Application of biomarker-based assessment of dietary patterns to nutritional epidemiology:

observational and interventional investigations with a focus on the Mediterranean diet and type 2 diabetes

Jakub Grzegorz Sobiecki

Jesus College

MRC Epidemiology Unit

University of Cambridge

December 2022

This thesis is submitted for the degree of Doctor of Philosophy

Abstract

Existing evidence on healthy dietary patterns suggests that they are modestly inversely associated with incidence of common non-communicable diseases, and in particular cardiometabolic diseases. Trials are rarely feasible to assess causality of these relationships which motivates leveraging observational data with improved methodological approaches. In this PhD thesis, I attempted to address the common limitation in nutritional epidemiology of subjective assessment of adherence to dietary patterns.

Chapter 2: I conducted a systematic review of the effects of Mediterranean diet interventions on nutritional biomarkers. I identified 29 trials reporting on 25 biomarkers eligible for meta-analysis (5-18 studies available per biomarker). Circulating carotenoids, vitamin C and fatty acids emerged as candidate biomarkers of compliance. Effect sizes were mostly small which likely reflected the multifaceted nature of whole diet interventions and insufficient validity of single analytes as biomarkers of adherence to the Mediterranean dietary pattern.

Chapter 3: I used the EPIC-InterAct case-cohort study which measured nutritional biomarkers at scale (~13,000 subcohort participants and ~9,000 incident type 2 diabetes cases) to evaluate the utility of combining them into biomarker scores predictive of adherence to dietary patterns, and to test their associations with incidence of type 2 diabetes. The available biomarkers included circulating carotenoids, vitamins C and 25(OH)D, fatty acid profiles, iron status biomarkers and cations. The dietary patterns of interest were the Mediterranean diet, alternative Healthy Eating Index-2010 and the Dietary Approaches to Stop Hypertension. The analyses showed modest correlations of the biomarker scores with their respective dietary patterns (r ~0.3) and statistically significant inverse associations with disease risk (hazard ratios ~0.8 per standard deviations of the scores).

Chapter 4: I established a collaboration with one of the randomised trials identified in Chapter 2, the MedLey trial, to address the limitations of internal derivation and validation of the biomarker scores. It compared the effects of a partial-feeding Mediterranean diet intervention with continuation of habitual diet in Australia on circulating carotenoids and fatty acids, 29 of which overlapped with those available in EPIC-InterAct. Using end-of-trial biomarker concentrations as predictors of the randomised assignment (n = 128), I developed a biomarker score which discriminated well between the trial arms (C-statistic = 0.88). It was robustly inversely associated with incidence of type 2 diabetes in EPIC-InterAct (hazard ratio 0.71 per standard deviation; 95% confidence and prediction intervals: 0.65-0.77 and 0.55-0.91).

Chapter 5: I additionally used the combined InterAct-MedLey data to test generalisability of biomarker scores of the Mediterranean diet. I derived a series of biomarker scores predictive of self-reported adherence to the Mediterranean diet in EPIC-InterAct countries, the MedLey trial baseline sample, and non-InterAct participants of the EPIC-Norfolk cohort (~5,000 with relevant biomarkers). Controlling for multiple testing, values of 8/13 biomarker scores were higher in the Mediterranean diet intervention than the control group of the MedLey trial, and 10/13 scores were inversely associated with incidence of type 2 diabetes in EPIC-InterAct.

Chapter 6: In the EPIC-Norfolk study, I investigated the impact of expanding a base set of predictors from circulating carotenoids and fatty acids with additional groups of nutritional biomarkers (urinary sodium, potassium and sugars, urinary and serum phytoestrogens, circulating vitamin C, iron status biomarkers, cations and stable isotopes) or using metabolomics on (i.) the correlations between self-reported Mediterranean diet and its biomarker scores and (ii.) the associations between the biomarker scores and incident cardiovascular disease, cancer, type 2 diabetes and mortality (n range ~500-11,000). The base set biomarker score had a moderate cross-validated correlation with self-report ($r = 0.40$) and the performance was similar or decreased with inclusion of additional nutritional biomarkers (r range: 0.30-0.41) and modestly improved with metabolomics ($r = 0.46$). Biomarker scores were inversely associated with disease and mortality outcomes, and use of different sets of biomarkers modified the strength of association for few diet-disease associations. Inverse associations were notably stronger for type 2 diabetes (hazard ratio ~0.80 per standard deviation of biomarker scores) than for other outcomes (range ~0.90-0.95).

Chapter 7: I conducted an outcome-wide analysis of 27 incident noncommunicable diseases in the EPIC-Norfolk study using as exposures the biomarker scores derived throughout the thesis (n range ~7,000-11,000) and dietary self-report of the Mediterranean diet (n ~22,000). Controlling for multiple testing, inverse associations were robustly detected across ≥2 of 4 methods of exposure assessment for type 2 diabetes, chronic obstructive pulmonary disease, and heart failure. At the nominal $\alpha = 0.05$, corresponding relationships were identified for ischaemic heart disease, renal disease, oesophageal and stomach cancers, and cataracts.

This PhD identified combinations of nutritional biomarkers as plausible biomarkers of the Mediterranean diet for application in epidemiological investigations. These findings contribute towards development of methods of objective assessment of diet and strengthen the evidence on the inverse relationship between the Mediterranean diet and cardiometabolic disease.

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Contributions section and specified in the text.

It is not substantially the same as any written work that I have submitted or is being concurrently submitted for a degree, diploma or other qualification at the University of Cambridge or any other University or similar institution.

This dissertation does not exceed the word limit of 60,000 words excluding references, tables, figures and appendices, as required by the Degree Committee of the Faculty of Clinical Medicine and Veterinary Medicine.

Jakub Sobiecki

December 2022

Acknowledgments

First and foremost, I would like to express my deep gratitude to my supervisors, Professor Nita Forouhi and Doctor Fumiaki Imamura. Thank you for being so generous with your time and scientific expertise, and for your overall support throughout this journey. Suffice to say, I have always left our meetings feeling more motivated and able to face the PhD challenges than before talking to you, even at times of low spirit. Thank you, Nita, for instilling in me a blueprint for an orderly scientific process and teaching me about the importance of big picture thinking for evaluation of the relevance and impact of my findings. Thank you, Fumiaki, for always motivating me to be the most accurate scientist possible in my analysis, writing, and interpretation of my results. I will, however, in the foreseeable future more realistically aim for being the second most accurate scientist (after Fumiaki). This PhD work was made possible by funding from the Medical Research Council and the MRC Epidemiology Unit.

I would like to give special thanks to Karen Murphy for agreeing to set up a collaboration and sharing with us the data from her elegant, partial-feeding randomised controlled trial of a Mediterranean diet intervention (MedLey). This had a pivotal role in shaping this PhD and allowed for generating more credible and impactful results. I am also very thankful to Stephen Sharp for his statistical input and to Albert Koulman for sharing his expertise on measurement of biomarkers. I further thank Nicola Kerrison and Abigail Britten for preparation of the datasets for my analyses and their patience in releasing data updates. I am grateful to Meriel Smith for all her support and help in organisational and administrative issues. I would like to thank my fellow PhD students and postdocs for the scientific discussions and general PhD and early-career scientist life chats. I especially thank Eirini Trichia for being so welcoming to me in the Nutritional Epidemiology group and the invaluable knowledge transfer from her as an outgoing PhD student on making the most of my PhD. Chunxiao Li, Divya Bhagtani and Daniel Ibsen, it has been great having you around as fellow nutritional epidemiology students.

I am indebted to my loving partner, Ula Rzeszutek, for enduring my long writing-up hours, her understanding, and support. I cannot thank enough my parents, Halina Sobiecka and Grzegorz Sobiecki for the stimulating upbringing and their encouragement to study and pursue research.

Multiple individuals had contributed to my development as a researcher in my pre-doctoral years which led to embarking on this PhD. I am forever grateful to Professor Tim Key and his group at the University of Oxford for hosting me as a visiting student and offering the invaluable research experience. I would not be the scientist that I am today without it. I thank

Professor Jonathan Wells from University College London for his unwavering support of my research endeavours. Dr Małgorzata Desmond and Professor Janusz Książyk are acknowledged as my first research mentors. I give credit to Małgorzata for sparking my interest in nutritional epidemiology. I thank Dr Teresa Norat for being an exceptional line manager and scientific supervisor in my first full-time academic position at Imperial College London. Finally, I would like to thank my professors from my first degree in Dietetics at the the Warsaw University of Life Sciences who motivated me to pursue research: Wojciech Roszkowski, Joanna Gromadzka-Ostrowska, Anna Brzozowska and Joanna Kałuża.

Completing this PhD may have not been possible at this time without the courage and tenacity of the Ukrainian nation in the face of the full-scale military invasion of their country. Many sleepless nights went by in early 2022 thinking about the victims of this atrocity and how the war might evolve. Had it not been for the successful defence of Kyiv which decisively stopped any prospects of the russian army advancing westwards towards my home country, Poland, I might have not been in the right space – both mentally and physically – to finish this thesis. I would like to express my heartfelt gratitude to the Ukrainian state under the leadership of president Volodymyr Zelenskyy, the Armed Forces of Ukraine under command of general Valerii Zaluzhnyi, the unknown soldier, and the unknown civilian for their fierce resistance against the terrorist attack launched on their country. The impact of their exceptional bravery extends far beyond the borders of Ukraine and has already protected the freedom and liberties of many individuals across the globe, mine included. For this I will be forever grateful. Слава Україні! *(Glory to Ukraine!)*

Contributions

Professor Nita Forouhi introduced me to the idea of combining nutritional biomarkers into composite measures reflecting dietary quality which enabled me to develop the research agenda pursued in my PhD thesis. This dissertation was completed under supervision of Professor Nita Forouhi and Dr Fumiaki Imamura. Stephen Sharp provided additional input on statistical analysis with regards to advice on using specific approaches and interpretation of results. Dr Albert Koulman shared his expertise on measurement of biomarkers and contributed to interpretation of results and design of the systematic review in Chapter 2.

I was not involved in the design or conduct of collection of data analysed within this dissertation, or any other activities related to day-to-day running of the associated studies, except for data extraction and dataset maintenance for the systematic review in Chapter 2. Datasets for analyses in the EPIC-InterAct (Chapters 3-5) and EPIC-Norfolk (Chapters 5-7) studies were provided by, respectively, Dr Nicola Kerrison and Dr Abigail Britten. Angela Mulligan prepared variables on disaggregated food and nutrient intakes estimated from food diary in EPIC-Norfolk (Chapter 5). Dr Karen Murphy from the University of South Australia provided the dataset from the MedLey trial (Chapters 4-6). Nutritional biomarkers were measured in multiple laboratories, as reported in the Methods sections of individual chapters. The most important contributions which enabled my PhD research included the development of high-throughput methods and measurement at scale of plasma phospholipid fatty acids by Dr Albert Koulman's laboratory and plasma carotenoids by VITAS AS (Dr Thomas Gundersen) in EPIC-InterAct (Chapters 3-5). Metabolomic assays were performed by Metabolon (Chapter 6), and this was co-ordinated by Professor Claudia Langenberg on behalf of the MRC Epidemiology Unit.

I have written the code for all analyses, with supervisory input from Dr Fumiaki Imamura, except for the code for figures depicting non-linear associations which were based on code adapted from a script provided by Dr Fumiaki Imamura. Part of the descriptive statistics on sociodemographic factors and medical history for the MedLey trial in Table 4.1 were provided by Dr Courtney Davis, one of the study authors, due to unavailability of some of the underlying variables in the data transferred from the University of South Australia to the MRC Epidemiology Unit.

Publications

From this thesis:

- Under the second round of peer-review in *PLoS Medicine*: **Sobiecki JG**, Imamura F, Davis CR, Sharp SJ, Koulman A, Hodgson J, Guevara M, Schulze MB, Zheng JS, Agnoli C, Bonet C, Colorado-Yohar SM, Fagherazzi G, Franks PW, Gundersen TE, Jannasch F, Kaaks R, Katzke V, Molina-Montes E, Nilsson PM, Palli D, Panico S, Papier K, Rolandsson O, Sacerdote C, Tjønneland A, Tong TYN, van der Schouw YT, Danesh J, Butterworth AS, Riboli E, Murphy KJ, Forouhi NG, Wareham NG. A nutritional biomarker score of the Mediterranean diet and incident type 2 diabetes: integrated analysis of experimental and observational data from the MedLey randomised controlled trial and the EPIC-InterAct case-cohort study.

Other work published as PhD student:

- Hudda MT, Wells JCK, Adair LS, Alvero-Cruz JRA, Ashby-Thompson M, Ballesteros-Vásquez MN, Barrera-Exposito J, Caballero B, Carnero EA, Cleghorn GJ; Davies PSW, Desmond M, Devakumar D, Gallagher D, Haschke F, Horlick M, Jemaa HB, Khan I, Mankai A, Monyeki A, Nashandi H, Ortiz-Hernández L, Plasqui G, Popkin BM, Reichert FF, Robles-Sardin AE, Rush E, Shypailo RJ, **Sobiecki JG**, ten Hoor G, Valdés J, Wickramasinghe VP, Wong WW, Riley RD, Owen CG, Whincup PH, Nightingale CM. External validation of a childhood fat mass prediction model: individual participant data meta-analysis of predictive performance in 19 countries. *BMJ* 2022;378:e071185
- Desmond MA, **Sobiecki JG**, Jaworski M, Płudowski P, Antoniewicz J, Shirley MK, Eaton S, Książyk J, Cortina-Borja M, De Stavola B, Fewtrell M, Wells JCK. Growth, body composition, and cardiovascular and nutritional risk of 5- to 10-y-old children consuming vegetarian, vegan, or omnivore diets. *American Journal of Clinical Nutrition* 2021;113:1565-1577.
- Abar L, **Sobiecki JG**, Cariolou M, Nanu N, Vieira AR, Stevens C, Aune D, Greenwood DC, Chan DSM, Norat T. Body size and obesity during adulthood, and risk of lympho-haematopoietic cancers: an update of the WCRF-AICR systematic review of published prospective studies. *Annals of Oncology* 2019;30:528-541.

Conference report

“Sobiecki JG on behalf of MedLey Study and EPIC-InterAct Study authors. The association between a biomarker score indicating adherence with the Mediterranean diet and incident type 2 diabetes: integrated analysis using experimental and observational data from the MedLey Study Trial and EPIC-InterAct Study.”

- Nutrition Futures annual student conference of the Nutrition Society
- London, 7 September 2021
- Award for the best in-person oral presentation

Abbreviations

AHA	American Heart Association
aHEI	Alternative Healthy Eating Index
aMED	Alternate Mediterranean diet
BIC	Bayesian information criterion
BLUP	Best linear unbiased prediction
BMI	Body mass index
CI	Confidence interval
COPD	Chronic obstructive pulmonary disease
CVD	Cardiovascular disease
DALY	Disability-adjusted life years
DASH	Dietary Approaches to Stop Hypertension
DNL	De novo lipogenesis
DQI	Diet Quality Index
EPIC	European Prospective Investigation into Cancer and Nutrition
FFQ	Food frequency questionnaire
HbA1c	Haemoglobin A1C
HR	Hazard ratio
HEI-2010	Healthy Eating Index-2010
IARC	International Agency for Research on Cancer
ICC	Intraclass correlation coefficient
IQR	Interquartile range
LFK index	Luis Furuya-Kanamori index
Lasso	Least absolute shrinkage and selection operator
LC-MS	Liquid chromatography-mass spectrometry
MICEpmm	Multiple imputation by chained equations with predictive mean matching
MDS	Mediterranean diet score
MDS-pyramid	Mediterranean diet score for adherence to pyramid-based guidelines
MUFA	Monounsaturated fatty acids
NCD	Non-communicable disease
NCEP	National Cholesterol Education Program
NHLBI	National Heart, Lung, and Blood Institute

ODMA	*o*-desmethylangolensin
OR	Odds ratio
PAF	Population attributable fraction
PI	Prediction interval
PUFA	Polyunsaturated fatty acids
QRILC	Quantile Regression Imputation of Left-Censored data
RCT	Randomised controlled trial
SD	Standard deviation
SE	Standard error
SFA	Saturated fatty acids
SMD	Standardised mean difference
TFA	Trans fatty acids
TMAO	Trimethylamine N-oxide
T2D	Type 2 diabetes
WHI	Women's Health Initiative
WHI-NPASS	WHI ancillary study: Nutrition and Physical Activity Assessment Study
25(OH)D	25-hydroxyvitamin D

Table of contents

Chapter 1. Introduction 1

1.1 Background and context 1
1.2 Dietary patterns 3
1.3 Nutritional biomarkers 5
1.4 Thesis aims 6

Chapter 2. Effects of the Mediterranean diet on nutritional biomarkers: systematic review and meta-analysis of experimental evidence 9

2.1 Background 11
2.2 Aim 11
2.3 Methods 11
2.3.1 Eligibility criteria 12
2.3.2 Information sources 13
2.3.3 Article screening, data extraction, and risk of bias assessment 15
2.3.4 Statistical analysis 16
2.3.5 Analytical issues in within-person trials 18
2.3.6 Sensitivity analyses 19
2.4 Results 19
2.4.1 Study characteristics 21
2.4.2 Corrections to data 35
2.4.3 Risk of bias assessment 35
2.4.4 Carotenoids, vitamins A and C, folate, tocopherols and phenolic compounds 38
2.4.5 Biomarkers of other vitamins, minerals, phytosterols and phenolic compounds 40
2.4.6 Saturated, mono- and trans-unsaturated fatty acids 42
2.4.7 Polyunsaturated fatty acids 43
2.4.8 Miscellaneous fatty acids reported on by ≤ 2 studies 47
2.4.9 Amino acids and amino-acid related analytes 49
2.4.10 Publication bias of biomarkers eligible for meta-analysis 50
2.4.11 Meta-analysis 53
2.4.12 Meta-regression and subgroup analysis 59
2.4.13 Sensitivity analyses 63
2.5 Discussion 69

2.5.1 Strengths and limitations 69
2.5.2 Comparison with previous research 72
2.5.3 What this study adds and implications of this research 75
2.5.4 Conclusions 76

Chapter 3. Nutritional biomarker scores of dietary patterns: derivation and associations with incident type 2 diabetes 77

3.1 Background 79
3.2 Aim 80
3.3 Methods: the EPIC-InterAct study 80
3.3.1 Nutritional and metabolic biomarkers 81
3.3.2 Dietary patterns 82
3.3.3 Covariates 85
3.3.4 Analytical samples and exclusions 85
3.4 Methods: statistical analysis 86
3.4.1 Pre-treatment of nutritional biomarkers and dietary pattern variables 86
3.4.2 Derivation of biomarker scores 87
3.4.3 Associations of biomarker scores with incident type 2 diabetes 89
3.4.4 Sensitivity analyses 91
3.5 Results 93
3.5.1 Background characteristics by country 93
3.5.2 Background characteristics by adherence to dietary patterns 96
3.5.3 Nutritional biomarker scores 100
3.5.4 Associations of biomarker scores with self-reported diet 103
3.5.5 Associations of biomarker scores and self-reported dietary patterns with incident T2D 106
3.5.6 Associations of biomarker scores with incident T2D: sensitivity and secondary analyses 111
3.6 Discussion 116
3.6.1 Strengths 116
3.6.2 Limitations: derivation of the biomarker scores 117
3.6.3 Comparison with previous research: biomarkers of dietary patterns 119
3.6.4 Comparison with previous research: diet-disease associations using biomarker scores 120
3.6.5 What this study adds and implications of this research 122
3.6.6 Conclusions 123

