139,140} These guidelines are based largely on evidence from randomized controlled trials of dietary intake of fats,^{141,142} but the effects of the different types of polyunsaturated fats (particularly omega-3 and omega-6), or of replacement of saturated fats with monounsaturated fats, remains unclear.^{141,143,144}

Fat consumption is known to affect circulating levels of fatty acids, which are the main constitutional component of circulating lipid classes and has been shown to modulate lipid metabolism.⁴⁶⁻⁴⁸ Although the importance of blood lipids, including low-density-lipoprotein cholesterol and triglyceride levels, to CHD risk is well established, the strength of the associations between circulating fatty acids levels and CHD risk remains unclear. In addition, the relevance of fatty acids levels independently of blood lipids has not been well described.⁴⁹ Understanding these associations and the underlying biological mechanisms between circulating fatty acids and CHD risk as an intermediate to dietary intake is important for the development of dietary guidelines, and may inform clinical trials targeting circulating levels of particular fatty acids.

Previous observational studies of the association of circulating fatty acid levels with CHD risk tended to be small in size, likely reflecting the challenges of storing and analysing blood samples at scale, and most lacked repeated measurement of fatty acids making them prone to underestimating the associations due to regression dilution bias.^{64,120} Furthermore, most of these studies failed to account adequately for the high correlations between some fatty acids, which may affect the interpretation of findings.^{82,145} The analyses in the present report use UK Biobank, a large-scale cohort

study, to quantify reliably the associations of CHD risk with the major circulating types of fatty acids, independently of each other and major lipid fractions.

7.2 Methods

The quantified plasma fatty acids represent a combination of fatty acids in lipid fractions (ie. triglycerides, phospholipids, or cholesterol esters) and free fatty acids (also called non-esterified fatty acids).¹⁴⁶ The fatty acid biomarkers included long-chain omega-3 DHA, omega-6 linoleic acid, total omega-3 PUFA, total omega-6 PUFA, total PUFA, total MUFA, total SFA, and total fatty acids. Both the concentration of each fatty acid and the corresponding percentage (by weight) of total fatty acids were estimated. The study followed the same criteria of study population selection, outcome ascertainment and processing of metabolites that described in Chapter 3.

Cox proportional-hazards models, stratified by sex and age (in 5-year groups), were used to derive hazard ratios for the associations of fatty acids with incident CHD; hazard ratios are reported per standard deviation higher level of each fatty acid. Models were first adjusted for ethnicity, education, region, Townsend Deprivation Index,¹⁴⁷ smoking, and alcohol intake, and then further adjusted for BMI. Hazard ratios were corrected for regression dilution bias (i.e. categorising people by their baseline fatty acid and estimating the long-term average mean fatty acid in each category using the correlation between re-survey and baseline measurements), and are therefore described as associations of usual fatty acids with CHD risk.^{64,120} The values of the regression dilution ratios are described in Chapter 5. Confidence intervals were calculated using the variance of the log risk, which appropriately attributes variance to all groups, including the reference.^{121,148}

To examine the independent association of each fatty acid, other fatty acids were progressively added to the model. To explore whether the association of fatty acids with CHD risk was independent of lipids, we further adjusted for plasma cholesterol (non-high-density lipoprotein [non-HDL] cholesterol and HDL cholesterol) and for plasma triglyceride levels. Changes in the likelihood ratio χ^2 statistic between nested models provide measures of the extent of variance explained by the fatty acid in addition to the other variables in the model. This statistic provides a significance test for the improvement in fit from including the fatty acid term.

Sensitivity analyses were conducted by excluding CHD events in the first two years of follow-up (to assess for potential reverse causality), and by further adjusting for other potential confounders, including dietary-related factors. Supplementary analyses on the ratio biomarkers (i.e. the percentage of different types of fatty acid to total fatty acids) were also assessed.

7.3 Association of fatty acids

The demographics of the study population are same as Chapter 6. As described in Chapter 4, the concentration of SFA at baseline was highly correlated with the concentration of MUFA (Spearman correlation, $r=0.93$). The concentration of omega-6 PUFA was moderately correlated with the concentration of SFA ($r=0.75$) and of MUFA ($r=0.73$), while omega-3 PUFA had lower correlations with the other types of fatty acids ($r=0.38$ to 0.47) (**Table 4.3**).

There were linear associations of usual levels of all the fatty acids with incident CHD. Both circulating SFA and MUFA had strong positive associations with incident CHD in fully adjusted models (HR per usual SD, 1.11 [95% CI 1.08-1.14] and 1.12 [1.09-1.15], respectively: **Table 7.1**), although the adjustment of BMI slightly attenuated the

associations (**S7.1**). Furthermore, the positive linear associations of SFA and MUFA remained largely unchanged when independently of other types of fatty acids (omega-3 and omega-6 PUFA) (**Figure 7.1** and **7.2**) (the associations of SFA and MUFA were not further adjusted for each other because of the very high correlation between these variables).

There were also positive associations of omega-6 PUFA and its major subtype, linoleic acid, with CHD risk, which was little altered by the adjustment of BMI (1.09 [1.06-1.13] for both) (**Table 7.1, S7.1**). However, the HR of omega-6 PUFA in the fully adjusted model was substantially attenuated to 1.04 (0.99-1.08) when further adjusting for SFA, and to 1.02 (0.98-1.07) when further adjusting for both SFA and MUFA (**Figure 7.1**); the large reduction (97%) in the LR χ^2 statistic (from 32.3 to 1.1) after further adjustment for SFA and MUFA indicates the positive association of omega-6 PUFA with CHD risk may be largely accounted by the effects from these other fatty acids. The independent associations of linoleic acid given other fatty acids followed a similar pattern with the independent associations of omega-6 PUFA (**Table 7.1, S7.2**).

In terms of omega-3 PUFA, there was no evidence of an association between the overall concentration with CHD risk in the fully adjusted model (1.00 [0.97-1.03]), and a slight inverse association (0.97 [0.94-1.00]) of DHA, a subtype of omega-3 PUFA (**Table 7.1**). However, the associations of omega-3 PUFA and DHA changed from null to significantly inverse in all models with progressive adjustments of other fatty acids (0.93 [0.90-0.97] and 0.95 [0.91, 0.98], respectively: **Table 7.1, Figure 7.1**). The plot of the independent association of omega-3 PUFA and DHA given other fatty acids was shown in **Figure 7.2** and **S7.2**.

7.4 Association of fatty acids independent of lipids

As described in Chapter 4, SFA and MUFA had high correlations with total triglycerides (0.87 and 0.93, respectively), and omega-6 PUFA had high correlation with non-HDL-C (0.82). The correlations between omega-3 PUFA and all lipid biomarkers were lower (0.16 to 0.40) (**Table 4.4**).

The changes of hazard ratios and LR χ^2 statistics after further adjusting for cholesterol or triglycerides are shown in **Figure 7.3**. The associations of SFA and MUFA with CHD risk were attenuated slightly after adjusting for non-HDL cholesterol and HDL cholesterol (HR per SD, 1.09 [1.04-1.14] for both), but were attenuated to null after adjusting for triglycerides (1.00 [0.92-1.08] and 1.00 [0.90-1.10], respectively). In contrast, there was no evidence that the independent association of omega-6 PUFA and omega-3 PUFA were altered following further adjustment, showing HR per usual SD of 1.05 (0.99, 1.10) and 0.93 (0.89, 0.96), respectively (**Figure 7.3**).

7.5 Sensitivity analyses

In sensitivity analyses, exclusion of the first two years of follow-up did not materially change the main associations, and neither did further adjustment for other potential confounders, including dietary habits (**S7.3**). The percentage composition of each fatty acid to total fatty acids with incident CHD was investigated in **S7.4** as a complementary analysis. The association of these percentage measures, including SFA, MUFA and omega-3 PUFA, were similar to the results of concentration biomarkers independently of other fatty acids (**S7.5**). The ratios of SFA and MUFA to total fatty acids remained associated with increased CHD risk (1.05 [1.02-1.09] and 1.14 [1.10-1.17], respectively), and higher percentage of omega-3 PUFA were associated with a decreased risk (0.94 [0.91-0.97]). However, in contrast to the results of concentration

biomarkers, an inverse association of omega-6 PUFA was observed when expressed as ratio biomarker relative to total fatty acids (0.91 [0.89-0.94]) (**S7.4**). The percentage composition of each fatty acid independently of cholesterol or triglycerides are shown in **S7.6**. Further adjustment of triglycerides attenuated the associations of both SFA and omega-6 PUFA compositions to null, and also modestly attenuated the ratio biomarker of MUFA. Similar to the analyses using concentration levels, the inverse association of omega-3 PUFA to total fatty acids remained unchanged after adjustment of cholesterol or triglycerides.

7.6 Discussion

In this large-scale prospective study, there was strong evidence of positive associations of circulating SFA and MUFA with CHD risk, independently of other fatty acids, and inverse associations of omega-3 PUFA (and DHA). The associations with omega-3 PUFA were independent of both triglyceride and cholesterol levels, while the associations of SFA and MUFA were completely attenuated following adjustment for circulating triglyceride levels. Omega-6 PUFA (and linoleic acid) showed an inverse association with CHD risk when measured as the percentage in total fatty acids, but little evidence of an association in concentration level independently of other fatty acids. This suggests that although the associations of SFA and MUFA are mostly attributed to their relationship with triglycerides, there are likely to be alternative mechanisms other than lipids lowering by which omega-3 PUFA is related to CHD risk.

This is one of the largest studies to date to quantify the associations of circulating fatty acids with CHD risk, including corrections for regression dilution bias, and assessing the strength of these associations independently of major lipid fractions. Many previous observational studies have assessed the associations of CVD risk with

circulating fatty acids expressed as ratios relative to total fatty acids (analyses using these measures are reported in the supplementary material for comparison)^{145,149,150}. However, it is challenging to infer the independent associations of fatty acids with CHD risk from such analyses, as it requires simultaneous consideration of the association of both the numerator and the denominator of the ratio. For example, in the present report we found omega-6 PUFA was unrelated to CHD risk independently of other fatty acids, but there was an inverse association with the ratio of omega-6 PUFA to total fatty acids. This finding may well be driven by the association of CHD risk with total fatty acids (largely composed of SFA and MUFA, which have a positive association with CHD risk), rather than any independent effect of omega-6 PUFA on CHD risk itself. However, both concentration- and percentage-based biomarkers have limitations in exploring the independent effect of fatty acids, and further understanding of the mechanism of fatty acids, especially for omega-6 PUFA, should consider the findings from both types of measurements, and triangulate the results from different study designs, including observational studies, trials, and Mendelian randomisation.

Our observation that both plasma SFA and MUFA were associated with higher CHD risk is consistent with a number of previous studies.^{52,82} However, a recent large meta-analysis of observational studies described a positive association of MUFA with CHD risk, but null association of SFA after adjustment for other fatty acids.⁸² The present report did not adjust the associations of SFA for MUFA, as there is evidence that levels of MUFA may be the potential mediator for the association between SFA and CHD risk, and this may explain the discrepancy in the findings; MUFAs constitute the major fatty acids stored in adipose tissue, and circulating MUFA are largely generated from desaturation of SFA.^{146,151}

In the present report, the weak positive associations of circulating omega-6 PUFA and linoleic acid with CHD risk were attenuated to the null after further controlling for other fatty acids, indicating that the association may be due to its correlation with SFA and MUFA. Similarly, a pooled analysis of six UK-based studies assessing plasma linoleic acid showed no evidence of association with CHD risk with and without adjusting for other fatty acids.⁸² Meta-analyses of RCTs and prospective cohort studies on replacing SFA dietary intake with PUFA have shown a CHD risk reduction.^{46,141,142,152,153} However, such studies which assess total PUFA have not been able to disaggregate the effects of changes in the different types of PUFA. Omega-6 linoleic acid is the most abundant dietary PUFA and increases in total dietary PUFA are likely to have resulted in high intake of omega-6 linoleic acid, but theoretically, a concurrent increase in omega-3 may have existed and complicate the interpretation of these findings.

Our study showed that circulating levels of omega-3 PUFAs were independently associated with lower CHD risk, which is consistent with numerous previous observational studies assessing independent risk of plasma fatty acids.^{82,145} Further adjustment of health lifestyle-related factors, such as dietary habit, did not materially change the protective associations in circulating level. Omega-3 PUFA has also long been hypothesized as the potential mediator of the inverse association of dietary fish intake with lower CHD risk.^{154,155} Randomized controlled trials on omega-3 PUFA supplementation, however, have not found consistent conclusions on the protective effect on CHD.¹⁵⁶⁻¹⁵⁸ Except for the concerns on patient selection criteria, the duration of treatment and the choice of placebo, the inconsistent conclusions also raised the hypothesis that the formulation of omega-3 PUFA supplementation (Eicosapentaenoic acid [EPA] plus DHA versus pure EPA) may be the key aspect of the effects to

cardiovascular event.^{49,50} Circulating levels of EPA and DHA are strongly influenced by dietary intake, and the results showed that higher circulating DHA was associated with decreased CHD risk in similar degree as omega-3 PUFA, which may not support this new hypothesis. RESPECT-EPA trial are ongoing to provide further evidence.⁵¹

Fatty acids are the main constitutional component of lipid classes,⁴⁷ and are known to modulate circulating lipids,⁴⁶ but current evidence is still unclear about how much cholesterol and triglycerides contribute to the relevance of fatty acids to CHD risk. A recent Mendelian randomisation analysis did not support a protective role of circulating PUFA with CVD risk after accounting for LDL-C, but the limitation of MR analysis made it difficult to separate the genetic determinate of PUFA from other dietary changes, and it is also difficult to avoid the bias from horizontal pleiotropy via lipoprotein-related traits.¹⁵⁹ On the other hand, it has been suggested in another study that mechanisms other than lipid-lowering may account for the association between omega-3 PUFA and CHD risk.⁴⁹ The findings showed that further adjustment for triglycerides or cholesterol did not attenuate the inverse association with omega-3 PUFA, indicating that mechanisms other than lipid-lowering may be relevant, which was consistent with the pathophysiologic hypothesis that a combination of various mechanisms contribute to the cardiovascular protection associated with omega-3 PUFA, including anti-inflammation, anti-thrombosis, plaque and membrane stabilization.^{49,160} A recent review of the cardiovascular impact of nutritional supplementation with omega-3 fatty

acids also concluded that omega-3 may have beneficial effects other than through triglyceride lowering.⁴⁹

Clinical Perspectives

Circulating levels of fatty acids, affected by fat and carbohydrate consumption, are one of the main constitutional components of circulating lipids. The results indicated the associations of CHD risk with fatty acids, independently of each other and major lipid fractions. Understanding these associations and the underlying biological mechanisms between circulating fatty acids and CHD risk is important to the development of dietary guidelines, and may inform clinical trials targeting circulating levels of particular fatty acids (including estimates of the epidemiologically-expected effect on disease risk of the different fatty acids, and the presence of linear associations throughout the fatty acid ranges examined) to better understand the atherosclerosis mechanism and the threshold of the associations. Furthermore, the results showed strong and robust associations with some types of fatty acids, which are likely to inform the development of risk prediction models to identify those at high risk.

Strengths and limitations

This study has a number of key strengths, including the large sample size, long follow-up, and reliable ascertainment of CHD events. It is the first large-scale observational study, to our knowledge, to assess the risk of each type of circulating fatty acid independently of major lipid fractions. The resurvey in the subset of study population allowed us to correct for regression dilution, enabling estimates of associations with long-term average levels of fatty acids, which was also not assessed before.

Furthermore, the baseline survey collected information on a wide range of factors to allow adjustment for major potential confounders.

Despite this, we cannot exclude the potential for residual confounding or reverse causality in observational studies, and the NMR platform did not include more subtypes of circulating fatty acids. In addition, blood samples in the UK biobank were taken in the non-fasting state, which may affect the stability of the measurements. However, a recent study found that fasting duration only accounts for a small proportion of variation in plasma fatty acids concentrations¹¹⁹, and the unchanged results after further adjusting for fasting time and dietary habits in the sensitivity analyses also proved the limited impact of postprandial states on the conclusion. Future analyses should further assess the causality of these associations, including using Mendelian randomisation, and explore the relevance of other fatty acid subtypes that are currently unmeasured in the cohort, such as EPA. Metabolomics data will also become available on the whole cohort in the near future and this will increase the precision of the estimated effects in this study, and permit exploration of effect modification of these associations by important characteristics.

7.7 Conclusion

This chapter quantifies the independent associations of circulating fatty acids with CHD risk. The findings suggest positive associations of circulating SFA and MUFA, inverse association of omega-3 PUFA, and no evidence of association of omega-6 PUFA with CHD risk. Although the associations of SFA and MUFA were mostly attributed to their relationship with triglyceride levels, the study indicates the inverse association with omega-3 PUFA is unlikely to be mediated by major lipid fractions.

Chapter 7 Tables and Figures

Tables	Title
Table 7.1	Association of fatty acids concentration with coronary heart disease

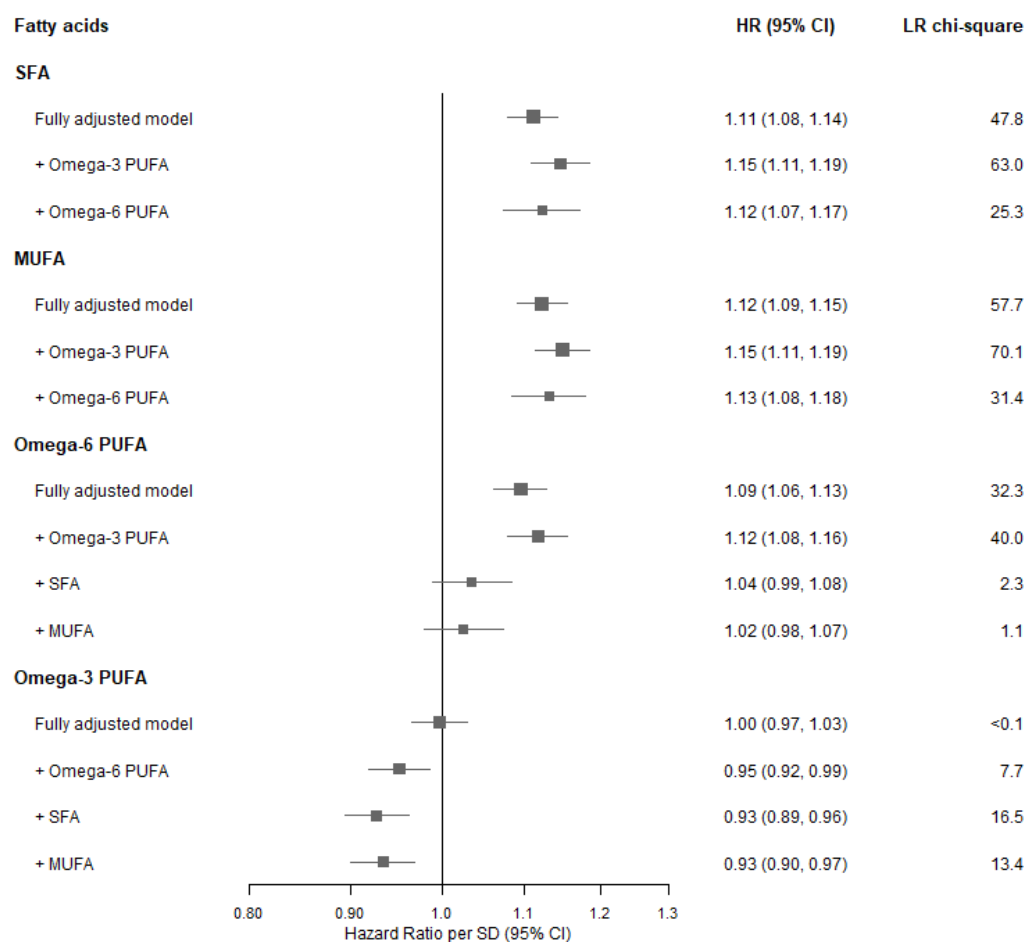
Figures	Title
Figure 7.1	Risk of coronary heart disease by usual fatty acids concentration, with progressively adjustment for other fatty acids
Figure 7.2	Fatty acids concentration vs coronary heart disease risk, with adjustment for other fatty acids
Figure 7.3	Risk of coronary heart disease by usual fatty acids concentration, with further adjustment for lipids

Table 7.1 Association of fatty acids concentration with coronary heart disease

Fatty acids	Fatty acids (mmol/L)		A. Adjusted for age and sex		B. Further adjusted for other confounders		C. Further mutually adjusted for other fatty acids	
	Baseline mean	Usual SD	HR (95% CI)	LR χ^2 *	HR (95% CI)	LR χ^2 *	HR (95% CI)	LR χ^2 *
SFA	4.1	0.7	1.16 (1.13,1.20)	103.5	1.11 (1.08,1.14)	47.8	1.12 (1.07,1.17)	25.3
MUFA	2.8	0.6	1.20 (1.16,1.23)	153.2	1.12 (1.09,1.15)	57.7	1.13 (1.08,1.18)	31.4
PUFA								
Omega-6 PUFA	4.5	0.5	1.10 (1.07,1.14)	37.5	1.09 (1.06,1.13)	32.3	1.02 (0.98,1.07)	1.1
Linoleic acids	3.5	0.5	1.09 (1.06,1.13)	31.6	1.09 (1.06,1.13)	31.7	1.02 (0.98,1.06)	0.7
Omega-3 PUFA	0.5	0.2	0.96 (0.93,0.99)	7.2	1.00 (0.97,1.03)	<0.1	0.93 (0.90,0.97)	13.4
DHA	0.2	0.1	0.89 (0.86,0.92)	49.2	0.97 (0.94,1.00)	4.3	0.95 (0.91,0.98)	9.1

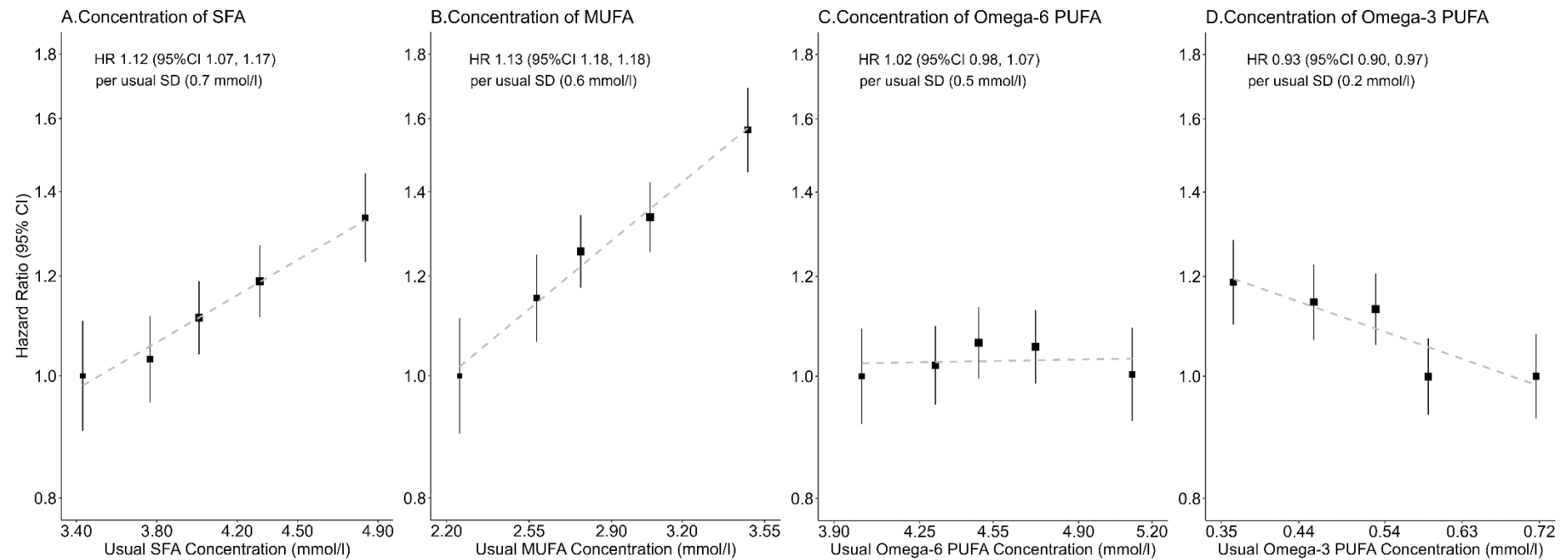
Hazard ratios (HR) per usual SD higher level of mean fatty acids concentration. *HRs calculated by Cox proportional-hazards models with: (A) stratification by age and sex; (B) stratification by age and sex, and adjusted for ethnicity, education, region, Townsend Deprivation Index, body-mass index, smoking, and alcohol; (C) model B with further mutually adjusted for other fatty acids (as last row of each type of fatty acids in Figure 7.1). †Likelihood ratio (LR) χ^2 improvement with the addition of the given factors to the model with stated adjustments. Likelihood ratio (LR) χ^2 improvement with the addition of the given factors to the model with stated adjustments. SFA=Saturated fatty acids; MUFA=Monounsaturated fatty acids; PUFA=Polyunsaturated fatty acids; DHA=Docosahexaenoic acid.

Figure 7.1 Risk of coronary heart disease by usual fatty acids concentration, with progressively adjustment for other fatty acids



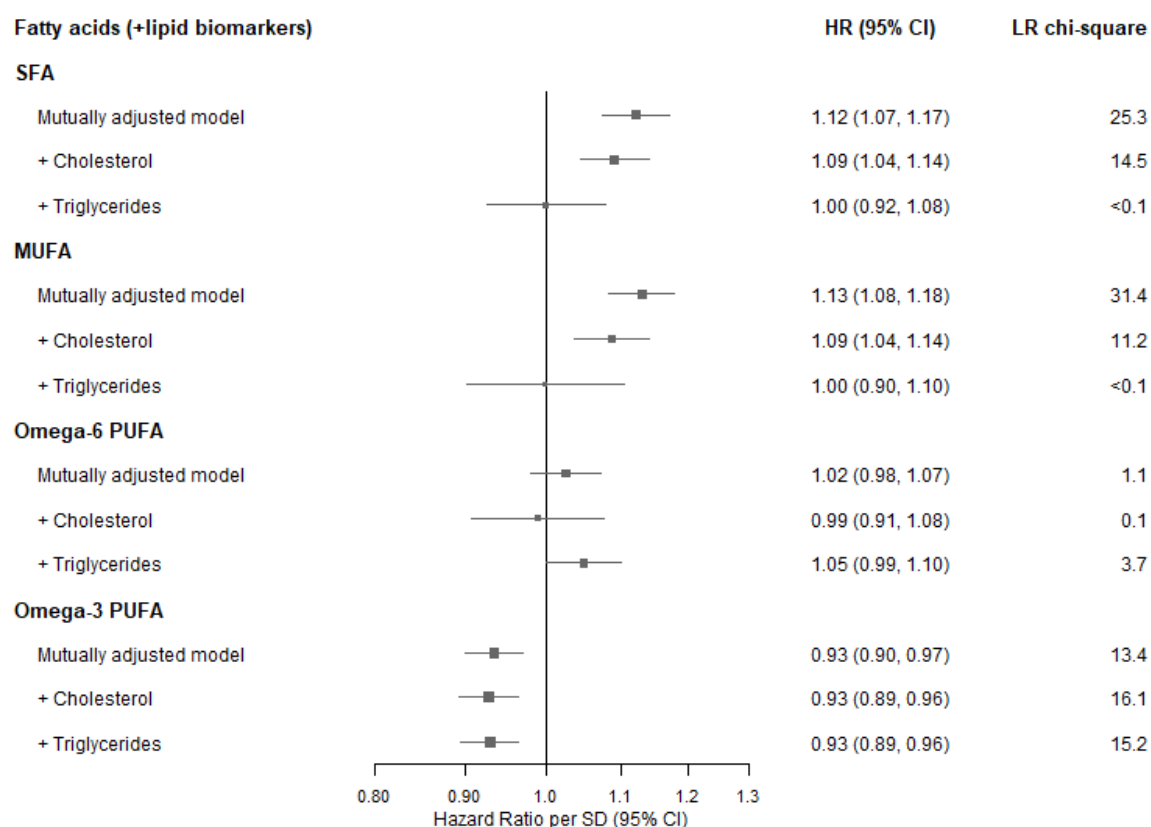
Hazard ratios (HR) per usual SD higher level of mean fatty acids ratio. HRs calculated by Cox proportional-hazards models with stratification by age and sex, and adjustment for ethnicity, education, region, Townsend Deprivation Index, smoking, alcohol, and body-mass index; and further adjustment for each fatty acids concentration progressively. Likelihood ratio (LR) χ^2 improvement with the addition of the given factors to the model with stated adjustments. SFA=Saturated fatty acids; MUFA=Monounsaturated fatty acids; PUFA=Polyunsaturated fatty acids.

Figure 7.2 Fatty acids concentration vs coronary heart disease risk, with adjustment for other fatty acids



Hazard ratios (HR) per usual SD higher level of fatty acids concentration. HRs calculated by Cox proportional-hazards models, stratified by age and sex, and adjusted for ethnicity, education, region, Townsend Deprivation Index, smoking, alcohol, body-mass index, and mutual adjusted for other fatty acids (as last row of each type of fatty acids in Figure 7.1). Area of the square is inversely proportional to the variance of the category-specific log risk. SFA= Saturated fatty acids; MUFA= Monounsaturated fatty acids; PUFA=Polyunsaturated fatty acids.

Figure 7.3 Risk of coronary heart disease by usual fatty acids concentration, with further adjustment for lipids



Hazard ratios (HR) per usual SD higher level of mean fatty acids concentration. The mutual adjusted model represented Cox proportional-hazards model with stratification by age and sex, and adjustment for ethnicity, education, region, Townsend Deprivation Index, smoking, alcohol, and body-mass index, and mutual adjusted for other fatty acids (as last row of each type of fatty acids in Figure 7.1), and further adjustment for non-HDL-cholesterol and HDL-cholesterol, or further adjustment for total triglycerides. Likelihood ratio (LR) χ^2 improvement with the addition of the given factors to the model with stated adjustments. SFA=Saturated fatty acids; MUFA=Monounsaturated fatty acids; PUFA=Polyunsaturated fatty acids.

CHAPTER 8: Prospective associations of non-lipid and non-fatty acids metabolic biomarkers with CHD risk

Chapter summary

There is little evidence from large-scale prospective studies on the long-term associations of non-lipid and non-fatty acids metabolic biomarkers in metabolite profiling, especially the low-molecular-weight metabolites, with cardiovascular risk. This chapter examines the associations of cholines, amino acids, ketone bodies, glycolysis-related, fluid balance-related, and inflammation-related metabolites with CHD risk, and further evaluates the associations independent of lipids. A strong association of glycoprotein acetyls (GlycA) with CHD risk was observed, which was 19% higher risk per usual SD. In addition to GlycA, increased levels of phenylalanine and glucose were associated with consistently higher CHD risk in both the fully adjusted model and the model independently of dietary habit and lipids, while glycine, histidine and albumin showed consistent inverse associations with CHD risk following the same adjustments. The analyses in this chapter indicate great potential of low-molecular-weight metabolites as predictors for CHD risk, but the low reproducibility of some biomarkers highlighted the importance of further research on their stability before their wide-application in clinical practice.

8.1 Background

Metabolite profiling (the extensive measurement of circulating metabolites across multiple biological pathways) is increasingly employed in clinical care, and has been successfully applied to identify biomarkers for the development of type 2 diabetes.¹⁶¹ However, few metabolites have been consistently associated with long-term cardiovascular risk across previous evidence.^{52,58,162} As a development in high-throughput technology, NMR spectroscopy currently allows for metabolic profiling of numerous metabolites across multiple biological pathways beyond the scope of the conventional chemistry method, and is being used in several large-scale prospective studies.³⁹

Previous chapters (Chapter 6 and 7) investigated the associations of plasma lipids, lipoproteins and fatty acids with risk of CHD, and compared the results with the evidence from multiple studies. However, the associations of other metabolic biomarkers in metabolite profiling, including cholines, amino acids, ketone bodies, glycolysis, fluid balance, and inflammation-related metabolites, remain unclear. Although some of these low-molecular-weight metabolites, such as branched-chain amino acids, glycoprotein acetyls and ketone bodies, have been shown to be associated with cardiovascular risk factors, the evidence of their associations with long-term cardiovascular risk among the general population is very limited, especially from large cohort studies.^{81,86,163} It is also unclear whether these low-molecular-weight metabolites are associated with cardiovascular risk through the same pathway as lipids or via an independent mechanism. Therefore, the aims of this chapter are to investigate the prospective associations of the circulating levels of each metabolic

biomarker in metabolic profiling with CHD risk, and to examine the associations of low-molecular-weight metabolites independent of lipids.

8.2 Methods

The study followed the same criteria of study population selection, outcome ascertainment and processing of metabolites described in Chapter 3. The association of each metabolic biomarker in metabolite profiling with CHD risk was quantified using Cox proportional-hazards models, stratified by sex and age, and adjusted for ethnicity, education, region, Townsend Deprivation Index, smoking, alcohol intake, and BMI. For the low-molecular-weight metabolites, concentrations of non-HDL cholesterol, HDL cholesterol and total triglycerides were further adjusted to assess their associations independent of lipids.

To account for regression dilution bias, the standard deviations of the associations were corrected to usual SD (by multiplying the baseline SD by the square root of the regression dilution ratio). The values of the regression dilution ratios are described in Chapter 5. The corrected associations are expressed per 'usual' SD higher level of each biomarker.¹²⁰ To account for multiple testing, p-values of all the associations were corrected for false discovery rate using the Benjamini-Hochberg method.¹⁶⁴ Associations were considered to be 'significant' if their corrected p-value was less than 0.05.

Sensitivity analyses of the associations were conducted which excluded CHD events in the first two years of follow-up, and which further adjusted for other potential confounders, including waist circumference, systolic blood pressure, diabetes and chronic kidney disease at baseline, fasting time, and spectrometer. Because of the low reproducibility in low-molecular-weight metabolites (and hence high regression

dilution bias), we compared the hazard ratios of biomarkers before and after the correction for regression dilution in sensitivity analyses. The correction of hazard ratios (same as the correction of SD, but more intuitive for comparison) was achieved by dividing the beta coefficient (and its standard error) of the original association by the regression dilution ratio.¹²⁰

8.3 Associations of other metabolic biomarkers

The demographics of the study population are the same as in Chapter 6. **Figure 8.1** described the association of each metabolic biomarker in NMR-derived metabolite profiling with CHD risk. Compared with lipids, lipoproteins and fatty acids related biomarkers, the overall associations of cholines and low-molecular-weight metabolites were weaker, and the change in CHD risk per usual SD was less than 10%. The only exception was glycoprotein acetyls (GlycA), which were associated with an increased CHD risk of 19% per usual SD (1.19 [95% CI 1.15, 1.23]) (**Table 8.1**).

Most amino acids remained significantly associated with CHD risk after false discovery rate correction (**Table 8.1**). Isoleucine and leucine of branched-chain amino acids (BCAA), as well as overall BCAA, were weakly positively associated with CHD risk [each HR of 1.04 (1.01, 1.07)]. Phenylalanine and tyrosine showed similar associations as BCAA, with HR (95%CI) of 1.05 (1.02, 1.08) and 1.03 (1.00, 1.07), respectively. In contrast, glycine and histidine, were inversely associated with CHD risk (0.90 [0.86, 0.93] and 0.95 [0.92, 0.98], respectively). Fluid balance-related metabolites were also significantly associated with CHD risk, although the associations of two metabolites in this category, albumin and creatinine, were in opposite directions (0.91 [0.89, 0.94] and 1.04 [1.01, 1.06], respectively). Glucose, acetoacetate and sphingomyelins showed 4% to 7% increased CHD risk, but other

cholines, glycolysis-related metabolites or ketone bodies showed null associations (**Table 8.1, S8.1**).

8.4 Associations of low-molecular-weight metabolites independent of lipids

After adjustment for lipids, the associations of BCAA (and its subtype metabolites) and creatinine attenuated to null, while other low-molecular-weight metabolites that associated with CHD risk in fully adjusted model remained significant after accounting for cholesterol and triglycerides. The association of GlycA with CHD risk remained the strongest, with HR of 1.16 (95% 1.11, 1.20). Acetone and β -hydroxybutyrate, two other types of ketone bodies, which were not associated with CHD risk in the fully adjusted model, were significantly associated with an increased risk after adjusting for lipid biomarkers, like that observed for acetoacetate (**Table 8.1**).

8.5 Sensitivity analyses

In sensitivity analyses, exclusion of CHD events in the first two years of follow-up attenuated the associations of BCAA (overall and subtypes), tyrosine, and acetoacetate. After further adjustment for other potential confounders (including baseline diabetes and systolic blood pressure) and baseline dietary habits, sphingomyelin, phenylalanine, and GlycA remained significantly associated with increased risk, while glycine, histidine, and albumin remained reversely associated with CHD risk (**S8.1**).

From Chapter 5, the within-person variability of biomarkers varied, and was much greater for some of the low-molecular-weight metabolites. Sensitivity analyses showed that accounting for regression dilution bias resulted in moderate increases in the strength of the associations (around 10% of changes in hazard ratios) among lipids,

lipoproteins and fatty acids, but higher increase in the association strength (20% or greater changes in hazard ratios) among most amino acids, ketone bodies and albumin (**S8.2**, regression dilution ratios were directly corrected in the estimates in the right plot rather than corrected in the standard deviations as the main tables and figures). Among the consistently valuable biomarkers identified in this chapter, sphingomyelin, GlycA and glycine had similar reproducibility (between baseline and repeat measurements) to blood pressure or lipids, with regression dilution ratios in the range of 0.55 to 0.60 (**Table 5.2**).

8.6 Discussion

This chapter compares the strength of associations across different types of metabolic biomarkers in NMR-derived metabolite profiling with CHD risk. In particular, the analysis highlights the potential of GlycA as a valuable biomarkers for CHD risk. It was found to have a similar strength of association as lipids and lipoproteins, and the association was only slightly attenuated after adjustment for lipids or dietary habits. Most amino acids were also strongly related to CHD risk but, except for glycine, histidine and phenylalanine, the associations of other amino acids attenuated to null in sensitivity analyses or after accounting for cholesterol and triglycerides. Albumin was another valuable biomarker showing a consistent inverse association with CHD risk, independent of lipids and dietary habits.

GlycA is a NMR-derived marker of systemic inflammation, and was found to be strongly associated with increased CHD risk. This finding is consistent with two recent studies, one based on 27,000 healthy women and the other based on 5,000 participants with history of CVD and cancer.^{81,165} These studies showed strong positive associations of GlycA with incident cardiovascular events before and after

adjusting for lipids, and after further adjusting for high-sensitivity C-reactive protein, a conventional biomarker of inflammation measured by the conventional chemistry method. The analyses in this chapter, based on the UK Biobank, provided evidence in a large sample size of healthy men and women. The results differed slightly from these previous studies, showing a much stronger association for GlycA than for high-sensitivity C-reactive protein, indicating GlycA as a potential valuable biomarker for cardiovascular risk prediction in a healthy population.

Albumin is another inflammation-related biomarker that is synthesized in the liver and declines during an inflammatory process. Evidence from meta-analyses of observational studies have demonstrated that a lower concentration of albumin is associated with higher risk of cardiovascular events, although albumin was measured in serum samples in these studies and not using the NMR method.^{166,167} A small case-cohort study among elderly healthy participants also showed that low albumin was associated with an increased mortality rate without evidence of inflammation.¹⁶⁸ This study found a similar inverse association between plasma albumin levels and CHD risk, independent of lipids and other cardiovascular risk factors. However, high within-person variability was observed in plasma albumin in the study, which may have been due to the lack of stability of albumin in the plasma sample (serum albumin by the chemistry method in this study was more stable [**Table 5.3**]) or the technical problem in the sample plating process in the UK Biobank (detailed discussion in Chapter 5). Further research on the stability of NMR-derived albumin will be necessary to confirm its value in clinical practice.

Amino acids serve as basic components for protein synthesis and a source of energy, and their metabolism may affect the development of atherosclerosis through chronic

inflammation or immune regulation in endothelium.¹⁶⁹ The evidence on prospective associations of circulating amino acids and cardiovascular risk are mostly focused on BCAAs. Three previous studies identified positive associations of BCAAs with incident cardiovascular disease, but showed diverse results on whether the associations were attenuated after adjustment for diabetes, dietary factors or lipids.^{52,86,170} The study on incident CHD also demonstrated higher risk among subjects with higher levels of BCAAs, but the associations attenuated to null after adjustment for baseline diabetes and dietary habits or lipids.

Phenylalanine, glycine and histidine are the three other amino acids that were found to be significantly associated with CHD risk, before and after adjusting for lipids and other cardiocascular risk factors. Phenylalanine was recently found to promote age-related cardiac dysfunction in mice models, while glycine and histidine may be related to the activation of nuclear factor-kappa B, which may further exhibit anti-inflammatory effects.^{171,172} However, the population-based evidence of the relevance of these amino acids to CHD risk is very limited. One study of metabolomic profiling in relation to incident CVD identified a significant association of phenylalanine before and after adjusting for lipids and other cardiovascular risk factors, but the significance in the associations did not remain for other amino acids.⁵² Conversely, both phenylalanine and histidine showed low reproducibility between baseline and repeat measurements leading to high regression dilution bias, so the interpretation of their strength of long-term associations may need to be cautious.

Sphingomyelins form one of the major classes of membrane phospholipids and contribute to membrane structure and functions.¹⁷³ As one of the major lipid constituents, sphingomyelins participate in several biological functions of lipids

including reverse cholesterol transport and regulation of plasma enzymes.¹⁷⁴ Previous prospective population-based studies found that sphingomyelins were positively associated with incident cardiovascular disease among people with diabetes,¹⁷⁵ while the results added consistent evidence in a large general population. Similarly, another mass spectrometry-based non-targeted metabolomics study in over 1000 subjects also identified sphingomyelin 28:1 as one of four metabolites associated with CHD events independently of main cardiovascular risk factors, and this study also suggested a causal role of sphingomyelin 28:1 to CHD by Mendelian randomisation analysis.¹⁷⁶

Ketone bodies are endogenous metabolites produced by the liver, with a series of bioenergetic and pleiotropic effects, including cardiac energy generation, endothelial function, oxidative stress, mitochondrial function, inflammation, cardiac remodeling, and other systemic extracardiac effects on body weight, blood pressure, glycaemia, and lipid profile.¹⁶³ There was limited evidence on long-term associations of ketone bodies with cardiovascular risk in a general population, and most of the evidence specifically focused on incident heart failure, showing a positive association.¹⁷⁷⁻¹⁷⁹ One recent study examined the prospective association of total ketone bodies, including acetoacetate, acetone, and β -hydroxybutyrate, with various cardiovascular outcomes, and found ketone bodies concentrations were only significantly associated with incident heart failure or overall CVD, but not with incident stroke or myocardial infarction.⁸⁸ The study assessed four types of ketone bodies (acetate, acetoacetate, acetone, and β -hydroxybutyrate) individually, and only acetoacetate showed a positive association with incident CHD, which was also attenuated to null after adjustment for blood pressure, baseline diabetes and dietary habits. Of note, ketone bodies in this

study also showed low reproducibility between the baseline and repeat measurements (and hence high regression dilution bias), which may affect the interpretation of long-term associations, but such variations were not reported in previous studies.

Clinical Perspectives

NMR spectroscopy enables to measure a wider range of metabolites out of the scope of the traditional chemistry method. The findings reported here from the UK Biobank indicated a few NMR-derived metabolites that were significantly associated with CHD risk independently of cardiovascular risk factors and lipids, and may have the potential to become valuable additional biomarkers of CHD risk. The performance of these metabolites in clinical practice, especially when incorporating with traditional biomarkers, remains to be further explored.

Strengths and limitations

This study has a number of key strengths. This is one of the few large-scale prospective studies to systematically investigate and compare the prospective associations across the whole metabolite profiling. On top of previous evidence, the study assessed the associations with further adjustment for several cardiovascular risk factors including dietary habits, and identified several biomarkers that remained associated with CHD risk independent of lipids. In terms of limitations, although the repeat assessment of each biomarker enabled the estimates of associations of long-term usual levels with CHD risk, the low reproducibility and hence high regression dilution bias of some low-molecular-weight metabolites indicated the necessity of assessing the stability of these biomarkers in other studies. Further studies are needed to elucidate the biological mechanisms underlying the associations with CHD and to clarify the clinical utility of these biomarkers to guide cardiovascular risk assessment.

8.7 Conclusion

This chapter quantifies the associations of other, non-lipid and non-fatty acids, biomarkers in NMR-derived metabolite profiling with CHD risk, and compares the association strength with lipids, lipoproteins and fatty acids. In summary, some of the amino acids, including glycine, histidine and phenylalanine, as well as albumin and GlycA, showed consistent significant associations with CHD risk, independent of cardiovascular risk factors or lipids. Except for GlycA, other low-molecular-weight metabolites did not have stronger associations than those of lipids with CHD risk.

Chapter 8 Tables and Figures

Tables	Title
Table 8.1	Low-molecular-weight metabolites vs coronary heart disease risk

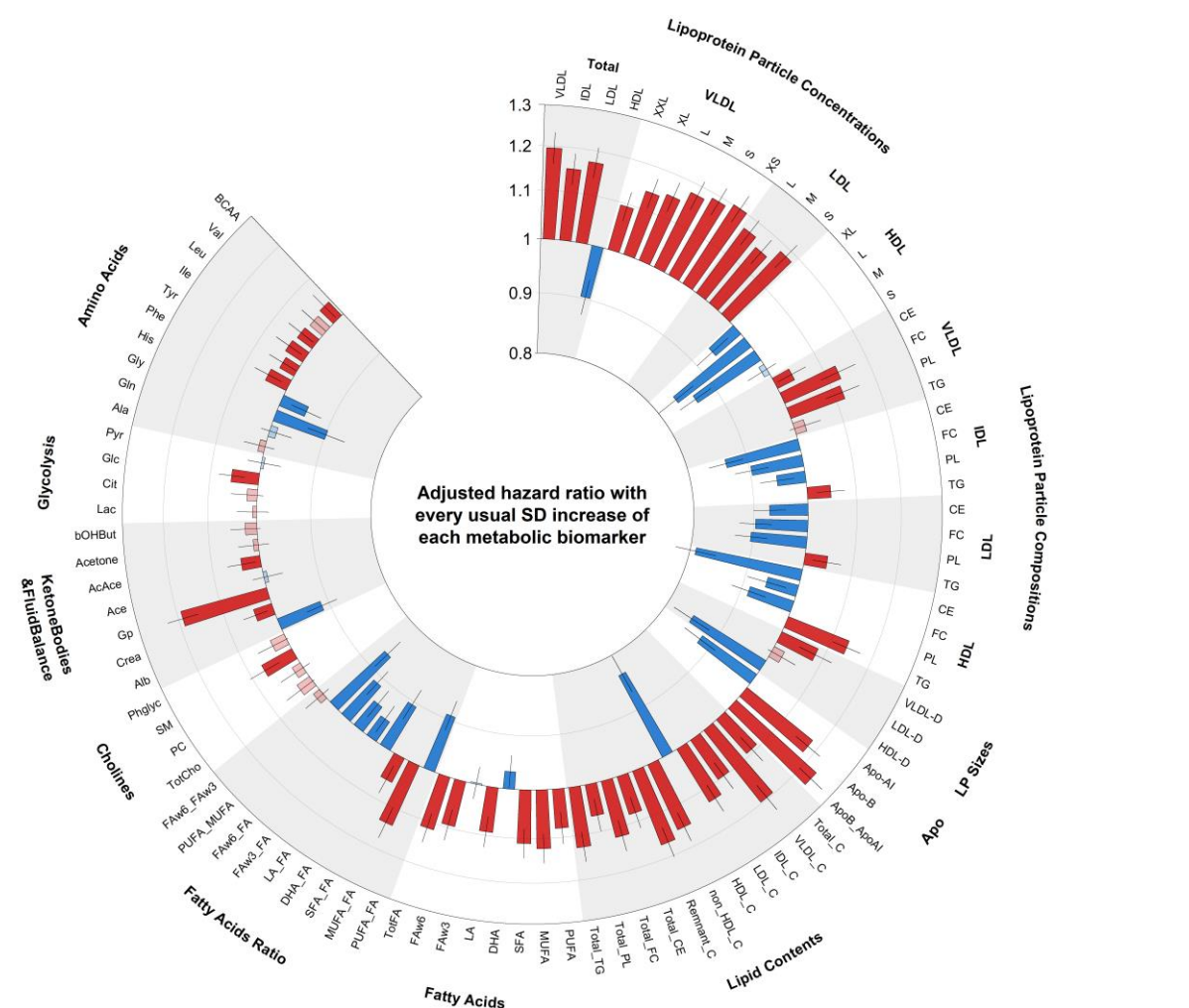
Figures	Title
Figure 8.1	Metabolite profiling vs coronary heart disease risk

Table 8.1 Low-molecular-weight metabolites vs coronary heart disease risk

	Concentration (mmol/L)		A. Fully adjusted model HR (95% CI)	B. Independent of lipids HR (95% CI)
	Baseline mean	Usual SD		
Amino acids				
Alanine	0.293	0.031	1.01 (0.98, 1.04)	1.00 (0.96, 1.03)
Glutamine	0.530	0.044	0.99 (0.96, 1.02)	0.99 (0.96, 1.02)
Glycine	0.166	0.050	0.90 (0.86, 0.93)*	0.90 (0.87, 0.94)*
Histidine	0.064	0.004	0.95 (0.92, 0.98)*	0.94 (0.91, 0.97)*
Isoleucine	0.049	0.008	1.04 (1.01, 1.07)*	1.02 (0.99, 1.06)
Leucine	0.099	0.015	1.04 (1.01, 1.07)*	1.01 (0.98, 1.05)
Valine	0.201	0.025	1.03 (1.00, 1.07)	1.00 (0.97, 1.03)
BCAA	0.350	0.048	1.04 (1.01, 1.07)*	1.00 (0.98, 1.04)
Phenylalanine	0.045	0.005	1.05 (1.02, 1.08)*	1.05 (1.02, 1.08)*
Tyrosine	0.061	0.008	1.03 (1.00, 1.07)*	0.99 (0.96, 1.03)
Glycolysis related				
Lactate	3.779	0.425	1.01 (0.98, 1.04)	0.99 (0.96, 1.03)
Citrate	0.062	0.007	1.02 (0.99, 1.05)	1.03 (0.99, 1.06)
Glucose	3.457	0.411	1.06 (1.03, 1.08)*	1.06 (1.03, 1.08)*
Pyruvate	0.078	0.013	1.00 (0.96, 1.03)	0.99 (0.96, 1.02)
Ketone bodies				
Acetate	0.015	0.004	0.99 (0.96, 1.02)	1.00 (0.98, 1.03)
Acetoacetate	0.012	0.004	1.04 (1.01, 1.07)*	1.06 (1.03, 1.09)*
Acetone	0.014	0.002	1.01 (0.98, 1.04)	1.05 (1.01, 1.07)*
β-hydroxybutyrate	0.057	0.027	1.02 (0.99, 1.05)	1.05 (1.02, 1.08)*
Fluid balance				
Albumin, g/L	39.977	1.839	0.91 (0.89, 0.94)*	0.89 (0.86, 0.92)*
Creatinine	0.065	0.009	1.04 (1.01, 1.06)*	1.01 (0.98, 1.04)
Inflammation				
GlycA	0.786	0.086	1.19 (1.15, 1.23)*	1.16 (1.11, 1.20)*

Hazard ratios (HR) per usual SD higher level of particle concentration. HRs calculated by Cox proportional-hazards models with: (A) stratification by age and sex, and adjusted for ethnicity, education, region, Townsend Deprivation Index, smoking, alcohol, and body-mass index; (B) model A further adjusted for non-HDL cholesterol, HDL cholesterol and total triglycerides. 'Usual SD' was obtained from the baseline SD through multiplication by square root of the regression dilution ratio, representing an unbiased estimate of the long-term average level in each baseline-defined group. * Represents $P < 0.05$ after correction of false discovery rate. BCAA=branched-chain amino acids; GlycA=Glycoprotein acetyls.

Figure 8.1 Metabolite profiling vs coronary heart disease risk



Metabolite	Usual SD	Metabolite	Usual SD	Metabolite	Usual SD	Metabolite	Usual SD	Metabolite	Usual SD
Lipoprotein Concentration, $\mu\text{mol/L}$		Lipoprotein Composition, molecules/particle		Lipid Contents, mmol/L		Fatty Acids Ratio		Fluid Balance, mmol/L	
VLDL_P	0.03	VLDL_CE	181.1	Total_C	0.65	PUFA_FA	2.71	Albumin,g/L	1.84
IDL_P	0.05	VLDL_FC	129.8	VLDL_C	0.19	MUFA_FA	2.00	Creatinine	0.01
LDL_P	0.21	VLDL_PL	272.2	IDL_C	0.15	SFA_FA	1.16	Glycolysis, mmol/L	
HDL_P	1.75	VLDL_TG	1415.8	LDL_C	0.31	DHA_FA	0.49	Lactate	0.42
XXL_VLDL_P	0.001	IDL_CE	138.0	HDL_C	0.27	LA_FA	2.36	Citrate	0.01
XL_VLDL_P	0.002	IDL_FC	48.5	non_HDL_C	0.60	FAw3_FA	1.16	Glucose	0.41
L_VLDL_P	0.004	IDL_PL	56.6	Remnant_C	0.30	FAw6_FA	2.60	Pyruvate	0.01
M_VLDL_P	0.009	IDL_TG	61.9	Total_CE	0.47	PUFA_MUFA	0.27	Amino Acids, mmol/L	
S_VLDL_P	0.010	LDL_CE	43.7	Total_FC	0.19	FAw6_FAw3	3.30	Alanine	0.031
XS_VLDL_P	0.010	LDL_FC	32.2	Total_PL	0.34	Cholines, mmol/L		Glutamine	0.044
L_LDL_P	0.13	LDL_PL	21.7	Total_TG	0.45	Total Choline	0.29	Glycine	0.050
M_LDL_P	0.05	LDL_TG	20.4	Fatty Acid, mmol/L		PC	0.27	Histidine	0.004
S_LDL_P	0.03	HDL_CE	7.8	PUFA	0.60	SM	0.05	Phenylalanine	0.005
XL_HDL_P	0.08	HDL_FC	2.2	MUFA	0.61	Phosphoglyc	0.28	Tyrosine	0.008
L_HDL_P	0.68	HDL_PL	6.9	SFA	0.69	Inflammation, mmol/L		Isoleucine	0.008
M_HDL_P	0.71	HDL_TG	2.4	DHA	0.06	Glycoprotein A	0.09	Leucine	0.015
S_HDL_P	0.89	Apolipoprotein, mmol/L		LA	0.47	Ketone Bodies, mmol/L		Valine	0.025
Lipoprotein Size, nm		ApoA-1	0.19	FAw3	0.17	Acetate	0.004	BCAA	0.048
VLDL_D	0.98	ApoB	0.15	FAw6	0.60	Acetoacetate	0.004		
LDL_D	0.05	ApoB_ApoA1	0.13	TotFA	1.77	Acetone	0.002		
HDL_D	0.18					bOHBut	0.027		

Adjusted for age, sex, region, Townsend deprivation index, ethnicity, education, smoking and alcohol status, and BMI. 'Usual SD' was obtained from the baseline SD through multiplication by square root of the regression dilution ratio, representing an unbiased estimate of long-term average level in each baseline-defined group. Dark color represents $P < 0.05$ after correction of false discovery rate. bOHBut= β -hydroxybutyrate; BCAA=branched-chain amino acids; PC= Phosphatidylcholine; SM= Sphingomyelin; Phosphoglyc= Phosphoglycerides.

CHAPTER 9: Predictive value of metabolic profiling in cardiovascular risk

Chapter summary

Metabolic biomarkers are increasingly employed in clinical care, such as in cancer screening and treatment monitoring. Although previous studies have examined the associations of several metabolic biomarkers with incidence and mortality of numerous common diseases, including CVD, there is little evidence on the benefit of metabolic profiling (the extensive measurement of numerous circulating metabolites across multiple biological pathways) in clinical practice to identify those at high risk of CVD. This chapter aims to evaluate whether adding circulating metabolites to an established risk score improves the prediction of 10-year ASCVD risk. Cox regression was used to compare model performance before and after addition of metabolites to QRISK3 (an established cardiovascular risk score used in primary care in England); analyses derived three models, with metabolites selected by association significance or by employing two different machine-learning approaches. However, adding selected metabolites did not significantly improve any measure of discrimination, reclassification or calibration in women or men in any of the methods used. These findings suggest that ASCVD prediction in primary care is unlikely to be improved by the addition of metabolic profiling as currently measured by nuclear magnetic resonance.

9.1 Background

Early identification of individuals at risk is important for primary prevention of major ASCVD. Several risk assessment algorithms have been developed, including the Framingham Risk Score, Systematic COronary Risk Evaluation (SCORE), and Pooled Cohort Equations^{22,25,27} (detailed description provided in **Chapter 1**). Among these established risk scores, QRISK3 is the one that widely used across England's primary health service.³¹ However, the discrimination of QRISK3 varies from 0.70 to 0.86 in different UK cohorts, and several studies suggested that QRISK3 may not perform very well in older and multi-morbid populations.¹⁸⁰⁻¹⁸² Polygenic risk score and lipoprotein(a) have been added to QRISK3 but showed modest improvement in the risk discrimination.^{183,184} Therefore, there is still considerable interest in finding new biomarkers to improve prediction accuracy.

Given the metabolic nature of atherosclerosis, circulating metabolic biomarkers are thought to have great potential to improve risk stratification.⁵⁶ However, current evidence on the predictive value of metabolites has only focused on a limited number of biomarkers with significant linear associations with ASCVD, which may not reflect the complex pathophysiology of atherosclerosis.^{52,53} NMR spectroscopy is a high-throughput technology used for metabolic profiling across multiple biological pathways, and is being used in large-scale prospective studies.³⁹ Therefore, when assessing the predictive value, the large number of metabolites measured through NMR and their complex interrelations need to be accounted for. Machine learning has been increasingly used for development of prediction models, with the strengths of incorporating highly correlated features and complex interactions that cannot be captured by traditional statistical models. Therefore, this chapter aims to evaluate whether adding high-throughput circulating metabolic profiling to a well-established

risk score using machine-learning methods improved the prediction of 10-year ASCVD risk.

9.2 Methods

To decrease the interference from some unstable biomarkers and to avoid the overfitting due to large number of lipids-related biomarkers, of the metabolites available, the main analyses only included 39 metabolites all measured with comparable validity to clinical chemistry, as the candidate biomarkers (**Table 3.1**), and followed the same criteria of processing of metabolites described in Chapter 3. In the sensitivity analyses, we expanded the candidate metabolites to a larger scope of NMR-derived metabolites that included all the metabolites in the association analyses in **Chapter 9**. The metabolites were then added to QRISK3, an established risk score widely used across England's primary health service.³¹ QRISK3 includes information on age, ethnicity, deprivation, systolic blood pressure, body mass index, total cholesterol to HDL cholesterol ratio, smoking status, family history of coronary heart disease, and medical history of a series of diseases, which were selected based on Bayes information criterion. In the sensitivity analyses, QRISK3 was replaced by SCORE2, which was another algorithm for risk prediction of CVD that is widely used in the European population, scoring by age, smoking status, SBP, and total and HDL cholesterol. Detailed definitions of QRISK3 and SCORE2 variables and mapping in the UK Biobank are provided in **S9.1** and **S9.2**.¹⁸⁵

In the analyses of this chapter, the end-point was extended to incident ASCVD in order to be consistent with the established risk score. Incident ASCVD was defined as the first-ever coronary heart disease, ischemic stroke, or transient ischemic attack, identified from Hospital Episode Statistics (including diagnostic codes and relevant procedures) and the Office for National Statistics cause of death data (**S9.3**).

The study population followed the same selection criteria as described in Chapter 3, and further excluded the participants with missing or outlying in QRISK3 variables (**S9.4**). Since the participants in the UK Biobank are overall healthier (with lower incidence of ASCVD) than the general UK population, QRISK3 was recalibrated by refitting the baseline survival function to the study population.

The candidate metabolites were selected in three ways: (1) adding the metabolites that were significantly associated with ASCVD (independently of QRISK3) to QRISK3; (2) adding all metabolites to QRISK3 and penalized by elastic-net; (3) adding the metabolites selected by Boruta SHapley Additive exPlanations (BorutaSHAP) based on Extreme gradient boosting algorithm (XGBoost) to QRISK3. Elastic-net is a regression method that performs regularization and variable selection simultaneously, with the strength of handling highly correlated variables.¹⁸⁶ XGBoost is a tree-based machine-learning method where new models are created that predict the residuals or errors of prior models and then added together to make the final prediction.^{187,188} It allows for including higher-order interactions and accounting for complex nonlinear relationships, and was chosen as our third model because of its modest computational cost and outstanding performance of risk prediction in recent studies involving a large number of proteins or metabolites.^{189,190} BorutaSHAP is a wrapper feature selection method to explain how much each factor in a model has contributed to the prediction, and the combination with Boruta feature selection algorithm ensures a faster and more stable feature selection.¹⁹¹ The detailed introduction of each model was provided below and in the appendix (**S9.5**). The hyperparameters were fine-tuned using five-fold cross-validation (**S9.6**). In all three cases, prediction performance was assessed using Cox proportional hazards regression w/o the metabolites. Bootstrapping (500 times) was applied to evaluate the optimism of the models.

Harrell's C-index was used to assess the discriminatory ability (how the model separates cases from controls) of each model. The improvement in reclassification after adding metabolites was evaluated by the integrated discrimination improvement (IDI) and net reclassification improvement (NRI). IDI summarises the extent a new model increases risk in events and decreases risk in non-events compared with the old model, while NRI quantifies the appropriateness of the change in predicted probabilities or categorised risk group when changing from old to new model. 10-year probability of event > 10% was categorized as high risk. The calibration, measuring how close the predicted probability is to the observed risk, was assessed with calibration plots at 10 years. All analyses followed suggestions from TRIPOD¹⁹², and all models were developed and evaluated separately for men and women in Python 3.9.12.

9.3 Prediction performance of the established risk score

After exclusions, 74,780 participants remained, with mean age of 55 years at study entry. Among the study population, 44% were men, 10% were current smokers, and 41% reported to have family history of heart disease. After a 10-year follow-up, 5,097 (6.8%) incident ASCVD events occurred, with about twice the rate in men than women (9.4% vs 4.8%). Compared to participants that did not have an incident ASCVD event, those with incident ASCVD were on average older, with higher BMI, systolic blood pressure and higher ratio of total cholesterol to HDL cholesterol, and more likely to be men and current moderate/heavy smokers. Participants who experienced ASCVD during follow-up were also more likely to have family heart disease history and baseline chronic disease history (**Table 9.1**).

The calibration plots by applying the original QRISK3 score showed overfitting in our study population (**S9.7**). We therefore refitted the baseline survival function from the

study population to obtain recalibrated predicted probabilities that were well-calibrated to the observed risk of each individual (**Figure 9.1**). The hazard ratio of the recalibrated QRISK3 score was 1.17 (95%CI, 1.15-1.18) per one point higher in women and 1.08 (1.07-1.09) in men (**Table 9.2**). Harrell's C-index of the recalibrated QRISK3 was 0.750 (95% CI, 0.739-0.763) for women and 0.706 (95% CI, 0.696-0.716) for men (**Table 9.3**).

9.4 Metabolites selected by different algorithms

Independently of QRISK3, twelve metabolites (HDL cholesterol, two apolipoproteins biomarkers, six fatty acids ratio biomarkers, histidine, albumin and glycoprotein acetyls) in women and five (VLDL cholesterol, ApoB to ApoA-1 ratio, omega-3 fatty acids concentration and its ratio to total fatty acids, albumin and glycoprotein acetyls) in men remained significantly associated with ASCVD. In the two machine-learning models of both sexes, fewer fatty acids were selected, but some amino acids and glycolysis-related metabolites were included as predictors. Compared with the selection criteria by association significance (first model), albumin and glycoprotein acetyls were also selected by the two machine-learning models for both sexes, while total triglycerides in women and glycine and leucine in men were newly selected by the two machine-learning models (**S9.8**).

9.5 Prediction performance after adding metabolites

Adding metabolites to QRISK3, in all three models, did not improve the discrimination in women (C-index of three models are 0.759 [0.747-0.772], 0.759 [0.746-0.770], and 0.759 [0.748-0.771], respectively), or men (0.710 [0.701-0.720], 0.710 [0.700-0.719], and 0.710 [0.701-0.719], respectively). The reclassification showed no improvement after adding the metabolites, with statistically significant relative IDI, but less than 0.5% in all three models of both sexes. Although the continuous NRI showed statistically

significant increase in most models, the categorical NRI (setting 10-year event probability >10% as high risk), which is a better measure of reclassification, showed no improvement in either men or women (**Table 9.3**). Calibration plots did not show any significant change either (**Figure 9.1**).

9.6 Sensitivity analyses

The hazard ratios (per one point higher) of the recalibrated SCORE2 were 1.12 (1.10-1.13) in women and 1.07 (1.06-1.07) in men (**S9.9**). Replacing QRISK3 by SCORE2 had limited impact on the metabolites selection in all three models, of which XGBoost selected the exactly same metabolites as using QRISK3 as the basic score (**S9.10**). Meanwhile, adding metabolites to SCORE2 did not significantly improve the overall prediction accuracy, although some slight improvements were observed in continuous NRI, which may be largely due to the poorer performance of SCORE2 in the study population (Harrell's C-index of SCORE2 were 0.731 [0.718-0.744] in women and 0.689 [0.679-0.699] in men) (**S9.11**, **S9.12**). Similarly, there was no evidence of prediction improvement when expanding the scope of the candidate metabolites. On the other hand, when expanding the scope of the candidate metabolites, albumin, glycoprotein acetyls and ApoB to ApoA-1 ratio were consistently selected by all algorithms in men, while albumin, glycoprotein acetyls and very small VLDL particle concentration were the common metabolites selected by all algorithms in women. The expansion of candidate metabolites scope enabled some lipoprotein-related biomarkers to be selected, including the lipoprotein composition and the subclass lipoprotein concentration, but no evidence of any low-molecule-weight metabolites that were out of the scope of the original 39 clinically-validated metabolites to be selected (**S9.13**, **S9.14**).

9.7 Discussion

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

Several previous studies have examined the value of metabolic profiling measured by NMR for the prediction of cardiovascular events or subclinical atherosclerosis (but neither study assessed the risk discrimination of adding metabolic profiling to established ASCVD scores).^{52,53,193} Two of these studies, both of which used traditional statistical algorithms, found moderate improvement in discrimination or reclassification, but neither included body mass index (BMI) as an established risk factor in the basic models. One other recent study used risk factors including BMI in the basic model, and observed very slight C-index improvement of coronary heart disease prediction (0.003 [0.001, 0.004]) and no improvement of cerebral stroke prediction (0.001 [-0.003, 0.005]) when adding metabolomics.¹⁹³ However, the basic model of this study still lacked detailed information on several major risk factors, such as family history of heart disease. By contrast, QRISK3 is a score developed from more comprehensive risk factors including BMI, cholesterol level, family history and aspects of medical history and medications. Similarly, when using the SCORE2 (a risk score not including BMI and medical history as risk factors) as the basic score in our

sensitivity analyses, adding metabolites showed a slight improvement in continuous NRI due to the poorer performance of the original SCORE2, however, the overall prediction accuracy that measured by C-index were not significantly improved.