Chapter 4. A nutritional biomarker score of the Mediterranean diet and incident type 2 diabetes: experimental and observational analysis from the MedLey randomised controlled trial and the EPIC-InterAct case-cohort study 125

4.1 Background 126
4.1.1 External derivation of biomarkers: study design considerations 126
4.1.2 Overview of previous research 128
4.2 Aim 129
4.3 Methods 129
4.3.1 The MedLey trial 129
4.3.2 The EPIC-InterAct study 130
4.3.3 Biomarkers 132
4.3.4 Measurement of covariates and self-reported diet 132
4.3.5 Methods for the EPIC-Norfolk study: impact of assays on the biomarker score 133
4.4 Statistical analysis 134
4.4.1 Derivation of the biomarker score 134
4.4.2 Effects of the Mediterranean diet intervention on biomarkers 136
4.4.3 Associations of the biomarker score with incident T2D 136
4.4.4 Sensitivity analyses 137
4.4.5 Associations of the biomarker score with self-reported diet 137
4.5 Results 138
4.5.1 Effects of the MedLey trial intervention on nutritional biomarkers 139
4.5.2 Derivation of the biomarker score 144
4.5.3 Association of the biomarker score with incident T2D 150
4.5.4 Impact of biomarker assays on reliability of the biomarker score and stability of its associations with incident T2D in EPIC-Norfolk 157
4.5.5 Associations of the biomarker score with self-reported diet 160
4.5.6 Secondary biomarker scores 162
4.6 Discussion 163
4.6.1 Effects of the Mediterranean diet on nutritional biomarkers 163
4.6.2 Strengths and limitations 166
4.6.3 Comparison with previous studies 168
4.6.4 What this study adds and implications of this research 170
4.6.5 Conclusions 172

Chapter 5. Generalisability of nutritional biomarker scores of the Mediterranean diet 173

5.1 Background **175**
5.2 Aim **176**
5.3 Methods **176**
5.3.1 The MedLey trial ... 177
5.3.2 The EPIC-InterAct study ... 177
5.3.3 The EPIC-Norfolk study ... 178
5.3.4 Literature-based biomarker score of the Mediterranean diet ... 180
5.4 Statistical analysis **180**
5.4.1 Additional information on derivation of the biomarker scores ... 181
5.5 Results **182**
5.5.1 Derivation and validation of biomarker scores ... 182
5.5.2 Associations of biomarker scores with incident T2D ... 194
5.6 Discussion **200**
5.6.1 Strengths and limitations ... 200
5.6.2 What this study adds and implications of this research ... 201
5.6.3 Conclusions ... 202

Chapter 6. Objective epidemiological assessment of the Mediterranean diet: nutritional biomarkers, metabolomics, or both? 203

6.1 Background **205**
6.2 Aim **206**
6.3 Methods **206**
6.3.1 The EPIC-Norfolk study ... 207
6.3.2 Nutritional biomarkers: counts and sampling schemes of participants ... 208
6.3.3 Nutritional biomarkers: laboratory assays ... 209
6.3.4 Metabolomics ... 211
6.3.5 Outcome ascertainment ... 212
6.3.6 Covariates ... 212
6.4 Statistical analysis **213**
6.4.1 Pre-treatment of biomarker variables ... 213
6.4.2 Estimation of 24-h urinary excretion of sodium and potassium ... 214
6.4.3 Imputation of metabolomic data ... 214
6.4.4 Derivation of biomarker and metabolomic scores ... 216
6.4.5 Associations of biomarker and metabolomic scores with incident outcomes ... 217
6.4.6 Sensitivity analyses ... 219
6.5 Results **220**

6.5.1 Background characteristics 220
6.5.2 Estimation of absolute excretion of sodium and potassium 220
6.5.3 Nutritional and metabolomic biomarker scores 227
6.5.4 Diet-disease associations: main results 232
6.5.5 Diet-disease associations: model assumptions 236
6.5.6 Sensitivity analyses 239
6.6 Discussion 243
6.6.1 Strengths 244
6.6.2 Limitations 244
6.6.3 Comparison with previous research: biomarkers of the Mediterranean diet 246
6.6.4 Diet-disease associations: implications and comparison with previous research 249
6.6.5 Conclusions 252

Chapter 7. Outcome-wide associations of the Mediterranean diet with chronic diseases 253

7.1 Background 255
7.2 Aim 255
7.3 Methods 255
7.4 Statistical analysis 258
7.5 Results 259
7.5.1 Primary results 259
7.5.2 Sensitivity analyses 263
7.5.3 Secondary nutritional biomarker scores 263
7.6 Discussion 264
7.6.1 Strengths and limitations 264
7.6.2 Implications and comparison with previous research 265
7.6.3 Conclusions 269

Chapter 8. Discussion 271

8.1 Summary of findings 271
8.1.1 Effects of the Mediterranean diet on nutritional biomarkers 271
8.1.2 Derivation of biomarker scores of dietary patterns 271
8.1.3 Associations of biomarker scores of dietary patterns with incident disease outcomes 273
8.2 Strengths and limitations 274
8.3 Interpretation of findings 275
8.4 Implications and directions for future research 277

8.5 Conclusions **279**

References **281**

Appendix **319**

List of figures

Figure 1.1 Thesis outline 7
Figure 2.1 PRISMA 2020 flow diagram depicting the selection process of publications 20
Figure 2.2 Doi plots and LFK indices of publication bias of the effects of the Mediterranean diet on circulating carotenoids and retinol 50
Figure 2.3 Doi plots and LFK indices of publication bias of the effects of the Mediterranean diet on circulating vitamin C, α-tocopherol, saturated, monounsaturated, and total polyunsaturated fatty acids 51
Figure 2.4 Doi plots and LFK indices of publication bias of the effects of the Mediterranean diet on circulating n-3 and n-6 fatty acids, and n-6:n-3 ratio 52
Figure 2.5 Effects of the Mediterranean diet on circulating carotenoids and retinol: standardised mean differences 54
Figure 2.6 Effects of the Mediterranean diet on circulating vitamin C, α-tocopherol, saturated, mono- and polyunsaturated fatty acids: standardised mean differences 55
Figure 2.7 Effects of the Mediterranean diet on circulating saturated and monounsaturated fatty acids: standardised mean differences 56
Figure 2.8 Effects of the Mediterranean diet on circulating total and individual n-3 fatty acids: standardised mean differences 57
Figure 2.9 Effects of the Mediterranean diet on circulating total and individual n-6 fatty acids: standardised mean differences 58
Figure 3.1 Associations between nutritional biomarker scores of dietary patterns and incidence of type 2 diabetes in the EPIC-InterAct case-cohort study 108
Figure 3.2 Non-linear associations between nutritional biomarker scores of dietary patterns and incidence of type 2 diabetes in EPIC-InterAct 109
Figure 4.1 Designs of the MedLey trial and the EPIC-InterAct case-cohort study and numbers of participants included in the analysis 131
Figure 4.2 Differences in standardised means of nutritional biomarkers between the Mediterranean and habitual diet groups in the MedLey trial at 6 months 143
Figure 4.3 Standardised differences in the biomarker score of the Mediterranean diet between the arms of the MedLey trial 147
Figure 4.4 Nutritional biomarker score of the Mediterranean diet: study-specific distribution by the MedLey trial arms post-intervention and the EPIC-InterAct subcohort by country at baseline. 148
Figure 4.5 Association between the nutritional biomarker score of the Mediterranean diet and incidence of type 2 diabetes in the EPIC-InterAct case-cohort study (n = 22,202). 152
Figure 5.1a Associations between internally derived nutritional biomarker scores of the Mediterranean diet and incidence of type 2 diabetes in the EPIC-InterAct case-cohort study 197

Figure 5.1b Associations between internally derived nutritional biomarker scores of the Mediterranean diet and incidence of type 2 diabetes in the EPIC-InterAct case-cohort study 198
Figure 5.2 Associations between externally derived nutritional biomarker scores of the Mediterranean diet and incidence of type 2 diabetes in the EPIC-InterAct case-cohort study 199
Figure 6.1 Non-linear association between a nutritional biomarker score of the Mediterranean diet and incidence of cardiovascular disease in the EPIC-Norfolk study 238

List of tables

Table 2.1 Characteristics of experimental studies of Mediterranean diet included in the systematic review 23
Table 2.2 Risk of bias assessment of controlled trials – National Heart, Lung, and Blood Institute Quality Assessment Tool 36
Table 2.3 Risk of bias assessment of pre-post studies without a control group – National Heart, Lung, and Blood Institute Quality Assessment Tool 37
Table 2.4 Standardised mean differences in biomarkers of carotenoids, vitamins A and C, folate, homocysteine and tocopherols after Mediterranean diet interventions relative to control diets 39
Table 2.5 Standardised mean differences in circulating and urinary biomarkers of vitamins, minerals, phytosterols and phenolic compunds after Mediterranean diet interventions relative to control diets reported by ≤ 2 studies 41
Table 2.6 Standardised mean differences in biomarkers of saturated, mono- and trans-unsaturated fatty acids after Mediterranean diet interventions relative to control diets 45
Table 2.7 Standardised mean differences in biomarkers of polyunsaturated fatty acids after Mediterranean diet interventions relative to control diets 46
Table 2.8 Standardised mean differences in biomarkers of fatty acids after Mediterranean diet interventions relative to control diets reported by ≤ 2 studies 48
Table 2.9 Standardised mean differences in biomarkers of amino acids and amino acid-related analytes after Mediterranean diet interventions relative to control diets 49
Table 2.10 Meta-regression analysis of the effects of Mediterranean versus control diet interventions on 25 nutritional biomarkers: results with p values for meta-regression coefficients < 0.10 61
Table 2.11 Subgroup analysis of the effects of Mediterranean versus control diet interventions on nutritional biomarkers: results with p values for heterogeneity < 0.10 and ≥5 studies per subgroup 62
Table 2.12 Sensitivity analyses of meta-analyses of the effects of Mediterranean versus control diet interventions on carotenoids and vitamins 65
Table 2.13 Sensitivity analyses of meta-analyses of the effects of Mediterranean versus control diet interventions on saturated and monounsaturated fatty acids 66
Table 2.14 Sensitivity analyses of meta-analyses of the effects of Mediterranean versus control diet interventions on total polyunsaturated and n-3 fatty acids 67
Table 2.15 Sensitivity analyses of meta-analyses of the effects of Mediterranean versus control diet interventions on n-6 fatty acids 68
Table 3.1 Scoring of dietary pattern indices in the EPIC-InterAct study 84
Table 3.2 Background characteristics and median concentrations of biomarkers in biomarker derivation samples: the EPIC-InterAct subcohort 94

Table 3.3 Baseline characteristics of the EPIC-InterAct subcohort participants by quintiles of dietary patterns assessed by self-report and nutritional biomarker scores 97
Table 3.4 Nutritional biomarker scores of dietary patterns derived in the EPIC-InterAct subcohort: standardised coefficients 102
Table 3.5 Correlations of biomarker scores of dietary patterns with dietary patterns assessed by self-report in the EPIC-InterAct study (minimum, maximum country-specific values) 104
Table 3.6 Correlations of biomarker scores of dietary patterns with food and nutrients components of dietary patterns assessed by self-report (minimum, maximum country-specific values) in the EPIC-InterAct subcohort (n = 12,455) 105
Table 3.7 Associations between biomarker scores of dietary patterns and incidence of type 2 diabetes in the EPIC-InterAct case-cohort study 107
Table 3.8 Associations between nutritional biomarker scores of dietary patterns and incidence of type 2 diabetes in the EPIC-InterAct case-cohort study: pooled hazard ratios (95% CI) after additional adjustment for circulating metabolic factors 110
Table 3.9 Nutritional biomarker scores of dietary patterns and incidence of type 2 diabetes in the EPIC-InterAct case-cohort study: sensitivity analyses 112
Table 3.10 Nutritional biomarker score of dietary patterns and incidence of type 2 diabetes in EPIC-InterAct: associations per 1 standard deviation by categories of covariates 113
Table 3.11 Nutritional biomarker scores of dietary patterns and incidence of type 2 diabetes in the EPIC-InterAct case-cohort study: adjustment for component biomarkers 114
Table 4.1 Baseline characteristics of participants of the MedLey randomised partial-feeding controlled trial of the Mediterranean diet and the EPIC-InterAct case-cohort study 140
Table 4.2 Medians (25th, 75th percentile) of circulating carotenoids and fatty acids: the MedLey trial post-intervention and the EPIC-InterAct subcohort at baseline 141
Table 4.3 Biomarker scores of discrimination between the Mediterranean and habitual diet in the MedLey trial 145
Table 4.4 Medians (25th, 75th percentile) of the nutritional biomarker score of the Mediterranean diet in the MedLey trial post-intervention and in the EPIC-InterAct subcohort 149
Table 4.5 Associations between the nutritional biomarker score of the Mediterranean diet with incidence of type 2 diabetes in the European Prospective Investigation into Cancer and Nutrition-InterAct (n = 22,202) 151
Table 4.6 Nutritional biomarker score of the Mediterranean diet derived in the MedLey trial and incidence of type 2 diabetes in EPIC-InterAct: sensitivity analyses 154
Table 4.7 Nutritional biomarker score of the Mediterranean diet derived in the MedLey trial and incidence of type 2 diabetes in EPIC-InterAct: associations per 1 standard deviation by categories of covariates 156

Table 4.8 Medians (25th, 75th percentile) and intraclass correlation coefficients of biomarkers by laboratory assays in subsamples of the EPIC-Norfolk study 158
Table 4.9 Associations between the nutritional biomarker score of the Mediterranean diet with incidence of type 2 diabetes in EPIC-Norfolk: comparison of results from different sets of biomarkers per 1 standard deviation higher score values 159
Table 4.10 Pearson correlation coefficients (minimum, maximum country-specific values in EPIC-InterAct) between a biomarker score of the Mediterranean diet derived in the Medley trial and the Mediterranean Diet Score estimated from self-report..... 161
Table 5.1 Scoring of dietary pattern indices..... 179
Table 5.2 Standardised coefficients of biomarker scores of self-reported Mediterranean Diet Score by sample of derivation..... 184
Table 5.3 Pearson correlations between biomarker scores of Mediterranean diet and Mediterranean diet score estimated from self-report in EPIC-InterAct subcohort: pooled and country-specific estimates..... 186
Table 5.4 Medians (interquartile ranges) of nutritional biomarker scores of Mediterranean diet in the MedLey trial post-intervention and EPIC-InterAct subcohort..... 187
Table 5.5 Differences in standardised means of nutritional biomarker scores between the Mediterranean and habitual diet groups in the MedLey trial at 6 months 188
Table 5.6 Mean differences in standardised biomarker scores of the Mediterranean diet between the habitual diet group and tertiles of the Mediterranean Diet Score in the Mediterranean diet group in the MedLey trial after 6 months of intervention..... 189
Table 5.7 Longitudinal associations between nutritional biomarker scores of Mediterranean diet and self-reported Mediterranean Diet Score in the MedLey trial, stratified by trial arm 190
Table 5.8 Pearson correlation coefficients (minimum, maximum country-specific values in EPIC-InterAct) between biomarker scores of Mediterranean diet, Mediterranean Diet Score, and its energy-adjusted food components..... 191
Table 5.9 Longitudinal associations between Mediterranean Diet Score and biomarker scores of Mediterranean diet in a subset of EPIC-Norfolk (n=432)..... 193
Table 5.10 Associations of biomarker scores of Mediterranean diet scores with incidence of type 2 diabetes in EPIC-InterAct (n = 22,202): pooled Hazard Ratios (95% CI)..... 196
Table 6.1 Baseline characteristics of EPIC-Norfolk participants included in the analyses of the Mediterranean diet and chronic disease risk and mortality by quintiles of adherence assessed by a nutritional biomarker score and food frequency questionnaire (n = 7,017)..... 221
Table 6.2 Baseline characteristics of EPIC-Norfolk participants included in the analyses of the Mediterranean diet and chronic disease risk and mortality by quintiles of adherence assessed by a metabolomic biomarker score and food frequency questionnaire (n = 11,518)..... 223

Table 6.3 Equations for estimations of 24-hour urinary output of sodium and potassium from spot urine samples 225
Table 6.4 Performance of prediction equations of 24-hour urinary output of sodium and potassium from spot urine samples in the EPIC-Norfolk study: comparison against repeated 24-h collections (n = 335) 226
Table 6.5 Standardised coefficients of biomarker scores of the Mediterranean diet in the EPIC-Norfolk study 229
Table 6.6 Standardised coefficients of metabolomic scores of the Mediterranean diet in the EPIC-Norfolk study 230
Table 6.7 Associations between the Mediterranean diet pyramid score and disease incidence per standard deviation of adherence: biomarker-based assessment and self-report in the EPIC-Norfolk study 234
Table 6.8 Associations between the Mediterranean diet pyramid score and mortality per standard deviation of adherence: biomarker-based assessment and self-report in the EPIC-Norfolk study 235
Table 6.9 Associations between biomarker scores of the Mediterranean diet and incident outcomes with violated proportional hazards assumption in the EPIC-Norfolk study 237
Table 6.10 Associations of biomarker scores of the Mediterranean diet with incidence of cardiovascular disease, cancer, type 2 diabetes, and mortality in the EPIC-Norfolk study: sensitivity analyses 241
Table 7.1 Outcome definitions and criteria for exclusion of prevalent cases 257

List of appendices

Appendix 2.1 Search strategies in Embase and Web of Science 319
Appendix 3.1 Multi-country (subcohort) and country-specific nutritional biomarker scores of the Mediterranean Diet Score in the EPIC-InterAct subcohort: standardised coefficients* 322
Appendix 3.2 Multi-country (subcohort) and country-specific nutritional biomarker scores of the alternative Healthy Eating Index-2010 in the EPIC-InterAct subcohort: standardised coefficients* 323
Appendix 3.3 Multi-country (subcohort) and country-specific nutritional biomarker scores of the DASH diet in the EPIC-InterAct subcohort: standardised coefficients* 324
Appendix 4.1 Differences in standardised means of nutritional biomarkers between the Mediterranean and habitual diet groups in the MedLey trial at 6 months 325
Appendix 4.2 Secondary biomarker scores of the Mediterranean diet derived in the MedLey trial and incidence of type 2 diabetes in EPIC-InterAct: associations per 1 standard deviation 326
Appendix 4.3 Secondary biomarker scores of the Mediterranean diet derived in the MedLey trial and incidence of type 2 diabetes in Denmark and comparison with associations of individual biomarkers comprising the scores: associations per 1 standard deviation 327
Appendix 5.1 Associations of biomarker scores of Mediterranean diet and incident type 2 diabetes in EPIC-InterAct per 1 standard deviation by categories of use of dietary supplements and age at baseline* 328
Appendix 5.2 Associations of biomarker scores of Mediterranean diet with incidence of type 2 diabetes in EPIC-InterAct: sensitivity analyses 329
Appendix 5.3 Associations of observational biomarker scores of Mediterranean diet with incidence of type 2 diabetes in EPIC-InterAct: effect of adjustment for component biomarkers 330
Appendix 6.1 Nutritional and metabolomic biomarkers correlated (absolute $r \geq 0.10$) with the Mediterranean diet pyramid score* after Bonferroni correction in the EPIC-Norfolk study 332

Chapter 1

Introduction

1.1 Background and context

Diabetes mellitus is a major and increasing public health problem globally in terms of observed and projected prevalence, contributions to mortality, as well as social and healthcare costs.[1] The global burden of diabetes is projected to rise from an estimated 537 million people in 2021 to 783 million by 2045, with type 2 diabetes (T2D) accounting for more than 90% of this burden.[1] There is an urgent need for tackling modifiable risk factors to stem the tide.

Diet plays a substantial contribution to mortality and morbidity, with an estimated 11 million deaths (22% of all deaths) and 255 million disability-adjusted years of life (DALYs) (15% of all DALYs) in 2017.[2,3] Of these, approximately 0.3 million deaths and 24 million DALYs were due to type 2 diabetes (T2D), whereas 10 million deaths and 207 million DALYs were due to cardiovascular disease – to which T2D contributes.[4] Notably, there has been an estimated 19% increase in deaths and a 16% increase in DALYs attributable to dietary factors between 2007 and 2017.[3]

Nutrition is aetiologically linked with T2D, at the very least through affecting energy balance and metabolic control.[5] Prospective observational studies have shown inverse associations between the level of adherence to healthy dietary patterns and risk of T2D independently from energy intake and adiposity,[6] thus suggesting that increasing dietary quality per se may be beneficial for diabetes prevention. However, reliable measurement and identification of specific dietary risk factors and quantification of their effects have been a major challenge, both in T2D research and nutritional epidemiology as an entire discipline.[7–10] Concurrently, interventional evidence has been limited. Lifestyle or health behavioural interventions of combined lower-fat dietary patterns, physical activity and weight loss have been shown to reduce the incidence of T2D in populations with impaired glucose tolerance.[11–14] The multifactorial nature of these interventions and confounding by weight loss preclude inference on changes in dietary quality alone, and application of such interventions to high-risk participants limits generalisability of the results to primary prevention.

Two landmark randomised controlled trials (RCT) of dietary pattern interventions have investigated T2D as a secondary outcome. The Women's Health Initiative (WHI) randomised 48,835 postmenopausal women in the USA to dietary counselling to reduce the intake of fat to 20% of energy and increase consumption of fruits, vegetable, and grains, or to continuation of habitual diet. The trial found no effect of the intervention on risk of self-reported treated diabetes during 8.1 years of follow-up (n cases = 3,342), as indicated by a hazard ratio of 0.96 with a 95% confidence interval (95% CI) of 0.90-1.03.[15] A secondary analysis has found a lower risk of initiation of insulin therapy (HR 0.74; 95% CI: 0.59-0.94) but not oral diabetes medications (HR 0.95; 95% CI: 0.59-0.94), and a lower risk of incident fasting glucose ≥100 mg/dL (odds ratio 0.75; 95% CI: 0.61-0.93) in the low-fat diet group compared to the control habitual diet.[16] A second study, the PREDIMED trial randomised 7,447 participants at high cardiovascular risk in Spain to provision of extra virgin olive oil or nuts with dietary counselling to increase adherence to the Mediterranean diet, or to a control dietary counselling group to follow a low-fat diet. Nearly half of the participants had prevalent diabetes at baseline and were excluded from the analysis. The risk of diabetes during 4.8 years of follow-up (273 cases) was lower with the Mediterranean diet interventions (HR 0.70; 95% CI: 0.54-0.92) and the Mediterranean diet with olive oil alone (HR 0.60; 95% CI: 0.43-0.85) compared to the control group.[17] The result for the Mediterranean diet with provision of nuts was compatible with both a decreased and an increased risk with the 95% confidence interval both below and above 1.0 (HR 0.82; 95% CI: 0.61-1.10). Notably, both the WHI[18] and PREDIMED[19] did not fully achieve the goals of reduction of fat intake in the low-fat diet groups. This decreased the magnitude of dietary differences between the intervention and control groups and may have biased the results towards the null.

Adequately powered RCTs of disease prevention with complex dietary interventions are rarely feasible. Thus, the relationship between adherence to dietary patterns and incidence of T2D or other disease outcomes is often investigated in prospective cohort studies based on dietary self-report. However, measurement error in self-reported dietary exposures has long been a major and valid criticism of nutritional epidemiology.[7–10] It has been used as one of the reasons for questioning whether nutritional science can at all generate useful dietary guidelines.[20] It includes both random and systematic error, which may bias the precision, magnitude, and even direction of diet-health associations.[21,22] Use of objectively measured nutritional biomarkers is one of the key recommendations for improving reliability in nutritional epidemiology.[23] It is an attractive prospect, potentially allowing for assessment of adherence to healthy diets without

the need to rely solely on self-report methods, which could be useful in both research and clinical settings. However, unlike nutrients and some foods, whole diets do not have biologically plausible specific biomarkers. This introduces a challenge of the need to combine data from multiple biomarkers. In this thesis, I aim to explore the methods to derive such composite biomarker measures of dietary patterns and to apply them to aetiological epidemiological investigations.

1.2 Dietary patterns

Isolated nutrients can be aetiologically important factors in the effects of diet on cardiometabolic disease risk, however, individuals make dietary choices with regards to foods, food groups or entire dietary patterns. In consequence, dietary guidelines for prevention of non-communicable diseases must be expressed in these 'units' for a comprehensible public health message, and so they need to be informed by nutritional research which is food- and dietary pattern-based.[24–26] Improving dietary quality, or increasing adherence to healthy dietary patterns, holds a great promise for public health as a cost-effective strategy for decreasing the burden of T2D and other non-communicable diseases.[27,28]

Dietary exposures can be viewed at different levels of reductionism – from dietary patterns through foods to nutrients or non-nutritive dietary constituents. Individual nutrients and foods have been a longstanding focus of epidemiological investigations. The last two decades of research have seen a major shift towards studying whole diets, as highlighted by the 2002 publication by Hu, entitled "Dietary pattern analysis: a new direction in nutritional epidemiology."[29] Though appropriately branded as 'new' in relation to development of modern methods of analysis, one of the earliest observations in nutritional epidemiology of non-communicable diseases (NCD) has in fact pertained to a dietary pattern. The Seven Countries Study conducted since the 1950s has found low rates of coronary heart disease in countries of the Mediterranean region.[30] Ecological analyses have identified nutrient patterns characteristic of the Mediterranean diet as inverses correlates of these rates.[30] A large body of observational evidence has since emerged, finding inverse associations between the Mediterranean diet and incidence of multiple NCDs, including cardiovascular diseases, diabetes, cancer, neurodegenerative diseases, as well as all-cause mortality, with low-to-moderate quality of evidence.[31,32] RCTs of the Mediterranean diet and its modified versions have reported reduced incidence of cardiovascular disease following interventions with this dietary pattern.[31]

Traditional exposure assessment of dietary patterns in observational research requires applying scores or indices that encapsulate information on estimated intakes of multiple correlated foods or nutrients into single variables representing the level of adherence to a given dietary pattern. These scoring algorithms typically aim to capture constructs relevant to underlying eating styles, disease prediction, or both. They can be used to investigate diet-disease associations as exposure variables, to control dietary confounding as covariates,[33] or for descriptive purposes of summarising dietary quality. Several approaches have been developed for dietary pattern analysis.[34] Selection of a given method is likely to impact on further steps in the analysis, such as associations with disease outcomes or, specific to this PhD thesis, derivation of biomarkers of adherence.

At the highest level, dietary pattern analysis can be divided into a priori and a posteriori methods.[34] A priori methods include pre-defined scoring algorithms or indices which aim to assess concordance with dietary guidelines, overall adequacy of nutrient intakes, or adherence to a named diet, such as the Mediterranean diet.[34] A posteriori methods are data-driven approaches that use a range of statistical approaches to identify sets of dietary variables based on their interrelatedness, for example using principal component analysis or cluster analysis.[34] Additionally, hybrid methods exist whereby variable selection and weighting is data-driven but informed by prior biological knowledge. For example, reduced rank regression can be used to identify a dietary pattern related to pre-selected biomarkers relevant to disease risk.[34] A disadvantage of a posteriori methods is their potentially limited generalisability outside of the derivation populations.[35,36] This is relevant to identification of targets for development of biomarkers in the current thesis. Valid biomarkers of exposure should be maximally generalisable to applications in different settings. Therefore, this PhD thesis focuses on pre-defined dietary patterns.

The diversity of approaches to derivation of dietary pattern indices, including multiplicity of scoring algorithms for the same named diets, has led to calls for standardisation of methods.[26] The Dietary Patterns Methods Project is an example of an initiative to systematically address this issue. It was established to inform development of Dietary Guidelines for Americans by conducting conduct standardised analyses of associations between dietary patterns and NCDs across multiple cohorts.[37] The project pre-selected the following dietary patterns based on their relevance to dietary recommendations and routine use in epidemiological research: the Mediterranean diet, Healthy Eating Index 2010, alternative Healthy Eating Index 2010, and Dietary Approaches to Stop Hypertension. I have applied this set of dietary patterns in my

research except for the Healthy Eating Index 2010 which aims to represent American dietary recommendations and thus may be of limited relevance to European populations used throughout the thesis.

1.3 Nutritional biomarkers

Nutritional biomarkers are commonly classified into biomarkers of dietary exposure and biomarkers of status.[38] The former group includes recovery and predictive biomarkers.[39] Biomarkers of dietary intake are sensitive, short-term intake biomarkers, and they have been established for a handful of dietary exposures. Recovery biomarkers include doubly-labelled water (energy intake), urinary nitrogen (protein), urinary potassium (potassium) and urinary sodium (sodium). Predictive biomarkers are a relatively new group of biomarkers which has been formed to accommodate such biomarkers as urinary glucose and fructose as biomarkers of total sugars intake.[40] They differ from recovery biomarkers in that only a small proportion of ingested amount is recovered, and they require applying a calibration equation to approximate intakes, which has so far only been developed in one study of 12 men in the UK.[41,42] In fact, it has been argued that recovery biomarkers can be ontologically thought of as a special case of predictive biomarkers free from bias.[41]

Nutritional biomarkers of status primarily consist in concentration biomarkers, which can be defined as biomarkers that have strong correlations with their respective nutrient intakes, but often lower than those observed for recovery biomarkers due to being affected by metabolism or personal characteristics.[39,43] However, it has recently been demonstrated that similarly to urinary sugars, calibration equations can be developed that provide valid estimated intakes.[44] Plasma carotenoids are an example of this group of biomarkers. Replacement biomarkers are a related group of biomarkers, and they can be defined by having biological properties of concentration biomarkers but lacking in or with poor-quality information in food composition tables. Examples include aflatoxins or some phytostrogens.[39]

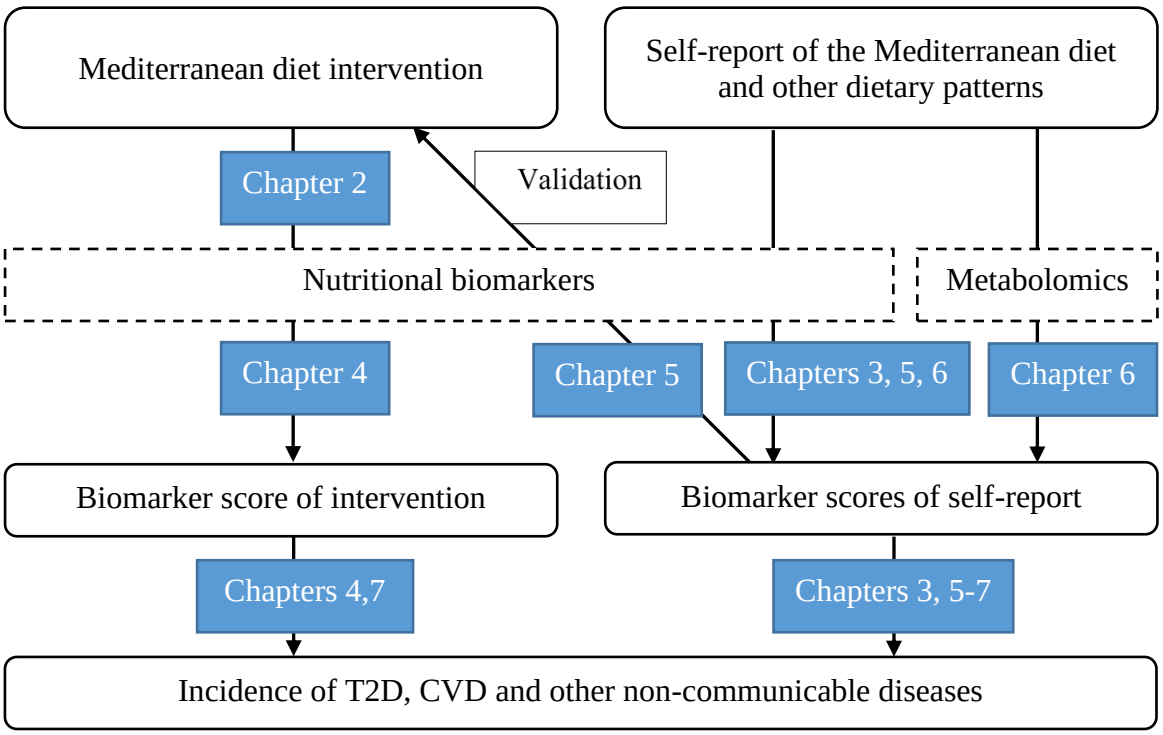
Nutritional biomarkers can complement self-reported assessment of dietary exposure. However, few established biomarkers of intake are available, and they have been largely limited to biomarkers of nutrients[38] or food groups, e.g. fruit and vegetables.[45] Their application to prospective associations with T2D has contributed considerably to elucidation of links with dietary intakes of energy and protein,[46] carotenoids and tocopherols,[47] dairy fat[48] and fruit and

vegetables.[49] Dietary patterns lack biologically plausible single biomarkers which necessitates combining multiple analytes into biomarker scores.[50,51] Previous research suggests that combinations of nutritional biomarkers can be used for assessment of complex dietary exposures. For example, profiles of serum phospholipid fatty acids perform well as biomarkers of habitual intake of carbohydrate and total saturated fatty acids in postmenopausal American women.[52] Group mean intakes of fruit and vegetable and ranking of individual intakes can be reliably estimated by combining information from circulating carotenoids, folate and vitamin C.[53]

1.4 Thesis aims

The totality of the effects of diet can only be reasonably expected to be captured at the level of dietary patterns as the exposure under analysis,[26] which motivates the search for biomarkers of adherence to overall diets of differing quality. As any single biomarker is unlikely to be sensitive and specific enough for such discrimination, multiple biomarkers have been used to identify circulating or urinary signatures of dietary patterns.[51,54,55] The research on this topic focusing on nutritional biomarkers so far has been limited,[50,51] perhaps because few studies concurrently measured multiple nutritional biomarkers. I hypothesised that combinations of these biomarkers could be combined to objectively characterise adherence to dietary patterns, and that such composite biomarkers, or nutritional biomarker scores, would be more strongly associated with incident T2D and other disease outcomes than adherence assessed using dietary self-report.

The overall aim of this PhD was to appraise the impact of dietary patterns on nutritional biomarkers, evaluate whether they can be usefully combined into composite measures reflecting dietary quality, and test their associations with incident disease outcomes by comparison with dietary self-report. This work focused predominantly on the Mediterranean diet and T2D. Outline of the thesis including chapter-specific aims is presented in **Figure 1.1**.



Aims:

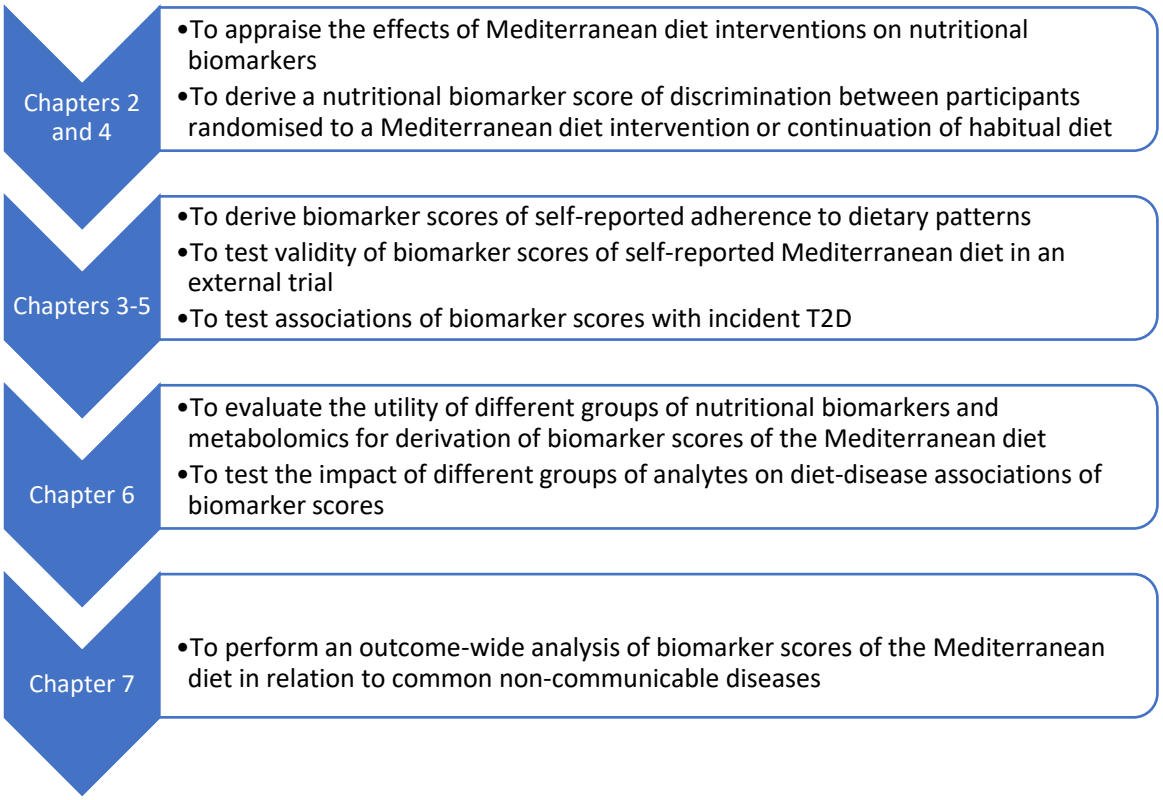


Figure 1.1 Thesis outline

Chapter 2

Effects of the Mediterranean diet on nutritional biomarkers: systematic review and meta-analysis of experimental evidence

Abstract

Background Assessment of adherence to interventions with the Mediterranean diet could be enhanced by using objectively measured nutritional biomarkers. However, their validity as markers of compliance to the Mediterranean dietary patterns has not been previously systematically evaluated. No reviews to date attempted to quantitatively synthesise the evidence on this topic.