Two other cohorts examined the predictive value of metabolites measured by mass spectrometry (comparison of this method with NMR was provided in **Chapter 1**).^{93,194} One study used traditional statistical algorithms and the other applied elastic-net and principal components analysis, and they both observed modest improvement in the prediction of coronary heart disease or subclinical CVD. However, similar as the previous evidence on NMR-derived metabolites, neither of the studies compared the prediction performance with any established risk score. Moreover, because mass spectrometry is more expensive and time-consuming than NMR, the sample size of both studies was relatively small (less than 3,000 individuals).

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

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

Clinical Perspectives

This study found no evidence of substantive improvement in prediction accuracy when adding metabolic profiling to a well-established CVD risk score (with information of cholesterol, blood pressure, BMI and medical history). This was despite the use of machine-learning methods to account for complex interactions of highly correlated metabolites. This finding indicates that cardiovascular risk prediction in clinical practice (or as part of national screening programmes) is unlikely to be improved by the addition of high-throughput circulating metabolites as currently measured by nuclear magnetic resonance.

Strengths and limitations

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

9.8 Conclusion

This large-scale prospective study provides evidence that compared with an established risk score with information on BMI, cholesterol and medical history, adding circulating metabolic profiling was not likely to have substantive improvement in cardiovascular risk prediction in primary care.

Chapter 9 Tables and Figures

Tables	Title
Table 9.1	Characteristics of baseline QRISK factors by 10-year incident ASCVD
Table 9.2	Associations of clinical metabolites independent from QRISK3
Table 9.3	Prediction performance of 10-year ASCVD risk w/o metabolites

Figures	Title
Figure 9.1	Calibration of risk prediction models for 10-year ASCVD risk

Table 9.1 Characteristics of baseline QRISK factors by 10-year incident ASCVD

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

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

Table 9.2 Associations of clinical metabolites independent from QRISK3

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

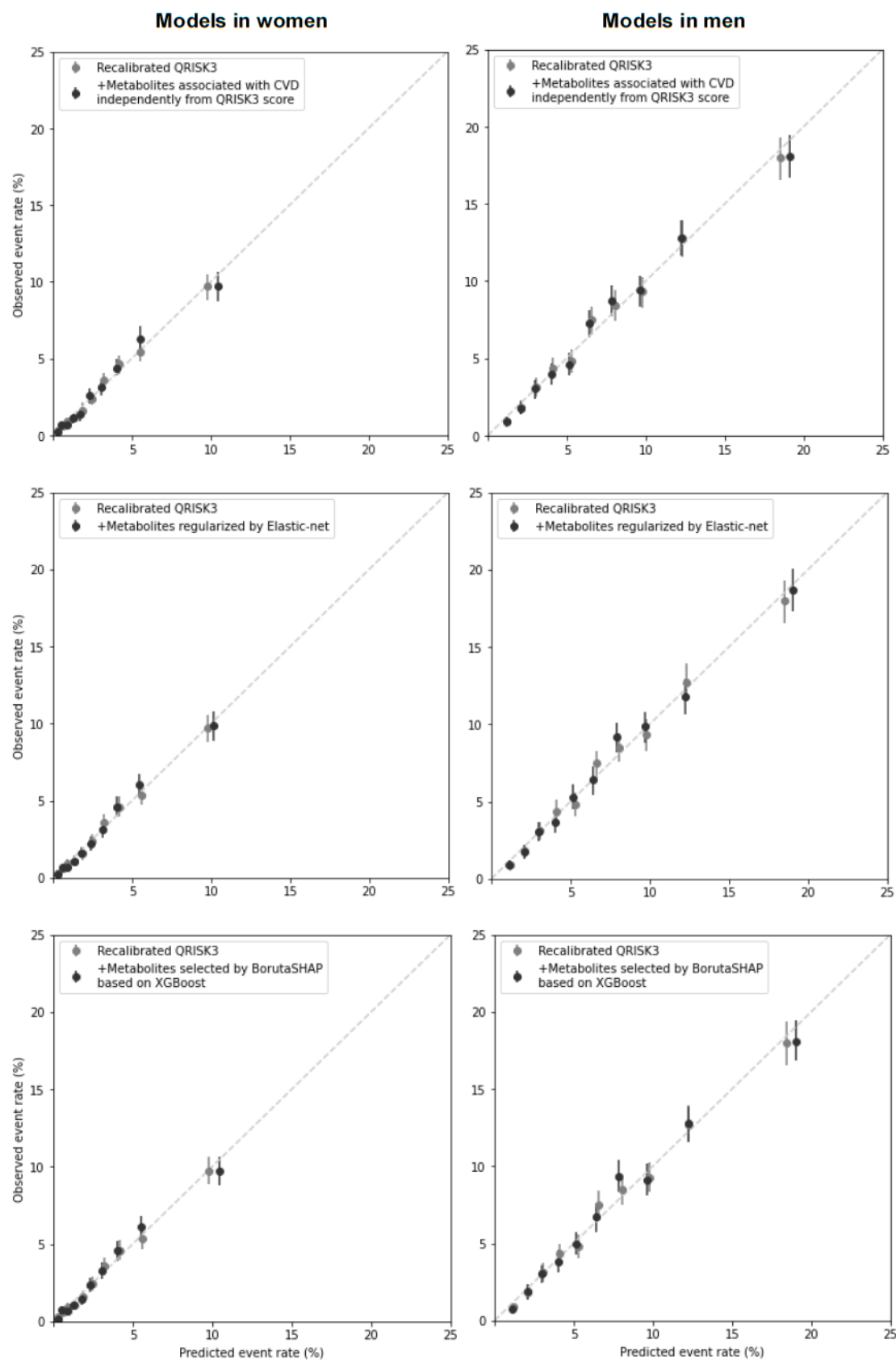
Hazard ratios (HR) per one score higher of concentration. HR of each metabolite was calculated by Cox proportional-hazards regression with adjustment of QRISK3 score. *Associations remained significant (p-value<0.01) by correction of false discovery rate using Benjamini-Hochberg method

Table 9.3 Prediction performance of 10-year ASCVD risk w/o metabolites

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

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

Figure 9.1 Calibration of risk prediction models for 10-year ASCVD risk



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

CHAPTER 10: Overall conclusions

Chapter summary

This thesis describes the prospective associations of NMR-derived plasma metabolites to CHD risk, which is the first large-scale prospective study accounting for within-person variability, and the first study to assess the potential benefit of adding metabolic profiling to the well-established cardiovascular risk scores. Using data from 86,499 individuals in the UK Biobank with NMR-metabolomics profiling, this thesis suggests that the likely mechanism by which plasma lipids are related to CHD risk is through the concentration of lipoprotein particles that transport them. It also indicates a few metabolites that may have the potential to become valuable additional biomarkers of CHD risk. Despite several strengths, the main limitations of the study include the lack of measurement of some important metabolites and the instability of some low-molecular-weight metabolites. Future research should further explore causality and mechanism pathways; and should include the assessment of the associations and predictive value in studies of more diverse populations that include repeated measurements to assess and account for within-person variability.

10.1 Main findings

Baseline distribution and cross-sectional correlations of metabolites

The main analyses included 92 metabolic biomarkers that were measured by NMR spectroscopy from the baseline plasma samples of the study population. The distribution of most of the metabolites was approximately normal, except for a few such as ketone bodies, phenylalanine, and creatinine, which were right-skewed. The average concentrations of most metabolites showed differences (in either direction) by sex, and getting gradually higher or showing no difference by age. For some of the biomarkers, including concentrations of VLDL, IDL and LDL particles, ApoB, non-HDL cholesterol, and total triglycerides, the level of these metabolites increased gradually with age in women, while remained stable or slightly decreased at older age in men. The correlations between the concentrations of lipids, lipoproteins, apolipoproteins, and some fatty acids were high, while the correlations of low-molecular-weight metabolites with other metabolites were lower. Overall, NMR-derived metabolites had high correlations with the same metabolites measured using clinical chemistry methods (**Chapter 4**).

Within-person variability of metabolites

There is limited evidence from large-scale prospective studies on the within-person variability of NMR-derived metabolites. Based on ~1,000 participants in UK Biobank with both baseline and repeat NMR assessment, the within-person variability of lipids, lipoproteins, apolipoproteins, fatty acids and cholines were similar, with the RDRs in the range of 0.5 to 0.7, while much greater variability was observed among most of the small-molecular-weight metabolites, except for glycine, creatinine and glycoprotein A. Most metabolites showed no or little difference in their within-person variability by sex, except for some lipid-related biomarkers (including concentration of IDL and LDL,

ApoB, ApoB to ApoA-1 ratio, and some lipids composition in HDL), which showed greater variability in men than women. These findings suggest the importance of considering within-person variability when assessing and interpreting the prospective associations of metabolites in the following chapters (**Chapter 5**).

Prospective associations of lipids and lipoproteins with CHD risk

The prospective analyses were conducted among 86,499 participants with NMR-metabolomics profiling and without CHD or statin use at baseline, of which 4,196 had an incident CHD event during a mean follow-up of 11.5 years. **Chapter 6** quantified the associations of lipids, lipoproteins and NMR-defined lipoprotein characteristics with CHD risk. The associations among traditional biomarkers of lipids and apolipoproteins when measured using clinical chemistry did not statistically differ from those among biomarkers measured by NMR. Among the lipoprotein characteristics (particle concentration, particle size, and lipid composition of each particle), CHD risk was most strongly related to particle concentrations of VLDL, IDL, LDL and HDL, and direct measurement of lipoprotein concentrations may be of greater value than their indirect measurement by ApoB, which was largely determined by LDL concentration alone. There was no evidence of associations with either LDL or HDL diameter after adjusting for lipoprotein concentrations, but an inverse association with VLDL diameter. Independently of lipoprotein particle concentrations, there was strong evidence of a positive association with average triglyceride levels per LDL particle, but little evidence of associations with total triglycerides or other lipid and lipoprotein fractions.

Prospective associations of other NMR-derived metabolites with CHD risk

As shown in **Chapter 7**, there was strong evidence of positive associations of circulating SFA and MUFA with CHD risk, independently of other fatty acids, and

inverse associations of omega-3 PUFA (and DHA). The associations with omega-3 PUFA were independent of both triglyceride and cholesterol levels, while the associations of SFA and MUFA were completely attenuated following adjustment for circulating triglyceride levels. Omega-6 PUFA (and linoleic acid) showed little evidence of an association with CHD risk, independently of other fatty acids. In terms of choline and other low-molecular-weight metabolites that were assessed in **Chapter 8**, GlycA was found to be a valuable metabolite with a similar strength of association as lipids and lipoproteins to CHD risk, and the association was slightly attenuated but remained significant after adjustment for lipids or dietary habits. Most amino acids were strongly related to CHD risk but, except for glycine, histidine and phenylalanine, the associations of the other amino acids attenuated to null in sensitivity analyses or after accounting for cholesterol and triglycerides. Albumin was another valuable biomarker showing a consistent inverse association with CHD risk, independent of lipids and dietary habits.

Predictive value of metabolic profiling in cardiovascular risk

The additional benefit of adding high-throughput circulating metabolites to an established risk score for the prediction of 10-year ASCVD risk was evaluated in the study population with 5,097 (6.8%) ASCVD events during the 10-year follow-up. The application of machine-learning approaches handled highly correlated variables and accounted for the complex interactions between metabolites in atherosclerosis, and provided insight into the predictive value of some low-molecular-weight metabolites (e.g. amino acids and glycolysis) that may have been previously overlooked in the traditional association analyses under linear assumptions. However, compared with an established cardiovascular risk score (QRISK3), there was no evidence of

substantive improvement in prediction of 10-year risk of ASCVD after adding the metabolic biomarkers (**Chapter 9**).

10.2 Strengths and limitations

Strengths

This thesis has a number of key strengths. This is the first large-scale observational study, to my knowledge, to assess robustly the independent associations of lipoprotein particle concentration, size and lipid composition with CHD risk, and also the first study to compare the additional value of metabolic profiling to a well-established risk score. The study included repeated measurement of metabolic biomarkers to account for within-person variability, which enabled estimation of the associations of metabolites in long-term average levels that have not previously been reported. Furthermore, linkage to NHS electronic health records and national death registries limited loss to follow-up and allowed reliable ascertainment of CHD events. In addition, the baseline survey collected information on a wide range of factors to allow adjustment for major potential confounders. Finally, in terms of building prediction algorithms, the use of different analytical methods with different assumptions showed that the results were robust against these different assumptions.

Limitations

Despite these strengths, it is important to acknowledge a number of potential limitations when interpreting the results presented in this thesis. Firstly, it is not possible to exclude the potential for residual confounding (due to measurement error in measured confounders and to lack of data on known and unknown confounders) or reverse causality in observational studies. Lack of measurement of some metabolites, such as EPA and many other non-lipids and non-fatty acids metabolic biomarkers, limited deeper explorations of mechanisms to atherosclerosis and pathways between

metabolites, which could be considered when additional data on metabolomics are released by UK Biobank. In addition, blood samples in UK biobank were taken in the non-fasting state, which may affect the stability of the measurements. The low reproducibility of some low-molecular-weight biomarkers highlighted the importance of further research on their stability before their wide-application in clinical practice. Finally, although the UK Biobank is not representative of the UK population, it has been shown that associations with health outcomes can be generalized. However, as about 95% of participants are white and all participants are aged 40-69 at recruitment in the UK Biobank, it is difficult to generalise the risk prediction results to other ethnicities or older individuals, and more studies are needed in diverse populations.

10.3 Clinical and etiologic implications

Clinical implications

NMR spectroscopy enables measurement of a wider range of metabolites beyond the scope of the traditional chemistry methods. The findings of this thesis indicated a few NMR-derived metabolites, including omega-3 fatty acids, some amino acids, albumin and GlycA, that were significantly associated with CHD risk independently of cardiovascular risk factors and lipids, and may have the potential to become valuable additional biomarkers of CHD risk. They also indicated that direct measurement of lipoprotein concentrations may be of greater value than their indirect measurement by ApoB concentration, and triglyceride level in LDL particles is likely to be a valuable additional biomarker of CHD risk. The cost-effectiveness of performing NMR spectroscopy and these metabolites in clinical practice, especially when incorporating with traditional biomarkers, remains to be explored. In addition, this thesis evaluated the benefit of adding metabolic profiling to the prediction of 10-year cardiovascular risk. However, the findings indicated that CVD prediction in primary care (or as part of

national screening programmes) is unlikely to be improved by the addition of high-throughput circulating metabolites as currently measured by nuclear magnetic resonance.

Etiological implications

In terms of lipids and lipoproteins, this thesis suggests that the likely mechanism by which plasma lipids are related to CHD risk is through the concentration of the lipoprotein particles that transport them. The finding of a positive association with triglyceride levels in LDL particles independently of lipoprotein concentrations could suggest either: elevated LDL triglycerides could be a marker of elevated triglyceride-rich remnant lipoproteins (related to insufficient activity of lipoprotein lipase) that are causally related to ASCVD; or, elevated LDL triglycerides may lead to low-grade chronic inflammation in the intima that leads to atherosclerosis with vulnerable plaques.^{128,129,138} The lack of any attenuation in the inverse association of omega-3 PUFA after adjustment for triglycerides or cholesterol indicates that mechanisms other than lipid-lowering may account for the association between omega-3 PUFA and CHD risk, consistent with the pathophysiologic hypothesis that a combination of various mechanisms contribute to the cardiovascular protection associated with omega-3 PUFA, including anti-inflammation, anti-thrombosis, plaque and membrane stabilization.^{49,160} The evaluation of other low-molecular-weight metabolites to CHD risk independently of lipids and other cardiovascular risk factors also provides insight into how these metabolites affect the development of atherosclerosis through complex biological pathways.

10.4 Future directions

Need for large-scale cohorts from diverse population

Although the current subset of the UK Biobank population is the largest cohort with NMR-derived metabolomics data, the limited cardiovascular events made it difficult to explore further subgroup analyses of the associations. NMR-derived metabolomics data became available on the whole cohort in July of this year (too late for this thesis). The larger sample size will increase the precision of the estimated effects in this study, and permit exploration of effect modifications of these associations by some important characteristics. In addition, future analyses should include studies in more diverse populations, especially non-white, and explore the associations and predictive value in both healthy individuals in primary care and high-risk individuals in secondary care.

Need for wider-range of metabolites and repeated measures

Future cohorts measuring metabolites by high-throughput methods should conduct repeated measurements and confirm the long-term stability of these metabolites, especially for low-molecular-weight metabolites, and account for within-person variability to get more accurate estimates for long-term average associations. In addition, the metabolic profiling measured by NMR only included 21 low-molecular-weight metabolites; the remainder (lipids, lipoproteins, fatty acids and cholines) are highly correlated with cholesterol, which limited the additional predictive value when adding the metabolic profiling to an established risk score. The untargeted metabolic profiling measured by mass spectrometry will become available by the end of this year, which will provide data on thousands of metabolites to enable further explorations of more types of metabolites, especially for those in very low molecular concentrations that cannot be well-captured by the NMR technique.

Need for more exploration of causality and mechanism pathways

Future analyses should further assess the causality of these associations, including using Mendelian randomisation. Furthermore, future direction of metabolomics

research will involve the comprehensive analysis of the metabolites network, exploring multiple physiological processes and pathways underlying atherosclerosis. Metabolomics is the omics discipline that is closest to phenotype, so future research of metabolomics will provide greater insight into understanding the pathways from genome, transcriptome to proteome, and how those biological and environmental factors interact with each other to affect information expression and lead to disease pathogenesis.

10.5 Conclusion

This large-scale prospective study reliably quantifies the prospective associations of NMR-derived circulating metabolites with CHD risk. Of the lipoprotein characteristics assessed, the associations of plasma lipids with CHD risk are likely to be accounted for largely by lipoprotein particle concentrations, and direct measurement of lipoprotein concentrations may be of greater value than their indirect measurement by ApoB. Independent of lipids and other cardiovascular risk factors, omega-3 fatty acids, some of the amino acids, as well as albumin and GlycA, remained significantly associated with CHD risk. However, compared with established cardiovascular risk scores, adding metabolic profiling measured by current NMR spectroscopy is unlikely to have substantive improvement in CVD risk prediction in primary care. Future research should further explore causality and mechanism pathways; and should assess the associations and predictive value in more diverse populations and in wider range of metabolites; and should include repeated measurements to account for within-person variability.

References

1. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet (London, England)*. 2018;392(10159):1736-1788.
2. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet (London, England)*. 2020;396(10258):1204-1222.
3. Timmis A, Vardas P, Townsend N, et al. European Society of Cardiology: cardiovascular disease statistics 2021. *European heart journal*. 2022;43(8):716-799.
4. Bhatnagar P, Wickramasinghe K, Wilkins E, Townsend N. Trends in the epidemiology of cardiovascular disease in the UK. *Heart (British Cardiac Society)*. 2016;102(24):1945-1952.
5. Institute for Health Metrics and Evaluation. GBD Compare. 2019; <https://vizhub.healthdata.org/gbd-compare/>, 2023.
6. Mendis S, Puska P, Norrving B, World Health O, World Heart F, World Stroke O. Global atlas on cardiovascular disease prevention and control / edited by: Shanthi Mendis ... [et al.]. In. Geneva: World Health Organization; 2011.
7. Malakar AK, Choudhury D, Halder B, Paul P, Uddin A, Chakraborty S. A review on coronary artery disease, its risk factors, and therapeutics. *Journal of cellular physiology*. 2019;234(10):16812-16823.
8. Ross R. Atherosclerosis--an inflammatory disease. *The New England journal of medicine*. 1999;340(2):115-126.
9. Severino P, D'Amato A, Pucci M, et al. Ischemic Heart Disease Pathophysiology Paradigms Overview: From Plaque Activation to Microvascular Dysfunction. *International journal of molecular sciences*. 2020;21(21).
10. Roth GA, Mensah GA, Johnson CO, et al. Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019: Update From the GBD 2019 Study. *Journal of the American College of Cardiology*. 2020;76(25):2982-3021.
11. Visseren FLJ, Mach F, Smulders YM, et al. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice. *European heart journal*. 2021;42(34):3227-3337.
12. Arnett DK, Blumenthal RS, Albert MA, et al. 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation*. 2019;140(11):e596-e646.

13. Smith SC, Jr., Allen J, Blair SN, et al. AHA/ACC guidelines for secondary prevention for patients with coronary and other atherosclerotic vascular disease: 2006 update: endorsed by the National Heart, Lung, and Blood Institute. *Circulation*. 2006;113(19):2363-2372.
14. Ioannidis JP, Tzoulaki I. Minimal and null predictive effects for the most popular blood biomarkers of cardiovascular disease. *Circulation research*. 2012;110(5):658-662.
15. Tzoulaki I, Siontis KC, Evangelou E, Ioannidis JP. Bias in associations of emerging biomarkers with cardiovascular disease. *JAMA internal medicine*. 2013;173(8):664-671.
16. Di Angelantonio E, Gao P, Pennells L, et al. Lipid-related markers and cardiovascular disease prediction. *Jama*. 2012;307(23):2499-2506.
17. Lin JS, Evans CV, Johnson E, Redmond N, Coppola EL, Smith N. Nontraditional Risk Factors in Cardiovascular Disease Risk Assessment: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *Jama*. 2018;320(3):281-297.
18. Burgess S, Ference BA, Staley JR, et al. Association of LPA Variants With Risk of Coronary Disease and the Implications for Lipoprotein(a)-Lowering Therapies: A Mendelian Randomization Analysis. *JAMA cardiology*. 2018;3(7):619-627.
19. Kamstrup PR, Tybjaerg-Hansen A, Nordestgaard BG. Extreme lipoprotein(a) levels and improved cardiovascular risk prediction. *Journal of the American College of Cardiology*. 2013;61(11):1146-1156.
20. Willeit P, Welsh P, Evans JDW, et al. High-Sensitivity Cardiac Troponin Concentration and Risk of First-Ever Cardiovascular Outcomes in 154,052 Participants. *Journal of the American College of Cardiology*. 2017;70(5):558-568.
21. Natriuretic Peptides Studies C, Willeit P, Kaptoge S, et al. Natriuretic peptides and integrated risk assessment for cardiovascular disease: an individual-participant-data meta-analysis. *The lancet Diabetes & endocrinology*. 2016;4(10):840-849.
22. D'Agostino RB, Sr., Vasan RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation*. 2008;117(6):743-753.
23. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation*. 1998;97(18):1837-1847.
24. Brindle P, Emberson J, Lampe F, et al. Predictive accuracy of the Framingham coronary risk score in British men: prospective cohort study. *BMJ (Clinical research ed)*. 2003;327(7426):1267.

25. Conroy RM, Pyörälä K, Fitzgerald AP, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *European heart journal*. 2003;24(11):987-1003.
26. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. *European heart journal*. 2021;42(25):2439-2454.
27. Goff DC, Jr., Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation*. 2014;129(25 Suppl 2):S49-73.
28. Yadlowsky S, Hayward RA, Sussman JB, McClelland RL, Min YI, Basu S. Clinical Implications of Revised Pooled Cohort Equations for Estimating Atherosclerotic Cardiovascular Disease Risk. *Annals of internal medicine*. 2018;169(1):20-29.
29. Assmann G, Cullen P, Schulte H. Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the prospective cardiovascular Münster (PROCAM) study. *Circulation*. 2002;105(3):310-315.
30. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. *BMJ (Clinical research ed)*. 2007;335(7611):136.
31. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ (Clinical research ed)*. 2017;357:j2099.
32. Hajifathalian K, Ueda P, Lu Y, et al. A novel risk score to predict cardiovascular disease risk in national populations (GloboRisk): a pooled analysis of prospective cohorts and health examination surveys. *The lancet Diabetes & endocrinology*. 2015;3(5):339-355.
33. World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions. *The Lancet Global health*. 2019;7(10):e1332-e1345.
34. Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. *Nature reviews Molecular cell biology*. 2016;17(7):451-459.
35. Jendoubi T. Approaches to Integrating Metabolomics and Multi-Omics Data: A Primer. *Metabolites*. 2021;11(3).
36. Ussher JR, Elmariah S, Gerszten RE, Dyck JR. The Emerging Role of Metabolomics in the Diagnosis and Prognosis of Cardiovascular Disease. *J Am Coll Cardiol*. 2016;68(25):2850-2870.

37. Emwas AH. The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. *Methods in molecular biology (Clifton, NJ)*. 2015;1277:161-193.
38. Emwas AH, Roy R, McKay RT, et al. NMR Spectroscopy for Metabolomics Research. *Metabolites*. 2019;9(7).
39. Soininen P, Kangas AJ, Würtz P, Suna T, Ala-Korpela M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circulation Cardiovascular genetics*. 2015;8(1):192-206.
40. Cohorts using Nightingale's data. 2021;
<https://research.nightingalehealth.com/cohorts-nightingale-data>.
41. Ruiz-Canela M, Hruby A, Clish CB, Liang L, Martínez-González MA, Hu FB. Comprehensive Metabolomic Profiling and Incident Cardiovascular Disease: A Systematic Review. *Journal of the American Heart Association*. 2017;6(10).
42. Ference BA, Kastelein JJP, Ray KK, et al. Association of Triglyceride-Lowering LPL Variants and LDL-C-Lowering LDLR Variants With Risk of Coronary Heart Disease. *Jama*. 2019;321(4):364-373.
43. Sarwar N, Sandhu MS, Ricketts SL, et al. Triglyceride-mediated pathways and coronary disease: collaborative analysis of 101 studies. *Lancet (London, England)*. 2010;375(9726):1634-1639.
44. Do R, Willer CJ, Schmidt EM, et al. Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nature genetics*. 2013;45(11):1345-1352.
45. Marston NA, Giugliano RP, Melloni GEM, et al. Association of Apolipoprotein B-Containing Lipoproteins and Risk of Myocardial Infarction in Individuals With and Without Atherosclerosis: Distinguishing Between Particle Concentration, Type, and Content. *JAMA cardiology*. 2022;7(3):250-256.
46. Sacks FM, Lichtenstein AH, Wu JHY, et al. Dietary Fats and Cardiovascular Disease: A Presidential Advisory From the American Heart Association. *Circulation*. 2017;136(3):e1-e23.
47. Tvrzicka E, Kremmyda LS, Stankova B, Zak A. Fatty acids as biocompounds: their role in human metabolism, health and disease--a review. Part 1: classification, dietary sources and biological functions. *Biomedical papers of the Medical Faculty of the University Palacky, Olomouc, Czechoslovakia*. 2011;155(2):117-130.
48. Sampath H, Ntambi JM. Polyunsaturated fatty acid regulation of genes of lipid metabolism. *Annual review of nutrition*. 2005;25:317-340.

49. Weinberg RL, Brook RD, Rubenfire M, Eagle KA. Cardiovascular Impact of Nutritional Supplementation With Omega-3 Fatty Acids: JACC Focus Seminar. *Journal of the American College of Cardiology*. 2021;77(5):593-608.
50. Cottin SC, Sanders TA, Hall WL. The differential effects of EPA and DHA on cardiovascular risk factors. *The Proceedings of the Nutrition Society*. 2011;70(2):215-231.
51. Nishizaki Y, Miyauchi K, Iwata H, et al. Study protocol and baseline characteristics of Randomized trial for Evaluation in Secondary Prevention Efficacy of Combination Therapy-Statin and Eicosapentaenoic Acid: RESPECT-EPA, the combination of a randomized control trial and an observational biomarker study. *American heart journal*. 2023;257:1-8.
52. Würtz P, Havulinna AS, Soininen P, et al. Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation*. 2015;131(9):774-785.
53. Würtz P, Raiko JR, Magnussen CG, et al. High-throughput quantification of circulating metabolites improves prediction of subclinical atherosclerosis. *European heart journal*. 2012;33(18):2307-2316.
54. McGarrah RW, Crown SB, Zhang GF, Shah SH, Newgard CB. Cardiovascular Metabolomics. *Circulation research*. 2018;122(9):1238-1258.
55. Cheng S, Shah SH, Corwin EJ, et al. Potential Impact and Study Considerations of Metabolomics in Cardiovascular Health and Disease: A Scientific Statement From the American Heart Association. *Circ Cardiovasc Genet*. 2017;10(2).
56. Shah SH, Kraus WE, Newgard CB. Metabolomic profiling for the identification of novel biomarkers and mechanisms related to common cardiovascular diseases: form and function. *Circulation*. 2012;126(9):1110-1120.
57. Griffin JL, Atherton H, Shockcor J, Atzori L. Metabolomics as a tool for cardiac research. *Nature reviews Cardiology*. 2011;8(11):630-643.
58. Holmes MV, Millwood IY, Kartsonaki C, et al. Lipids, Lipoproteins, and Metabolites and Risk of Myocardial Infarction and Stroke. *Journal of the American College of Cardiology*. 2018;71(6):620-632.
59. Higgins JPT TJ, Chandler J, Cumpston M, Li T, Page MJ, Welch VA. Cochrane Handbook for Systematic Reviews of Interventions version 6.3 (updated February 2022). 2022; www.training.cochrane.org/handbook. Accessed May, 2023.
60. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*. 2015;4:1.

61. Paynter NP, Balasubramanian R, Giulianini F, et al. Metabolic Predictors of Incident Coronary Heart Disease in Women. *Circulation*. 2018;137(8):841-853.
62. Schrimpe-Rutledge AC, Codreanu SG, Sherrod SD, McLean JA. Untargeted Metabolomics Strategies-Challenges and Emerging Directions. *Journal of the American Society for Mass Spectrometry*. 2016;27(12):1897-1905.
63. Dunn WB, Broadhurst DI, Atherton HJ, Goodacre R, Griffin JL. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chemical Society reviews*. 2011;40(1):387-426.
64. Clarke R, Shipley M, Lewington S, et al. Underestimation of risk associations due to regression dilution in long-term follow-up of prospective studies. *American journal of epidemiology*. 1999;150(4):341-353.
65. Akinkuolie AO, Paynter NP, Padmanabhan L, Mora S. High-density lipoprotein particle subclass heterogeneity and incident coronary heart disease. *Circ Cardiovasc Qual Outcomes*. 2014;7(1):55-63.
66. Arsenault BJ, Lemieux I, Després JP, et al. Comparison between gradient gel electrophoresis and nuclear magnetic resonance spectroscopy in estimating coronary heart disease risk associated with LDL and HDL particle size. *Clin Chem*. 2010;56(5):789-798.
67. Chandra A, Neeland IJ, Das SR, et al. Relation of black race between high density lipoprotein cholesterol content, high density lipoprotein particles and coronary events (from the Dallas Heart Study). *Am J Cardiol*. 2015;115(7):890-894.
68. El Harchaoui K, Arsenault BJ, Franssen R, et al. High-density lipoprotein particle size and concentration and coronary risk. *Annals of internal medicine*. 2009;150(2):84-93.
69. El Harchaoui K, van der Steeg WA, Stroes ES, et al. Value of low-density lipoprotein particle number and size as predictors of coronary artery disease in apparently healthy men and women: the EPIC-Norfolk Prospective Population Study. *J Am Coll Cardiol*. 2007;49(5):547-553.
70. Joshi R, Wannamethee SG, Engmann J, et al. Triglyceride-containing lipoprotein sub-fractions and risk of coronary heart disease and stroke: A prospective analysis in 11,560 adults. *European journal of preventive cardiology*. 2020;27(15):1617-1626.
71. Kettunen J, Holmes MV, Allara E, et al. Lipoprotein signatures of cholesteryl ester transfer protein and HMG-CoA reductase inhibition. *PLoS biology*. 2019;17(12):e3000572.
72. Mackey RH, Greenland P, Goff DC, Jr., Lloyd-Jones D, Sibley CT, Mora S. High-density lipoprotein cholesterol and particle concentrations, carotid atherosclerosis, and coronary events: MESA (multi-ethnic study of atherosclerosis). *J Am Coll Cardiol*. 2012;60(6):508-516.

73. Mora S, Buring JE, Ridker PM. Discordance of low-density lipoprotein (LDL) cholesterol with alternative LDL-related measures and future coronary events. *Circulation*. 2014;129(5):553-561.
74. Parish S, Offer A, Clarke R, et al. Lipids and lipoproteins and risk of different vascular events in the MRC/BHF Heart Protection Study. *Circulation*. 2012;125(20):2469-2478.
75. Pichler G, Amigo N, Tellez-Plaza M, et al. LDL particle size and composition and incident cardiovascular disease in a South-European population: The Hortega-Liposcale Follow-up Study. *International journal of cardiology*. 2018;264:172-178.
76. Steffen BT, Guan W, Remaley AT, et al. Use of lipoprotein particle measures for assessing coronary heart disease risk post-American Heart Association/American College of Cardiology guidelines: the Multi-Ethnic Study of Atherosclerosis. *Arterioscler Thromb Vasc Biol*. 2015;35(2):448-454.
77. van der Steeg WA, Holme I, Boekholdt SM, et al. High-density lipoprotein cholesterol, high-density lipoprotein particle size, and apolipoprotein A-I: significance for cardiovascular risk: the IDEAL and EPIC-Norfolk studies. *J Am Coll Cardiol*. 2008;51(6):634-642.
78. Kuller L, Arnold A, Tracy R, et al. Nuclear magnetic resonance spectroscopy of lipoproteins and risk of coronary heart disease in the cardiovascular health study. *Arterioscler Thromb Vasc Biol*. 2002;22(7):1175-1180.
79. Tsai MY, Steffen BT, Guan W, et al. New automated assay of small dense low-density lipoprotein cholesterol identifies risk of coronary heart disease: the Multi-ethnic Study of Atherosclerosis. *Arterioscler Thromb Vasc Biol*. 2014;34(1):196-201.
80. Balling M, Afzal S, Davey Smith G, et al. Elevated LDL Triglycerides and Atherosclerotic Risk. *Journal of the American College of Cardiology*. 2023;81(2):136-152.
81. Akinkuolie AO, Buring JE, Ridker PM, Mora S. A novel protein glycan biomarker and future cardiovascular disease events. *Journal of the American Heart Association*. 2014;3(5):e001221.
82. Borges MC, Schmidt AF, Jefferis B, et al. Circulating Fatty Acids and Risk of Coronary Heart Disease and Stroke: Individual Participant Data Meta-Analysis in Up to 16 126 Participants. *J Am Heart Assoc*. 2020;9(5):e013131.
83. Heianza Y, Ma W, DiDonato JA, et al. Long-Term Changes in Gut Microbial Metabolite Trimethylamine N-Oxide and Coronary Heart Disease Risk. *J Am Coll Cardiol*. 2020;75(7):763-772.