Methods This review was prospectively registered in the International Prospective Register of Systematic Reviews (registration code: CRD42020168862). Three databases were searched up to February 2020 for interventions with the Mediterranean diet which assayed any nutritional biomarkers (Medline, Embase, Web of Science). Random-effects meta-analysis was undertaken to estimate standardised mean differences (SMD) if ≥5 trials reported on any given biomarker.

Results Forty-five publications from 29 trials were identified, reporting primarily on circulating fatty acids, vitamins and pro-vitamins. Twenty-one trials were assessed to be of poor quality, primarily over inadequate reporting of key study procedures to ascertain risk of bias. The most frequently used control diets were habitual diets (11 trials), followed by lower-fat diets (8 trials), and healthy dietary patterns (5 trials). Fatty acids were most frequently assayed in plasma or serum (up to 10 out of 18 trials), and otherwise heterogeneous blood fractions were used, including phospholipids, erythrocytes, cholesterol esters and triglycerides. Twenty-five biomarkers were eligible for meta-analysis. Heterogeneity was moderate-to-high (I^2 >50%) for 21 comparisons. Relative to control diets, the Mediterranean diet increased circulating concentrations of β-carotene, lycopene, retinol, vitamin C, and several monounsaturated and n-3 polyunsaturated fatty acids. It decreased the levels of total saturated and several n-6 polyunsaturated fatty acids and the n-6:n-3 ratio. Effect sizes were small-to-moderate (absolute SMD ≤ 0.67). Only the results for β-carotene and arachidonic acid were robust to all sensitivity analyses. Between-trial heterogeneity was sufficiently low to suggest reproducibility of the inverse effect of the Mediterranean diet on circulating arachidonic acid

in future trials (pooled SMD = -0.19; 95% confidence interval: -0.27, -0.11; 95% prediction interval -0.30, -0.06). The narrative synthesis without meta-analysis additionally identified decreased γ-tocopherol as a potential biomarker of the Mediterranean diet.

Conclusion Multiple nutritional biomarkers can be affected by the Mediterranean diet; however, single biomarkers are unlikely to capture adherence to an overall dietary pattern. Future research should consider combining information from multiple biomarkers into composite biomarkers to improve assessment of compliance to Mediterranean diet interventions. Results of this systematic review should be interpreted with caution, given the unclear risk of bias of most included trials, and heterogeneity of control diets and blood fractions used for measurement of fatty acids.

2.1 Background

The Mediterranean diet has been inversely associated with multiple noncommunicable disease outcomes (NCD).[31,32] It remains one of the few dietary exposures which have been tested against hard clinical endpoints for their preventative role.[56,57] Multiple trials have evaluated short-term effects of the Mediterranean diet on intermediate markers of cardiometabolic risk such as lipids and measures of glycaemia, suggesting a beneficial effect on metabolic health.[58] Despite being a dietary pattern of major scientific interest and the large body of evidence underpinning its application to dietary guidelines,[59] the research on biomarkers of adherence to this diet has been limited and largely confined to metabolomic biomarkers.[51,60]

Trials of the Mediterranean diet had historically assayed circulating nutritional biomarkers as measures of compliance to the intervention, in particular fatty acids.[61–63] This practice has continued till present; however, there has been little justification provided for using this group of biomarkers and the selection of particular sets of compunds.[64,65] Some components of the Mediterranean diet have established biomarkers of intake. For example, circulating carotenoids can be used as objective measures for fruit and vegetable consumption[45], and long-chain n-3 fatty acids for marine fish.[66] However, the validity of nutritional biomarkers as markers of the whole Mediterranean diet remains unknown.

Moreover, some nutritional biomarkers like circulating fatty acids are of scientific interest as risk factors for cardiovascular disease and type 2 diabetes.[67] Appraisal of their regulation by the Mediterranean diet could provide insights into the mechanisms by which this dietary pattern influences disease risk.

2.2 Aim

The aim of this chapter was to identify nutritional biomarkers of compliance to Mediterranean diet interventions, and to quantify the pooled effects of the Mediterranean diet on these outcomes.

2.3 Methods

This systematic review was prospectively registered in the International Prospective Register of Systematic Reviews (PROSPERO) as part of a broader review on biomarkers indicative of

adherence to dietary patterns (registration code: CRD42020168862). The broader review included all dietary patterns that are constructs of overall dietary quality, and it aimed to appraise the evidence from interventional and observational studies. It considered as candidate biomarkers both classical nutritional biomarkers, as well as metabolites measured using omics approaches. Anticipating a large number of eligible studies, the PROSPERO protocol of the broader review specified the Mediterranean diet and interventional evidence as the key areas for prioritisation of data synthesis. I have restricted the current chapter to the interventional study designs due to the limitations of construct validity of the Mediterranean diet scores and the multiplicity of such scores used in observational research.[68] Additionally, I have restricted the eligible biomarkers to nutritional biomarkers given the focus of the subsequent thesis chapters and emergence in the literature of reviews on metabolomic biomarkers of dietary patterns.[60,69]

I followed the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) 2020 statement[70] and the Synthesis Without Meta-Analysis (SWiM) guidelines[71] in conducting and reporting of the review.

2.3.1 Eligibility criteria

The systematic review was restricted to studies conducted in adults (≥ 18 years old at entry). I did not apply any additional exclusion criteria relating to the characteristics of the source populations.

Studies using any interventional designs with the Mediterranean diet were eligible for inclusion, i.e., randomised and non-randomised trials of parallel or cross-over design, as well as pre-post comparisons without a control group. I considered trials to have used the Mediterranean diet if the investigators described the intervention using the terms including, but not limited to "Mediterranean", "Mediterranean-type" or "Mediterranean-style" diet. Studies assessing postprandial responses to single meals representative of the Mediterranean diet were excluded. Pre-post studies were excluded if they combined the Mediterranean diet with another intervention, and controlled trials were excluded if they applied another intervention differentially to the Mediterranean and control diet groups. I did not consider the use of dietary supplements to constitute a separate intervention per se; however, use of dietary supplements was an exclusion criterion if an investigated biomarker was a known or plausible biomarker of a given nutrient included in the supplement. I used data from the longest duration of the active

intervention from trials reporting on biomarkers measured at multiple timepoints during the follow-up. In instances where biomarkers were measured at later timepoints in targeted subsamples (beyond the loss to follow-up), I used the full sample size data based on the shorter duration.

Any non-Mediterranean diet intervention, including continuation of baseline diet, was eligible as the control group. For studies which used multiple non-Mediterranean diet arms, I selected as the control group the intervention which most closely resembled the habitual diet of study participants. If none of the control groups were eligible, controlled trials were considered for inclusion in the review as pre-post studies without a control group using data from the Mediterranean diet arm only. For pre-post studies, baseline values of the outcomes were used as the reference to assess the effects of intervention.

All nutritional biomarkers of intake or exposure measured in any tissue using standard, non-omic laboratory methods were considered as eligible outcomes.[39] I followed the classic definition by Potischman of nutritional biomarker as "any biological specimen that is an indicator of nutritional status with respect to intake or metabolism of dietary constituents."[38] I prioritised inclusion of biomarkers that are chemically distinct compounds. Total sums of groups of biomarkers (e.g., total carotenoids) were eligible for inclusion if they were reported by multiple studies without separate reporting on all component biomarkers by at least one study.

2.3.2 Information sources

I searched MEDLINE, Embase and Web of Science since their inceptions until 19.02.2020. No language restrictions were applied. I consulted the search strategy with a librarian from the Medical Library of the University of Cambridge. The search string for MEDLINE is presented below and search strings for the remaining databases are available in **Appendix 2.1**. I used backward and forward citation searching as additional strategies to identify eligible publications. I used the "Cited by" option of Google Scholar for the forward searching. In cases of >100 citations, I filtered the Google Scholar search results with the keywords "trial" and "intervention".

- MEDLINE search via Pubmed:

#1 “dietary pattern”[Title/Abstract] OR “dietary patterns”[Title/Abstract] OR “diet pattern”[Title/Abstract] OR “diet patterns”[Title/Abstract] OR “diet quality”[Title/Abstract] OR “dietary quality”[Title/Abstract] OR “food pattern”[Title/Abstract] OR “food patterns”[Title/Abstract] OR “diet score”[Title/Abstract] OR “diet scores”[Title/Abstract] OR “dietary score”[Title/Abstract] OR “dietary scores”[Title/Abstract] OR “diet index”[Title/Abstract] OR “diet indices”[Title/Abstract] OR “dietary index”[Title/Abstract] OR “dietary indices”[Title/Abstract] OR “eating index”[Title/Abstract] OR “eating indices”[Title/Abstract] OR “eating patterns”[Title/Abstract] OR “eating pattern”[Title/Abstract] OR “healthy diet”[Title/Abstract] OR “healthy diets”[Title/Abstract] OR “food score”[Title/Abstract] OR “foods score”[Title/Abstract] OR “diet diversity”[Title/Abstract] OR “dietary diversity”[Title/Abstract] OR “Mediterranean diet”[Title/Abstract] OR "dietary approaches to stop hypertension"[Title/Abstract] OR “healthy eating index”[Title/Abstract] OR “DASH”[Title/Abstract] OR “HEI”[Title/Abstract] OR “AHEI”[Title/Abstract] OR “Nordic diet"[Title/Abstract]

#2 plasma[Title/Abstract] OR serum[Title/Abstract] OR circulating[Title/Abstract] OR blood[Title/Abstract] OR urin*[Title/Abstract] OR excret*[Title/Abstract]

#3 vitamin*[Title/Abstract] OR mineral*[Title/Abstract] OR ascorbate[Title/Abstract] OR acid[Title/Abstract] OR acids[Title/Abstract] OR caroten*[Title/Abstract] OR lycopene[Title/Abstract] OR cryptoxanthin[Title/Abstract] OR lutein[Title/Abstract] or zeaxanthin[Title/Abstract] OR folate[Title/Abstract] OR tocopherol*[Title/Abstract] OR polyphenol*[Title/Abstract] OR phytochemical*[Title/Abstract] OR nitrogen[Title/Abstract] OR potassium[Title/Abstract] OR sodium[Title/Abstract]

#4 biomarkers[MeSH Terms] OR biomarker*[Title/Abstract] OR metabolomic*[Title/Abstract] OR metabonomic*[Title/Abstract] OR lipidomic*[Title/Abstract] OR proteomic*[Title/Abstract] OR omic*[Title/Abstract] OR isotop*[Title/Abstract] OR “metabolic profile”[Title/Abstract] OR “metabolic profiles”[Title/Abstract] OR “metabolite profile”[Title/Abstract] OR “metabolite profiles”[Title/Abstract] OR “metabolic signature”[Title/Abstract] OR “metabolic signatures”[Title/Abstract] OR “lipid signature”[Title/Abstract] OR “lipid signatures”[Title/Abstract] OR VOC*[Title/Abstract] OR volatile[Title/Abstract]

#5 microbiota[MeSH Terms] OR gastrointestinal microbiome[MeSH Terms] OR urin*[Title/Abstract] OR plasma[Title/Abstract] OR serum[Title/Abstract] OR

blood[Title/Abstract] OR hair[Title/Abstract] OR “adipose tissue”[Title/Abstract] OR toenail*[Title/Abstract] OR fingernail*[Title/Abstract] OR “metabolic profile”[Title/Abstract] OR “metabolic profiles”[Title/Abstract] OR “metabolic signature”[Title/Abstract] OR “metabolic signatures"[Title/Abstract] OR Microbiota[Title/Abstract] OR “lipid signature”[Title/Abstract] OR “lipid signatures”[Title/Abstract] OR microbiome*[Title/Abstract] OR microflora*[Title/Abstract] OR microbiota*[Title/Abstract] OR microbial[Title/Abstract] OR gut flora*[Title/Abstract] OR intestinal flora*[Title/Abstract] OR intestine flora*[Title/Abstract] OR fecal[Title/Abstract] OR faecal[Title/Abstract] OR faeces[Title/Abstract] OR breath[Title/Abstract]

#6 animal[MeSH Terms] NOT human[MeSH Terms]

#7 #1 AND ((#2 AND #3) OR (#4 AND #5)) NOT #6

2.3.3 Article screening, data extraction, and risk of bias assessment

I undertook the data extraction and risk of bias assessment as the sole reviewer. Search results were initially deduplicated in Endnote X9. I used the Rayyan online software for screening the eligibility of articles for inclusion. The following information was extracted from each included trial: first author, year of publication, study name, study design, including duration and presence and extent of a feeding component; participants’ characteristics, country, central tendency measures or ranges of baseline age and body mass index (BMI), percentage of women, biomarkers, tissue or blood fraction in which they were measured, names of the dietary interventions and concise qualitative statements describing them (if available), details of the intervention goals and mode of delivery, p values for differences in biomarker concentrations; and by trial arm: numbers of participants, means and standard deviations of biomarkers (or between-arm difference and its standard deviation), and any other measures of central tendency, spread or precision. The feeding aspect of the study design was assigned to the following categories based on the provision of foods to the participants from the Mediterranean diet arm: no feeding, minor feeding component (single foods provided), partial-feeding (~10-50% of target energy intake provided), major feeding component (>50% of target energy intake) and full-feeding design. I used the WebPlotDigitizer version 4.5 to read values from figures in publications which did not report exact values. Trial duration times reported in weeks were converted to months and rounded to nearest integer.

For reporting and data presentation, I categorised biomarkers into groups based on compound similarity: (i) vitamins and pro-vitamins, (ii) saturated and monounsaturated fatty acids, (iii) polyunsaturated fatty acids and (iv) amino acids and amino acid-related compounds. I presented separately the results on biomarkers reported by ≤ 2 studies for vitamins and pro-vitamins and fatty acids, as well as cations and polyphenols.

I categorised the control diets into four groups: habitual, healthy, lower fat diet, or diet high in saturated fatty acids (high-SFA). Habitual diets were defined as either lack of a dietary intervention or a diet typical for the source population. Healthy diets included any dietary interventions which were not specifically focused on decreasing total fat or increasing SFA intake relative to Mediterranean diets, and which shared some of the characteristics of the Mediterranean diet with the goal of improving dietary quality, e.g., increased fruit and vegetable intake.

I used the National Heart, Lung, and Blood Institute Quality Assessment Tools to evaluate the risk of bias.[72]

2.3.4 Statistical analysis

All analyses were conducted using Stata version 15.1.

I estimated standardised mean differences (SMD) in nutritional biomarker concentrations between the Mediterranean diet and control interventions for all biomarkers reported by the included studies. I applied the Hedges' correction to account for small sample sizes.[73] Post-intervention means, standard deviations (SD) and numbers of participants by trial arm were prioritised as the data sources for estimation of the SMDs. If unavailable, means of change from baseline and their SDs were used, or post-intervention means were estimated from medians[74] and geometric means.[75] For multi-arm trials, I pooled separately the relevant intervention and control groups prior to estimating SMDs using the Cochrane Handbook formulae for combining summary statistics from multiple groups.[76] I estimated p values for the null hypothesis of no difference between the Mediterranean and control diets in biomarker concentrations post-intervention using the Z-test in cases where they were not reported.[77] Two-sided $\alpha = 0.05$ was used as the threshold for determining statistical significance of the results.

I performed random-effects meta-analysis for biomarkers for which at least five trials with sufficient summary statistics were available as a pre-specified criterion in the PROSPERO

registration.[78] I selected the random-effects model a priori on the basis of the anticipated differences in dietary compositions of the Mediterranean diet interventions, different types of control diets, diverse baseline diets and possible differences in the levels of baseline nutritional biomarkers due to non-dietary factors. I combined the post-intervention biomarker values and changes in biomarker concentrations from baseline in the meta-analysis.[79] I estimated the 95% confidence intervals (CI) and the 95% prediction intervals of the pooled effect sizes. In addition to the prediction intervals, heterogeneity was assessed based on the I^2 values and their associated 95% CIs.[78]

In cases of studies reporting results for the same biomarker measured in multiple tissues or blood fractions, I selected for inclusion into the meta-analysis the results for the specimen most reported on by all studies. For fatty acids, such decisions were additionally guided by suitability of different fractions as biomarkers of dietary intake if ties occurred in terms of the numbers of studies reporting on a given combination of a biomarker and blood fraction.[80] In the PREDIMED trial, total plasma measurements were prioritised, followed by VLDL-phospholipids over VLDL-triglycerides.[81,82] In the Mediterranean Eating Study, results from total plasma were selected for inclusion in the meta-analysis over plasma phospholipids.[83]

I evaluated publication bias based on Doi plots and LFK indices.[84] This novel method consists in plotting effect sizes against absolute z-score values of effect sizes and quantifying the geometric asymmetry of the plot. I investigated sources of heterogeneity using meta-regression for continuous variables and subgroup analysis for categorical variables. The Cochran's Q statistic for heterogeneity was used to evaluate effect modification by categorical covariates. Similarly as in the main analyses, subgroup analyses were conducted only if at least five trials were available in each subgroup as a pre-specified criterion in the PROSPERO registration. A priori study characteristics to investigate as the potential sources of heterogeneity included: study duration, baseline age, BMI, percent of female participants, type of control diet (habitual versus other), presence of a feeding component in the study design (dietary advice only or minor feeding component versus major feeding component or full-feeding design), and for fatty acids, blood fraction of measurement (plasma versus other). Dichotomous groupings of the categorical characteristics were derived a posteriori. I identified the following additional factors upon the initial data extraction: mean baseline biomarker concentrations, negative energy balance (present or not), and geographical location (Mediterranean countries, i.e., Spain, Italy or Greece, versus the remaining countries and the USA versus the remaining countries). Due to the outcome-wide nature of the current review, and thus multiplicity of such

secondary comparisons, I report only results with p values < 0.10 for subgroup heterogeneity or meta-regression coefficients.

2.3.5 Analytical issues in within-person trials

Estimation of treatment effects in cross-over RCTs and pre-post comparisons without a control group requires accounting for intra-individual variation. This aspect of study design can be addressed in the analysis of summary statistics by incorporating in it the correlations between the outcome after the treatment and control periods.[85,86] Studies often do not report the summary statistics in sufficient detail to reliably estimate the correlation coefficients internally[76], which necessitates their imputation, typically by using values from other trials.[85,86] All four cross-over RCTs included in the current review[87–90] and two meta-analysed pre-post studies[91,92] required the use of imputed values.