84. Ottosson F, Brunkwall L, Smith E, et al. The gut microbiota-related metabolite phenylacetylglutamine associates with increased risk of incident coronary artery disease. *J Hypertens*. 2020;38(12):2427-2434.
85. Ottosson F, Ericson U, Almgren P, et al. Dimethylguanidino Valerate: A Lifestyle-Related Metabolite Associated With Future Coronary Artery Disease and Cardiovascular Mortality. *J Am Heart Assoc*. 2019;8(19):e012846.
86. Tobias DK, Lawler PR, Harada PH, et al. Circulating Branched-Chain Amino Acids and Incident Cardiovascular Disease in a Prospective Cohort of US Women. *Circ Genom Precis Med*. 2018;11(4):e002157.
87. Santos JL, Ruiz-Canela M, Razquin C, et al. Circulating citric acid cycle metabolites and risk of cardiovascular disease in the PREDIMED study. *Nutrition, metabolism, and cardiovascular diseases : NMCD*. 2023;33(4):835-843.
88. Shemesh E, Chevli PA, Islam T, et al. Circulating ketone bodies and cardiovascular outcomes: the MESA study. *European heart journal*. 2023;44(18):1636-1646.
89. Floegel A, Kühn T, Sookthai D, et al. Serum metabolites and risk of myocardial infarction and ischemic stroke: a targeted metabolomic approach in two German prospective cohorts. *European journal of epidemiology*. 2018;33(1):55-66.
90. Ganna A, Salihovic S, Sundstrom J, et al. Large-scale metabolomic profiling identifies novel biomarkers for incident coronary heart disease. *PLoS Genet*. 2014;10(12):e1004801.
91. Wang Z, Zhu C, Nambi V, et al. Metabolomic Pattern Predicts Incident Coronary Heart Disease. *Arteriosclerosis, thrombosis, and vascular biology*. 2019;39(7):1475-1482.
92. Ward-Caviness CK, Xu T, Aspelund T, et al. Improvement of myocardial infarction risk prediction via inflammation-associated metabolite biomarkers. *Heart (British Cardiac Society)*. 2017;103(16):1278-1285.
93. Cruz DE, Tahir UA, Hu J, et al. Metabolomic Analysis of Coronary Heart Disease in an African American Cohort From the Jackson Heart Study. *JAMA cardiology*. 2022;7(2):184-194.
94. Cavus E, Karakas M, Ojeda FM, et al. Association of Circulating Metabolites With Risk of Coronary Heart Disease in a European Population: Results From the Biomarkers for Cardiovascular Risk Assessment in Europe (BiomarCaRE) Consortium. *JAMA Cardiol*. 2019;4(12):1270-1279.
95. Holmes MV, Millwood IY, Kartsonaki C, et al. Lipids, Lipoproteins, and Metabolites and Risk of Myocardial Infarction and Stroke. *J Am Coll Cardiol*. 2018;71(6):620-632.

96. Klepacki J, Klawitter J, Klawitter J, et al. Amino acids in a targeted versus a non-targeted metabolomics LC-MS/MS assay. Are the results consistent? *Clinical biochemistry*. 2016;49(13-14):955-961.
97. Roberts LD, Souza AL, Gerszten RE, Clish CB. Targeted metabolomics. *Curr Protoc Mol Biol*. 2012;Chapter 30:Unit 30 32 31-24.
98. Yu Z, Kastenmüller G, He Y, et al. Differences between human plasma and serum metabolite profiles. *PloS one*. 2011;6(7):e21230.
99. Teahan O, Gamble S, Holmes E, et al. Impact of analytical bias in metabonomic studies of human blood serum and plasma. *Analytical chemistry*. 2006;78(13):4307-4318.
100. Rhee EP, Ho JE, Chen MH, et al. A genome-wide association study of the human metabolome in a community-based cohort. *Cell metabolism*. 2013;18(1):130-143.
101. Shin SY, Fauman EB, Petersen AK, et al. An atlas of genetic influences on human blood metabolites. *Nat Genet*. 2014;46(6):543-550.
102. Fearnley LG, Inouye M. Metabolomics in epidemiology: from metabolite concentrations to integrative reaction networks. *International journal of epidemiology*. 2016;45(5):1319-1328.
103. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-209.
104. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine*. 2015;12(3):e1001779.
105. UK Biobank Biomarker assay quality procedures. 2019; Version 1.2:https://biobank.ndph.ox.ac.uk/showcase/ukb/docs/biomarker_issues.pdf, Assessed 2023.
106. Nightingale Health Metabolic Biomarkers: Phase 1 Release. 2020; https://biobank.ndph.ox.ac.uk/showcase/ukb/docs/nmrm_companion_doc.pdf. Accessed May 31, 2022.
107. UK Townsend Deprivation Scores from 2011 Census Data. 2017; http://s3-eu-west-1.amazonaws.com/statistics.digitalresources.jisc.ac.uk/dkan/files/Townsend_Deprivation_Scores/UK%20Townsend%20Deprivation%20Scores%20from%202011%20census%20data.pdf, Accessed 2023.
108. Holmes MV, Ala-Korpela M. What is 'LDL cholesterol'? *Nature reviews Cardiology*. 2019;16(4):197-198.
109. Julkunen H, Cichońska A, Tiainen M, et al. Atlas of plasma NMR biomarkers for health and disease in 118,461 individuals from the UK Biobank. *Nature communications*. 2023;14(1):604.

110. Hill PG. The measurement of albumin in serum and plasma. *Annals of clinical biochemistry*. 1985;22 (Pt 6):565-578.
111. Ritchie SC, Surendran P, Karthikeyan S, et al. Quality control and removal of technical variation of NMR metabolic biomarker data in ~120,000 UK Biobank participants. *Scientific data*. 2023;10(1):64.
112. Lewington S, Clarke R, Qizilbash N, Peto R, Collins R. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet (London, England)*. 2002;360(9349):1903-1913.
113. Lewington S, Thomsen T, Davidsen M, Sherliker P, Clarke R. Regression dilution bias in blood total and high-density lipoprotein cholesterol and blood pressure in the Glostrup and Framingham prospective studies. *Journal of cardiovascular risk*. 2003;10(2):143-148.
114. Lewington S, Whitlock G, Clarke R, et al. Blood cholesterol and vascular mortality by age, sex, and blood pressure: a meta-analysis of individual data from 61 prospective studies with 55,000 vascular deaths. *Lancet (London, England)*. 2007;370(9602):1829-1839.
115. Emberson JR, Whincup PH, Morris RW, Walker M. Re-assessing the contribution of serum total cholesterol, blood pressure and cigarette smoking to the aetiology of coronary heart disease: impact of regression dilution bias. *European heart journal*. 2003;24(19):1719-1726.
116. MacMahon S, Peto R, Cutler J, et al. Blood pressure, stroke, and coronary heart disease. Part 1, Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet (London, England)*. 1990;335(8692):765-774.
117. Rosner B, Willett WC, Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics in medicine*. 1989;8(9):1051-1069; discussion 1071-1053.
118. Frost C, Thompson SG. Correcting for regression dilution bias: comparison of methods for a single predictor variable. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 2000;163(2):173-189.
119. Li-Gao R, Hughes DA, le Cessie S, et al. Assessment of reproducibility and biological variability of fasting and postprandial plasma metabolite concentrations using ¹H NMR spectroscopy. *PLoS One*. 2019;14(6):e0218549.
120. Clarke R, Emberson JR, Breeze E, et al. Biomarkers of inflammation predict both vascular and non-vascular mortality in older men. *Eur Heart J*. 2008;29(6):800-809.

121. Plummer M. Improved estimates of floating absolute risk. *Statistics in medicine*. 2004;23(1):93-104.
122. Sniderman AD, Islam S, Yusuf S, McQueen MJ. Discordance analysis of apolipoprotein B and non-high density lipoprotein cholesterol as markers of cardiovascular risk in the INTERHEART study. *Atherosclerosis*. 2012;225(2):444-449.
123. Welsh C, Celis-Morales CA, Brown R, et al. Comparison of Conventional Lipoprotein Tests and Apolipoproteins in the Prediction of Cardiovascular Disease. *Circulation*. 2019;140(7):542-552.
124. Parish S, Offer A, Clarke R, et al. Lipids and lipoproteins and risk of different vascular events in the MRC/BHF Heart Protection Study. *Circulation*. 2012;125(20):2469-2478.
125. Mora S, Otvos JD, Rifai N, Rosenson RS, Buring JE, Ridker PM. Lipoprotein particle profiles by nuclear magnetic resonance compared with standard lipids and apolipoproteins in predicting incident cardiovascular disease in women. *Circulation*. 2009;119(7):931-939.
126. Ikezaki H, Lim E, Cupples LA, Liu CT, Asztalos BF, Schaefer EJ. Small Dense Low-Density Lipoprotein Cholesterol Is the Most Atherogenic Lipoprotein Parameter in the Prospective Framingham Offspring Study. *Journal of the American Heart Association*. 2021;10(5):e019140.
127. McNamara JR, Small DM, Li Z, Schaefer EJ. Differences in LDL subspecies involve alterations in lipid composition and conformational changes in apolipoprotein B. *Journal of lipid research*. 1996;37(9):1924-1935.
128. Hirayama S, Miida T. Small dense LDL: An emerging risk factor for cardiovascular disease. *Clinica chimica acta; international journal of clinical chemistry*. 2012;414:215-224.
129. Miller M. Low-Density Lipoprotein Triglycerides: Widening the Atherogenic Landscape in CVD Risk Assessment. *Journal of the American College of Cardiology*. 2018;72(2):170-172.
130. Di Angelantonio E, Sarwar N, Perry P, et al. Major lipids, apolipoproteins, and risk of vascular disease. *Jama*. 2009;302(18):1993-2000.
131. Varbo A, Benn M, Tybjaerg-Hansen A, Jørgensen AB, Frikke-Schmidt R, Nordestgaard BG. Remnant cholesterol as a causal risk factor for ischemic heart disease. *Journal of the American College of Cardiology*. 2013;61(4):427-436.
132. Joshi PH, Khokhar AA, Massaro JM, et al. Remnant Lipoprotein Cholesterol and Incident Coronary Heart Disease: The Jackson Heart and Framingham Offspring Cohort Studies. *Journal of the American Heart Association*. 2016;5(5).

133. McQueen MJ, Hawken S, Wang X, et al. Lipids, lipoproteins, and apolipoproteins as risk markers of myocardial infarction in 52 countries (the INTERHEART study): a case-control study. *Lancet (London, England)*. 2008;372(9634):224-233.
134. Richardson TG, Sanderson E, Palmer TM, et al. Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: A multivariable Mendelian randomisation analysis. *PLoS medicine*. 2020;17(3):e1003062.
135. Ference BA, Kastelein JJP, Ginsberg HN, et al. Association of Genetic Variants Related to CETP Inhibitors and Statins With Lipoprotein Levels and Cardiovascular Risk. *Jama*. 2017;318(10):947-956.
136. Sniderman AD, Thanassoulis G, Glavinovic T, et al. Apolipoprotein B Particles and Cardiovascular Disease: A Narrative Review. *JAMA cardiology*. 2019;4(12):1287-1295.
137. Nordestgaard BG, Varbo A. Triglycerides and cardiovascular disease. *Lancet (London, England)*. 2014;384(9943):626-635.
138. Saeed A, Feofanova EV, Yu B, et al. Remnant-Like Particle Cholesterol, Low-Density Lipoprotein Triglycerides, and Incident Cardiovascular Disease. *Journal of the American College of Cardiology*. 2018;72(2):156-169.
139. Organization WH. *Diet, nutrition, and the prevention of chronic diseases: report of a joint WHO/FAO expert consultation*. Vol 916: World Health Organization; 2003.
140. Piepoli MF, Hoes AW, Agewall S, et al. 2016 European Guidelines on cardiovascular disease prevention in clinical practice: The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts) Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). *European heart journal*. 2016;37(29):2315-2381.
141. Hooper L, Martin N, Jimoh OF, Kirk C, Foster E, Abdelhamid AS. Reduction in saturated fat intake for cardiovascular disease. *The Cochrane database of systematic reviews*. 2020;5(5):Cd011737.
142. Mozaffarian D, Micha R, Wallace S. Effects on coronary heart disease of increasing polyunsaturated fat in place of saturated fat: a systematic review and meta-analysis of randomized controlled trials. *PLoS medicine*. 2010;7(3):e1000252.
143. Ramsden CE, Zamora D, Leelarthaepin B, et al. Use of dietary linoleic acid for secondary prevention of coronary heart disease and death: evaluation of recovered data from the Sydney Diet Heart Study and updated meta-analysis. *BMJ (Clinical research ed)*. 2013;346:e8707.

144. Ramsden CE, Zamora D, Majchrzak-Hong S, et al. Re-evaluation of the traditional diet-heart hypothesis: analysis of recovered data from Minnesota Coronary Experiment (1968-73). *BMJ (Clinical research ed)*. 2016;353:i1246.
145. Del Gobbo LC, Imamura F, Aslibekyan S, et al. ω -3 Polyunsaturated Fatty Acid Biomarkers and Coronary Heart Disease: Pooling Project of 19 Cohort Studies. *JAMA internal medicine*. 2016;176(8):1155-1166.
146. Hodson L, Skeaff CM, Fielding BA. Fatty acid composition of adipose tissue and blood in humans and its use as a biomarker of dietary intake. *Progress in lipid research*. 2008;47(5):348-380.
147. Yousaf S, Bonsall A. UK Townsend Deprivation Scores from 2011 Census Data. 2017; http://s3-eu-west-1.amazonaws.com/statistics.digitalresources.jisc.ac.uk/dkan/files/Townsend_Deprivation_Scores/UK%20Townsend%20Deprivation%20Scores%20from%202011%20census%20data.pdf. Accessed 13 Dec 2021.
148. Easton DF, Peto J, Babiker AG. Floating absolute risk: an alternative to relative risk in survival and case-control analysis avoiding an arbitrary reference group. *Stat Med*. 1991;10(7):1025-1035.
149. Marklund M, Wu JHY, Imamura F, et al. Biomarkers of Dietary Omega-6 Fatty Acids and Incident Cardiovascular Disease and Mortality. *Circulation*. 2019;139(21):2422-2436.
150. Chowdhury R, Warnakula S, Kunutsor S, et al. Association of dietary, circulating, and supplement fatty acids with coronary risk: a systematic review and meta-analysis. *Annals of internal medicine*. 2014;160(6):398-406.
151. Garaulet M, Pérez-Llamas F, Pérez-Ayala M, et al. Site-specific differences in the fatty acid composition of abdominal adipose tissue in an obese population from a Mediterranean area: relation with dietary fatty acids, plasma lipid profile, serum insulin, and central obesity. *The American journal of clinical nutrition*. 2001;74(5):585-591.
152. Farvid MS, Ding M, Pan A, et al. Dietary linoleic acid and risk of coronary heart disease: a systematic review and meta-analysis of prospective cohort studies. *Circulation*. 2014;130(18):1568-1578.
153. Jakobsen MU, O'Reilly EJ, Heitmann BL, et al. Major types of dietary fat and risk of coronary heart disease: a pooled analysis of 11 cohort studies. *The American journal of clinical nutrition*. 2009;89(5):1425-1432.
154. Fats and fatty acids in human nutrition. Report of an expert consultation. *FAO food and nutrition paper*. 2010;91:1-166.

155. Rimm EB, Appel LJ, Chiuve SE, et al. Seafood Long-Chain n-3 Polyunsaturated Fatty Acids and Cardiovascular Disease: A Science Advisory From the American Heart Association. *Circulation*. 2018;138(1):e35-e47.
156. Hu Y, Hu FB, Manson JE. Marine Omega-3 Supplementation and Cardiovascular Disease: An Updated Meta-Analysis of 13 Randomized Controlled Trials Involving 127 477 Participants. *Journal of the American Heart Association*. 2019;8(19):e013543.
157. Aung T, Halsey J, Kromhout D, et al. Associations of Omega-3 Fatty Acid Supplement Use With Cardiovascular Disease Risks: Meta-analysis of 10 Trials Involving 77 917 Individuals. *JAMA cardiology*. 2018;3(3):225-234.
158. Nicholls SJ, Lincoff AM, Garcia M, et al. Effect of High-Dose Omega-3 Fatty Acids vs Corn Oil on Major Adverse Cardiovascular Events in Patients at High Cardiovascular Risk: The STRENGTH Randomized Clinical Trial. *Jama*. 2020;324(22):2268-2280.
159. Borges MC, Haycock PC, Zheng J, et al. Role of circulating polyunsaturated fatty acids on cardiovascular diseases risk: analysis using Mendelian randomization and fatty acid genetic association data from over 114,000 UK Biobank participants. *BMC medicine*. 2022;20(1):210.
160. Mozaffarian D, Wu JH. Omega-3 fatty acids and cardiovascular disease: effects on risk factors, molecular pathways, and clinical events. *Journal of the American College of Cardiology*. 2011;58(20):2047-2067.
161. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. *Nature medicine*. 2011;17(4):448-453.
162. Rankin NJ, Preiss D, Welsh P, et al. The emergence of proton nuclear magnetic resonance metabolomics in the cardiovascular arena as viewed from a clinical perspective. *Atherosclerosis*. 2014;237(1):287-300.
163. Yurista SR, Chong CR, Badimon JJ, Kelly DP, de Boer RA, Westenbrink BD. Therapeutic Potential of Ketone Bodies for Patients With Cardiovascular Disease: JACC State-of-the-Art Review. *Journal of the American College of Cardiology*. 2021;77(13):1660-1669.
164. Benjamini Y, Hochberg Y. Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing. *J Royal Statist Soc, Series B*. 1995;57:289-300.
165. Gruppen EG, Riphagen IJ, Connelly MA, Otvos JD, Bakker SJ, Dullaart RP. GlycA, a Pro-Inflammatory Glycoprotein Biomarker, and Incident Cardiovascular Disease: Relationship with C-Reactive Protein and Renal Function. *PloS one*. 2015;10(9):e0139057.

166. Danesh J, Collins R, Appleby P, Peto R. Association of fibrinogen, C-reactive protein, albumin, or leukocyte count with coronary heart disease: meta-analyses of prospective studies. *Jama*. 1998;279(18):1477-1482.
167. Pignatelli P, Farcomeni A, Menichelli D, Pastori D, Violi F. Serum albumin and risk of cardiovascular events in primary and secondary prevention: a systematic review of observational studies and Bayesian meta-regression analysis. *Internal and emergency medicine*. 2020;15(1):135-143.
168. Reuben DB, Ferrucci L, Wallace R, et al. The prognostic value of serum albumin in healthy older persons with low and high serum interleukin-6 (IL-6) levels. *Journal of the American Geriatrics Society*. 2000;48(11):1404-1407.
169. Nitz K, Lacy M, Atzler D. Amino Acids and Their Metabolism in Atherosclerosis. *Arteriosclerosis, thrombosis, and vascular biology*. 2019;39(3):319-330.
170. Ruiz-Canela M, Toledo E, Clish CB, et al. Plasma Branched-Chain Amino Acids and Incident Cardiovascular Disease in the PREDIMED Trial. *Clinical chemistry*. 2016;62(4):582-592.
171. Czibik G, Mezdari Z, Murat Altintas D, et al. Dysregulated Phenylalanine Catabolism Plays a Key Role in the Trajectory of Cardiac Aging. *Circulation*. 2021;144(7):559-574.
172. Hasegawa S, Ichiyama T, Sonaka I, et al. Cysteine, histidine and glycine exhibit anti-inflammatory effects in human coronary arterial endothelial cells. *Clinical and experimental immunology*. 2012;167(2):269-274.
173. Barenholz Y, Thompson TE. Sphingomyelin: biophysical aspects. *Chemistry and physics of lipids*. 1999;102(1-2):29-34.
174. Martínez-Beamonte R, Lou-Bonafonte JM, Martínez-Gracia MV, Osada J. Sphingomyelin in high-density lipoproteins: structural role and biological function. *International journal of molecular sciences*. 2013;14(4):7716-7741.
175. Jensen PN, Fretts AM, Hoofnagle AN, et al. Circulating ceramides and sphingomyelins and the risk of incident cardiovascular disease among people with diabetes: the strong heart study. *Cardiovascular diabetology*. 2022;21(1):167.
176. Ganna A, Salihovic S, Sundström J, et al. Large-scale metabolomic profiling identifies novel biomarkers for incident coronary heart disease. *PLoS genetics*. 2014;10(12):e1004801.
177. Voorrips SN, Westenbrink BD. Circulating ketone bodies as signals for cardiovascular disease and mortality. *European heart journal*. 2023;44(18):1647-1649.

178. Flores-Guerrero JL, Westenbrink BD, Connelly MA, et al. Association of beta-hydroxybutyrate with development of heart failure: Sex differences in a Dutch population cohort. *European journal of clinical investigation*. 2021;51(5):e13468.
179. Jauhiainen R, Jauhiainen M, Vangipurapu J, et al. Novel biomarkers associated with incident heart failure in 10 106 Finnish men. *ESC heart failure*. 2021;8(1):605-614.
180. Livingstone SJ, Guthrie B, Donnan PT, Thompson A, Morales DR. Predictive performance of a competing risk cardiovascular prediction tool CRISK compared to QRISK3 in older people and those with comorbidity: population cohort study. *BMC medicine*. 2022;20(1):152.
181. Parsons RE, Liu X, Collister JA, Clifton DA, Cairns BJ, Clifton L. Independent external validation of the QRISK3 cardiovascular disease risk prediction model using UK Biobank. *Heart (British Cardiac Society)*. 2023.
182. Livingstone S, Morales DR, Donnan PT, et al. Effect of competing mortality risks on predictive performance of the QRISK3 cardiovascular risk prediction tool in older people and those with comorbidity: external validation population cohort study. *The lancet Healthy longevity*. 2021;2(6):e352-e361.
183. Elliott J, Bodinier B, Bond TA, et al. Predictive Accuracy of a Polygenic Risk Score-Enhanced Prediction Model vs a Clinical Risk Score for Coronary Artery Disease. *Jama*. 2020;323(7):636-645.
184. Trinder M, Uddin MM, Finneran P, Aragam KG, Natarajan P. Clinical Utility of Lipoprotein(a) and LPA Genetic Risk Score in Risk Prediction of Incident Atherosclerotic Cardiovascular Disease. *JAMA cardiology*. 2020;6(3):1-9.
185. Parsons RE, Colopy GW, Clifton DA, Clifton L. Clinical Prediction Models in Epidemiological Studies: Lessons from the Application of QRISK3 to UK Biobank Data. *Journal of Data Science*. 2022;20(1):1-13.
186. Zou H, Hastie T. Regularization and Variable Selection Via the Elastic Net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 2005;67(2):301-320.
187. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016; San Francisco, California, USA.
188. Brownlee J. A Gentle Introduction to XGBoost for Applied Machine Learning. 2016; <https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/>.
189. Unterhuber M, Kresoja KP, Rommel KP, et al. Proteomics-Enabled Deep Learning Machine Algorithms Can Enhance Prediction of Mortality. *Journal of the American College of Cardiology*. 2021;78(16):1621-1631.

190. Khera R, Haimovich J, Hurley NC, et al. Use of Machine Learning Models to Predict Death After Acute Myocardial Infarction. *JAMA cardiology*. 2021;6(6):633-641.
191. Keany E. BorutaShap : A wrapper feature selection method which combines the Boruta feature selection algorithm with Shapley values. *Zenodo*. 2020.
192. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Annals of internal medicine*. 2015;162(1):W1-73.
193. Buergel T, Steinfeldt J, Ruyoga G, et al. Metabolomic profiles predict individual multidisease outcomes. *Nature medicine*. 2022;28(11):2309-2320.
194. Murthy VL, Reis JP, Pico AR, et al. Comprehensive Metabolic Phenotyping Refines Cardiovascular Risk in Young Adults. *Circulation*. 2020;142(22):2110-2127.

Appendix

No.	Title
S2.1	Search terms used for systematic review
S3.1	Baseline characteristics in UK Biobank versus in study population
S4.1	Histograms of NMR-derived metabolites
S4.2	Histograms of NMR-derived metabolites after log-transform
S4.3	Means of NMR-derived metabolites by age and sex
S5.1	Distribution of metabolites at baseline versus resurvey
S5.2	Within-person variability of metabolites by methods
S5.3	Within-person variability of metabolites
S6.1	Correlations coefficients of lipoprotein particle concentrations with selected NMR metabolites at baseline
S6.2	Sensitivity analyses of lipoprotein characteristics vs coronary heart disease risk
S6.3	Improvement of model fit by adding different measures
S6.4	Discordance analysis of total triglycerides and triglycerides in LDL
S7.1	Association of fatty acids concentration with coronary heart disease
S7.2	Subtype of polyunsaturated fatty acids concentration vs coronary heart disease risk, with adjustment for other fatty acids
S7.3	Sensitivity analyses of fatty acids and coronary heart disease risk
S7.4	Association of fatty acids ratios with coronary heart disease
S7.5	Fatty acids ratio vs coronary heart disease risk in fully adjusted model
S7.6	Risk of coronary heart disease by usual fatty acids ratio, with further adjustment for lipids
S8.1	Sensitivity analyses of other metabolites and coronary heart disease risk
S8.2	Metabolite profiling vs coronary heart disease risk w/o correcting for regression dilution bias on hazard ratios
S9.1	QRISK3 variables and mapping in UK Biobank
S9.2	Disease and medication codes of QRISK3 variables in UK biobank
S9.3	ICD-10 and operation code of cardiovascular disease

S9.4	Flowchart of exclusion criteria for study population in risk prediction
S9.5	Key parameters of applied algorithms
S9.6	Fine-tuning of hyper-parameters of applied algorithms
S9.7	Calibration plot of risk prediction by original QRISK3 and SCORE2
S9.8	List of selected novel metabolites by different methods (QRISK3)
S9.9	Associations of clinical metabolites independent from SCORE2
S9.10	List of selected metabolites using different methods (SCORE2)
S9.11	Comparing prediction performance of 10-year CVD risk w/o metabolites (SCORE2)
S9.12	Calibration of risk prediction models for 10-year CVD risk (SCORE2)
S9.13	Prediction performance of 10-year ASCVD risk w/o metabolites (QRISK3 and wider scope of candidate metabolites)
S9.14	Calibration of risk prediction models for 10-year ASCVD risk (QRISK3 and wider scope of candidate metabolites)

S2.1 Search terms used for systematic review

	Search term of Medline	Search term of Embase
#1 CHD	exp Myocardial Ischemia/ (isch?emi\$ adj3 (heart or myocard*)).tw. (coronary adj3 disease).tw. angina.tw. myocardial infarct\$.tw. heart infarct\$.tw. (heart adj3 disease).tw. (cardiac adj3 disease).tw. chd.tw. cad.tw.	exp ischemic heart disease/ exp heart infarction/ (isch?emi\$ adj3 (heart or myocard*)).tw. (coronary adj3 disease).tw. angina.tw. myocardial infarct\$.tw. heart infarct\$.tw. (heart adj3 disease).tw. (cardiac adj3 disease).tw. chd.tw. cad.tw.
#2 Metabolomics	exp Metabolomics/ exp Proteomics/ metabolom*.tw. metabonom*.tw. metabolit*.tw. lipidom*.tw. proteom*.tw.	exp metabolomics/ exp metabonomics/ exp Proteomics/ metabolom*.tw. metabonom*.tw. metabolit*.tw. lipidom*.tw. proteom*.tw.
#3 MS/NMR method	exp Mass Spectrometry/ exp Magnetic Resonance Spectroscopy/ nuclear magnetic resonance.tw. NMR.tw. mass spectromet*.tw. MS.tw.	exp Mass Spectrometry/ exp nuclear magnetic resonance/ nuclear magnetic resonance.tw. NMR.tw. mass spectromet*.tw. MS.tw.
#4 Specific subtypes of metabolites	exp Cholesterol/ exp Lipids/ exp Ketone Bodies/ exp Glycolysis/ Cholester*.tw. Triglycerid*.tw. phospholipid*.tw. lipoprotein*.tw. apolipoprotein*.tw. Total lipid*.tw. fatty acid*.tw. amino acid*.tw. ketone bod*.tw. phosphatidylcholin*.tw. lysophosphatidylcholin*.tw. sphingomyelin*.tw. glycolys*.tw. glycoprotein*.tw. inflamm*.tw. oxidat*.tw.	exp Cholesterol/ exp Lipids/ exp Amino Acid/ exp Ketone Bodies/ exp glucose metabolism/ Cholester*.tw. Triglycerid*.tw. phospholipid*.tw. lipoprotein*.tw. apolipoprotein*.tw. Total lipid*.tw. fatty acid*.tw. amino acid*.tw. ketone bod*.tw. phosphatidylcholin*.tw. lysophosphatidylcholin*.tw. sphingomyelin*.tw. glycolys*.tw. glycoprotein*.tw. inflamm*.tw. oxidat*.tw.
#5 Animal study	exp animals/ not humans/ #1 AND (#2 OR (#3 AND #4)) NOT #5	exp animals/ not humans/
Total	5096	8774
Total (drop duplicates)	11088	

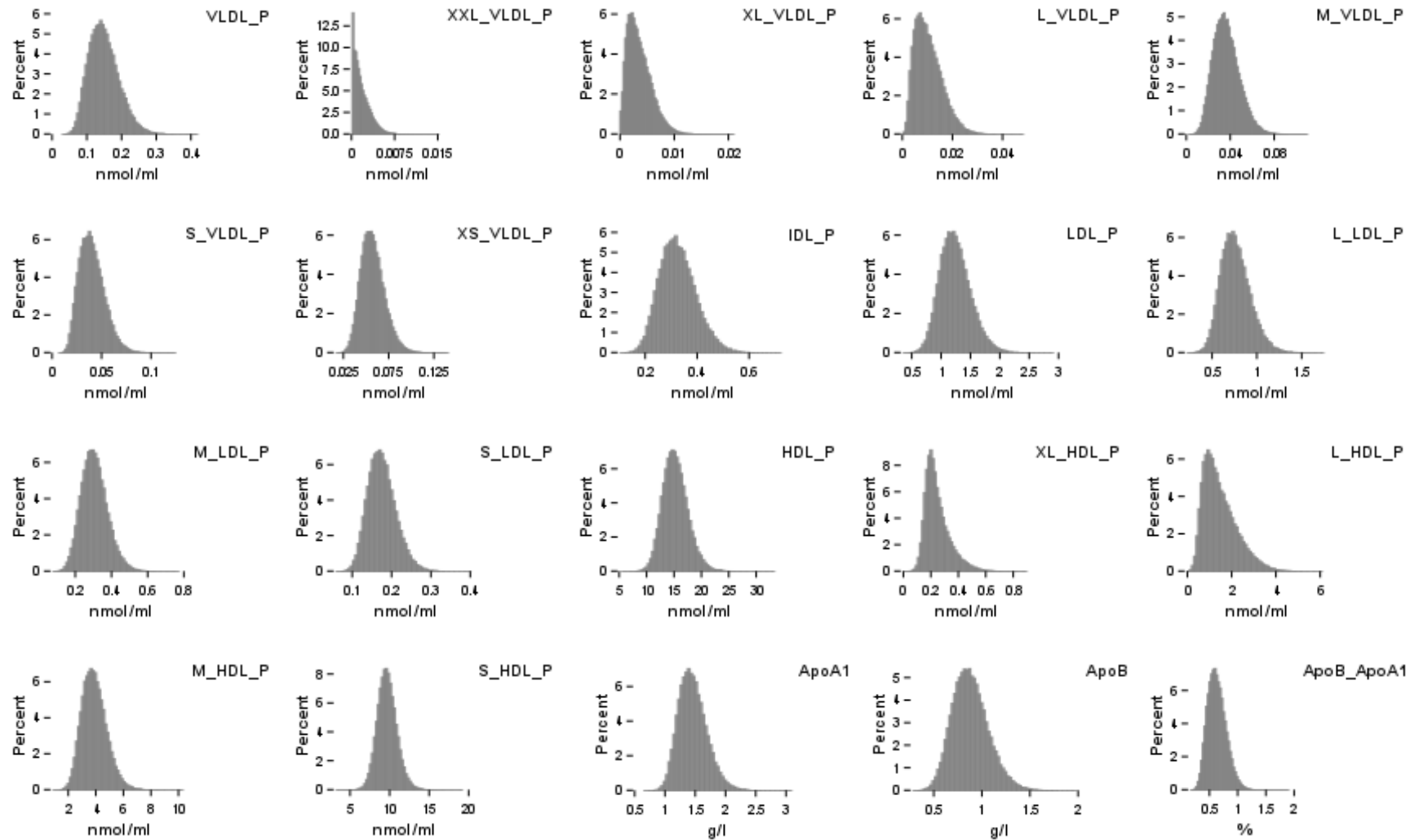
S3.1 Baseline characteristics in UK Biobank versus in study population

	UK Biobank population		Study Population	
	Women	Men	Women	Men
No. of participants	233,743	171,166	49,489	37,010
Age and socioeconomic factors				
Baseline Age, years	55.5 (8.0)	55.3 (8.2)	55.4 (7.9)	55.2 (8.2)
White, %	94.4	94.1	95.2	95.1
University education, %	38.9	42.3	39.2	42.3
Townsend deprivation index*	-1.4 (3.0)	-1.3 (3.1)	-1.5 (2.9)	-1.6 (2.9)
Lifestyle factors				
Current smoker, %	8.7	12.6	8.8	12.1
Current regular alcohol drinker, %	64.2	78.4	64.6	78.6
Anthropometry				
Body Mass Index, kg/m ²	26.7 (5.0)	27.4 (4.0)	26.7 (4.9)	27.4 (4.0)
Waist circumference, cm	83.7 (12.1)	95.6 (10.9)	83.5 (11.9)	95.5 (10.8)
Waist to hip ratio	0.81 (0.07)	0.93 (0.06)	0.8 (0.1)	0.9 (0.1)
Lipids and apolipoproteins measured by clinical chemistry				
LDL cholesterol, mmol/l	3.7 (0.8)	3.7 (0.8)	3.7 (0.8)	3.7 (0.8)
HDL cholesterol, mmol/l	1.6 (0.4)	1.3 (0.3)	1.6 (0.4)	1.3 (0.3)
Total triglycerides, mmol/l	1.5 (0.8)	1.9 (1.1)	1.5 (0.8)	1.9 (1.1)
Apolipoprotein A-I, g/l	1.6 (0.3)	1.4 (0.2)	1.6 (0.3)	1.4 (0.2)
Apolipoprotein B, g/l	1.1 (0.2)	1.1 (0.2)	1.0 (0.2)	1.1 (0.2)
Blood pressure and diabetes				
Systolic BP, mmHg	134.4 (19.2)	140.4 (17.3)	134.0 (18.9)	140.3 (17.3)
Baseline diabetes, %	1.8	2.6	1.7	2.5
Fasting time, h	3.0 (2.0)	3.0 (2.0)	3.0 (2.0)	3.0 (2.0)

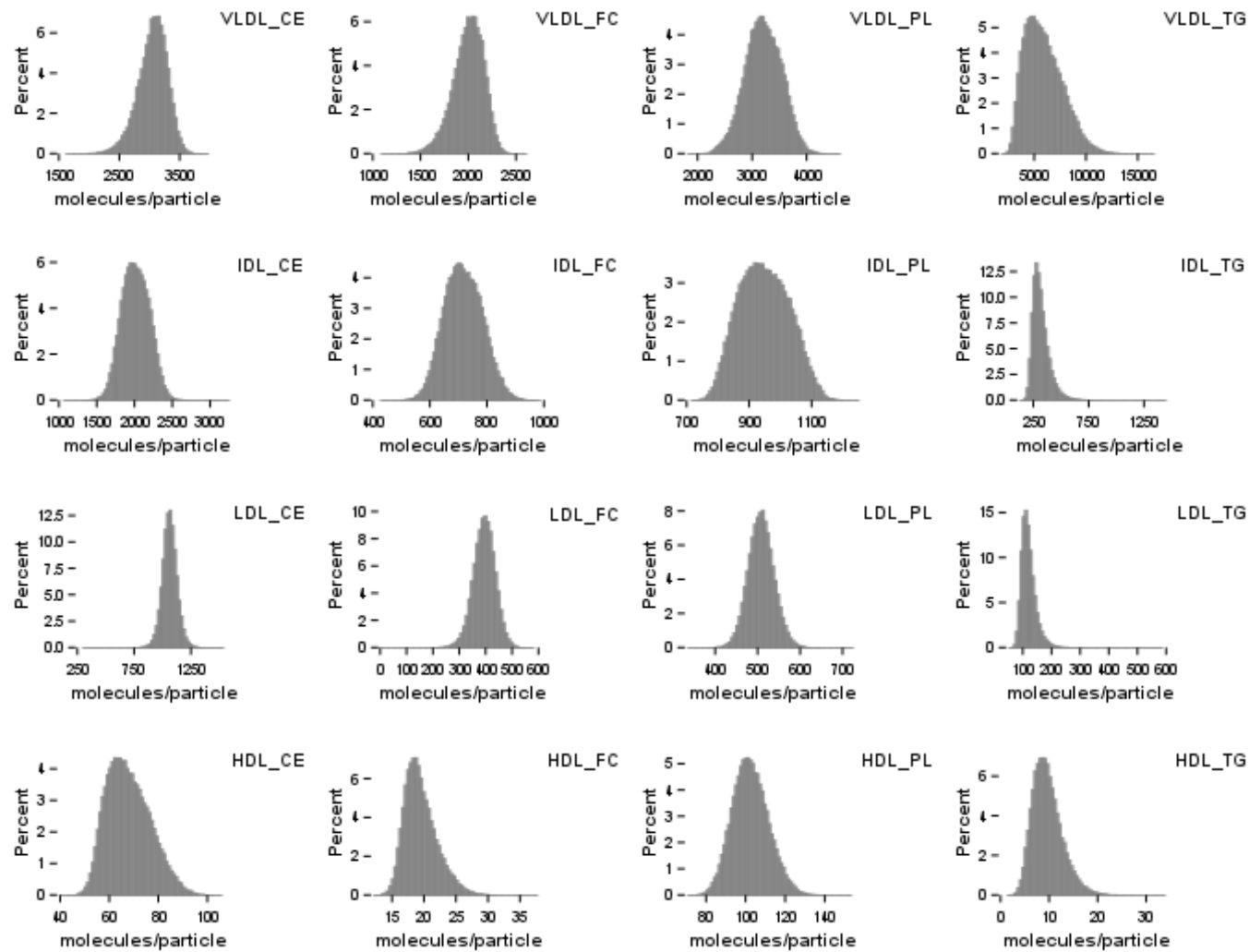
Continuous variables are presented as mean (standard deviation) (except for skew variable fasting time, presented as median [quartile range]), and categorical variables are presented as column percentages. *Area-level measure of material deprivation. BP=blood pressure; LDL=low-density lipoproteins; HDL=high-density lipoproteins; SD=standard deviation.

S4.1 Histograms of NMR-derived metabolites

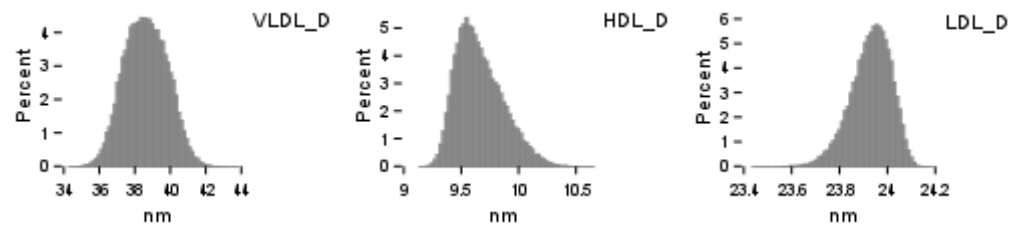
Concentration of lipoproteins and apolipoproteins



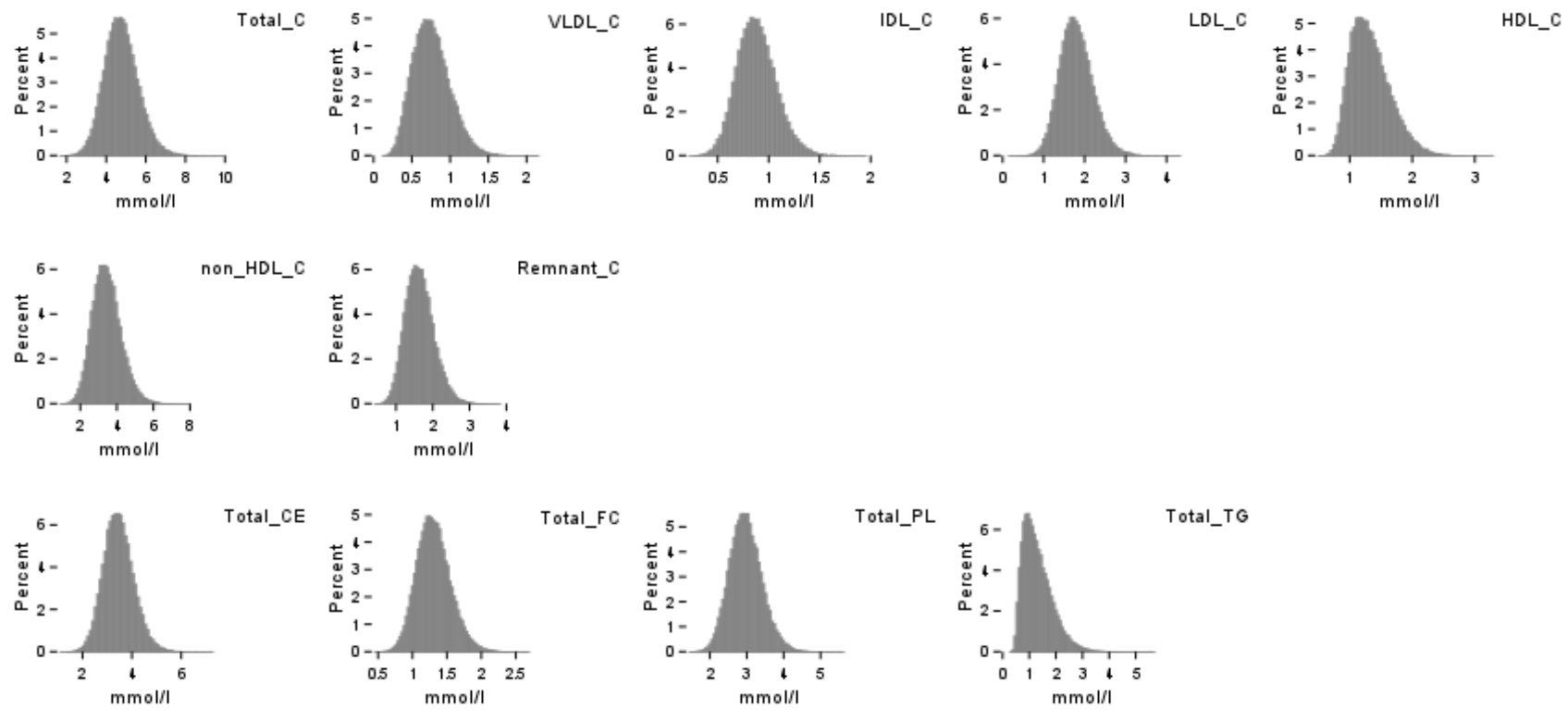
Composition of lipoproteins



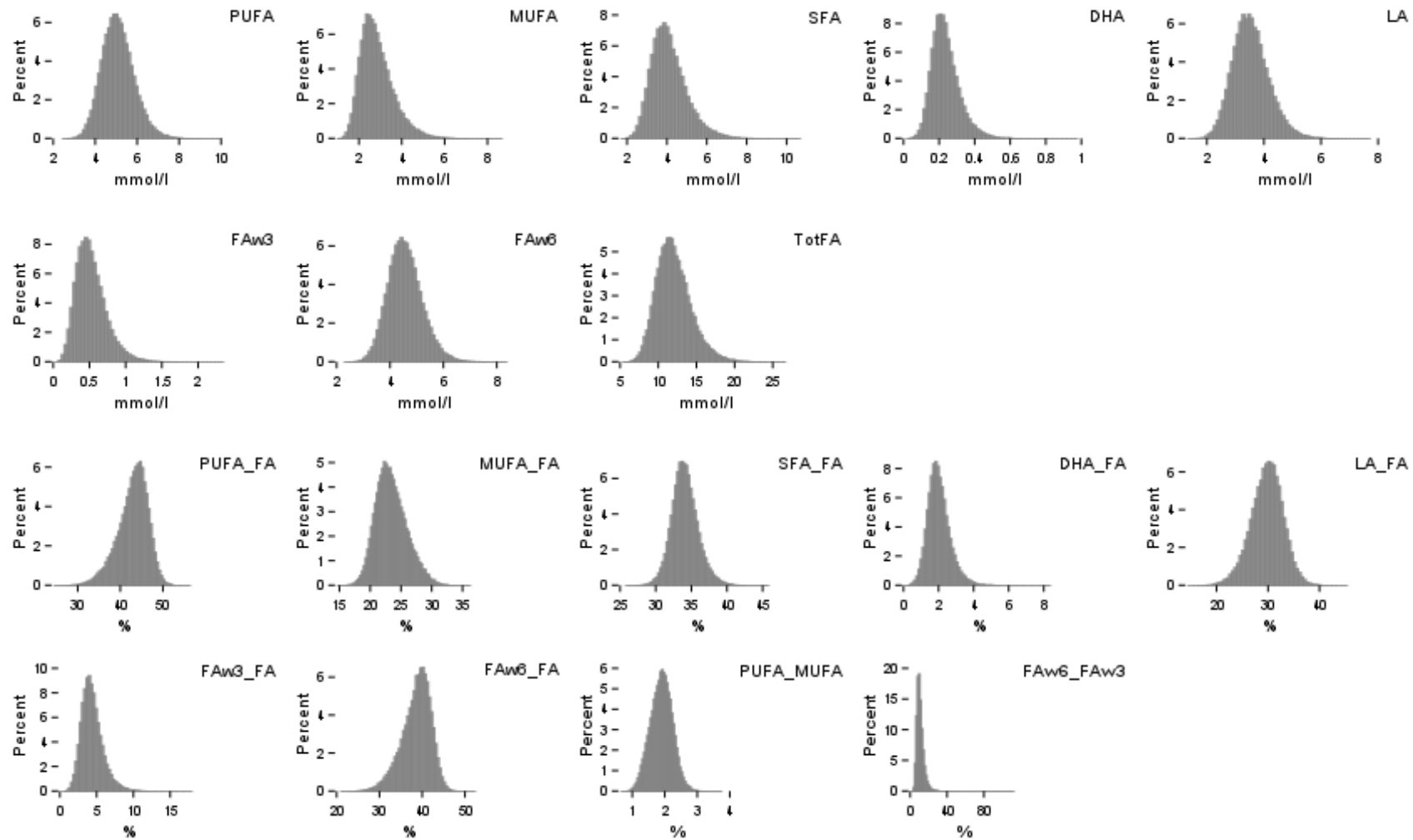
Size of lipoproteins



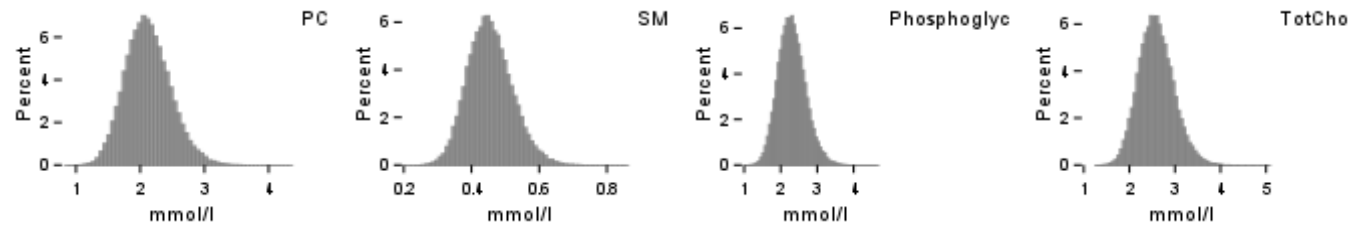
Total lipids



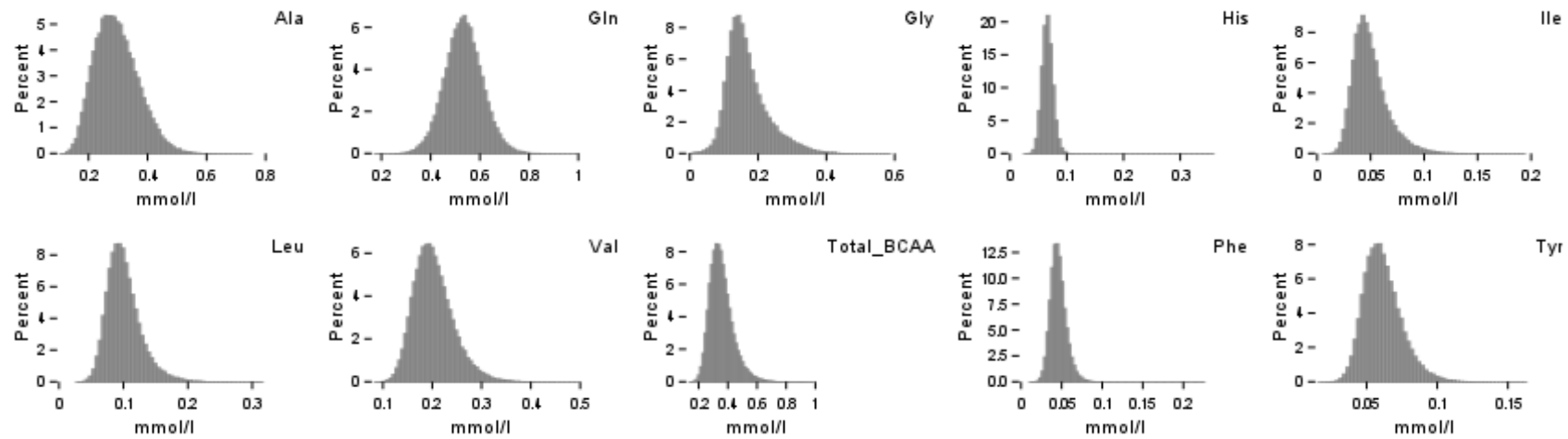
Fatty acids



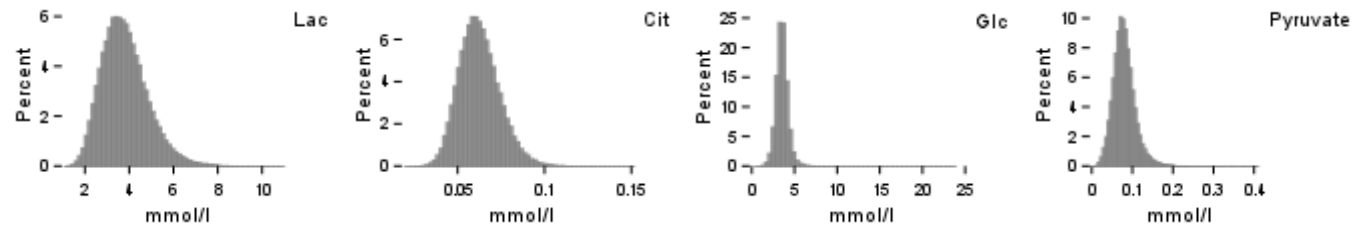
Cholines



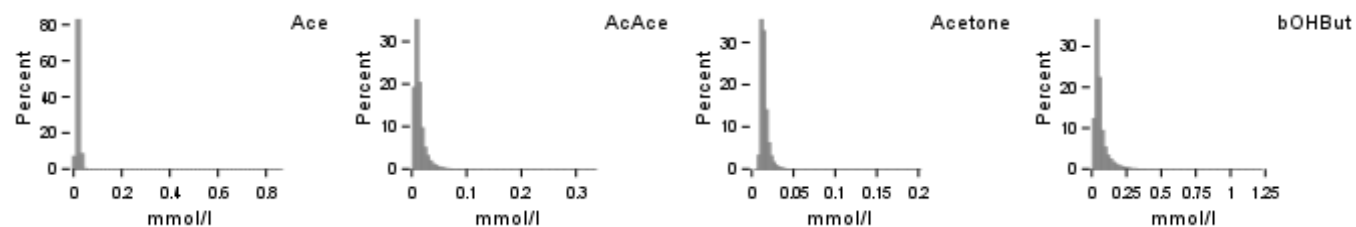
Amino acids



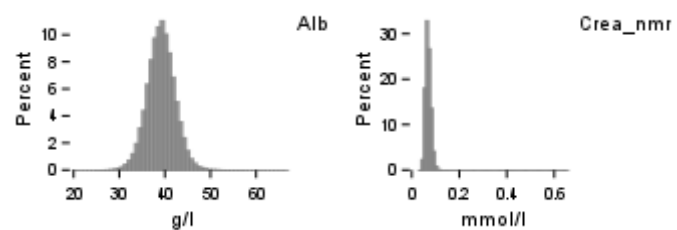
Glycolysis related metabolites



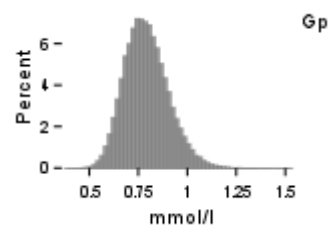
Ketone bodies



Fluid balance

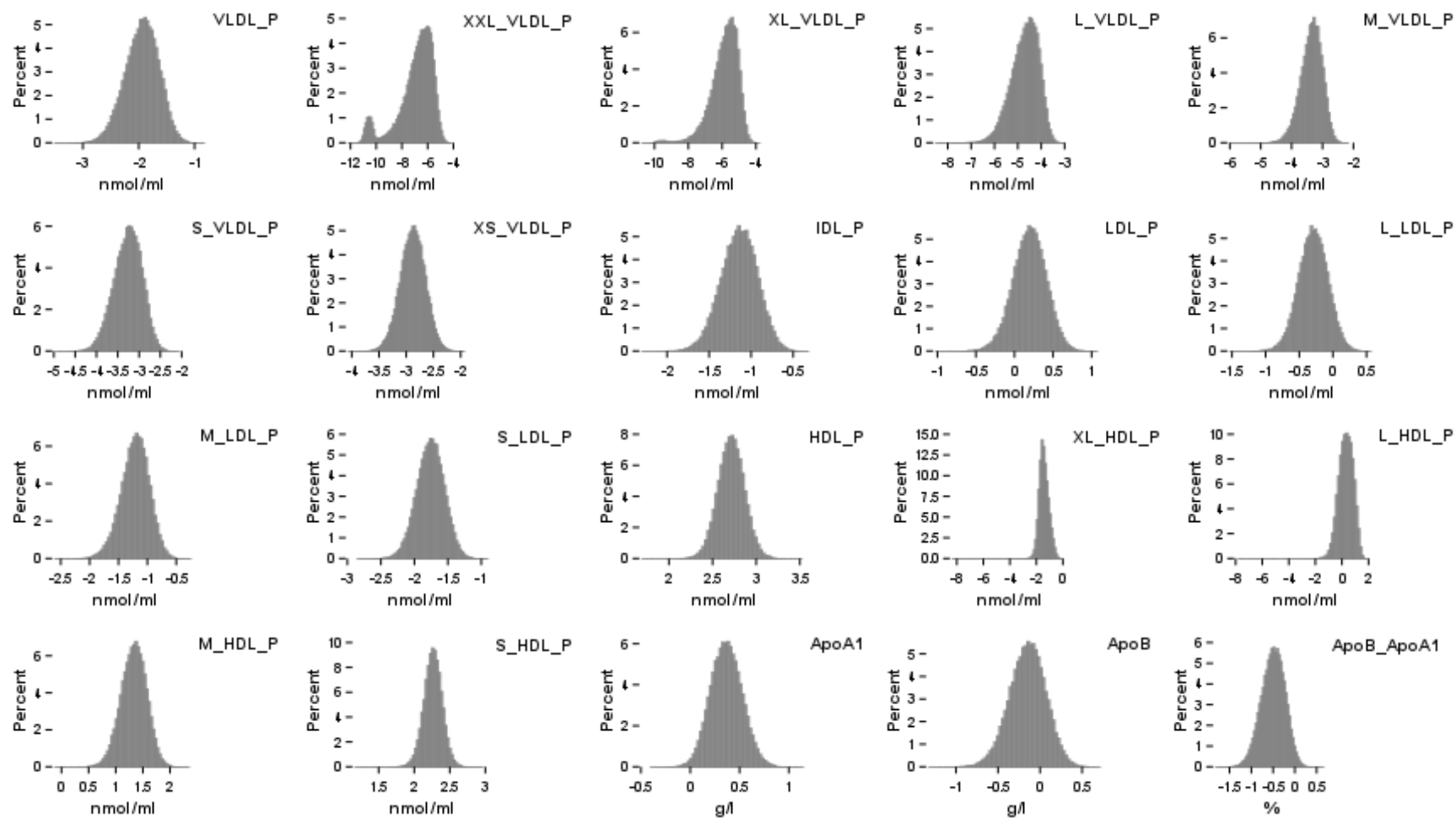


Inflammation

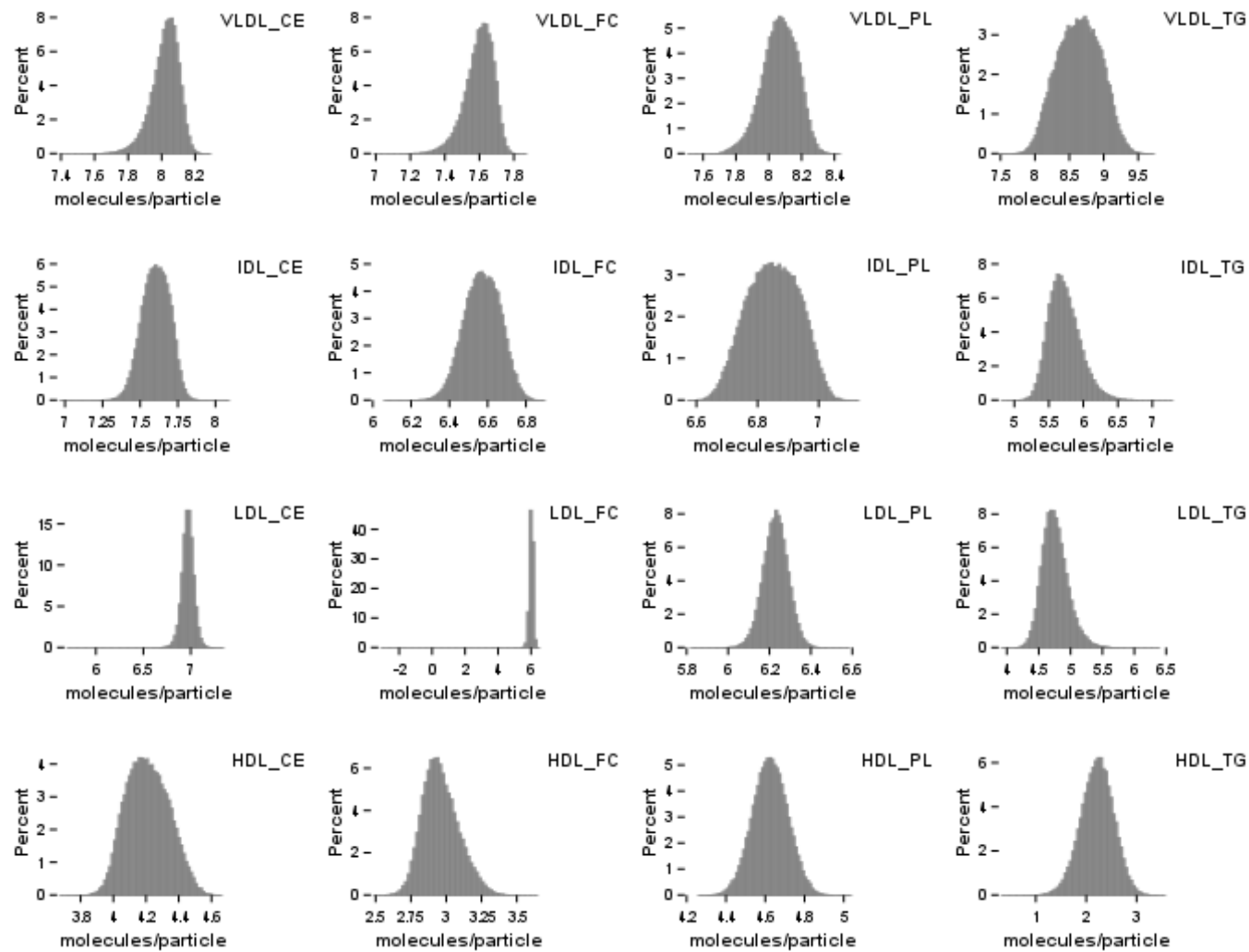


S4.2 Histograms of NMR-derived metabolites after log transformation

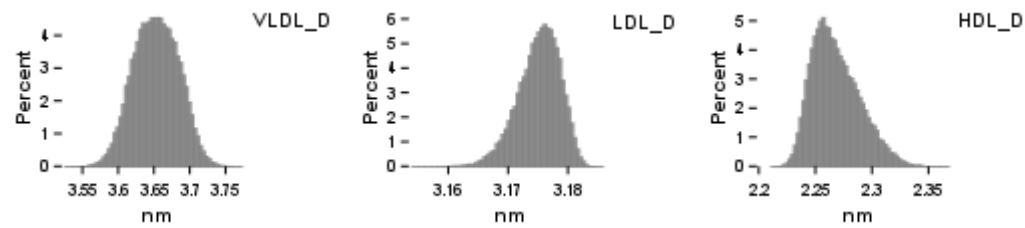
Concentration of lipoproteins and apolipoproteins