One of the parallel RCTs included in the current review, the MedLey Study, deposited individual participant data[64] which I used as a source of the correlation coefficients. The MedLey Study compared the effects of the Mediterranean diet and continuation of habitual diet on the following nutritional biomarkers included in the current review after three and six months of the intervention:

- plasma carotenoids – total carotenoids, α-carotene, β-carotene, β-cryptoxanthin, lycopene, sum of lutein and zeaxanthin
- erythrocyte fatty acids – SFA, trans (TFA), monounsaturated (MUFA), and polyunsaturated fatty acids (PUFA); total n-3, C18:3n-3, C20:5n-3, C22:5n-3, C22:6n-3, total n-6, C20:4n-6, C18:2n-6, n-6/n-3 ratio
- 24-h urinary cations – sodium, potassium, calcium, magnesium

Using data from the intervention arm only (n = 69 for most of the biomarkers), I calculated the Pearson correlation coefficients between the measurements at baseline and three months, and the baseline and six months. I applied the former to within-person trials with duration <4.5 months, and the latter to such studies lasting ≥4.5 months. Median correlation values were applied to biomarkers not overlapping with those measured in the MedLey Study. I considered fatty acids to be overlapping with those measured in the MedLey Study irrespective of blood fraction used in a given assay. At month 3, the correlation coefficients ranged from 0.08 for

lycopene to 0.88 for β-cryptoxanthin, with a median value of 0.66. The minimum, maximum and median values at month 6 were 0.16 (lycopene), 0.90 (total n-3) and 0.70.

2.3.6 Sensitivity analyses

As sensitivity analyses to test robustness of findings, I estimated (i) pooled effects of interventions as weighted (not standardised) mean differences; pooled effects as SMDs with the (ii) minimum and (iii) maximum correlation coefficient values from the MedLey Study applied to within-person trials, (iv) excluding studies which required imputed correlation coefficients, (v) non-randomised designs and (vi) cross-over RCTs. For the analysis of weighted mean differences, the results for fatty acids reported in absolute concentrations were converted to weight% if the mean total of all fatty acids could be derived. Studies were otherwise excluded from this analysis.

2.4 Results

The search identified 2,336 records in Medline, 3,049 in Embase, and 3,769 in the Web of Science. I detected 3,042 duplicates via manual identification in EndNote, and the algorithm of Rayyan identified an additional 170 duplicate records, yielding a total of 5,117 articles for screening. I excluded 4,342 publications based on titles and abstracts. I reviewed 775 full-text articles reporting on any dietary patterns and diet-related biomarkers. Of these, 663 were excluded from the current review based on non-interventional design or intervention other than Mediterranean diet, and 63 were excluded due ineligible biomarkers. I identified seven additional Mediterranean diet interventions reporting on nutritional biomarkers through forward and backward citation searching (**Figure 2.1**).

Among publications on Mediterranean diet interventions, I excluded three trials reporting on circulating fatty acids: one RCT due to a joint intervention with meditation[93], one pre-post study without a control group due to a joint intervention with physical activity[94], and one RCT[95] on the basis of using plant sterols with the Mediterranean diet which may affect the concentrations of fatty acids.[96] Additionally, I excluded one RCT reporting on serum 25(OH) D due to a control diet that was not clearly defined and dynamically changing throughout the intervention[97] (**Figure 2.1**). Forty five publications from 29 trials published between 1994 and 2019 were included in the review.[61,62,89–92,98–103,63,104–113,64,114–123,65,124–128,81–83,87,88]

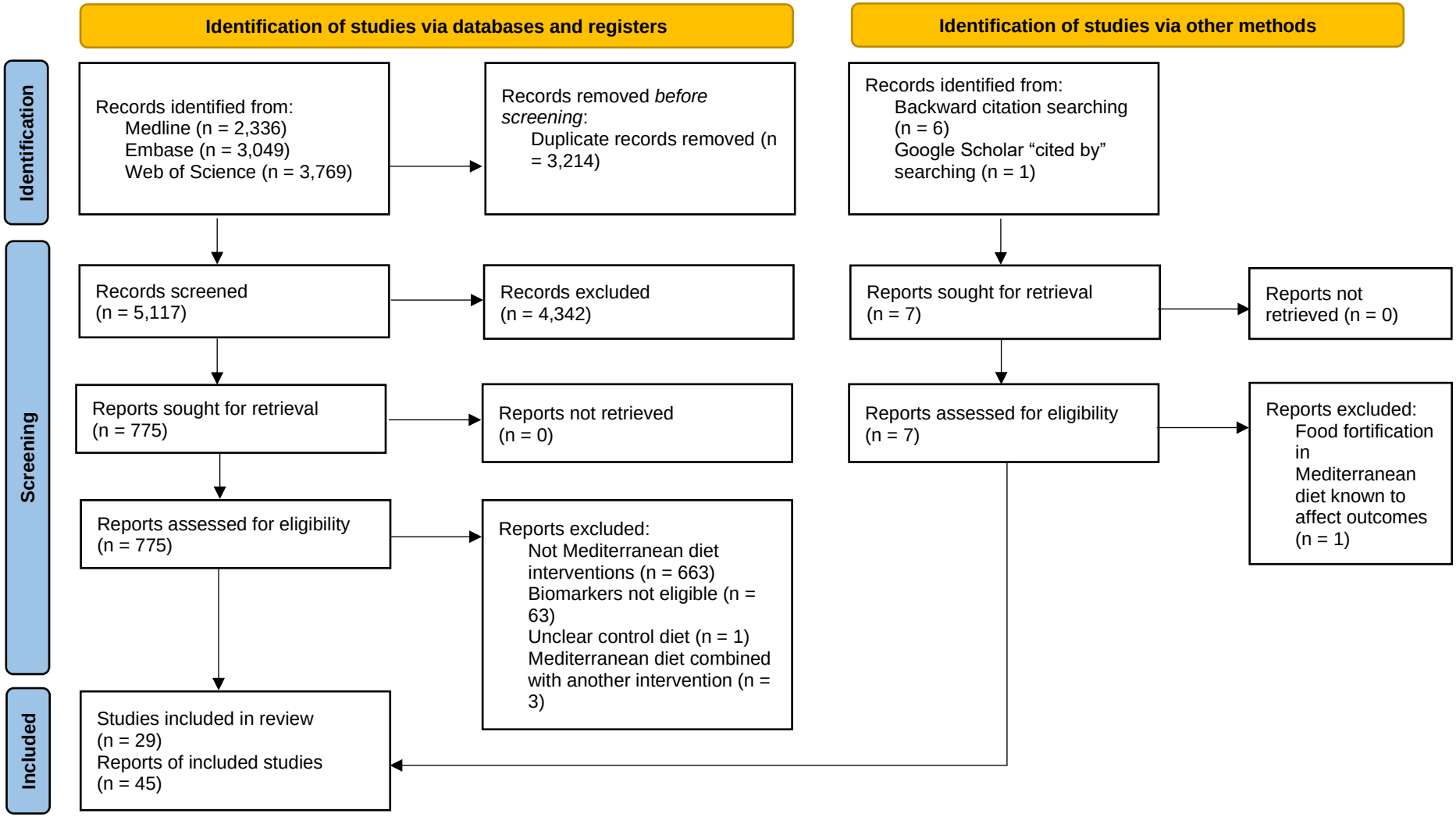


Figure 2.1 PRISMA 2020 flow diagram depicting the selection process of publications

2.4.1 Study characteristics

The review identified 19 parallel RCTs, four cross-over RCTs, two controlled trials without randomisation and four pre-post studies without a control group (**Table 2.1**). Per each of these study designs one trial was not eligible for meta-analysis due to insufficient numbers of other studies reporting on the same biomarkers. Two of the pre-post studies were intervention arms from controlled trials with ineligible control groups.[65,91] One RCT applied a 2x2 factorial design to compare the effects on the Mediterranean diet and n-3 fatty acids supplementation against placebo[105] Only the Mediterranean diet and placebo arms were included in the review. All but one trial[124] reporting on fatty acids used relative concentrations as percentages of total fatty acids measured or provided sufficient data to estimate the denominator and convert absolute concentrations to relative measures.

Seven trials were conducted in Italy, Spain, or Greece, 12 in other European countries, eight in North America, and two in Australia (**Table 2.1**). Eleven trials enrolled participants in general good health, 11 recruited from diverse patient populations with mostly cardiometabolic diseases, and seven included participants at high cardiometabolic risk. Central tendency measures of baseline age ranged from 22-77 years, with a median of 54 years. The range of baseline BMI values was 22.9-33.5 kg/m^2 with a median of 27. The numbers of participants ranged from 15-831, with a median of 90. The median duration of the intervention was 3 months, with a minimum of two weeks and a maximum of 5 years.

The trials were heterogeneous in terms of the implementation of dietary interventions, definitions of the Mediterranean diet and types of control diets used (**Table 2.1**). Three trials used a full-feeding outpatient design. Eight studies used only dietary counselling, six additionally provided plant oils, nine additionally supplied several major components of the interventions, and two additionally provided most of the foods. Seventeen out of the 29 studies utilised Mediterranean diet interventions described by the authors as modified Mediterranean diets. The modifications typically consisted in using local alternatives to olive oil and restricted total fat and SFA intake goals in line with local guidelines on cardiovascular disease prevention (**Table 2.1**). The most frequently used control diets were habitual diets (11 trials), followed by lower fat diets (eight trials). Five studies applied other dietary patterns associated with lower risk of NCDs or aimed at improving dietary quality. The interventions were calorie-restricted in five trials (**Table 2.1**).

Two trials compared the Mediterranean diet to multiple control diets. The cross-over RCT by Marin et al.[90] used a lower fat and a high-SFA diet. The parallel RCT by Parcina at al.[111] used a traditional German diet, as well as a "fast-food diet." I selected the lower fat and the German diet, respectively, as the comparator.

Table 2.1 Characteristics of experimental studies of Mediterranean diet included in the systematic review

Study, country	Participants	Design, duration	Mediterranean diet details	Comparator	Δ Energy	Biomarkers
Barona, 2012[91] USA	Women with metabolic syndrome, 77 ± 7 y (n = 15)	Parallel RCT, 3 months; used as a pre-post study without a control group	"Mediterranean-style low glycemic-load diet" Further information NR	Habitual, baseline diet of participants in the intervention arm; control arm not eligible for inclusion	Estimated ↓875 kcal/day	Plasma carotenoids
Bemelmans, 2002[98] Mediterranean Alpha-linolenic Enriched Groningen Dietary Intervention (MARGARIN) The Netherlands	Men and women at high risk of CVD, 67 ± 6 y (n = 265)	Controlled trial without randomisation, 12 months; minor feeding component Provision of ALA- or LA-rich margarine , random within each arm (amount equivalent to consumption of added fat)	"Mediterranean-type diet" based on the Lyon Diet Heart Study Three 2-h group nutritional counselling sessions and provision of written materials Goals: ↑study margarine /↓habitual added fat; bread 5-7 slices, vegetable 400 g, fruit 2 svgs, 1-2 alcoholic beverages (prevalent consumers only), low-fat dairy 2-3 svgs per day; fish 2 svgs/wk; ↓red meat, ↑poultry, ↓fatty cheese, ↓eggs, ↑ALA-rich foods	Postal leaflet with national dietary guidelines Goal: ↑study margarine /↓habitual added fat	↑0.4 kg/m^2 in the intervention and ↑0.6 kg/m^2 in the control group at 24 months	Cholesteryl ester fatty acids
Bihuniak, 2016[92] USA	Postmenopausal women in good general health consuming non-Mediterranean diets, 77 ± 7 y (n = 16)	Pre-post study without a control group, 3 months; partial-feeding Provision of EVOO, n-3 rich fish and walnuts	"Mediterranean-style diet" based on the PREDIMED trial Three counselling sessions with a dietitian and provision of written materials Goals: EVOO 3 tbs, walnuts 43 g, fruit and vegetable 5 svgs per day; fish 3-5 svgs/wk; ↑whole/↓processed grains	N/A	↓0.2 kg compared to baseline	Serum fatty acids

Study, country	Participants	Design, duration	Mediterranean diet details	Comparator	Δ Energy	Biomarkers
Davis, 2017[64] MedLey Study Australia	Non-smoking men and women in good general health, 71 ± 5 y (n = 137)	Parallel RCT, 6 months; partial-feeding Provision of estimated 30-35% of energy requirements in the intervention arm as EVOO, low fat Greek yogurt, unsalted nuts and canned legumes and tuna; in the control arm provision of monetary vouchers to local grocery stores (monetary value not reported)	"Based on a literature review to determine approximate food and nutrient content of the Mediterranean diet" Fortnightly sessions with a dietitian and availability of dietitian's advice over phone or email between visits; provision of written materials Goals: EVOO 1-3 tbs, vegetable 5-6 svgs, fruit 2- svgs, grains 4-6 svgs, potatoes ≤ 1, red wine ≤200 mL, skim milk ≤200 mL per day; nuts 4-6 svgs, Greek yogurt 6 svgs, cheese 3-4 svgs, poultry or pork 1-3 svgs, fish 3 svgs, legumes 3 svgs, red meat ≤1 svg, eggs ≤ 6, discretionary foods ≤ 3 svgs per week	Request to maintain habitual diet	Active maintenance of baseline body weight in the intervention arm; ↓0.2 kg/m^2 in both arms	Serum carotenoids, erythrocyte fatty acids, urinary cations
Djuric, 2009[83] Mediterranean Eating Study USA	Non-obese, non-smoking women without high intakes of MUFA and fruit and vegetable, 44 (25-59) y (n = 69)	Parallel RCT, 6 months; minor feeding component	"Modified Mediterranean diet" Provision of 3 l of olive oil and in-person dietary counselling at baseline and 3 months; weekly telephone counselling between baseline and 3 months, and fortnightly thereafter; provision of written materials Goals: PUFA:SFA:MUFA ratio of 1:2:5 while maintaining baseline total fat and energy	Written materials on increasing insufficient nutrient intakes (<67% of RDA estimated from 7-day food records) and provision of the National Cancer Institute's Action Guide to Healthy Eating	↓1.2 kg in the intervention and ↑0.2 kg in the control group	Plasma carotenoids and tocopherols, total plasma and phospholipid fatty acids

Study, country	Participants	Design, duration	Mediterranean diet details	Comparator	Δ Energy	Biomarkers
Fuentes, 2008[87] Spain	Non-obese male university students, 23 ± 2 y (n = 20)	Crossover RCT, 1 month; full feeding All meals provided to participants and consumed onsite	Goals: 15% protein, 47% CHO, 38% fat, 24% MUFA, 10% SFA, 4% PUFA (0.4% ALA); olive oil stated as a source of MUFA, further details NR	Lower fat diet: 15% protein, 55% CHO, 30% fat, 12% MUFA, 10% SFA, 8% PUFA (2% ALA); walnuts stated as the main source of ALA, further details NR High-SFA (22%) diet as a second comparison	No data reported	LDL cholesteryl ester fatty acids; fasting and after oral fat challenge
Hagfors, 2003[99] & 2005[100] Sweden	Men and women with rheumatoid arthritis not consuming a Mediterranean-like diet, 58 y (n = 51)	Parallel RCT, 3 months; partial-feeding Provision of arm-specific lunches and dinners on weekdays for the first 3 weeks as part of an outpatient rehabilitation programme	"Mediterranean-type diet" based on the Lyon Diet Heart Study and modified to a Swedish setting Six group sessions with a dietitian and provision of written materials; availability of dietitian thereafter every 3 weeks in group sessions and weekly over the phone; provision of frozen vegetables, tea, olive oil, canola oil and canola oil margarines throughout the intervention (amounts and frequency not specified) Goals: ↑fruit, vegetable, pulses, cereals, fish (n-3-rich), nuts and seeds (ALA-rich); ↓processed meat/↑poultry, fish or vegetarian meals, ↑olive and canola oils and canola oil margarines; ↓high-fat/↑low-fat dairy, ↑black tea	Habitual diet No explicit information provided on whether or not the outpatient rehabilitation programme includes nutritional counselling	No data reported	Plasma carotenoids, vitamins A and C, tocopherols, and serum phospholipid fatty acids

Study, country	Participants	Design, duration	Mediterranean diet details	Comparator	Δ Energy	Biomarkers
Healthy Eating for Colon Cancer Prevention Study Griffin, 2019[101]; Li, 2015[102]; Porenta, 2013[103]; Sen, 2013[104] USA	Men and women with family history of colon cancer, and fruit, vegetable and fat intake non-compliant with dietary guidelines, 52 ± 12 y (n = 94)	Parallel RCT, 6 months	Weekly nutritional counselling over telephone during the first month, fortnightly during the second and third month, and once a month thereafter; fortnightly study newsletter including dietary advice and motivational content	Healthy Eating diet based on the US Healthy People 2010 guidelines; large overlap of dietary goals with the intervention; analogous counselling as in the intervention	Diets designed to be isocaloric with baseline energy intakes	Serum and colon carotenoids, tocopherols, fatty acids; choline, betaine and related metabolites
Hjerkinn, 2006[105] Diet and Omega-3 Intervention Trial Norway	Men at high risk of CVD, 70 (65-75) y (n = 114)	Parallel RCT, 36 months; minor feeding component	"Mediterranean type",[129] "Mediterranean-like"[130] diet; target macronutrient composition according to standard CVD prevention guidelines, including restriction on total dietary fat Rapeseed oil and margarine provided at study visits (amounts and frequency NR); dietary counselling by a nutritionist at baseline and 3-month visits, and thereafter via telephone or in-person every 6 months	Usual care Not stated whether usual care includes any nutritional counselling	↓0.2 kg/m^2 in the intervention and ↑0.4 kg/m^2 in the control group	Serum PUFA (absolute concentrations)
Itsiopoulos, 2011[88] Australia	Men and women with well-controlled type 2 diabetes previously not exposed to a Mediterranean diet, 47-77 y (n = 27)	Crossover RCT, 3 months; major feeding component	"Reconstruction of the traditional Cretan Mediterranean diet" Provision of 70% of foods: frozen meals, wholegrain bread, olives, dried fruit, nuts, Greek coffee and herbal tea, and olive oil; counselling by a dietitian	Habitual diet	30.0 kg/m^2 after the intervention and 30.3 kg/m^2 after the and control period	Plasma carotenoids, retinol, phospholipid fatty acids and homocysteine