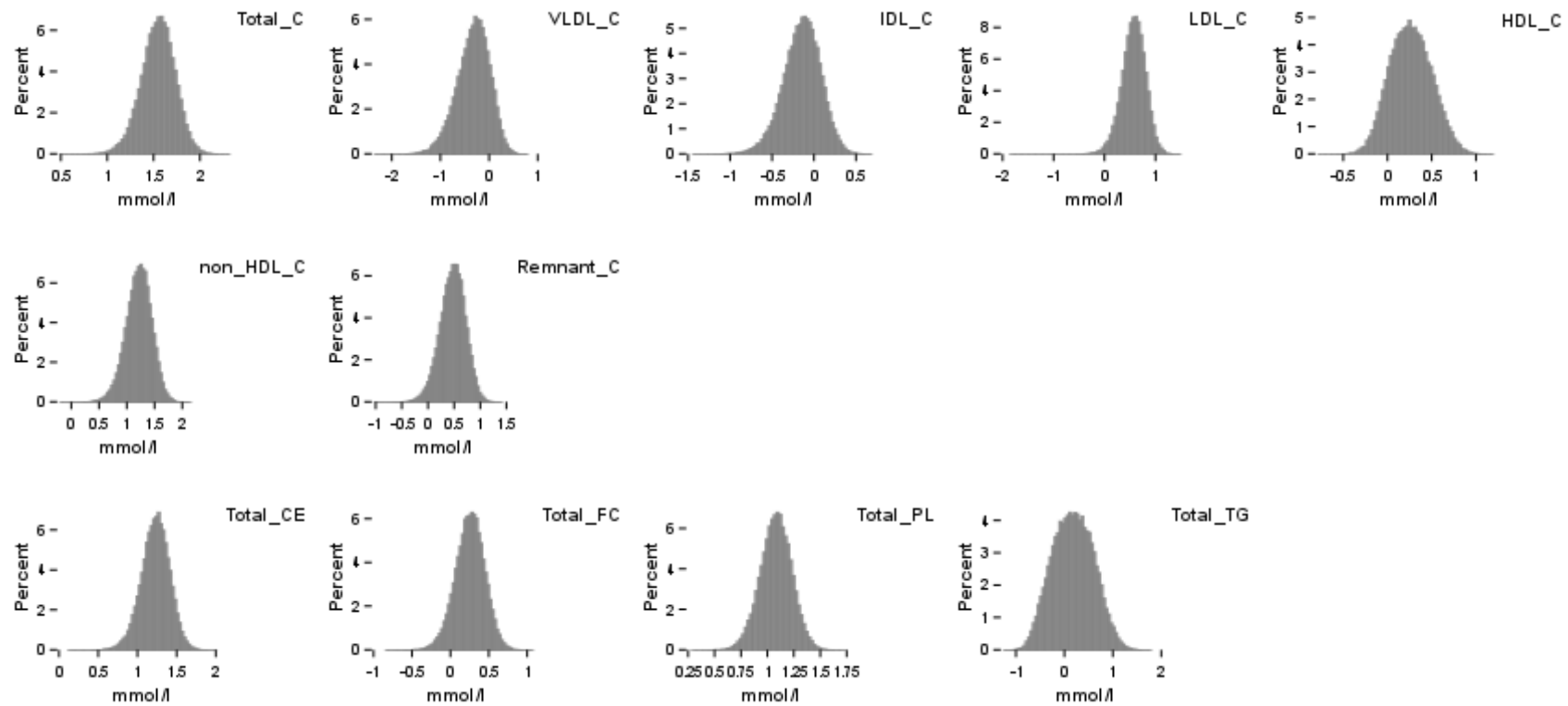
Composition of lipoproteins



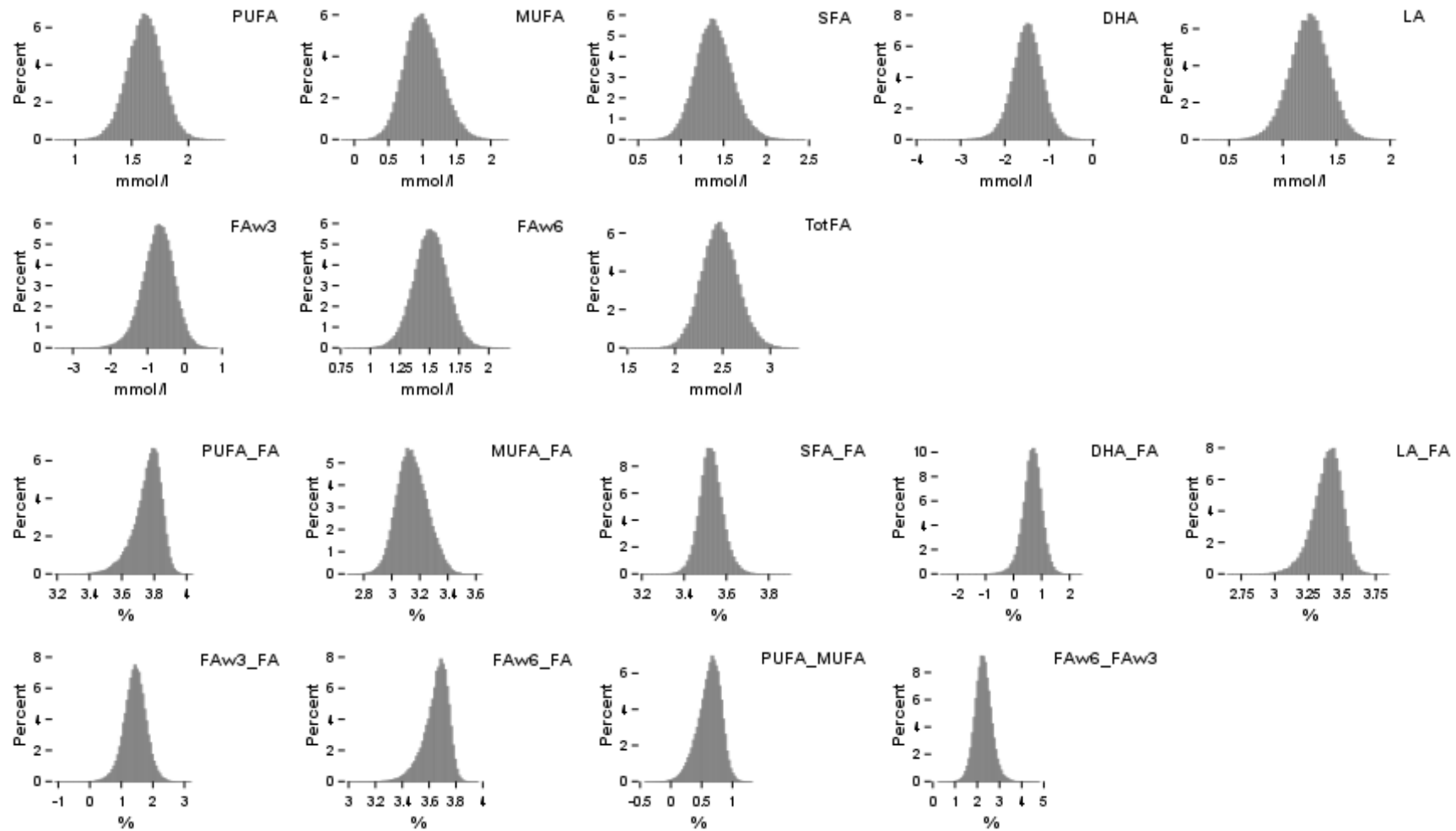
Size of lipoproteins



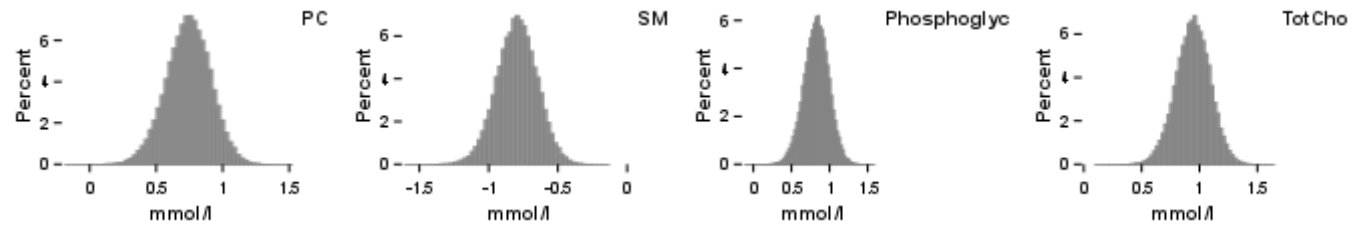
Total lipids



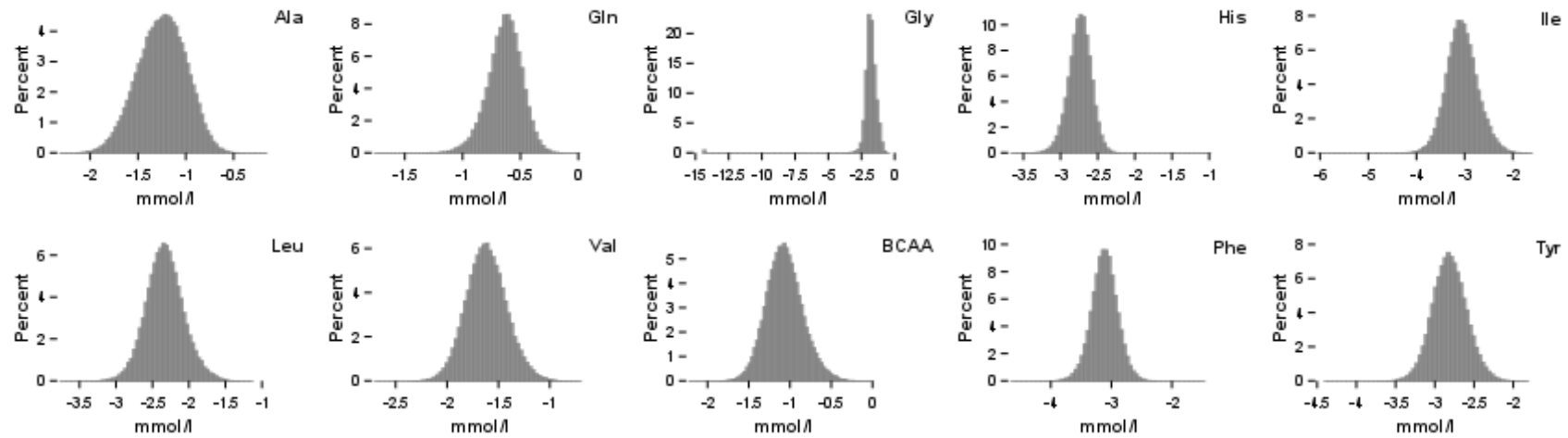
Fatty acids



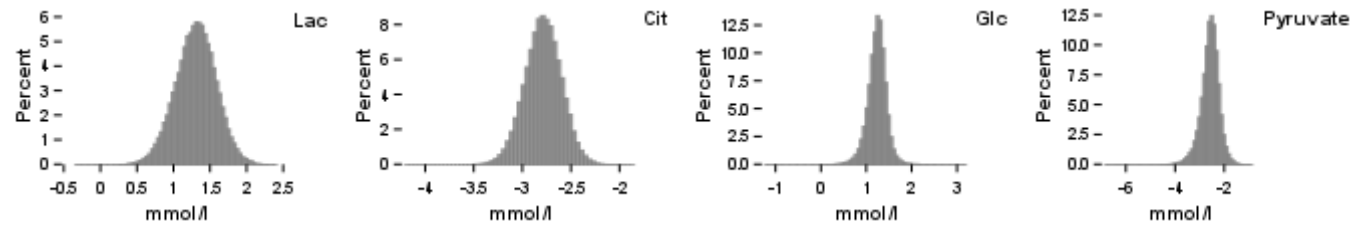
Cholines



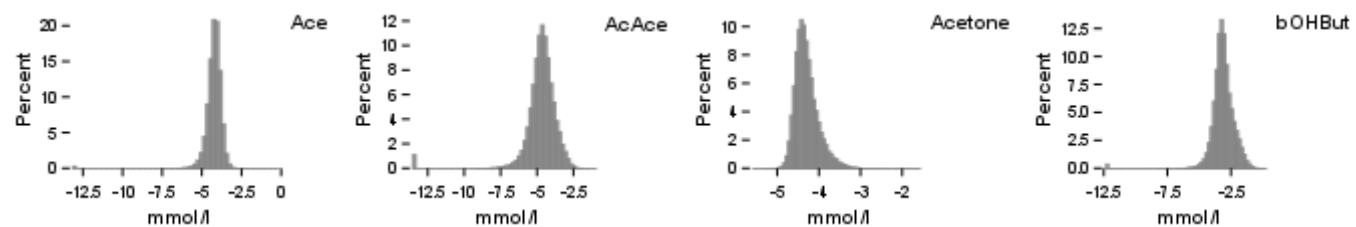
Amino acids



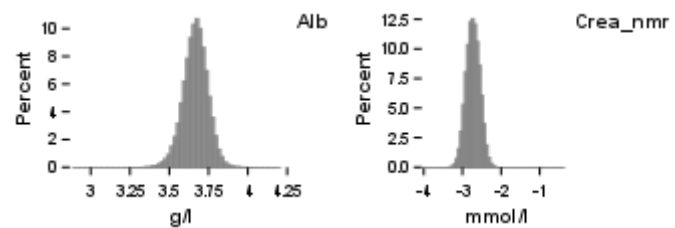
Glycolysis related metabolites



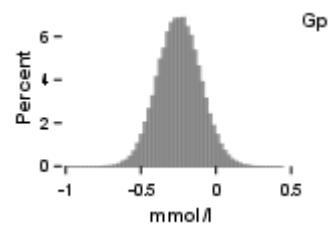
Ketone bodies



Fluid balance

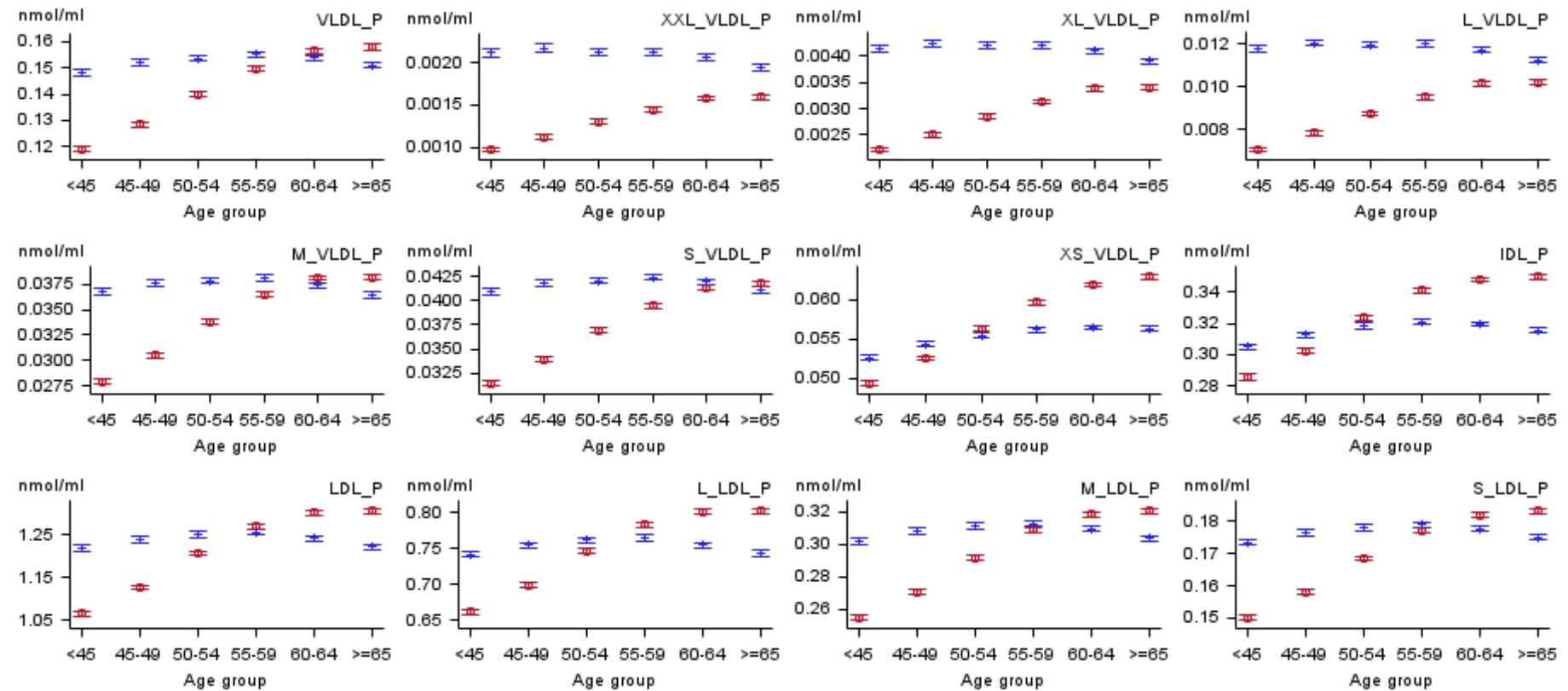


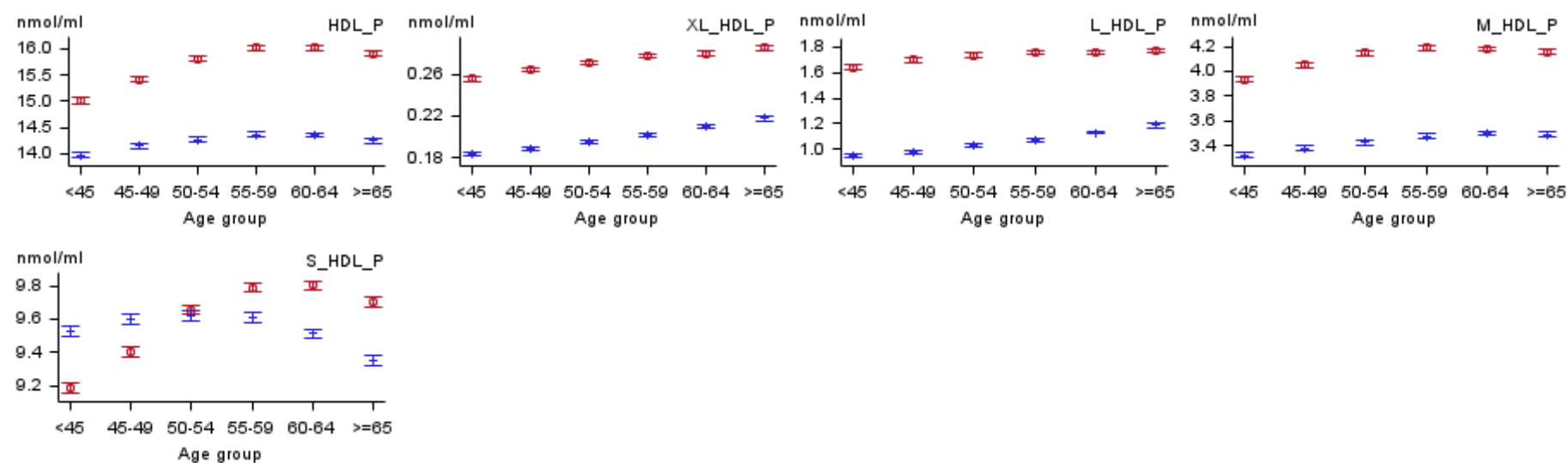
Inflammation



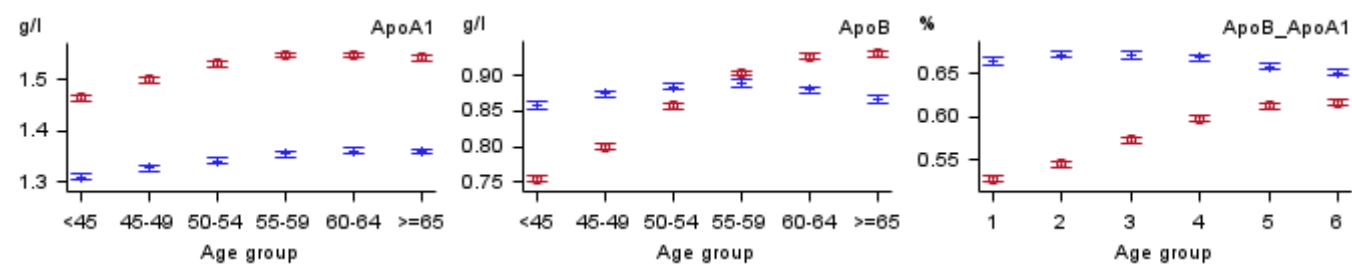
S4.3 Means of NMR-derived metabolites by age and sex

Lipoprotein particle concentration

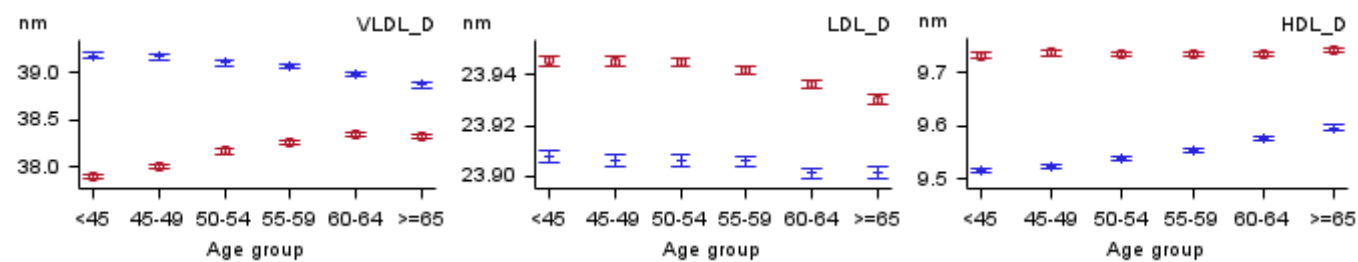




Apolipoprotein

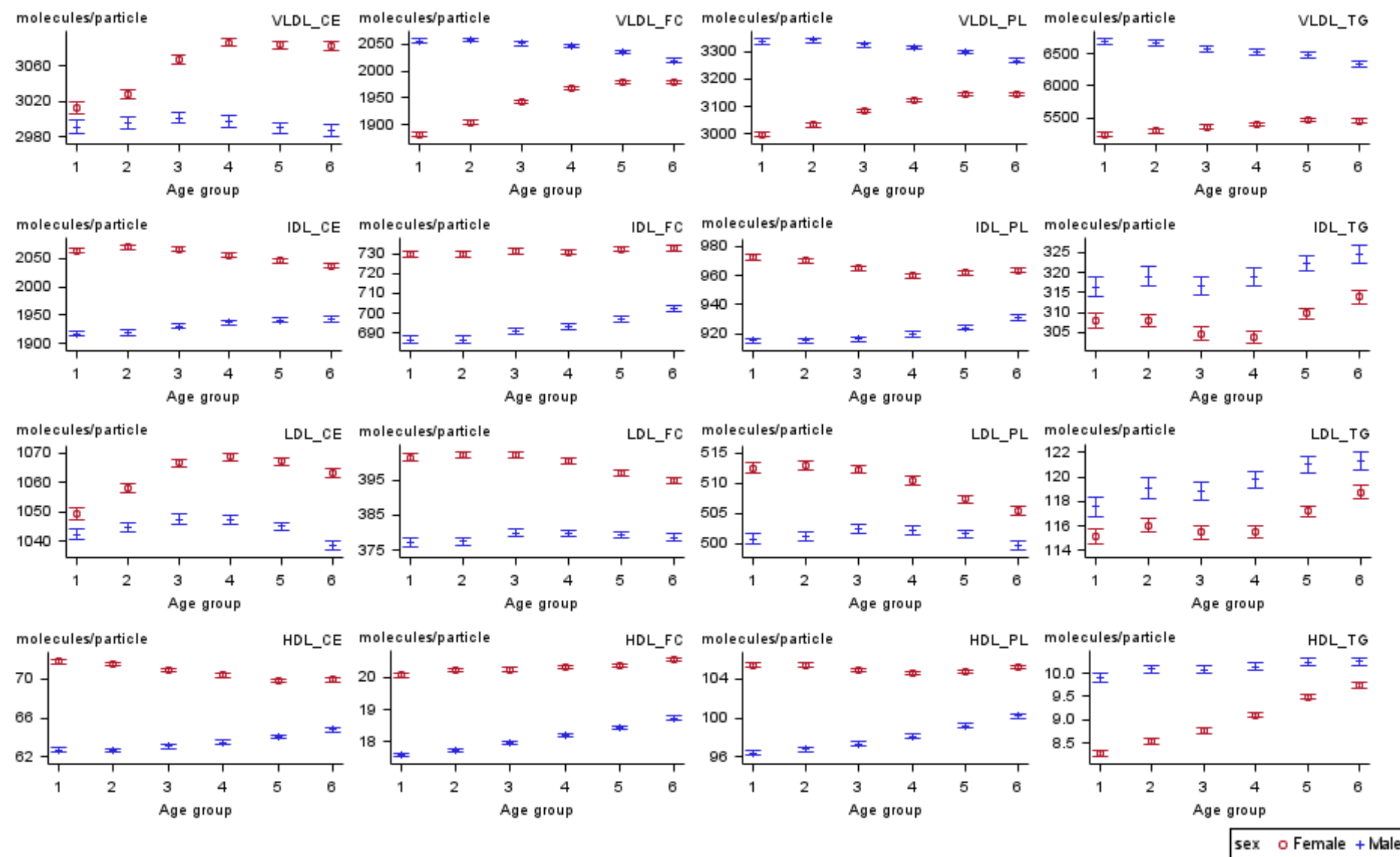


Lipoprotein particle diameter

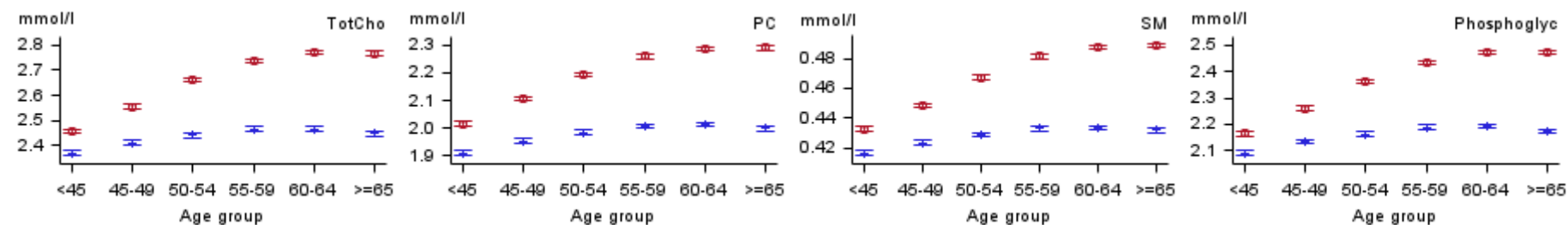


sex ○ Female + Male

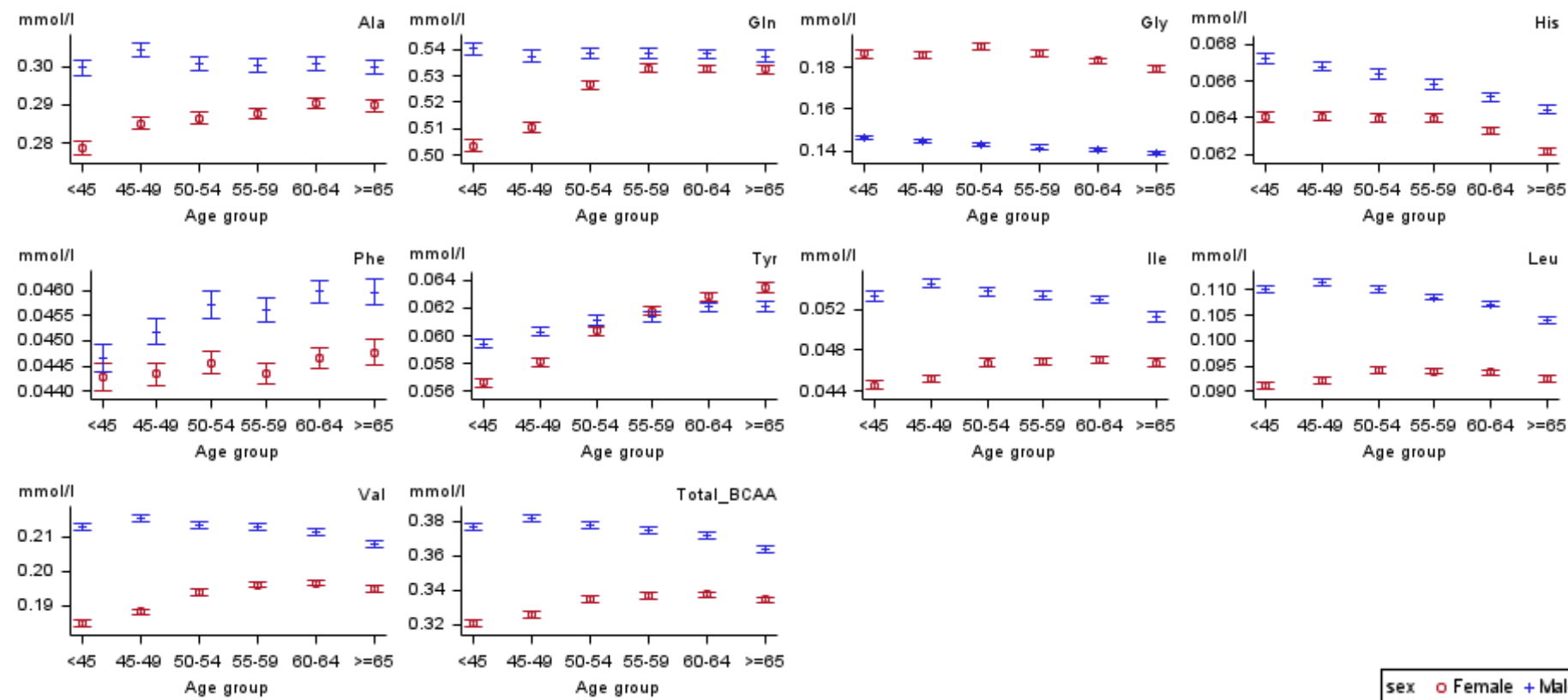
Lipoprotein composition



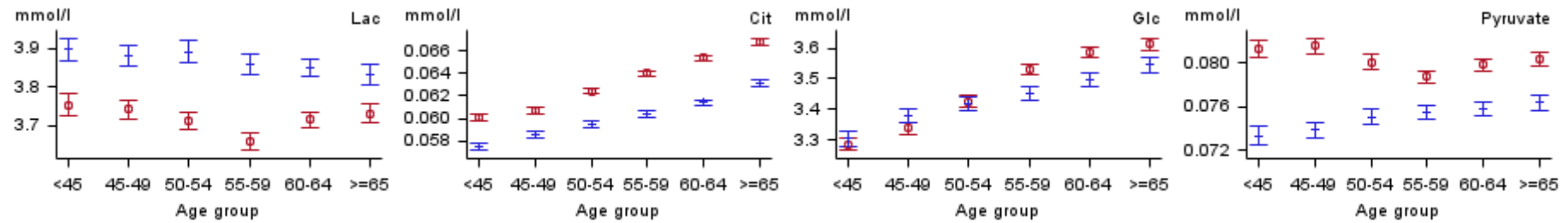
Cholines



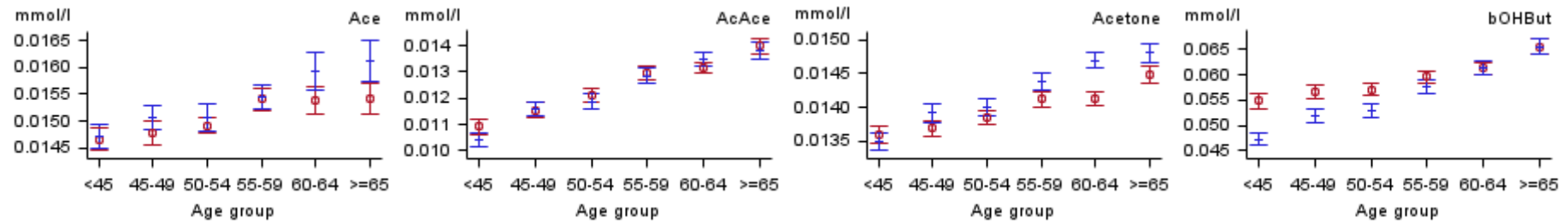
Amino acids



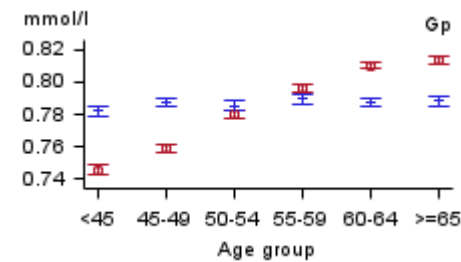
Glycolysis related metabolites



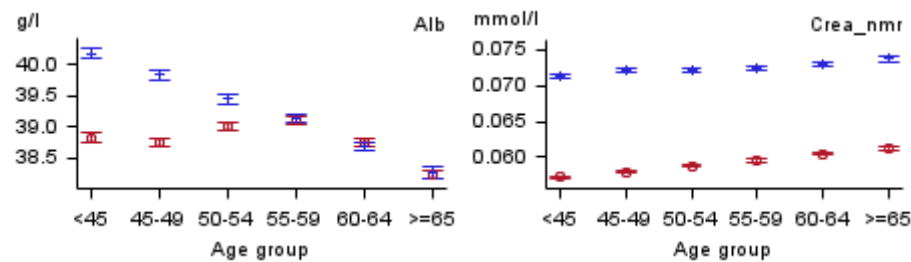
Ketone bodies



Inflammation

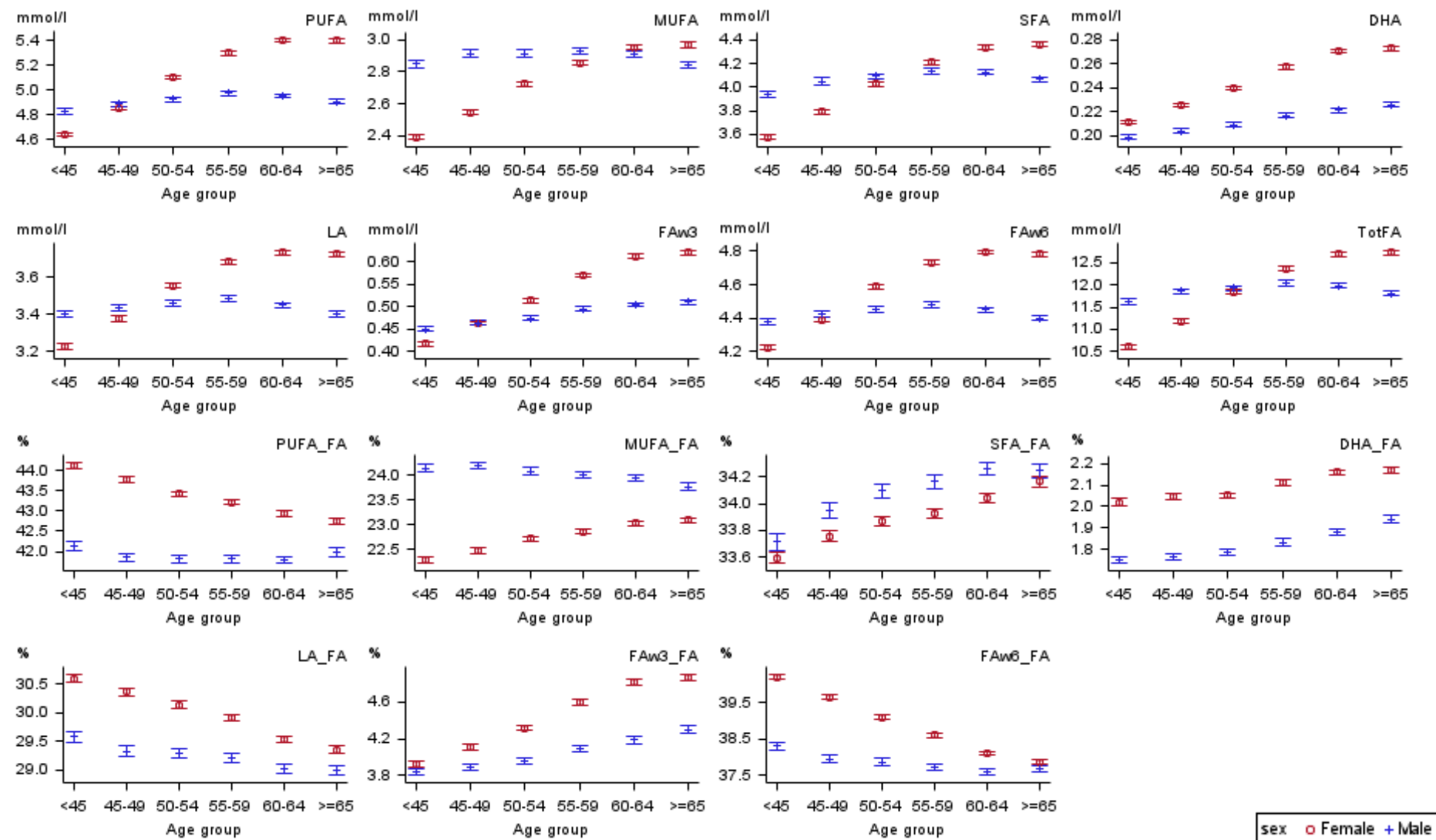


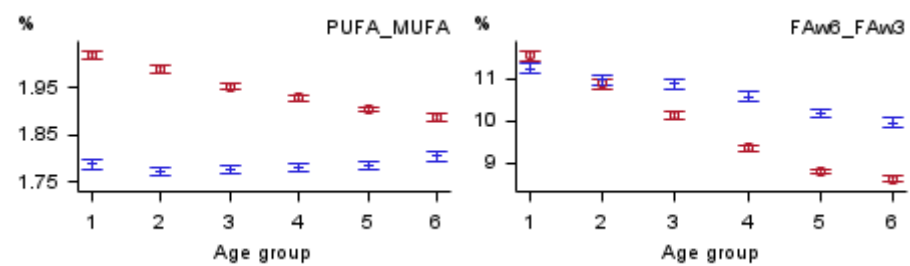
Fluid balance



sex ○ Female + Male

Fatty acids





sex ○ Female + Male

S5.1 Distribution of metabolites at baseline versus resurvey

A. Concentration of lipoproteins and apolipoproteins

	Baseline Mean	Baseline SD	Resurvey Mean	Resurvey SD
Lipoproteins (*10⁻⁵), mmol/L				
VLDL	14.6	4.4	14.7	4.4
Chylomicrons & extremely large	0.2	0.1	0.2	0.1
Very large	0.3	0.2	0.3	0.2
Large	1.0	0.5	1.0	0.5
Medium	3.6	1.3	3.6	1.2
Small	3.9	1.3	4.0	1.3
Very small	5.6	1.3	5.6	1.4
IDL	32.3	7.3	31.4	7.3
LDL	122.9	27.1	122.0	27.8
Large	75.4	16.9	74.7	17.3
Medium	30.2	7.4	30.0	7.5
Small	17.3	3.6	17.3	3.7
HDL	1512.3	222.2	1528.8	237.4
Very large	24.6	9.0	25.4	9.8
Large	148.2	71.9	154.8	75.5
Medium	384.3	84.1	391.5	87.2
Small	955.2	127.2	957.1	132.6
Apolipoproteins				
Apolipoprotein A-I, g/L	1.5	0.2	1.5	0.2
Apolipoprotein B, g/L	0.9	0.2	0.9	0.2
Apolipoprotein B to A-1 ratio	0.6	0.2	0.6	0.2

VDLD=very-low-density lipoproteins; IDL=intermediate-density lipoproteins; LDL=low-density lipoproteins; HDL=high-density lipoproteins.