Study, country	Participants	Design, duration	Mediterranean diet details	Comparator	Δ Energy	Biomarkers
Jaacks, 2018[106] USA	Overweight and obese, non-smoking men and women with stable weight and habitual diet, 55 y (n = 20)	Parallel RCT (pilot), 2 months; partial-feeding	"Prototypical Mediterranean diet" Provision of three meals with beverages and two snacks per day during the first four weeks; weekly in-person or telephone counselling by research nutritionist; provision of written materials	No intervention	↓2.2 kg in the intervention and ↑3.9 kg in the control group	Plasma cysteine and cystine
Jula, 2002[107] Finland	Men with previously untreated hypercholesterolaemia, 48 ± 6 y (n = 120)	Parallel RCT, 6 months; within-group crossover randomisation to simvastatin vs placebo; minor feeding component	"Modified Mediterranean-type diet" Provision of rapeseed margarine and oil, oat bran (20 g/day) and frozen berries (50 g/day); one individual and two group counselling sessions with a nutritionist at baseline, and monthly group sessions thereafter	No intervention	Normocaloric diet	Plasma α-tocopherol, β-carotene and vitamin C, and erythrocyte folate
Lyon Diet Heart Study de Lorgeril, 1994[61] & 1998[62]; Renaud, 1995[63] France	Clinically stable male (> 90%) and female survivors of a recent myocardial infarction, 53.5 ± 10 y (n range: 250 -483)	Subsets of a parallel RCT, 2 and 12 months; minor feeding component	"Mediterranean alpha-linolenic acid-rich diet" Provision of canola oil margarine for participants and their families (amounts and frequency not reported); counselling by research cardiologist and dietitian at baseline, 2 months, and once a year thereafter	Usual care Nutritional counselling by attending physician expected in line with standard CVD prevention guidelines (not ascertained)	↑1 kg in the intervention and ↑1.6 kg in the control group	Plasma α-tocopherol , β-carotene, vitamin C and plasma fatty acids

Study, country	Participants	Design, duration	Mediterranean diet details	Comparator	Δ Energy	Biomarkers
Marin, 2011[90] Spain	Non-smoking men and women in general good health, >65 y (n = 20)	Crossover RCT, 1 month; feeding component of unclear extent Olive oil and walnuts likely provided; full-feeding intervention also plausible	"Mediterranean diet enriched in MUFA" Olive oil stated as a source of 80% of MUFA; sample daily menu indicated a predominantly plant-based diet low in processed foods with poultry, eggs and low-fat dairy; nutritional counselling on eating outside of home and provision of written materials	Lower fat diet enriched in ALA Olive oil in the intervention replaced with biscuits, jam, cereals, low-fat cakes, bread and walnuts Very high-SFA (22%) diet used as a second comparison	Estimated intakes of 1,960 kcal/day after the intervention and 1,982 kcal/day after the control period	Plasma α-tocopherol and β-carotene
Muzsik, 2019[108] Poland	Non-smoking postmenopausal women with or at risk of metabolic syndrome, 61 ± 5 y (n = 119)	Parallel RCT, 4 months; partial-feeding Provision of estimated 35% of energy requirements as ready-to-eat main meals	Mediterranean diet based on the Mediterranean Diet Foundation guidelines Olive used or recommended in every meal and 5-7 nuts/day provided; 14-day menus, recipes and written instructions	Lower fat diet in line with standard CVD prevention guidelines Staple centre-European foods used as fibre sources: oatmeal, barley, pulses, root and cruciferous vegetable	↓7.7 kg in the intervention and ↓7.5 kg in the control group	Erythrocyte fatty acids (absolute concentrations)
NU-AGE Jennings, 2019[109]; O'Neill, 2015[110] Italy, France, UK, The Netherlands, Poland	Elderly men and women in general good health, 71 ± 4 y (n range: 140-831)	Parallel RCT, 12 months; partial-feeding	"Mediterranean-style" diet Provision of food items every 4 months (amounts of foods NR): wholegrain pasta, EVOO, low-fat low-salt cheese, high-PUFA margarine, vitamin D supplement (10 mcg/d); 10 sessions with a dietitian in person or over the phone and provision of written materials	Request to continue usual diet and provision of a leaflet on current dietary guidelines	Normocaloric diet	Plasma fatty acids (UK only), urinary Na & K

Study, country	Participants	Design, duration	Mediterranean diet details	Comparator	Δ Energy	Biomarkers
Parcina, 2015[111] Germany	Non-obese, non-smoking men without chronic disease not consuming food additives, 30 ± 6 y (n = 27)	Parallel RCT, 2 weeks; major feeding component or full feeding (unclear); 2 week run-in period with the control diet Three freshly prepared meals/day consumed onsite	Goals: fruit 3-4 svgs/d; ↑pasta, rice, couscous, potatoes, nuts, legumes, vegetable, olive oil; moderate amounts of wine, fish, dairy, poultry; ↓red meat and eggs Information on dietary counselling NR	"Traditional German" Goals: ↑pork, sausages, butter; moderate amounts of wholegrain bread, potatoes, fruit, vegetable, dairy, fish, eggs; moderate/high amounts of beer, sweet foods, sugar Fast-food diet used as a second comparison	Normocaloric diet (2,500 kcal/day) changed to hypercaloric (2,900 kcal/day) on day 7 to prevent dropouts from the fast-food arm	Plasma or serum vitamins A, E, 25(OH) D, B_1 and B_6, magnesium, selenium, zinc and homocysteine
Pérez-Jiménez, 2001[89] Spain	Male and female university students after one month of run-in high-SFA (20%) diet feeding, 23 ± 2 y (n = 59)	Crossover RCT, 1 month; full-feeding, all meals consumed onsite Similar nutrient content as target goals confirmed in diet duplicates	"Mediterranean diet enriched in olive oil" Goals: 15% protein, 47% CHO, 38% fat, 22% MUFA (75% from olive oil), 10% SFA, 6% PUFA, dietary cholesterol 115 mg/1,000, fibre 30 g/d	"Low fat, high carbohydrates diet" Goals: ↑biscuits, bread and jam/↓olive and palm oil; 15% protein, 57% CHO, 28% fat, 12% MUFA, 10% SFA, 6% PUFA, dietary cholesterol 115 mg/1,000, fibre 30 g/d	Mean weight 66.2 kg at the end of the intervention period and 66.3 kg after the control period	LDL cholesteryl ester fatty acids
Pontifical Catholic University of Chile (PCUC) study Leighton, 1999[112]; Mezzano, 2003[113]; Urquiaga, 2004[114] & 2010[115]	Male university students, 22 ± 3 y (n = 42)	Partially (81%) randomised controlled trial, 1-2 months; partial-feeding Provision of lunches and dinners; daily soft	"Mediterranean-type diet" Goals: fruit and vegetable 675 g, olive oil 32 ml, poultry 157 g, red meat 59 g, cereals and legumes 311 g per day; protein 17.6 %E, CHO 55.1 %E, fat 27.3 %E, PUFA/n-3/MUFA/SFA 12.5/2.13/49.2/29.5 % of total fat	"High-fat Western-type diet" Goals: fruit and vegetable 246 g, sunflower oil 32 ml, poultry 74 g, red meat 209 g, cereals and legumes 261 g per day;	Both diets designed to provide 2,565 kcal/d	Plasma carotenoids, fatty acids, vitamins C and E, polyphenols; serum vitamin B12 and folate; urinary polyphenols

Study, country	Participants	Design, duration	Mediterranean diet details	Comparator	Δ Energy	Biomarkers
Chile		drink in month 1 and red wine (240 ml/d) in month 2	Breakfast request: tea or coffee, skimmed milk, yogurt, whole wheat bread, marmalade and avocado.	protein 17.5 %E, CHO 42.8 %E, fat 39.9 %E, PUFA/n-3/MUFA/SFA 28.4/0.85/31.5/31.8 % of total fat Breakfast request: tea or coffee with whole milk, white bread, butter, and cheese		
Prevención con Dieta Mediterránea (PREDIMED) trial Bullo, 2009[116]; Estruch, 2018[117]; Guasch-Ferré, 2017[118]; Medina-Remón, 2014[119]; Perona, 2010[82]; Mayneris-Perxachs, 2014[81]; Ruiz-Canela, 2016[120] Spain	Men and women at high risk of CVD, 67 ± 6 y (n range: 32-750)	Subsets of a parallel RCT, 3 months-5 years; minor feeding component Deviations from randomised assignment; extent or bias in subsets with nutritional biomarkers NR	Provision of either 15 l of EVOO or 4 kg of nuts per 3 months (personal intake goal of 30 g/d of nuts) per participant's household Quarterly individual and group dietary counselling by dietitians Goals: olive oil ≥4 tbsp, fruit ≥3 svgs, vegetable ≥2 svgs, soda drinks <1, spread fats <1 svg, red and processed meat <1 svg per day; nuts ≥3 svgs, fish (fatty preferred) and seafood ≥3 svgs, legumes ≥3 svgs, sofrito ≥2 svgs, sweets <2 svgs, wine ≥7 glasses (in drinkers) per week; ↑poultry/↓red meat	"Low-fat diet" Baseline dietary counselling and yearly leaflets Goals: low-fat dairy ≥3 svgs, fruit ≥3 svgs, vegetable ≥2 svgs, grains ≥3 svgs, vegetable oils ≤2 tbsp per day; lean fish and seafood ≥3 svgs, sweets ≤1 svg, nuts and fried snacks ≤1 svg, red and processed fatty meats ≤1 svg, fatty fish ≤1 svg, spread fats ≤1 svg, sofrito ≤2 svgs per week; ↓visible fat in meats and soups	At 12 months intervention with EVOO feeding vs control -0.006 kg, intervention with nuts feeding vs control 0.124 kg[131]	Plasma and VLDL phospholipid fatty acids, serum amino acids and related metabolites, serum and urinary calcium, urinary hydroxytyrosol

Study, country	Participants	Design, duration	Mediterranean diet details	Comparator	Δ Energy	Biomarkers
Richard, 2012[121] Canada	Men with metabolic syndrome without a history of CVD or type 2 diabetes, 51 ± 11 (18-65) y (n = 19)	Pre-post study without a control group, 5 weeks after a 5 week run-in period with a North American diet; full-feeding	Goals (per 2,500 kcal/d): whole grains 5.4 svgs, nuts 0.9, fruit and vegetable 16.1 svgs, svgs, low-fat dairy 2 svgs, poultry 0.9 svg, red meat 0.2 svg, red wine 2.6 svg per day; legumes 3.6 svgs, fish 8.8 svgs, eggs 2.2 svgs, sweets 2 svgs, olive oil 302.8 g per week	Goals: whole grains 1.2 svgs, nuts 0.5 svgs, fruit and vegetable 6.6 svgs, , dairy 2 svgs, poultry 1 svg, red meat 1.9 svg, red wine 1 svg per day; legumes 0.6 svgs, fish 1 svg, eggs 2.6 svgs, sweets 13 svgs, olive oil 4.5 g per week	↓1.2 kg compared to baseline	Plasma phytosterols
Skouroliakou, 2018[122] Greece	Women with stage I-IIIA breast cancer without severe coexisting diseases not taking any dietary supplements, 52 y (n = 50)	Parallel RCT, 6 months Physical activity recommendations provided to participants; details NR	"Personalised dietary intervention based on Mediterranean diet"; further details NR	Ad libitum diet based on the American Cancer Society guidelines for cancer prevention; further details NR	↓3.77 kg in the intervention and ↑2.16 kg in the control group	Serum vitamin C, retinol and α-tocopherol
Sofi, 2018[123] Cardiovascular Prevention with Vegetarian Diet (CARDIVEG) Italy	Overweight men and women at low-to-moderate CVD risk, 50 (21–75) y (n = 103)	Crossover RCT, 3 months In-person nutritional counselling (details NR) and arm-specific written materials	Goals: red meat ≤1 svg, poultry ≤3 svg, fish 2-3 svgs per week; further details on food goals NR; protein 15-20%, CHO 50-55 %E, fat 25-30 %E, SFA ≤7 %E, dietary cholesterol <200 mg/d	Vegetarian diet Goals: abstinence from flesh of any animal; further details on food goals NR; macronutrient goals as in the intervention arm	↓1.8 kg after the intervention and ↓1.9 kg after the control phase	Serum haemoglobin, haematocrit, folate, vitamin B_{12}, ferritin, iron, sodium, potassium, calcium, magnesium

Study, country	Participants	Design, duration	Mediterranean diet details	Comparator	Δ Energy	Biomarkers
Sotos-Prieto, 2019[65] Feeding America's Bravest USA	Mostly (95%) male firefighters from the control arm of a previous RCT of Mediterranean diet vs no intervention, 48y (n = 20)	Parallel RCT, 6 months; used as a pre-post study; partial-feeding Provision of EVOO, nuts and wholegrain pasta to the workplace (amounts NR), grocery discounts on staple Mediterranean diet foods for participants and their families	Group educational sessions and provision of written materials Goals: olive oil ≥4 tbs (EVOO encouraged), fruit ≥3 svgs, vegetable ≥2 svgs, fresh herbs and allium ≥2 svgs, yogurt ≤2 svgs, soda drinks <1 svg, spread fats <1 svg per day; fish and seafood ≥3 svgs, legumes ≥3 svgs, sofrito ≥2 svgs, poultry 2-3 svgs/↓ red meat, wine (in drinkers) ≥7 glasses, red and processed meat ≤2 svgs, sweets <3 svgs, fast-food ≤1 svg per week; ↑whole grains/↓refined grains	Not eligible for inclusion in the current review (self-sustained Mediterranean diet after the end of the same active intervention as in the current intervention arm)	No data reported	Plasma fatty acids, tyrosol and hydroxytyrosol in the first spot morning urine
Stachowska, 2005[124] Poland	Non-smoking male and female kidney graft recipients on long-term low-fat diets, 43 y (n = 37)	Parallel RCT, 6 months	"Mediterranean-type diet" Counselling by dietitian (duration and frequency NR); provision of 4-week menus Goals: nuts and seeds ~30 g/d; ↑cereals, pulses, wholegrain bread, vegetable, oat flakes, spaghetti; protein 15 %E, CHO 47 %E, fat 38 %E; estimated fibre intake 47 g/d	"Low-fat diet" (usual care in kidney graft recipients), isocaloric with the intervention; dietary counselling NR Goals: bread, potatoes, rice as main source of CHO; protein 17 %E, CHO 57 %E, fat 26 %E; estimated fibre intake 24 g/d	Energy intake goals 2,500 kcal/d in men and 2,000 kcal/d in women	Plasma α-tocopherol and triglyceride fatty acids
Thomazella, 2011[125] Brazil	Non-smoking, non-diabetic, clinically stable consecutive hospital patients with	Controlled trial without randomisation, 3 months	Goals: whole grains 4-6 svgs, vegetable and legumes 2-3 svgs, EVOO 30 ml, low-fat dairy 1-2 svgs, nuts 10g per day; fish 3-4 svgs, poultry 3-4 svgs, eggs 0-4,	Lower fat diet (NCEP) Goals: ↑fruit, vegetable, legumes, whole grains, low-fat	↓1.6 kg in the intervention and ↓1.7 kg in the control group; weight loss	Plasma arginine

Study, country	Participants	Design, duration	Mediterranean diet details	Comparator	Δ Energy	Biomarkers
	history of coronary events, 55 ± 5 (45-65) y (n = 40)	Assignment to trial arm by the investigators "on the basis of previous cultural and dietary habits and 4-day food records"	red meat 1 svg per week; protein 12-17 %E, CHO 45-50 %E, fat 33-38 %E, SFA ≤8 %E, PUFA ≤10 %E, n-3 PUFA >0.75 %E, MUFA 20-25 %E, dietary cholesterol <200 mg/d	dairy; moderate intake of lean meat, fish, or poultry; protein 15 %E, CHO 55-60 %E, fat 25-30 %E, SFA ≤7 %E, PUFA ≤10 %E, MUFA ≤20 %E, dietary cholesterol <200 mg/d, plant stanols 2 g/d, viscous fibre 10-15 g/d	encouraged in overweight participants	
Tutino, 2018[126] NUTRItion and Ac-TiviTy (NUTRIATT) Italy	Overweight men and women with moderate or severe non-alcoholic fatty liver disease without major CVD or type 2 diabetes, 54 ± 10 y (n = 32)	Parallel RCT, 3 months Arm-specific lists of foods by requested frequency of consumption; participants followed by dietitians, details NR	Three arms: (i) "Low glycemic index Mediterranean diet" (LGIMD), (ii) LGIMD and aerobic training, (iii) LGIMD and resistance training Goals: ↑wholegrain bread and pasta, vegetable and seasonal fruit, legumes, nuts, oily fish, EVOO; "white meats" in moderation, low-fat cheese and eggs weekly	3 arms: (i) Italian dietary guidelines, (ii) no dietary intervention and aerobic training, (iii) no dietary intervention and resistance training Details NR on the Italian dietary guidelines diet	Change in energy intakes or expenditure not targeted by either dietary intervention arm; details NR on physical activity-only arms	Erythrocyte fatty acids
Tuttle, 2008[127] The Heart Institute of Spokane Diet Intervention and Evaluation Trial (THIS-DIET) USA	Male and female recent MI survivors, 58 ± 10 y (n = 94)	Parallel RCT, 6 months Individual counselling sessions with a dietitian in months 1 (2x), 3 and 6; ≥6 arm-specific group sessions	"Mediterranean-style diet" Goals: fruit and vegetable ≥5 svgs/d, fish 3-5 svgs/wk; ↑whole grains, olive, canola and soybean oils; protein 10-20 %E, CHO 50 %E, SFA ≤7 %E, n-3 PUFA 0.75 %E, MUFA 20-25 %E, dietary cholesterol ≤200 mg/d	"Low-fat diet (the AHA Step II diet)" Goals: fruit and vegetable ≥5 svgs/d; protein 10-20 %E, CHO 55-60 %E, SFA ≤7 %E, n-3 PUFA 0.3-0.45 %E, MUFA 10-15 %E, dietary cholesterol ≤200 mg/d	↓1 kg/m^2 in each arm; weight loss encouraged in overweight participants	Plasma fatty acids

Study, country	Participants	Design, duration	Mediterranean diet details	Comparator	Δ Energy	Biomarkers
Vincent-Baudry, 2005[128] Medi-RIVAGE France	Men and women at high risk of CVD not treated with hypolipidaemic or hypoglycaemic medications, 57 (18-70) y Western-type diet consumed at baseline (n = 169)	Parallel RCT, 3 months; partial-feeding or minor feeding component (amounts of foods NR) in the intervention arm only Nutritional counselling by physicians and dietitians (details NR) and provision of written materials	"Mediterranean diet adapted from the traditional model" Provision of oat bran-enriched pasta, tomato sauce, and olive oil. Goals: ↑nuts, wholegrain bread, cereals, fruit, vegetable, legumes, olive oil; fish 4 svgs, red meat 1 svg, red wine ≤2 glasses per week; sheep and poultry as recommended meat types, sheep and goat cheese as recommended cheese types; total fat ~37%E, SFA ~9%E, MUFA ~18%E, PUFA ~9%E, fibre 25 g/d, carotenoids 7 mg/d, Ca ≤800 mg/d (limited dairy)	"Low-fat AHA-type diet" Goals: ↑fruit, vegetable, low-fat dairy, vegetable oils, ↑poultry/↓ mammal meat, ↓offal and SFA-rich animal products; fish 3 svgs/wk, alcohol avoidance; fat 30%E, SFA 10%E, MUFA 10%E, PUFA 10%E, fibre 20 g/d	↓1.5 kg/m^2 in the intervention and ↓1.2 kg/m^2 in the control group	Plasma carotenoids, fatty acids, phenolic compounds, folate and vitamin B_{12}

Abbreviations: %E – percent of energy; AHA – American Heart Association; CHO – carbohydrate; d – day; EVOO – extra-virgin olive oil; MUFA – monounsaturated fatty acids; NCEP – National Cholesterol Education Program; MI – myocardial infarction; n – number of participants; NR – not reported; PUFA – polyunsaturated fatty acids; RCT – randomised controlled trial; SFA – saturated fatty acids; svgs – servings; y - years

Numbers of participants are given for those with biomarkers measured at baseline, excluding any trial arms not eligible for inclusion.

2.4.2 Corrections to data

There were several inconsistencies which required corrections to the extracted data. One of the three publications from the Lyon Diet Heart Study[61–63] reported substantially larger standard errors than the two other articles.[62] I confirmed them to be standard deviations by estimating standard deviations from the standard errors in the two remaining articles, which yielded similar values for the biomarkers overlapping between the three publications. Three articles from two studies reported implausible serum or plasma concentration units. I corrected them so that the concentrations would be of the same order of magnitude as the remaining studies as follows: pg/mL was replaced with ng/mL for carotenoids[104], nmol/l was replaced with mmol/l for tocopherols[102], and mmol/l was replaced with mcmol/l for a range of antioxidants.[99] The study which I used as the source of information for imputation of correlation coefficients for the analysis of within-person trials reported serum β-carotene levels[64] which were among the highest reported in the literature.[132] I contacted the authors of the trial, and they confirmed the reported β-carotene values as correct. One study mislabelled C18:2n-6 and C18:3n-3[65] as was already confirmed with the authors by another systematic review,[133] and I swapped the statistics reported for these two fatty acids.

2.4.3 Risk of bias assessment

Among controlled trials, I assigned an overall "good" rating to one study in the assessment of risk of bias,[64] followed by four studies with a "fair" rating,[61–63,108–110,123] and the remainder marked as "poor". The main drivers of the latter were non-randomised designs,[98,125] irregularities in the implementation of randomisation in two RCTs,[81,82,120,112–119] and lack of or inadequate reporting on randomisation and concealment to allocation (**Table 2.2**). In addition to the domains included in the quality assessment tool, cross-over RCTs were at risk of bias due to carry-over effects.[85,86] Neither of the included cross-over RCTs used a washout period between the treatment and control interventions or tested for presence of this effect.[134] Among pre-post studies without a control group, three studies were rated as "fair" and one as "poor" (**Table 2.3**).

Table 2.2 Risk of bias assessment of controlled trials – National Heart, Lung, and Blood Institute Quality Assessment Tool

Study (first author, year or study name)	Described as randomised	Randomisation adequate	Allocation concealed	Participants blinded	Assessors blinded	Baseline balance	Overall dropout <20%	Dropout difference <15%	High adherence	Other treatments avoided	Valid outcome assessment	Adequate power	Outcomes pre-specified	Intention-to-treat analysis	**Overall rating**
Bemelmans, 2002[98] (MARGARIN)	No	N/A	N/A	No	NR	CD	Yes	Yes	Yes	Yes	Yes	Yes	NR	Yes	**Poor**
Davis, 2017[64] (MedLey Study)	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	**Good**
Djuric, 2009[83]	Yes	NR	NR	No	NR	Yes	Yes	No	NR	Yes	Yes	No	NR	Yes	**Poor**
Fuentes, 2008[87]	Yes	NR	NR	No	NR	NR	Yes	Yes	NR	Yes	Yes	NR	NR	Yes	**Poor**
Hagfors, 2003[99] & 2005[100]	Yes	Yes	NR	No	NR	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	**Poor**
HECCPS [101–104]	Yes	NR	NR	No	NR	Yes	No	Yes	CD	Yes	Yes	No	No	Yes	**Poor**
Hjerkinn, 2006[105]	Yes	Yes	NR	No	NR	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	**Poor**
Itsiopoulos, 2011[88]	Yes	NR	NR	No	NR	NR	Yes	Yes	CD	Yes	Yes	CD	NR	Yes	**Poor**
Jaacks, 2018[106]*	Yes	NR	NR	No	NR	Yes	Yes	Yes	Yes	No	Yes	No	NR	Yes	**Poor**
Jula, 2002[107]	Yes	NR	NR	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	NR	Yes	**Poor**
Lyon Diet Heart Study[61–63]	Yes	NR	Yes	No	Yes	No	NR	NR	Yes	Yes	Yes	Yes	No	Yes	**Fair**
Marin, 2011[90]	Yes	NR	NR	No	NR	Yes	Yes	Yes	Yes	Yes	Yes	NR	NR	Yes	**Poor**
Muzsik, 2019[108]	Yes	Yes	NR	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	**Fair**
NU-AGE[109,110]	Yes	Yes	NR	No	Yes	Yes	Yes	Yes	Yes	No	CD	Yes	Yes	Yes	**Fair**
Parcina, 2015[111]*	Yes	NR	NR	No	NR	CD	Yes	Yes	Yes	No	Yes	NR	NR	Yes	**Poor**
Pérez-Jiménez, 2001[89]	Yes	NR	NR	No	NR	NR	Yes	Yes	Yes	Yes	Yes	NR	NR	Yes	**Poor**
PCUC study[112–115]	Yes	No	NR	No	NR	NR	Yes	Yes	Yes	No	Yes	NR	NR	Yes	**Poor**
PREDIMED[81,82,116–120]	Yes	CD	NR	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	CD	Yes	**Poor**
Skouroliakou, 2018[122]	Yes	NR	NR	No	NR	No	Yes	Yes	Yes	Yes	Yes	NR	CD	Yes	**Poor**
Sofi, 2018[123] (CARDIVEG)	Yes	CD	CD	No	Yes	Yes	Yes	Yes	Yes	Yes	CD	Yes	Yes	Yes	**Fair**
Stachowska, 2005[124]	Yes	NR	NR	No	NR	Yes	Yes	Yes	NR	Yes	Yes	NR	CD	Yes	**Poor**
Thomazella, 2011[125]	No	N/A	N/A	No	NR	No	Yes	Yes	Yes	Yes	Yes	NR	NR	Yes	**Poor**
Tutino, 2018[126] (NUTRIATT)	Yes	Yes	NR	No	Yes	Yes	Yes	Yes	NR	Yes	Yes	NR	Yes	Yes	**Poor**
Tuttle, 2008[127] (THIS-DIET)	Yes	NR	Yes	No	CD	Yes	Yes	Yes	Yes	Yes	Yes	Yes	NR	Yes	**Fair**
Vincent-Baudry, 2005[128] (Medi-RIVAGE)	Yes	Yes	NR	No	NR	Yes	No	No	CD	Yes	Yes	Yes	NR	No	**Poor**

Abbreviations: N/A – not applicable; NR – not reported; CD – could not determine; HECCPS – Healthy Eating for Colon Cancer Prevention Study; PCUC – Pontifical Catholic University of Chile

Table 2.3 Risk of bias assessment of pre-post studies without a control group – National Heart, Lung, and Blood Institute Quality Assessment Tool

Item	Barona, 2012[91]	Bihuniak, 2016[92]	Richard, 2012[121]	Sotos-Prieto, 2019[65]
Objective clear	Yes	Yes	Yes	Yes
Eligibility pre-specified	No	Yes	Yes	Yes
Representative	NR	No	CD	Yes
All eligible enrolled	NR	Yes	NR	No
Sample size adequate	No	Yes	Yes	NR
Intervention clearly described	No	Yes	Yes	Yes
Valid outcome assessment	Yes	NR	Yes	Yes
Assessors blinded	NR	NR	NR	NR
Dropout <20%	No	No	Yes	Yes
Pre-post statistical methods	Yes	Yes	Yes	Yes
Repeat follow-up measures	No	No	No	Yes
Overall rating	**Poor**	**Fair**	**Fair**	**Fair**

Abbreviations: N/A – not applicable; NR – not reported; CD – could not determine

The domain of accounting for group interventions in the analysis is not shown as it was not applicable to any of the trials.

2.4.4 Carotenoids, vitamins A and C, folate, tocopherols and phenolic compounds

The post-intervention SMDs between the Mediterranean and control diets in carotenoids, vitamins A and C, folate, tocopherols and phenolic compounds are shown in **Table 2.4**. Circulating biomarkers were measured in all studies, predominantly in plasma. Additionally, one study measured carotenoids and tocopherols in colonic tissue which yielded mostly similar SMD values as those obtained from serum.[102,104]

Circulating total carotenoids were statistically significantly increased in two out of three studies, α-carotene and β-cryptoxanthin in one out of five, β-carotene in four out of ten, and lutein and zeaxanthin in one out of four each (**Table 2.4**). Lycopene was significantly increased in two and decreased in one out of seven studies. Effect sizes for these results were mostly moderate or large (absolute SMD range: 0.29, 1.46). The majority of point estimates for carotenoids were positive among studies which did not report statistically significant results. Mostly negative values were reported by one trial which used a control diet based on dietary guidelines, with a large overlap of dietary goals with the Mediterranean intervention.[102,104] Neither of the trials reported increased vitamin A concentrations, however, the point estimates were consistently positive with small to moderate effect sizes (SDM range: 0.08, 0.39). Two studies reported very large effect sizes for plasma vitamin C (SMDs of 1.82 and 1.59), while three reported statistically null results with point estimates close to zero (SMD range: -0.04, 0.24). Increased circulating folate was reported by one out of four trials with a large effect size (SMD = 1.02), and homocysteine by one out of three trials (SMD = 0.27). In contrast to biomarkers described thus far, point estimates for tocopherols generally tended to have negative values. Statistically significant reductions in plasma α-tocopherol were reported in two out of ten, and two out of four studies for γ-tocopherol with moderate to large effect sizes (SMD range: -1.03, -0.23; **Table 2.4**).

Large, statistically significant effect sizes (SMD > 0.9) were reported almost exclusively by two parallel trials with limited sample sizes (total n ≤50),[112,113,122] one of which was only partially-randomised,[112,113] and a trial which was eligible for inclusion only as a pre-post study without a control group (n =15).[91]

Table 2.4 Standardised mean differences in biomarkers of carotenoids, vitamins A and C, folate, homocysteine and tocopherols after Mediterranean diet interventions relative to control diets

Study (first author, year or study name)	Control diet	Nt/Nc	Tissue	Time, mo.	Total carotenoids	α-carotene	β-carotene	β-cryptoxanthin	Lycopene	Lutein	Zeaxanthin	Vitamin A	Vitamin C	Folate	Homocysteine	α-tocopherol	γ-tocopherol
Barona, 2012[91]	N/A	15/-	Plasma	0.5			0.91‡		-0.39*	0.61‡	0.40						
Davis, 2017[64]	Habitual	70/66	Serum	6	0.38*	0.44	0.48‡	0.10	0.51†								
Djuric, 2009[83]	Habitual	27/33	Plasma	6	0.53*	0.49*	0.68*	1.46*	0.25	-0.08	0.29*					0.29	-0.36*
Hagfors, 2003[99]	Habitual	26/23	Plasma	3			0.20		0.64*			0.17	-0.04			-0.11	-0.25
HECCPS[102,104]	Healthy	47/47	Serum	6	-0.13	-0.68	-0.04	0.03	-0.18	0.15	0.03					-0.02	-0.36
			Colon			-0.52*	-0.17*	-0.11	-0.43	-0.41	-0.27					0.07	-0.24
Itsiopoulos, 2011[88]	Habitual	27 cross.	Plasma	3		0.17		-0.06				0.19			0.27*	-0.11	-0.49*
		26 cross.	Plasma				0.20		0.51†								
Jula, 2002[107]	Habitual	60/60	Plasma	6			0.11						0.03		0.14	-0.23*	
			Erythr.											0.02			
Lyon Diet Heart Study[61]	Habitual	122/128	Plasma	12								0.18	0.24			0.28	
Marin, 2011[90]	Lower fat	20 cross.	Plasma	1			0.46									-0.03	
Parcina, 2015[111]	Habitual	14/13	Plasma	0.5								0.39		0.45			
PCUC study[112,113]	High-SFA	21/21	Plasma	2			1.40‡						1.82‡			-1.03†	
				1										1.02†			
Skouroliakou, 2018[122]	Healthy	26/24	Serum	6								0.08	1.59*			0.56	
Sofi, 2018[123]	Healthy	103 cross	Serum	3										0.05			
Stachowska, 2005[124]	Lower fat	21/16	Plasma	6												-0.82	
Vincent-Baudry, 2005[128]	Lower fat	88/81	Plasma	3		0.14	0.22	0.07	0.47	-0.19	0.00				0.07		

Abbreviations: cross. – crossover; HECCPS – Healthy Eating for Colon Cancer Prevention Study; Nt/Nc – numbers of participants in the intervention/control group; mo. – months; N/A – not applicable; PCUC – Pontifical Catholic University of Chile

Blue colour indicates negative values, and red colour indicates positive values with intensity proportional to effect size. Results for Barona, 2012 were estimated relative to baseline values. Study names were used for studies reporting across multiple publications in the current review. Total carotenoids for Davis, 2017 were calculated from raw data deposited with the publication as the sum of α-carotene, β-carotene, β-cryptoxanthin, lycopene, and the sum of lutein and zeaxanthin.

* p difference < 0.05; † p difference < 0.01; ‡ p difference < 0.001

2.4.5 Biomarkers of other vitamins, minerals, phytosterols and phenolic compounds

Twenty plasma or serum biomarkers of vitamins, minerals, phytosterols and phenolic compounds were reported on by ≤2 studies: cyclolycopene, vitamins B_1, B_6 and B_{12}, homocysteine, vitamins E and 25(OH) D, calcium, ferritin, haematocrit, haemoglobin, iron, magnesium, potassium, sodium, selenium, zinc, the sum of campesterol and β-sitosterol, lathosterol and total phenolic compounds. Out of the 23 available estimates (**Table 2.5**), statistically significant differences after interventions with Mediterranean diet relative to control diets were reported for vitamin B_{12} (SMD = 0.24 relative to a healthy vegetarian diet[123]), lathosterol (SMD = -0.31 relative to participants' habitual diet in Canada[121]), and phenolic compounds in one[128] out of two studies[115,128] (SMD = 0.44 relative to a lower fat diet[128]). Six of the null estimates were contributed by a single study with duration of two weeks and a sample size of 27 participants for the relevant interventions, which experienced a deviation in the protocol of dietary intervention and isocalorically increased energy intakes in the second week (**Table 2.1**).[111]

For 24-hour urinary excretion of cations, there was no evidence of effects of Mediterranean diet interventions on calcium, potassium and sodium (**Table 2.5**). One study reported increased excretion of magnesium with Mediterranean diet compared to continuation of participants' habitual diet in Australia (SMD = 0.37).[64]

Among studies reporting on urinary excretion of polyphenols in spot urine samples, only subsamples of the PREDIMED trial reported positive effects of Mediterranean diet interventions relative to a lower fat diet on hydroxytyrosol (SMD = 0.18) and total phenolic compounds with a very large effect size (SMD = 1.85; **Table 2.5**). Two trials with small sample sizes reported null effects on hydroxytyrosol (n = 16) and tyrosol (n = 9)[65], and total phenolic compounds (n = 42).[119]

Table 2.5 Standardised mean differences in circulating and urinary biomarkers of vitamins, minerals, phytosterols and phenolic compunds after Mediterranean diet interventions relative to control diets reported by ≤ 2 studies

Biomarker	Study (first author, year or study name)	Control diet	Nt/Nc	Time, mo.	SMD
Plasma or serum					
Vitamins and vitamin-related					
Cyclolycopene	Djuric, 2009[83]	Habitual	27/33	6	0.27
Vitamin B_1	Parcina, 2015[111]	Habitual	14/13	0.5	-0.38
Vitamin B_6					0.02
Vitamin B_{12}	Sofi, 2018[123]	Healthy	103 cross.	3	0.24*
	Vincent-Baudry, 2005[128]	Lower fat	88/81	3	0.08
Homocysteine	Itsiopoulos, 2011[88]	Habitual	27 cross.	3	0.27*
	Parcina, 2015[111]	Habitual	14/13	0.5	0.07
Vitamin E					-0.21
Vitamin 25(OH) D					0.51
Minerals and mineral-related					
Calcium	Sofi, 2018[123]	Healthy	103 cross.	3	0.01
	PREDIMED[116]	Lower fat	143/59	12	0.02
Ferritin	Sofi, 2018[123]	Healthy	103 cross.	3	0.05
Haematocrit					0.09
Haemoglobin					0.10
Iron					-0.03
Magnesium	Parcina, 2015[111]	Habitual	14/13	0.5	0.28
	Sofi, 2018[123]	Healthy	103 cross.	3	0.01
Potassium					0.08
Sodium					-0.11
Selenium	Parcina, 2015[111]	Habitual	14/13	0.5	-0.02
Zinc					-0.65
Phytosterols					
Campesterol + β-sitosterol	Richard, 2012[121]	N/A	19/-	1	0.03
Lathosterol					-0.31*
Phenolic compounds	PCUC study[115]	High-SFA	21/21	2	0.09
	Vincent-Baudry, 2005[128]	Lower fat	88/81	3	0.44*
Urinary					
24 h excretion of minerals					
Calcium/creatinine	Perticone, 2019[97]	Lower fat	143/59	12	-0.09
Calcium	Davis, 2017[64]	Habitual	70/65	6	0.07
Magnesium					0.37*
Potassium	Davis, 2017[64]	Habitual	70/65	6	0.20
	NU-AGE[109]	Habitual	407/424	12	0.13
Sodium	Davis, 2017[64]	Habitual	70/65	6	-0.10
	NU-AGE[109]	Habitual	407/424	12	-0.04
Polyphenols (spot urine)					
Hydroxytyrosol	PREDIMED[117]	Lower fat	500/250	60	0.18*
	Sotos-Prieto, 2019[65]	N/A	16/-	6	-0.16
Tyrosol	Sotos-Prieto, 2019[65]	N/A	9/-	6	0.22
Total phenolic compounds	PCUC study[115]	High-SFA	21/21	1	0.06
	PREDIMED[119]	Lower fat	131/69	12	1.85‡

Abbreviations: cross – crossover; mo. – months; N/A – not applicable; Nt/Nc – numbers of participants in the intervention/control group; PREDIMED – Prevención con Dieta Mediterránea; PCUC – Pontifical Catholic University of Chile; SFA – saturated fatty acids; SMD – standardised mean difference

Blue colour indicates negative values, and red colour indicates positive values with intensity proportional to effect size. Results for studies without control diets were estimated relative to baseline values. Study names were used for studies reporting across multiple publications in the current review.