B. Size and composition of lipoproteins

	Baseline Mean	Baseline SD	Resurvey Mean	Resurvey SD
Lipoprotein diameter, nm*				
VLDL	38.5	1.2	38.6	1.2
LDL	23.9	0.1	23.9	0.1
HDL	9.7	0.2	9.7	0.2
Lipoprotein composition (*10), molecules/particle †				
VLDL				
Esterified cholesterol	303.3	23.8	299.4	26.4
Free cholesterol	198.2	16.4	198.7	15.5
Phospholipids	316.8	33.5	318.2	31.6
Triglycerides	580.3	18.1	593.6	18.3
IDL				
Esterified cholesterol	201.5	17.2	201.1	17.0
Free cholesterol	72.0	6.6	73.0	6.5
Phospholipids	95.0	7.9	96.3	7.7
Triglycerides	30.8	7.5	32.1	7.8
LDL				
Esterified cholesterol	105.5	6.6	105.0	6.8
Free cholesterol	39.2	4.1	38.6	4.0
Phospholipids	50.6	3.0	50.2	2.9
Triglycerides	11.6	2.5	11.8	2.4
HDL				
Esterified cholesterol	6.8	0.9	6.9	0.8
Free cholesterol	2.0	0.3	2.0	0.3
Phospholipids	10.2	0.9	10.3	0.9
Triglycerides	0.9	0.3	0.9	0.3

*Diameter of IDL particles not given in UK Biobank. †Lipoprotein composition derived by dividing the lipidic contents by the particle concentration. VLDL=very-low-density lipoproteins; IDL=intermediate-density lipoproteins; LDL=low-density lipoproteins; HDL=high-density lipoproteins.

C. Total Lipids and fatty acids

	Baseline Mean	Baseline SD	Resurvey Mean	Resurvey SD
Total plasma lipids, mmol/L				
Total Cholesterol	4.7	0.9	4.7	0.9
VLDL Cholesterol	0.7	0.2	0.7	0.2
IDL Cholesterol	0.9	0.2	0.9	0.2
LDL Cholesterol	1.8	0.4	1.8	0.4
HDL Cholesterol	1.3	0.3	1.4	0.3
Non-HDL Cholesterol*	3.4	0.8	3.3	0.8
Remnant Cholesterol	1.6	0.4	1.6	0.4
Total Esterified Cholesterol	3.4	0.6	3.4	0.7
Total Free Cholesterol	1.3	0.3	1.3	0.3
Total Phospholipids	3.0	0.4	3.0	0.5
Total Triglycerides	1.3	0.5	1.3	0.5
Fatty acids concentration, mmol/L				
Polyunsaturated fatty acids	5.1	0.8	5.1	0.8
Monounsaturated fatty acids	2.8	0.8	2.8	0.7
Saturated fatty acids	4.0	0.9	4.1	0.9
Docosaehaenoic acid	0.2	0.1	0.2	0.1
Linoleic acid	3.5	0.7	3.5	0.7
Omega-3 fatty acids	0.5	0.2	0.5	0.2
Omega-6 fatty acids	4.5	0.7	4.5	0.7
Total fatty acids	11.9	2.3	12.0	2.3
Fatty acids ratio to total*				
Polyunsaturated fatty acids	43.1	3.4	42.7	3.3
Monounsaturated fatty acids	23.0	2.4	23.4	2.3
Saturated fatty acids	33.9	1.8	33.9	1.9
Docosaehaenoic acid	2.0	0.7	2.0	0.6
Linoleic acid	29.8	3.1	29.2	3.1
Omega-3 fatty acids	4.4	1.6	4.4	1.5
Omega-6 fatty acids	38.7	3.3	38.3	3.1
Other fatty acids ratio				
Omega-6 to Omega-3 fatty acids	9.9	4.1	9.7	3.7
Poly- to mono-unsaturated fatty acids	1.9	0.3	1.9	0.3

*Total cholesterol minus HDL cholesterol; *Ratio of concentration of given type of fatty acid to concentration of total fatty acid.

D. Other low-molecular-weight metabolites

	Baseline Mean	Baseline SD	Resurvey Mean	Resurvey SD
Cholines concentration, mmol/L				
Phosphatidylcholine	2.11	0.34	2.11	0.36
Sphingomyelin	0.45	0.07	0.45	0.07
Phosphoglycerides	2.28	0.37	2.29	0.38
Total cholines	2.57	0.38	2.58	0.40
Amino acids, mmol/L				
Alanine	0.29	0.07	0.30	0.07
Glutamine	0.53	0.07	0.56	0.08
Glycine	0.16	0.06	0.16	0.06
Histidine	0.07	0.01	0.07	0.01
Isoleucine	0.05	0.02	0.05	0.02
Leucine	0.10	0.03	0.09	0.02
Valine	0.20	0.04	0.20	0.04
Branched-chain amino acids	0.36	0.08	0.34	0.07
Phenylalanine	0.05	0.01	0.04	0.01
Tyrosine	0.06	0.01	0.06	0.01
Glycolysis related metabolites, mmol/L				
Lactate	3.68	0.94	2.48	0.78
Citrate	0.06	0.01	0.06	0.01
Glucose	3.49	0.74	4.33	0.66
Pyruvate	0.07	0.02	0.05	0.02
Ketone bodies, mmol/L				
Acetate	0.015	0.006	0.029	0.011
Aceto acetate	0.013	0.011	0.015	0.013
Acetone	0.014	0.005	0.013	0.004
Beta-hydroxybutyrate	0.055	0.053	0.050	0.046
Fluid balance				
Albumin, g/L	39.16	3.05	30.65	3.25
Creatinine, mmol/L	0.065	0.012	0.066	0.012
Inflammation, mmol/L				
Glycoprotein Acetyls	0.77	0.11	0.79	0.11

S5.2 Within-person variability of metabolites by methods

A. Concentration of lipoproteins and apolipoproteins

	MacMahon-Peto	Rosner	Self-correlation
Lipoproteins, mmol/L			
VLDL	0.62	0.62	0.63
Chylomicrons & extremely large	0.63	0.58	0.64
Very large	0.65	0.63	0.67
Large	0.64	0.63	0.65
Medium	0.60	0.61	0.61
Small	0.61	0.62	0.64
Very small	0.62	0.61	0.61
IDL	0.58	0.57	0.58
LDL	0.63	0.61	0.60
Large	0.65	0.60	0.60
Medium	0.56	0.55	0.54
Small	0.61	0.61	0.59
HDL	0.65	0.55	0.56
Very large	0.89	0.83	0.75
Large	0.89	0.82	0.79
Medium	0.69	0.60	0.61
Small	0.49	0.51	0.50
Apolipoproteins			
Apolipoprotein A-I, g/L	0.71	0.61	0.61
Apolipoprotein B, g/L	0.64	0.61	0.61
Apolipoprotein B to A-1 ratio	0.64	0.60	0.66

VLDL=very-low-density lipoproteins; IDL=intermediate-density lipoproteins; LDL=low-density lipoproteins; HDL=high-density lipoproteins.

B. Size and composition of lipoproteins

	MacMahon-Peto	Rosner	Self-correlation
Lipoprotein diameter, nm*			
VLDL	0.67	0.64	0.66
LDL	0.34	0.30	0.32
HDL	0.83	0.79	0.78
Lipoprotein composition, molecules/particle[†]			
VLDL			
Esterified cholesterol	0.60	0.58	0.54
Free cholesterol	0.60	0.60	0.63
Phospholipids	0.64	0.63	0.65
Triglycerides	0.64	0.59	0.59
IDL			
Esterified cholesterol	0.61	0.56	0.57
Free cholesterol	0.55	0.51	0.52
Phospholipids	0.52	0.49	0.50
Triglycerides	0.62	0.56	0.57
LDL			
Esterified cholesterol	0.50	0.46	0.41
Free cholesterol	0.58	0.57	0.56
Phospholipids	0.50	0.49	0.49
Triglycerides	0.61	0.54	0.57
HDL			
Esterified cholesterol	0.81	0.78	0.79
Free cholesterol	0.82	0.78	0.73
Phospholipids	0.63	0.58	0.57
Triglycerides	0.58	0.56	0.57

*Diameter of IDL particles not given in UK Biobank. [†]Lipoprotein composition derived by dividing the lipidic contents by the particle concentration. VLDL=very-low-density lipoproteins; IDL=intermediate-density lipoproteins; LDL=low-density lipoproteins; HDL=high-density lipoproteins.

C. Total Lipids and fatty acids

	MacMahon-Peto	Rosner	Self-correlation
Total plasma lipids, mmol/L			
Total Cholesterol	0.63	0.59	0.57
VLDL Cholesterol	0.62	0.62	0.63
IDL Cholesterol	0.63	0.59	0.59
LDL Cholesterol	0.56	0.56	0.56
HDL Cholesterol	0.79	0.72	0.70
Non-HDL Cholesterol*	0.62	0.59	0.59
Remnant Cholesterol	0.63	0.60	0.60
Total Esterified Cholesterol	0.63	0.58	0.57
Total Free Cholesterol	0.64	0.60	0.58
Total Phospholipids	0.64	0.58	0.57
Total Triglycerides	0.61	0.60	0.63
Fatty acids concentration, mmol/L			
Polyunsaturated fatty acids	0.60	0.58	0.59
Monounsaturated fatty acids	0.59	0.58	0.60
Saturated fatty acids	0.57	0.54	0.56
Docosahexaenoic acid	0.52	0.48	0.54
Linoleic acid	0.56	0.54	0.53
Omega-3 fatty acids	0.58	0.55	0.61
Omega-6 fatty acids	0.59	0.56	0.56
Total fatty acids	0.59	0.57	0.58
Fatty acids ratio to total*			
Polyunsaturated fatty acids	0.58	0.56	0.57
Monounsaturated fatty acids	0.62	0.62	0.63
Saturated fatty acids	0.43	0.42	0.38
Docosahexaenoic acid	0.51	0.47	0.52
Linoleic acid	0.56	0.56	0.55
Omega-3 fatty acids	0.53	0.51	0.56
Omega-6 fatty acids	0.59	0.57	0.60
Other fatty acids ratio			
Omega-6 to Omega-3 fatty acids	0.59	0.59	0.62
Poly- to mono-unsaturated fatty acids	0.59	0.51	0.57

*Total cholesterol minus HDL cholesterol; *Ratio of concentration of given type of fatty acid to concentration of total fatty acid.

D. Other low-molecular-weight metabolites

	MacMahon-Peto	Rosner	Self-correlation
Cholines concentration, mmol/L			
Phosphatidylcholine	0.64	0.56	0.56
Sphingomyelin	0.65	0.57	0.55
Phosphoglycerides	0.63	0.55	0.55
Total cholines	0.64	0.56	0.55
Amino acids, mmol/L			
Alanine	0.23	0.19	0.17
Glutamine	0.34	0.33	0.32
Glycine	0.68	0.63	0.60
Histidine	0.18	0.18	0.16
Isoleucine	0.24	0.18	0.24
Leucine	0.31	0.23	0.30
Valine	0.35	0.31	0.38
Branched-chain amino acids	0.30	0.27	0.35
Phenylalanine	0.26	0.28	0.26
Tyrosine	0.33	0.32	0.36
Glycolysis related metabolites, mmol/L			
Lactate	0.11	0.13	0.16
Citrate	0.39	0.35	0.32
Glucose	0.22	0.27	0.23
Pyruvate	0.20	0.19	0.22
Ketone bodies, mmol/L			
Acetate	0.25	0.23	0.14
Acetoacetate	0.16	0.16	0.09
Acetone	0.21	0.15	0.15
Beta-hydroxybutyrate	0.19	0.16	0.21
Fluid balance			
Albumin, g/L	0.33	0.31	0.32
Creatinine, mmol/L	0.67	0.58	0.57
Inflammation, mmol/L			
Glycoprotein Acetyls	0.55	0.55	0.57

216

S6.1 Correlations coefficients of lipoprotein particle concentrations with selected NMR metabolites at baseline

Spearman Correlations	VLDL concentration	IDL concentration	LDL concentration	HDL concentration
Apolipoprotein				
ApoA-1	-0.12	0.09	-0.02	0.96
ApoB	0.87	0.91	0.99	0.14
Lipoprotein concentration				
VLDL	1.00	0.71	0.85	-0.004
IDL	0.71	1.00	0.87	0.20
LDL	0.85	0.87	1.00	0.14
HDL	-0.004	0.20	0.14	1.00
Lipoprotein particle size				
VLDL	0.67	0.14	0.38	-0.24
LDL	-0.19	0.08	0.09	0.16
HDL	-0.43	-0.16	0.31	0.51
Cholesterol				
Non-HDL-C	0.81	0.92	0.97	0.30
LDL-C	0.72	0.86	0.95	0.35
HDL-C	-0.32	0.02	-0.11	0.87

Spearman correlations of lipoprotein particle concentrations with selected NMR metabolites. VLDL=very-low-density lipoproteins; LDL=low-density lipoproteins; HDL=high-density lipoproteins; Apo=apolipoprotein; C=cholesterol.

S6.2 Sensitivity analyses of lipoprotein characteristics vs coronary heart disease risk

Particle type	Original adjusted model		Adjustment for other potential confounders		Exclusion of events in first two years of follow-up	
	HR (95% CI)	LR χ^2	HR (95% CI)	LR χ^2	HR (95% CI)	LR χ^2
Particle concentration ($\mu\text{mol/L}$)						
VLDL	1.21 (1.74, 1.25)	141.4	1.19 (1.16, 1.23)	118.0	1.21 (1.17, 1.24)	122.0
LDL	1.20 (1.17, 1.24)	134.8	1.20 (1.16, 1.24)	128.1	1.19 (1.15, 1.23)	108.2
HDL	0.87 (0.84, 0.91)	54.9	0.85 (0.82, 0.88)	75.9	0.89 (0.85, 0.92)	39.9
Particle size (nm)						
VLDL	0.92 (0.87,0.97)	9.9	0.91 (0.86,0.95)	13.5	0.91 (0.87,0.97)	10.7
LDL	1.04 (0.99,1.07)	3.7	1.04 (1.00,1.08)	5.6	1.03 (0.99,1.07)	2.0
HDL	0.97 (0.92,1.02)	1.7	1.00 (0.95,1.05)	<0.1	0.97 (0.92,1.02)	1.8
Triglyceride (molecules per particle)						
VLDL	0.93 (0.89,0.98)	7.2	0.92 (0.88,0.97)	9.3	0.92 (0.87, 0.97)	9.4
LDL	1.17 (1.11,1.23)	33.9	1.16 (1.10,1.22)	29.9	1.15 (1.09, 1.22)	26.2
HDL	1.17 (1.07,1.28)	12.1	1.20 (1.09,1.31)	15.3	1.15 (1.05, 1.27)	8.9

Hazard ratios (HR) per usual SD higher level of mean lipoprotein trait. 'Original model' is the stratified Cox proportional-hazards model by age and sex, with difference adjustment, included ethnicity, education, region, Townsend Deprivation Index, smoking, alcohol, and body mass index, and further adjusted for concentrations of VLDL, LDL and HDL particles (further mutually adjusted for other particle concentrations for the analyses of lipoprotein concentrations). 'Further adjustments for other potential confounders' included waist circumference, systolic blood pressure, diabetes and chronic kidney disease at baseline, fasting time, and spectrometer. Further exclusion of events in first two years of follow-up excluded 368 CHD event. Likelihood ratio (LR) χ^2 improvement with the addition of the given particle concentration to the model with stated adjustments. VLDL=very-low-density lipoproteins; LDL=low-density lipoproteins; HDL=high-density lipoproteins.

S6.3 Improvement of model fit by adding different measures

Model (+ add measures)	LR χ^2	Improvement in LR χ^2
A: Model containing non-lipid confounders, non-HDL cholesterol and HDL cholesterol	829.5	-
Adding triglycerides-related measures into the model:		
A + Total triglycerides	830.7	1.2
A + Mean triglycerides molecules per LDL particle	854.9	25.4
B: Model containing non-lipid confounders, and concentrations of VLDL, LDL and HDL particle	834.9	-
Adding triglycerides-related measures into the model:		
B + Total triglycerides	836.3	1.4
B + Mean triglycerides molecules per LDL particle	868.2	33.3

The goodness-of-fit of model by adding each biomarker. Model containing non-lipid confounders indicates the stratified Cox proportional-hazards model by age and sex, and further adjusted for ethnicity, education, region, Townsend Deprivation Index, smoking, alcohol, and body mass index. Improvement in LR χ^2 by the addition of the listed lipid/lipoprotein biomarker to the basic model (A/B) represent the change in the LR χ^2 between models with and without the listed lipid/lipoprotein biomarker. LR=likelihood ratio; VLDL=very-low-density lipoproteins; LDL=low-density lipoproteins; HDL=high-density lipoproteins; ApoB=apolipoprotein B.

S6.4 Discordance analysis of total triglycerides and triglycerides in LDL

Model with different adjustment	Total triglycerides	Triglycerides in LDL
	HR (95% CI)	HR (95% CI)
Adjust for non-HDL cholesterol and HDL cholesterol	1.04 (0.94, 1.14)	1.16 (1.07, 1.25)
Adjust for concentrations of VLDL, LDL and HDL particles	1.03 (0.87, 1.23)	1.20 (1.10, 1.32)

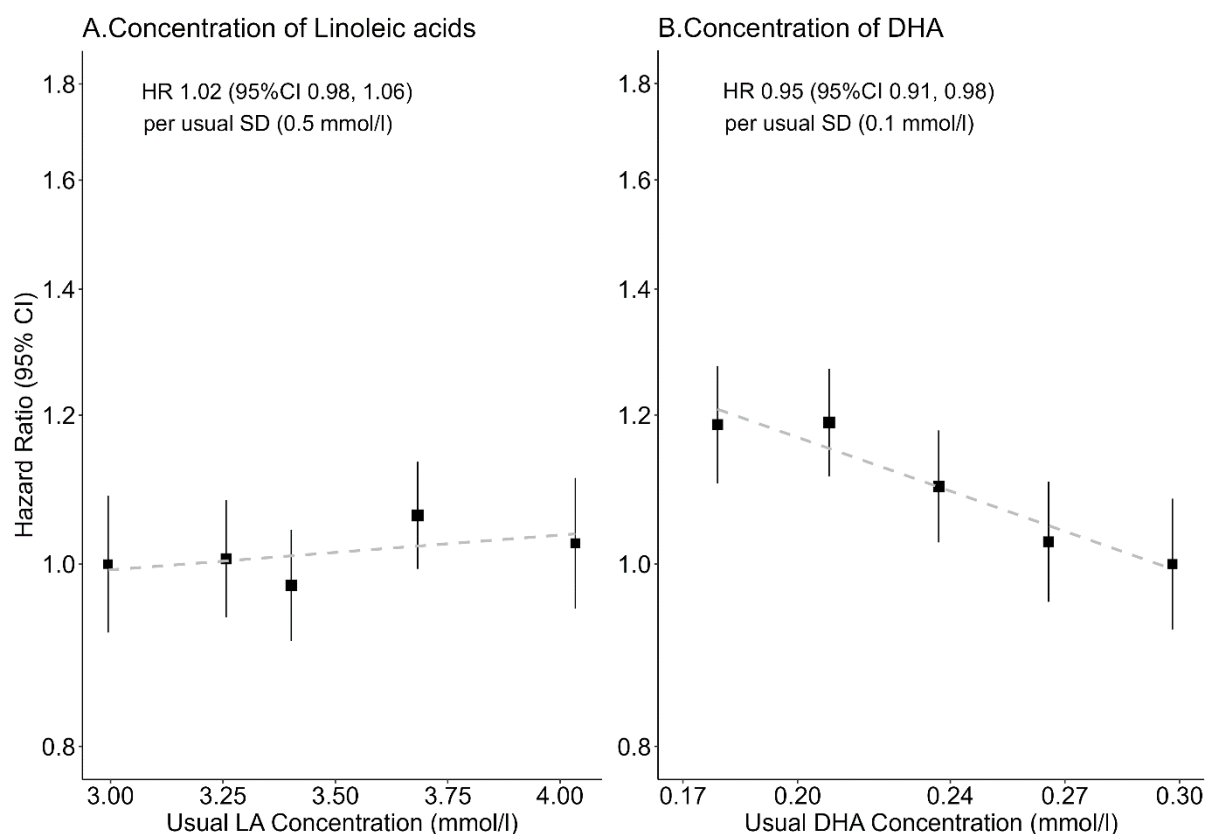
Hazard ratios (HR) of two highly correlated measurements, total triglycerides and average triglycerides level per LDL, were compared among 33,314 (38.6%) participants who are discordant in total triglycerides and average triglycerides in LDL particle (those with an absolute difference of >20 percentage points between percentiles of total triglycerides and average triglycerides in LDL particle). HRs calculated by Cox proportional-hazards models stratified by age and sex, adjusted for ethnicity, education, region, Townsend Deprivation Index, smoking, alcohol, and body-mass index, and further adjusted for different lipids or lipoproteins in different rows.

S7.1 Association of fatty acids concentration with coronary heart disease

Fatty acids	Fatty acids (mmol/L)		A. Adjusted for age and sex		B. Further adjusted for social-economic and lifestyle factors		C. Further adjusted for body-mass index	
	Baseline mean	Usual SD	HR (95% CI)	LR χ^2 *	HR (95% CI)	LR χ^2 *	HR (95% CI)	LR χ^2 *
SFA	4.1	0.7	1.16 (1.13,1.20)	103.5	1.15 (1.11,1.18)	85.3	1.11 (1.08,1.14)	47.8
MUFA	2.8	0.6	1.20 (1.16,1.23)	153.2	1.16 (1.13,1.19)	106.3	1.12 (1.09,1.15)	57.7
PUFA								
Omega-6 PUFA	4.5	0.5	1.10 (1.07,1.14)	37.5	1.10 (1.07,1.14)	39.5	1.09 (1.06,1.13)	32.3
Linoleic acids	3.5	0.5	1.09 (1.06,1.13)	31.6	1.10 (1.06,1.13)	34.9	1.09 (1.06,1.13)	31.7
Omega-3 PUFA	0.5	0.2	0.96 (0.93,0.99)	7.2	1.00 (0.97,1.03)	<0.1	1.00 (0.97,1.03)	<0.1
DHA	0.2	0.1	0.89 (0.86,0.92)	49.2	0.95 (0.92,0.98)	10.3	0.97 (0.94,1.00)	4.3

Hazard ratios (HR) per usual SD higher level of mean fatty acids concentration. *HRs calculated by Cox proportional-hazards models with: (A) stratification by age and sex; (B) stratification by age and sex, and adjusted for ethnicity, education, region, Townsend Deprivation Index, smoking, and alcohol; (C) model B with further adjustment for body-mass index. Likelihood ratio (LR) χ^2 improvement with the addition of the given factors to the model with stated adjustments. SFA=Saturated fatty acids; MUFA=Monounsaturated fatty acids; PUFA=Polyunsaturated fatty acids; DHA= Docosahexaenoic acid.

S7.2 Subtype of polyunsaturated fatty acids concentration vs coronary heart disease risk, with adjustment for other fatty acids



Hazard ratios (HR) per usual SD higher level of fatty acids concentration. HRs calculated by Cox proportional-hazards models, stratified by age and sex, and adjusted for ethnicity, education, region, Townsend Deprivation Index, smoking, alcohol, body-mass index, and mutual adjusted for other fatty acids (adjust for omega-3 FA, saturated FA and monounsaturated FA for the analysis of linoleic acids; adjust for omega-6 FA, saturated FA and monounsaturated FA for the analysis of DHA). Area of the square is inversely proportional to the variance of the category-specific log risk. DHA= Docosahexaenoic acid.

S7.3 Sensitivity analyses of fatty acids and coronary heart disease risk

Fatty acids concentration (mmol/L)	Original mutually adjusted model		Exclusion of events in first two years of follow-up		Further adjustment for other potential confounders		Further adjustment for dietary habit factors	
	HR (95% CI)	LR χ^2	HR (95% CI)	LR χ^2	HR (95% CI)	LR χ^2	HR (95% CI)	LR χ^2
SFA	1.12 (1.07,1.17)	25.3	1.12 (1.07, 1.18)	24.0	1.08 (1.04, 1.13)	11.8	1.09 (1.04, 1.14)	12.7
MUFA	1.13 (1.08,1.18)	31.4	1.13 (1.08, 1.18)	28.5	1.10 (1.05, 1.15)	18.0	1.10 (1.05, 1.15)	18.0
PUFA								
Omega-6	1.02 (0.98,1.07)	1.1	1.02 (0.97, 1.07)	0.8	1.04 (0.99, 1.09)	2.4	1.03 (0.98, 1.08)	1.4
Linoleic acids	1.02 (0.98,1.06)	0.7	1.02 (0.97, 1.06)	0.5	1.03 (0.99, 1.08)	2.3	1.03 (0.98, 1.07)	1.5
Omega-3	0.93 (0.90,0.97)	13.4	0.94 (0.90, 0.97)	11.3	0.94 (0.90, 0.98)	10.8	0.94 (0.90, 0.98)	8.2
DHA	0.95 (0.91,0.98)	9.1	0.95 (0.91, 0.99)	7.6	0.95 (0.92, 0.99)	7.6	0.95 (0.92, 0.99)	5.5

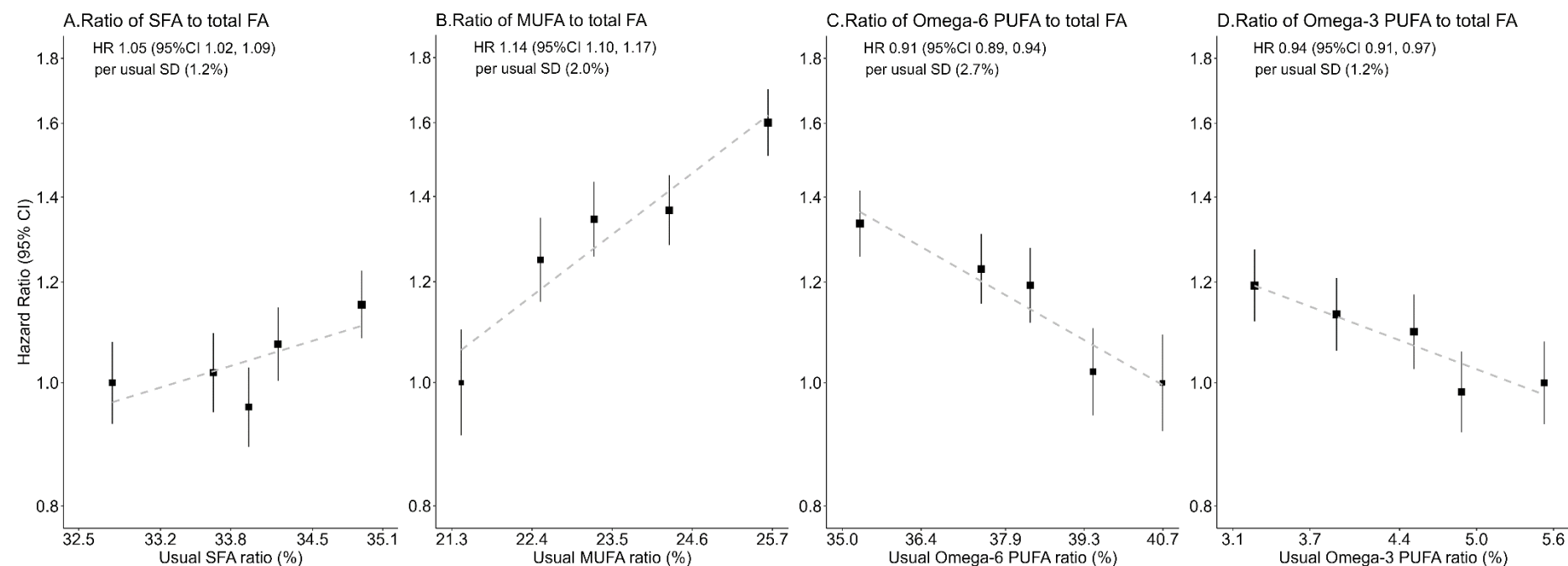
Hazard ratios (HR) per usual SD higher level of fatty acids concentration or ratio. HRs calculated by Cox proportional-hazards models, stratified by age and sex, 'Original mutually adjusted model' adjusted for ethnicity, education, region, Townsend Deprivation Index, smoking, alcohol, and body mass index, and mutual adjusted for other fatty acids (as last row of each type of fatty acids in Figure 7.1). 'Exclusion of events in first two years of follow-up' excluded 368 CHD events. 'Further adjustments for other potential confounders' included further adjustment for waist circumference, systolic blood pressure, diabetes and chronic kidney disease at baseline, fasting time, and spectrometer; 'Further adjustments for dietary habit factors' included intake frequency of whole grains, fruit, vegetables, cheese, red meat, oily fish, and intake of dairy product, among 83,746 participants with complete data. Likelihood ratio (LR) χ^2 improvement with the addition of the given factors to the model with stated adjustments. SFA=Saturated fatty acids; MUFA=Monounsaturated fatty acids; PUFA=Polyunsaturated fatty acids; DHA= Docosahexaenoic acid.

S7.4 Association of fatty acids ratios with coronary heart disease

Fatty acids ratio, relative to total fatty acids	Fatty acids ratio (%)		A. Adjusted for age and sex		B. Further adjusted for social-economic and lifestyle factors		C. Further adjusted for body-mass index	
	Baseline mean	Usual SD	HR (95% CI)	LR χ^2 *	HR (95% CI)	LR χ^2 *	HR (95% CI)	LR χ^2 *
SFA (%)	34.0	1.2	1.10 (1.07,1.13)	40.1	1.09 (1.05,1.12)	30.3	1.05 (1.02,1.09)	12.0
MUFA (%)	23.3	2.0	1.25 (1.22,1.29)	217.9	1.19 (1.15,1.23)	121.8	1.14 (1.10,1.17)	58.9
PUFA (%)								
Omega-6 PUFA (%)	38.4	2.7	0.85 (0.83,0.88)	112.8	0.87 (0.85,0.90)	80.3	0.91 (0.89,0.94)	32.9
Linoleic acids (%)	29.6	2.4	0.89 (0.86,0.92)	59.4	0.91 (0.88,0.94)	36.4	0.95 (0.93,0.99)	8.3
Omega-3 PUFA (%)	4.3	1.2	0.87 (0.85,0.90)	67.8	0.93 (0.90,0.96)	20.6	0.94 (0.91,0.97)	13.6
DHA (%)	2.0	0.5	0.82 (0.79,0.85)	150.4	0.87 (0.85,0.90)	65.8	0.91 (0.88,0.94)	34.3

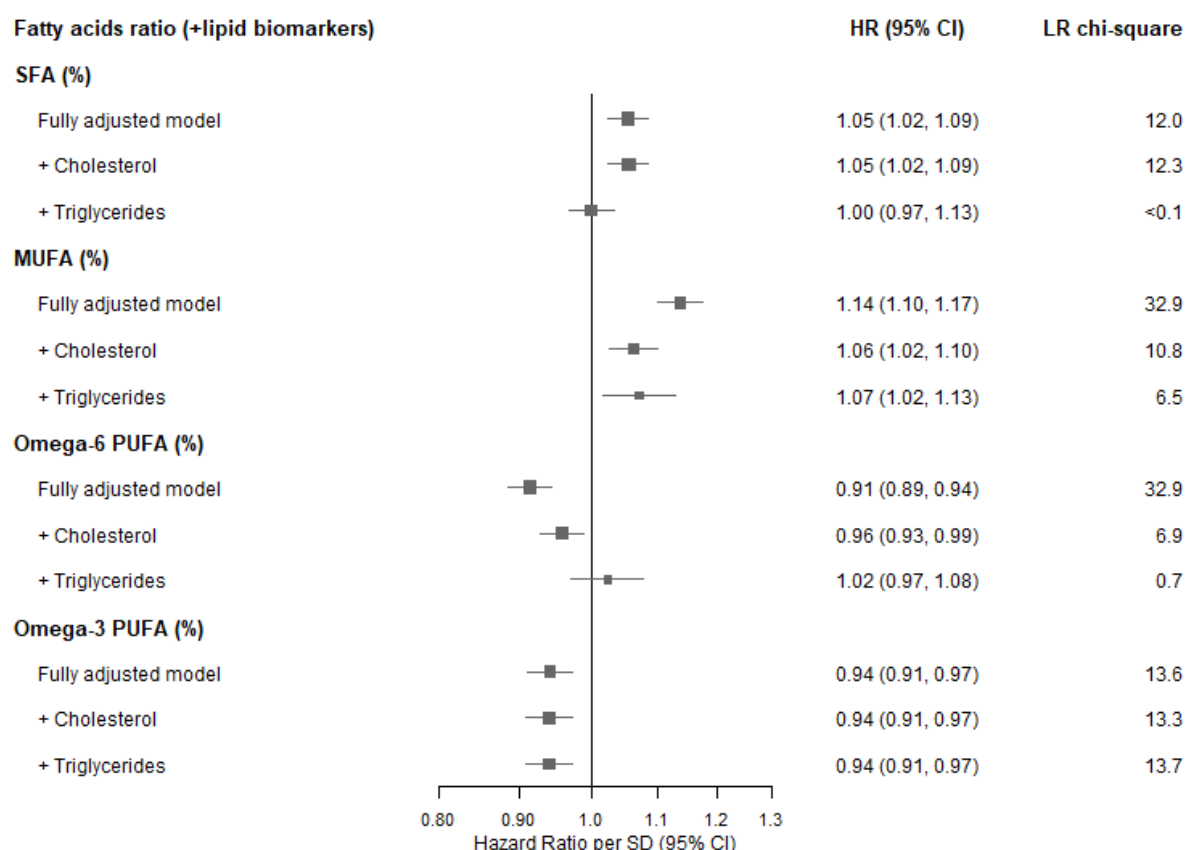
Hazard ratios (HR) per usual SD higher level of mean fatty acids ratio. *HRs calculated by Cox proportional-hazards models with: (A) stratification by age and sex; (B) stratification by age and sex, and adjusted for ethnicity, education, region, Townsend Deprivation Index, smoking, and alcohol; (C) model B with further adjustment of body-mass index. Likelihood ratio (LR) χ^2 improvement with the addition of the given factors to the model with stated adjustments. SFA=Saturated fatty acids; MUFA=Monounsaturated fatty acids; PUFA=Polyunsaturated fatty acids; DHA= Docosahexaenoic acid.

S7.5 Fatty acids ratio vs coronary heart disease risk in fully adjusted model



Hazard ratios (HR) per usual SD higher level of fatty acids ratio). HRs calculated by Cox proportional-hazards models, stratified by age and sex, and adjusted for ethnicity, education, region, Townsend Deprivation Index, smoking, alcohol, and body-mass index. Area of the square is inversely proportional to the variance of the category-specific log risk. SFA=Saturated fatty acids; MUFA=Monounsaturated fatty acids; PUFA=Polyunsaturated fatty acids.

S7.6 Risk of coronary heart disease by usual fatty acids ratio, with further adjustment for lipids



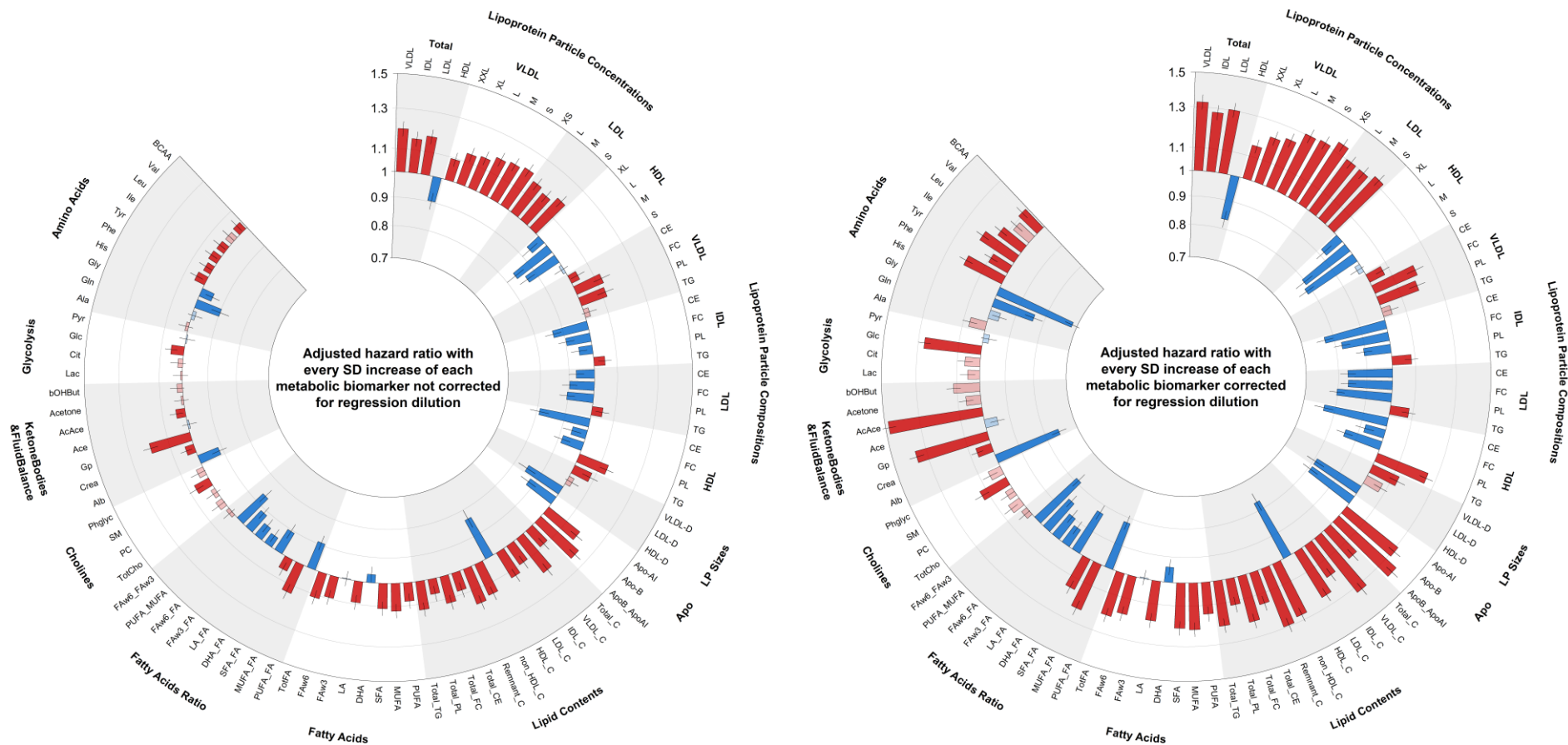
Hazard ratios (HR) per usual SD higher level of mean fatty acids ratio. The fully adjusted model represented Cox proportional-hazards model with stratification by age and sex, and adjustment for ethnicity, education, region, Townsend Deprivation Index, smoking, alcohol, and body-mass index (as last column of each type of fatty acids in S7.4), and further adjustment for non-HDL-cholesterol and HDL-cholesterol, or further adjustment for total triglycerides. Likelihood ratio (LR) χ^2 improvement with the addition of the given factors to the model with stated adjustments. SFA=Saturated fatty acids; MUFA=Monounsaturated fatty acids; PUFA=Polyunsaturated fatty acids.

S8.1 Sensitivity analyses of other metabolites and coronary heart disease risk

	Concentration (mmol/L)		Original Fully adjusted model	Exclusion of events in first two years of follow-up	Further adjustment for other confounders and dietary habit
	Baseline mean	Usual SD			
Cholines					
Phosphatidylcholine	2.112	0.268	1.02 (0.99, 1.05)	1.02 (0.99, 1.06)	1.00 (0.96, 1.03)
Sphingomyelins	0.453	0.050	1.07 (1.04, 1.11)*	1.07 (1.03, 1.11)*	1.06 (1.02, 1.10)*
Phosphoglycerides	2.285	0.284	1.03 (1.00, 1.07)	1.04 (1.00, 1.07)	1.01 (0.97, 1.04)
Total cholines	2.573	0.291	1.03 (1.00, 1.07)	1.04 (1.00, 1.07)	1.01 (0.97, 1.04)
Amino acids					
Alanine	0.293	0.031	1.01 (0.98, 1.04)	1.02 (0.99, 1.05)	1.01 (0.97, 1.05)
Glutamine	0.530	0.044	0.99 (0.96, 1.02)	0.99 (0.95, 1.02)	1.00 (0.97, 1.03)
Glycine	0.166	0.050	0.90 (0.86, 0.93)*	0.90 (0.86, 0.93)*	0.91 (0.87, 0.95)*
Histidine	0.064	0.004	0.95 (0.92, 0.98)*	0.95 (0.92, 0.98)*	0.95 (0.91, 0.98)*
Isoleucine	0.049	0.008	1.04 (1.01, 1.07)*	1.04 (1.01, 1.07)	1.04 (1.01, 1.07)
Leucine	0.099	0.015	1.04 (1.01, 1.07)*	1.03 (1.00, 1.07)	1.04 (1.01, 1.07)
Valine	0.201	0.025	1.03 (1.00, 1.07)	1.03 (1.00, 1.07)	1.03 (0.99, 1.07)
BCAA	0.350	0.048	1.04 (1.01, 1.07)*	1.03 (1.00, 1.07)	1.04 (1.00, 1.07)
Phenylalanine	0.045	0.005	1.05 (1.02, 1.08)*	1.05 (1.01, 1.09)*	1.05 (1.02, 1.08)*
Tyrosine	0.061	0.008	1.03 (1.00, 1.07)*	1.03 (0.99, 1.06)	1.04 (1.00, 1.07)
Glycolysis related					
Lactate	3.779	0.425	1.01 (0.98, 1.04)	1.02 (0.98, 1.05)	0.99 (0.96, 1.02)
Citrate	0.062	0.007	1.02 (0.99, 1.05)	1.03 (0.99, 1.06)	1.01 (0.98, 1.05)
Glucose	3.457	0.411	1.06 (1.03, 1.08)*	1.05 (1.02, 1.08)*	1.01 (0.98, 1.04)
Pyruvate	0.078	0.013	1.00 (0.96, 1.03)	1.00 (0.97, 1.04)	0.98 (0.95, 1.02)
Ketone bodies					
Acetate	0.015	0.004	0.99 (0.96, 1.02)	1.00 (0.97, 1.03)	1.00 (0.97, 1.02)
Acetoacetate	0.012	0.004	1.04 (1.01, 1.07)*	1.03 (1.00, 1.06)	1.03 (1.00, 1.06)
Acetone	0.014	0.002	1.01 (0.98, 1.04)	1.00 (0.97, 1.04)	1.01 (0.97, 1.03)
β-hydroxybutyrate	0.057	0.027	1.02 (0.99, 1.05)	1.02 (0.97, 1.04)	1.01 (0.98, 1.04)
Fluid balance					
Albumin, g/L	39.977	1.839	0.91 (0.89, 0.94)*	0.92 (0.89, 0.95)*	0.90 (0.97, 0.93)*
Creatinine	0.065	0.009	1.04 (1.01, 1.06)*	1.04 (1.01, 1.07)*	1.03 (1.00, 1.06)
Inflammation					
GlycA	0.786	0.086	1.19 (1.15, 1.23)*	1.20 (1.16, 1.24)*	1.16 (1.13, 1.20)*

Hazard ratios (HR) per usual SD higher level of particle concentration. HRs calculated by Cox proportional-hazards models with stratification by age and sex, and adjusted for ethnicity, education, region, Townsend Deprivation Index, smoking, alcohol, and body-mass index; 'Exclusion of events in first two years of follow-up' excluded 368 CHD events. 'Further adjustments for other potential confounders' included further adjustment for waist circumference, systolic blood pressure, diabetes and chronic kidney disease at baseline, fasting time, and spectrometer; 'Further adjustments for dietary habit factors' included intake frequency of whole grains, fruit, vegetables, cheese, red meat, oily fish, and intake of dairy product, among 83,746 participants with complete data. 'Usual SD' was obtained from the baseline SD through multiplication by square root of the regression dilution ratio, representing an unbiased estimate of the long-term average level in each baseline-defined group. * Represents P<0.05 after correction of false discovery rate. BCAA=branched-chain amino acids; GlycA=Glycoprotein acetyls.

S8.2 Metabolite profiling vs coronary heart disease risk w/o correcting for regression dilution bias on hazard ratios



Adjusted for age, sex, region, Townsend deprivation index, ethnicity, education, smoking and alcohol status, and BMI. The hazard ratios in left plot were not corrected for regression dilution ratios, while the hazard ratios in right plot were corrected for regression dilution ratios. Both plots based on the baseline SD (see Table 4.1). Dark color represents P<0.05 after correction of false discovery rate.

S9.1 QRISK3 variables and mapping in UK Biobank

- Age at study entry (years)
- Ethnic origin (White or not state; Indian; Pakistani; Bangladeshi; Other Asian; Black Caribbean; Black African; Chinese; Other ethnic group)
- Deprivation (as measured by the Townsend score, where higher values indicate higher levels of material deprivation)
- Systolic blood pressure (SBP) (mmHg)
- Measure of systolic blood pressure variability (standard deviation of repeated measures): *UK Biobank does not include information on variability in SBP. Our study derived this variable by the standard deviation between two automated or manual SBP readings at baseline (Variable ID 4080 and 93).*
- Body mass index (kg/m²)
- Total cholesterol-to-high density lipoprotein cholesterol ratio
- Smoking status (non-smoker, former smoker, light smoker (1-9/day), moderate smoker (10-19/day), or heavy smoker (≥ 20 /day)):
- Family history of coronary heart disease in a first-degree relative aged less than 60 years: *UK Biobank includes illnesses in father (Variable ID 20107), illnesses in mother (Variable ID 20110), and illnesses of siblings (Variable ID 20111), but does not have information on age at diagnosis. Our study assumed age less than 60 years at diagnosis.*
- Diabetes (type 1, type 2, or no diabetes)
- Treated hypertension (diagnosis of hypertension and treatment with at least one antihypertensive drug)
- Rheumatoid arthritis (diagnosis of rheumatoid arthritis, Felty's syndrome, Caplan's syndrome, adult onset Still's disease, or inflammatory polyarthropathy not otherwise specified)
- Atrial fibrillation (including atrial fibrillation, atrial flutter, and paroxysmal atrial fibrillation)
- Chronic kidney disease (stage 3, 4 or 5) and major chronic renal disease (including nephrotic syndrome, chronic glomerulonephritis, chronic pyelonephritis, renal dialysis, and renal transplant)
- Diagnosis of migraine (including classic migraine, atypical migraine, abdominal migraine, cluster headaches, basilar migraine, hemiplegic migraine, and migraine with or without aura)
- Corticosteroid use (including oral or parenteral prednisolone, betamethasone, cortisone, depo-medrone, dexamethasone, deflazacort, efcortisol, hydrocortisone, methylprednisolone, or triamcinolone)
- Systemic lupus erythematosus (including diagnosis of SLE, disseminated lupus erythematosus, or Libman-Sacks disease)
- Second generation "atypical" antipsychotic use (including amisulpride, aripiprazole, clozapine, lurasidone, olanzapine, paliperidone, quetiapine, risperidone, sertindole, or zotepine)
- Diagnosis of severe mental illness (including psychosis, schizophrenia, or bipolar affective disease)
- Diagnosis of erectile dysfunction or treatment for erectile dysfunction (including alprostadil, phosphodiesterase type 5 inhibitors, papaverine, or phentolamine)

S9.2 Disease and medication codes of QRISK3 variables in UK biobank

	ICD-10 & ICD-9 code	Verbal interview or questionnaire code	Medication code or other measurement
Diabetes			
Type 1	E10; 25001, 25011	1222; Variable ID 2443=1 & age≤20	If recode both type1 & type2, then categorize as type1
Type 2	E11;E13;E14; 25000, 25009; 25010, 25019; 2502, 2503, 2504, 2505, 2509	1220;1223; Variable ID 2443=1 & age>20	If recode not specific, then categorize as type2; HbA1c≥48 & ≤184 mmol/mol 1140868902, 1140874646, 1140874674, 1140874718, 1140874744, 1140883066, 1140884600, 1141152590, 1141157284, 1141168660, 1141171646, 1141173882, 1141189090
Chronic kidney disease (stage 3, 4, 5)	N183; N184; N185; N180; 585; 586	1193	eGFR <60 ml/min
Atrial fibrillation	I48; 4273	1471;1483	
Hypertension treatment		Variable ID 6177, 6153 =2	1140860192, 1140860292, 1140860696, 1140860728, 1140860750, 1140860806, 1140860882, 1140860904, 1140861088, 1140861190, 1140861276, 1140866072, 1140866078, 1140866090, 1140866102, 1140866108, 1140866122, 1140866138, 1140866156, 1140866162, 1140866724, 1140866738, 1140868618, 1140872568, 1140874706, 1140874744, 1140875808, 1140879758, 1140879760, 1140879762, 1140879802, 1140879806, 1140879810, 1140879818, 1140879822, 1140879826, 1140879830, 1140879834, 1140879842, 1140879866, 1140884298, 1140888552, 1140888556, 1140888560, 1140888646, 1140909706, 1140910442, 1140910614, 1140916356, 1140923272, 1140923336, 1140923404, 1140923712, 1140926778, 1140928226, 1141145660, 1141146126, 1141152998, 1141153026, 1141164276, 1141165470, 1141166006, 1141169516, 1141171336, 1141180592, 1141180772, 1141180778, 1141184722, 1141193282, 1141194794, 1141194810
Migraines	G43; 346	1265	
Rheumatoid arthritis	M05; M06; 714	1464	
Systemic lupus erythematosus	M32; 7100	1381	
Severe mental illness*	F20; F31; F331; F332; F333; 295; 296	1289;1291; Variable ID 20126=1,2,3,4	
Atypical antipsychotic medication			1140867420, 1140867444, 1140927956, 1140928916, 1141152848, 1141153490, 1141169714, 1141195974
Regular steroid tablets			1140874790, 1140874816, 1140874896, 1140874930, 1140874976, 1141145782, 1141173346
Erectile dysfunction	N484; 6078	1518	1141168936, 1141168948, 1141168944, 1141168946, 1140869100, 1140883010

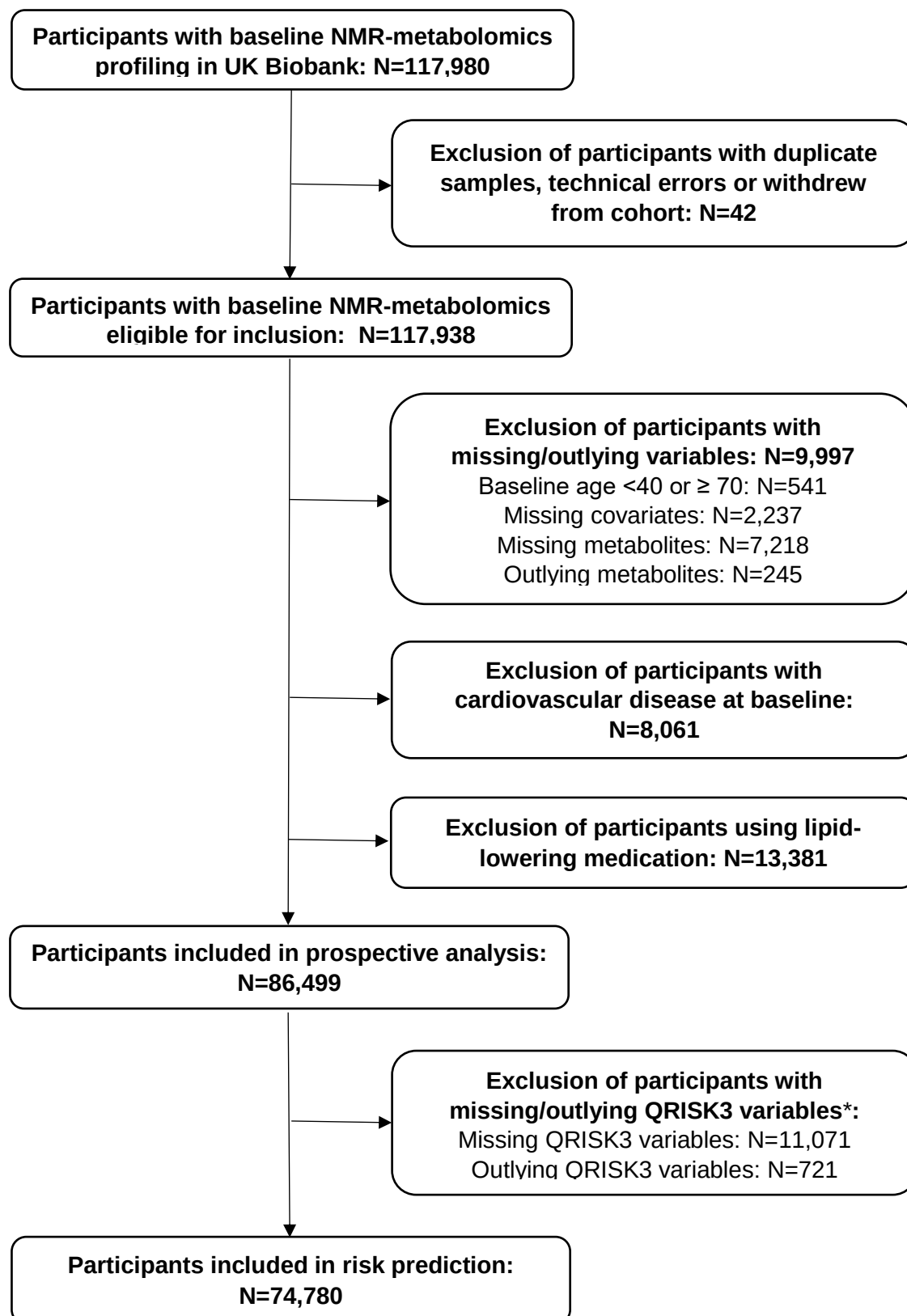
* Includes schizophrenia, bipolar disorder and moderate/severe depression.

S9.3 ICD-10 and operation code of cardiovascular disease

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

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

S9.4 Flowchart of exclusion criteria for study population in risk prediction



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

S9.5 Key parameters of applied algorithms

Elastic-net

Elastic-net was applied by the package of *sksurv.linear_model.CoxnetSurvivalAnalysis*. The key parameters include:

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

XGBoost

XGBoost was applied by the Python package of *xgboost*. The key parameters include:

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

BorutaSHAP

BorutaSHAP was applied by the Python package of *BorutaShap*, The key parameters include:

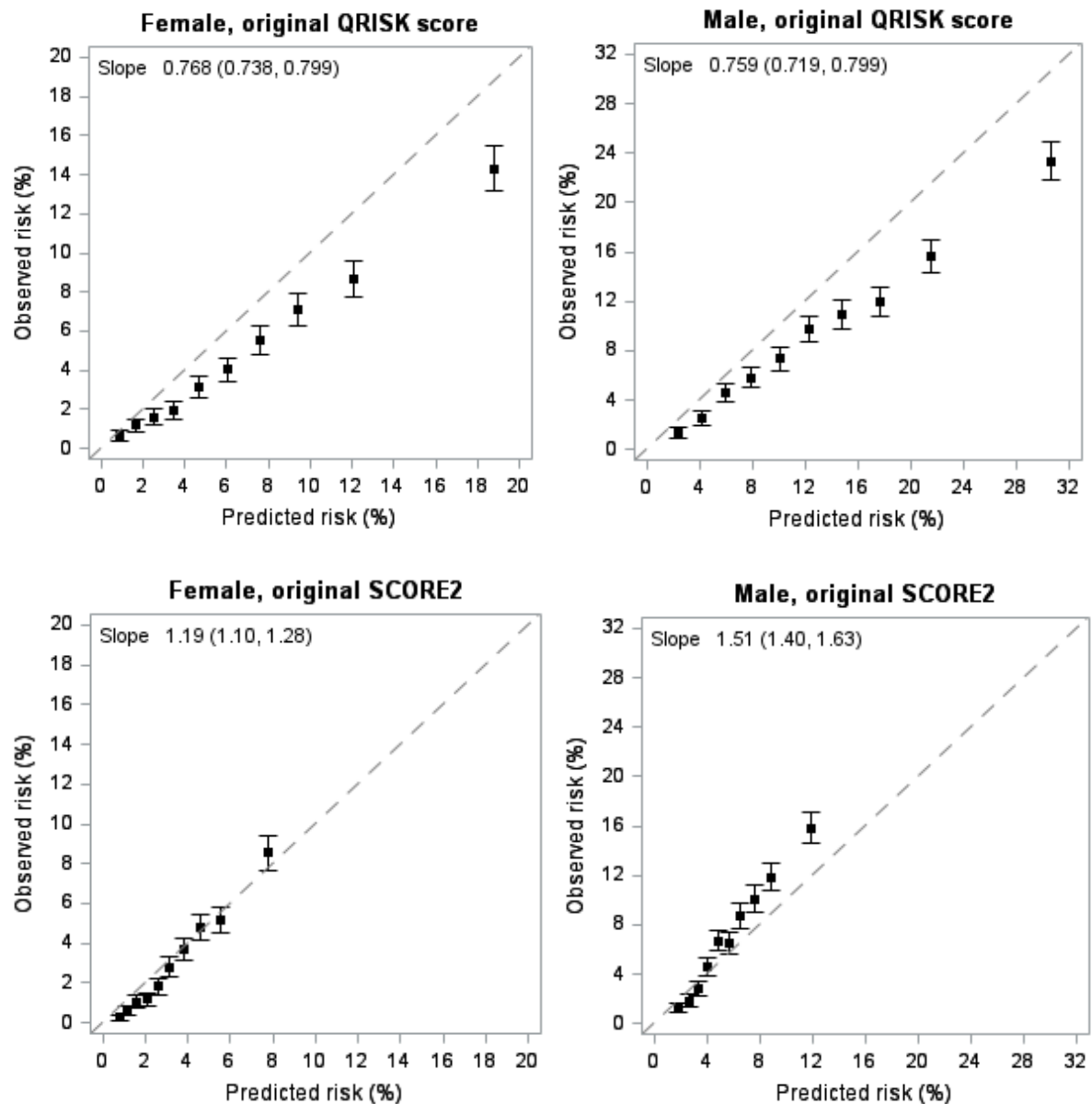
- ✓ `importance_measure` ("shap", "gain" or "permutation", default: "shap"): BorutaShap object
- ✓ `n_trials` (range: (0,∞], default: 100): Number of iterations for Boruta algorithm

S9.6 Fine-tuning of hyper-parameters of applied algorithms

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

*The meaning of each hyperparameter is explained in S9.5; † Tuning the hyperparameter from the lowest value to the highest value in the tuning range, with increase of the tuning step each time, and selected the hyperparameter with the best performance.

S9.7 Calibration plot of risk prediction by original QRISK3 and SCORE2



S9.8 List of selected novel metabolites by different methods (QRISK3)

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

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

S9.9 Associations of clinical metabolites independent from SCORE2

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

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

S9.10 List of selected metabolites using different methods (SCORE2)

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

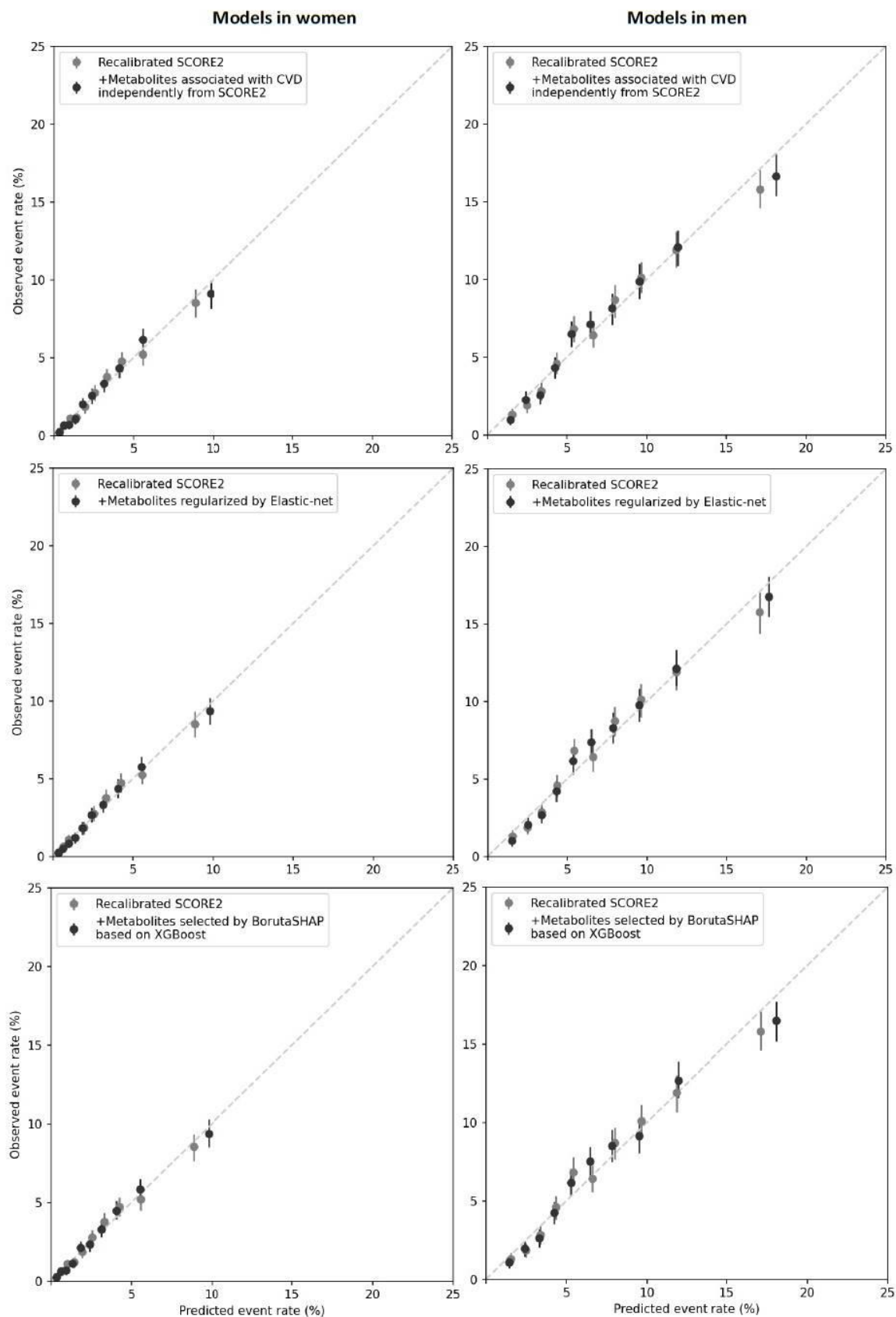
*Association was calculated using Cox proportional-hazards regression with adjustment of established risk factors, including age, education, region, townsend deprivation index, smoking, alcohol intake, body mass index, systolic blood pressure, and baseline diabetes; Significant association was defined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

S9.11 Comparing prediction performance of 10-year CVD risk w/o metabolites (SCORE2)

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

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

S9.12 Calibration of risk prediction models for 10-year CVD risk (SCORE2)



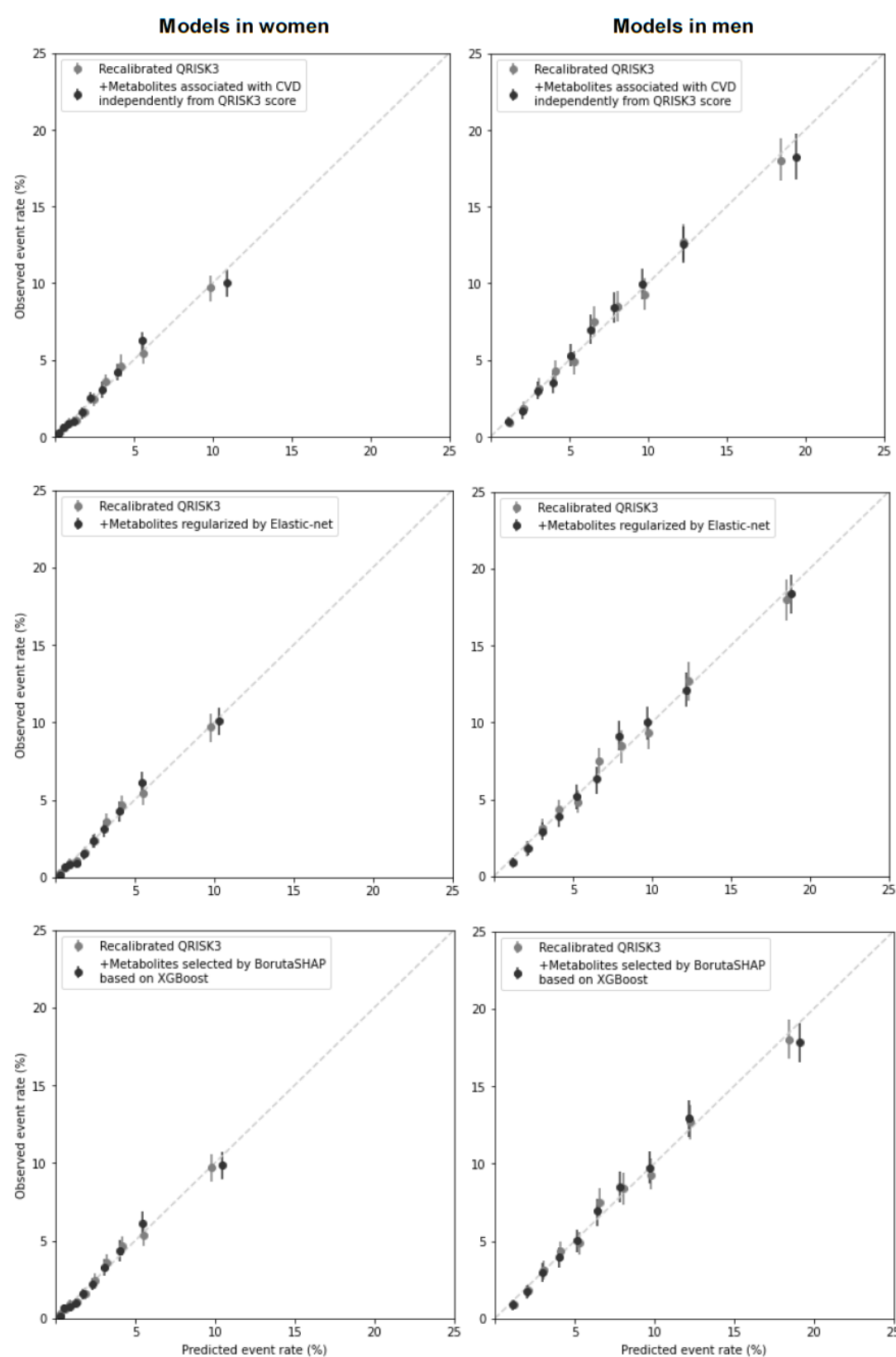
Calibration of risk prediction models for 10-year CVD risk. For each model, the observed and predicted CVD event rates are shown for each of 10 equally sized groups of absolute predicted risk. Vertical lines represent 95% CIs (bootstrap percentile confidence interval, bootstrap for 500 times).

S9.13 Prediction performance of 10-year ASCVD risk w/o metabolites (QRISK3 and wider scope of candidate metabolites)

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

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

S9.14 Calibration of risk prediction models for 10-year ASCVD risk (QRISK3 and wider scope of candidate metabolites)



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