* p difference < 0.05; † p difference < 0.01; ‡ p difference < 0.001

2.4.6 Saturated, mono- and trans-unsaturated fatty acids

The systematic review identified 12 saturated, mono- or trans-unsaturated fatty acids or fatty acid groupings which were reported on by more than two studies (**Table 2.6**). Up to 19 comparisons of Mediterranean and control diets from 14 trials were available for C18:1n-9c, and C16:1n-9c and C18:1n-9t was reported on by two studies each. Mostly positive point estimates of SMDs for Mediterranean relative to control diets were reported for total MUFA and C18:1n-9c, whereas mostly negative point estimates were reported for the remaining individual cis-MUFAs and total and individual SFAs (**Table 2.6**). Point estimates for trans-fatty acids (total and C18:1n-9t) were directionally inconsistent and neither of the trials observed a statistically significant effect of Mediterranean diet on these fatty acids.

Three studies measured the same fatty acids in multiple blood fractions or tissues (**Table 2.6**). Djuric et al.[83] reported statistically significantly higher C18:1n9c concentrations in plasma (SMD = 1.09), but not in plasma phospholipids following six months of active counselling on Mediterranean diet compared to baseline provision of written materials on dietary guidelines. Subsamples of the PREDIMED trial of unknown overlap reported on plasma fatty acids after 12 months of intervention[81], and on fatty acids in phospholipids and triglycerides of very low-density lipoproteins (VLDL) after three months.[82] Among fatty acids measured in the three fractions, C14:0 was not affected by the Mediterranean diet intervention in plasma and VLDL phospholipids, but it was increased in VLDL triglycerides (SDM = 0.99; **Table 2.6**). C16:0 in plasma was also not affected by the intervention, whereas a negative response in both VLDL fractions was observed (SMDs of -0.81 and -0.64, respectively). By contrast, C18:1n-9c was increased in plasma (SMD = 0.39), but not in VLDL. C16:1n-7c was not measured or reported on from plasma, and it was statistically significantly decreased in VLDL phospholipids only (SMD = -0.72; **Table 2.6**). In the Healthy Eating for Colon Cancer Prevention study, the results were null and materially similar between assays of plasma and colonic tissue for total SFA and MUFA (**Table 2.6**).[103] Where available, only plasma measurements from the above results are considered in the next paragraph.

Total SFAs were statistically significantly decreased in three out of 12 Mediterranean-control diet comparisons (SMD range: -0.94, -0.35), and individual SFAs as follows: C14:0 in one out of five (SMD = -0.88), C16:0 in three out of 12 (negative SMD range: -0.81, 0.64) with one study reporting an increase (SMD = 0.34), C17:0 in one out of three (SMD = -0.55), and C18:0 in one out of 12 (SMD = -0.41; **Table 2.6**). Total MUFAs were statistically significantly

increased in three out of nine Mediterranean-control diet comparisons (SMD range: -0.94, -0.35), and one study reported a decrease (SMD = -0.77). Eight trials reported on C16:1n-7c including one positive (SMD = 0.28) and one negative result (SMD = -0.72), and two trials reported on small decreases in C16:1n-9c, one of which was statistically significant (SMD = -0.11). Three trials reported no effect of Mediterranean diet on C18:1n-7c. C18:1n-9c was statistically significantly and substantially higher in Mediterranean versus control diet arms in four out of 14 comparisons (SMD range: 0.80, 1.96; **Table 2.6**). One comparison from Italy reported a negative effect (SMD = -0.82) against a diet based on dietary guidelines using a limited number of participants (n = 33), which may have been in part affected by baseline imbalance in the outcome or loss to follow-up.[126] A second comparison from the same trial of Mediterranean diet with continuation of habitual diet and background physical activity programmes in both dietary interventions (n = 75) yielded a positive point estimate (SMD = 0.31) without formal statistical evidence to suggest an increase in C18:1n-9c (**Table 2.6**).

Very large point estimates (absolute SMDs >1.3) were reported exclusively by two trials which measured fatty acids in LDL cholesteryl esters[89] and triglycerides.[124] However, one other study which assayed fatty acids in LDL cholesteryl esters yielded SMDs of similar magnitude as the remainder of the trials (**Table 2.6**).[87]

2.4.7 Polyunsaturated fatty acids

The systematic review identified 13 polyunsaturated fatty acids or fatty acid groupings and the ratio of n-6:n-3 which were reported on by more than two studies (**Table 2.7**). Up to 20 comparisons of Mediterranean and control diets from 14 trials were available for C20:4n-6 and C20:5n-3 each, and C22:4n-6 was reported on by two studies. Overall, the point estimates of SMDs for Mediterranean relative to control diets were highly heterogeneous, and negative values tended to be reported for n-6 and positive for n-3 fatty acids (**Table 2.7**).

The same three studies measured fatty acids in multiple blood fractions or tissues as previously described for SFAs and MUFAs (**Table 2.7**). Djuric et al.[83] reported statistically significantly lower C18:2n-6 concentrations in plasma (SMD = -0.48), but not in plasma phospholipids. The results for C18:3n-3, C20:4n-6 and C20:5n-3 were null in both fractions. The subsamples of the PREDIMED trial reported no effect Mediterranean diet on plasma or VLDL triglyceride C18:2n-6 and a moderate increase in VLDL phospholipids (SMD = 0.65). C18:3n-3 was increased in both plasma (SMD = 0.48) and VLDL phospholipids (SMD = 1.09), but not in

VLDL triglycerides. C20:4n-6 was decreased in plasma (SMD = -0.17) but not in the VLDL fractions, whereas C20:5n-3 was decreased in VLDL phospholipids (SMD = -0.75) but not in plasma (not assayed in VLDL triglycerides). There were no statistically significant effects detected in either fraction for C20:3n-6 and C22:6n-3. (**Table 2.7**). In the Healthy Eating for Colon Cancer Prevention study, the results were null and materially similar between assays of plasma and colonic tissue for total n-6 and n-3 fatty acids, n-6:n-3 ratio, C18:3n-3 and C20:4n-6, whereas C18:2n-6 was increased in plasma only (SMD = -0.44; **Table 2.7**).[103] Where available, only plasma measurements from the above results are considered in the next paragraph.

Total PUFAs were statistically significantly decreased in two out of eight Mediterranean-control diet comparisons (SMDs of -0.51 and -1.02), total n-6 fatty acids in two out of nine (SMDs of -0.32 and -1.65), total n-3 fatty acids were increased in two out of ten (SMDs of 0.93 and 1.26), and the n-6:n-3 ratio was decreased in four out of 11 comparisons (SMD range: -1.87, -0.56; **Table 2.7**). C18:2n-6 was statistically significantly decreased in seven out of 18 Mediterranean-control diet comparisons (SMD range: -2.21, -0.30), and two studies reported an increase (SMDs of 0.65 and 0.86). In turn, C18:3n-3 was statistically significantly increased in five out of 16 trials (SMD range: 0.40, 1.09). For C18:3n-6, one out of three studies reported a negative effect (SMD = -0.33), and one out of three detected an increase in C20:2n-6 (SMD = 0.48), whereas there were no statistically significant results or consistency in directionality of point estimates for C20:3n-6. There were two negative results among 16 comparisons of C20:4n-6 (SMDs of -0.30 and -0.17). Among long chain n-3 fatty acids, five out of 17 results were positive for C20:5n-3 (SMD range: 0.36, 1.18) and two out of 13 were positive for C22:6n-3. Results for C22:5n-3 were directionally inconsistent with one trial reporting a negative effect (SMD = -0.37) and another a positive effect (SMD = 0.32) out of seven Mediterranean-control diet comparisons (**Table 2.7**).

Table 2.6 Standardised mean differences in biomarkers of saturated, mono- and trans-unsaturated fatty acids after Mediterranean diet interventions relative to control diets

Study (first author, year or study name)	Control diet	Nt/Nc	Blood fraction or tissue	Time, mo.	SFA	C14:0	C16:0	C17:0	C18:0	MUFA	C16:1n-7c	C16:1n-9c	C18:1n-7c	C18:1n-9c	TFA	C18:1n-9t
Bihuniak, 2016[92]	N/A	16/-	Plasma	6		-0.88†	-0.73‡	-0.55†	0.12							
Bemelmans, 2002[98]	Habitual	97/154	CE	12		0.04	0.34‡		0.05		0.28*			0.13		
Davis, 2017[64]	Habitual	69/65	Erythr.	6	-0.79‡					0.76‡					-0.26	
Djuric, 2009[83]	Habitual	27/33	PL	6	-0.35					0.44				0.46		
			Plasma											1.09*		
Fuentes, 2008[87]	Lower fat	20 cross.	LDL-CE	1		0.14	0.28		-0.26		-0.36			0.89		
Hagfors, 2005[100]	Habitual	26/25	PL	6	-0.28		-0.30		-0.42	-0.35	-0.12		0.09	-0.44		
HECCPS[103]	Healthy	47/47	Plasma	6	-0.20					0.51						
			Colon		0.00					0.44						
Itsiopoulos, 2011[88]	Habitual	27 cross.	PL	3	-0.35*					0.85‡					-0.46	
Lyon Diet Heart Study[62,63]	Habitual	219/213	Plasma	2	-0.17		-0.06		-0.41‡					0.80‡	0.30	0.19
Muzsik, 2019[108]	Healthy	59/58	Erythr.	4	-0.05	-0.31	-0.08	0.04	-0.03	-0.02	-0.27		-0.17	-0.06	0.03	-0.19
NU-AGE[110]	Habitual	70/70	Plasma	12			-0.08		0.09		-0.06			-0.01		
PCUC study[114]	High-SFA	21/21	Plasma	2	0.04					0.92†						
Pérez-Jiménez, 2001[89]	Lower fat	59 cross.	LDL-CE	1			-1.36		0.16		-1.56			1.55*		
PREDIMED[81,82]	Lower fat	283/141	Plasma	12		-0.05	-0.10	-0.01	-0.11					0.39‡		
		35/15	VLDL-PL	3		-0.10	-0.81*		-0.44		-0.72*		0.00	0.20		
			VLDL-TG			0.99†	-0.64*		-0.03		-0.25	-0.29	-0.09	0.16		
Sotos-Prieto, 2019[65]	N/A	20/-	Plasma	6	-0.94†									0.07		
Stachowska, 2005[124]	Lower fat	21/16	TG	6										1.96‡		
Tutino, 2018[126]	Healthy	18/15	Erythr.	3	0.64		0.62		0.70	-0.77*				-0.82*		
	Habitual	46/29			-0.22		-0.18		0.09	0.35				0.31		
Tuttle, 2008[127]	Lower fat	47/43	Plasma	6	0.15						-0.25			-0.17		
Vincent-Baudry, 2005[128]	Lower fat	88/81	Plasma	3			-0.13		-0.32			-0.11‡		0.33		

Abbreviations: CE – cholesteryl esters; cross. – crossover; Erythr. – erythrocytes; HECCPS – Healthy Eating for Colon Cancer Prevention Study; LDL – low-density lipoprotein; mo. – months; MUFA – monounsaturated fatty acids; N/A – not applicable; Nt/Nc – numbers of participants in the intervention/control group; PCUC – Pontifical Catholic University of Chile; PL – phospholipids; PREDIMED – Prevención con Dieta Mediterránea; SFA – saturated fatty acids; TFA – trans-saturated fatty acids; TG – triglycerides; VLDL – very low-density lipoprotein

Blue colour indicates negative values, and red colour indicates positive values with intensity proportional to effect size. Results for studies without control diets were estimated relative to baseline values. Study names were used for studies reporting across multiple publications in the current review.

* p difference < 0.05; † p difference < 0.01; ‡ p difference < 0.001

Table 2.7 Standardised mean differences in biomarkers of polyunsaturated fatty acids after Mediterranean diet interventions relative to control diets

Study (first author, year or study name)	Control diet	Nt/Nc	Blood fraction or tissue	Time, mo.	PUFA	n-6	n-3	n-6:n-3	C18:2n-6	C18:3n-3	C18:3n-6	C20:2n-6	C20:3n-6	C20:4n-6	C20:5n-3	C22:5n-3	C22:6n-3
Bihuniak, 2016[92]	N/A	16/-	Plasma	6				-0.85*	0.86‡	0.87‡		0.00		-0.23	0.59*	-0.37*	1.19‡
Bemelmans, 2002[98]	Habitual	97/154	CE	12				-0.56*	-0.30*	0.40‡				-0.30*	0.36†		0.24
Davis, 2017[64]	Habitual	69/65	Erythr.	6	0.05	-0.05	0.09	-0.20	0.18	0.13				-0.15	0.00	-0.23	0.31
Djuric, 2009[83]	Habitual	27/33	PL	6	-0.02				-0.56	-0.06				0.24	-0.22		0.00
			Plasma	6					-0.48*	0.16				-0.25	-0.36		-0.08
Fuentes, 2008[87]	Lower fat	20 cross.	LDL-CE	1					-1.62*	-0.49				-0.34	0.09		
Hagfors, 2005[100]	Habitual	26/25	PL	3	0.00	-0.32*	0.93*	-1.87*		0.00	-0.33*	-0.12		-0.14	1.01‡	0.28	0.56‡
HECCPS[103,104]	Healthy	47/43	Plasma	6		-0.30	0.08	-0.03	-0.44*	-0.18				0.16			
			Colon			0.00	-0.22	0.03	-0.04	-0.09				-0.07			
Hjerkinn, 2006[105]	Habitual	58/56	Serum	36		0.27	0.12	-0.07	0.34	0.22				-0.17	0.08		0.09
Itsiopoulos, 2011[88]	Habitual	27 cross.	PL	3	-0.51*	-0.32	-0.10	0.00	-0.21*	-0.30‡					-0.13*		
Lyon Diet Heart Study[62,63]	Habitual	219/213	Plasma	2					-0.45	1.07‡				-0.34	0.42‡	0.32‡	0.17
Muzsik, 2019[108]	Healthy	59/58	Erythr.	4	0.05	0.04	0.05	-0.17	-0.06	-0.23	-0.17	0.48*	0.22	0.02	0.04	0.10	0.03
NU-AGE[110]	Habitual	70/70	Plasma	12		0.00	0.17	-0.03	-0.03	0.21			0.07	0.10	0.21	0.01	0.17
PCUC study[113,114]	High-SFA	21/21	Plasma	1-2	-1.02†	-1.65‡	1.26‡	-1.37‡							1.18‡		
Pérez-Jiménez, 2001[89]	Lower fat	59 cross.	LDL-Ce	1					-0.61*								
PREDIMED[81,82]	Lower fat	283/141	Plasma	12					-0.18	0.48‡	-0.07	0.00	-0.28	-0.17†	0.20		-0.20
		35/15	VLDL-PL	3					0.65*	1.09‡			0.31	0.43	-0.75*	-0.34	0.19
			VLDL-TG						0.22	0.32			-0.66	0.30			
Sotos-Prieto, 2019[65]	N/A	20/-	Plasma	6			0.20		0.39	0.20							
Stachowska, 2005[124]	Lower fat	21/16	TG	6					-2.21‡								
Tutino, 2018[126]	Healthy	18/15	Erythr.	3	-0.22								-0.39	-0.27	-0.16		0.00
	Habitual	46/29			-0.17								0.18	-0.42	-0.05		0.25
Tuttle, 2008[127]	Lower fat	47/43	Plasma	6		-0.15	0.23	-0.18	-0.11	0.00				0.14	0.17		0.24
Vincent-Baudry, 2005[128]	Lower fat	88/81	Plasma	3					-0.19	-0.07			-0.15	-0.06	0.19		0.36

Abbreviations: CE – cholesteryl esters; cross. – crossover; Erythr. – erythrocytes; HECCPS – Healthy Eating for Colon Cancer Prevention Study; LDL – low-density lipoprotein; mo. – months; MUFA – monounsaturated fatty acids; N/A – not applicable; Nt/Nc – numbers of participants in the intervention/control group; PCUC – Pontifical Catholic University of Chile; PL – phospholipids; PREDIMED – Prevención con Dieta Mediterránea; SFA – saturated fatty acids; TFA – trans-saturated fatty acids; TG – triglycerides; VLDL – very low-density lipoprotein

Blue colour indicates negative values, and red colour indicates positive values with intensity proportional to effect size. Results for studies without control diets were estimated relative to baseline values. Study names were used for studies reporting across multiple publications in the current review.
* p difference < 0.05; † p difference < 0.01; ‡ p difference < 0.001

2.4.8 Miscellaneous fatty acids reported on by ≤ 2 studies

The systematic review identified results from Mediterranean diet trials on further 23 circulating fatty acids and ten fatty acid ratios (**Table 2.8**). Of these, 17 were reported on as null results exclusively by one weight-loss parallel RCT from Poland which compared the effects of Mediterranean diet with a dietary guidelines-based diet on absolute concentrations of erythrocyte fatty acids.[108] It detected an approximately two-fold decrease in total erythrocyte fatty acids from baseline within each arm, and a corresponding effect on most individual fatty acids. However, it found little difference in post-intervention fatty acid profiles between the groups, reporting three statistically significant effects across 48 comparisons.[108] Among the remaining fatty acids identified by the search, the Mediterranean diet in the Lyon Diet Heart Study RCT affected post-intervention C16:1n-7t (SMD = 0.37), total n-3 excluding C18:3n-3 (SMD = 0.33) and total n-6 excluding C18:2n-6 (SMD = -0.47) relative to lack of dietary intervention beyond usual care (**Table 2.8**). Additionally, Bihuniak et al.[92] found decreased C22:4n-6 (SMD = -0.96) and the (n-6+SFA):n-3 ratio (SMD = -1.08) in a pre-post study without a control group. There was no evidence of an effect of Mediterranean diet on the remainder of the multiple fatty acids presented in **Table 2.8**.

Table 2.8 Standardised mean differences in biomarkers of fatty acids after Mediterranean diet interventions relative to control diets reported by ≤ 2 studies

Biomarker	Study (first author, year or study name)	Control diet	Nt/Nc	Blood fraction	Time, mo.	SMD
C14:1n-5	Muzsik, 2019[108]	Healthy	59/58	Erythr.	4	0.04
C14:1n-9						-0.02
C15:0						0.06*
C15:1						0.08
C16:1n-7t	Lyon Diet Heart Study[62]	Habitual	219/213	Plasma	2	0.37‡
C17:1	Muzsik, 2019[108]	Healthy	59/58	Erythr.	4	0.21
C18:1n-5c						-0.25
C18:1n-7t						0.19
C18:2n-6t						0.18
C18:3n-3t						0.24
C18:4	Vincent-Baudry, 2005[128]	Lower fat	88/81	Plasma	3	0.00
C20:0	Hagfors, 2005[100]	Habitual	26/25	PL	3	0.16
C20:3n-3	Muzsik, 2019[108]	Healthy	59/58	Erythr.	4	-0.05
C20:3n-9						0.03
C20:6n-3	Fuentes, 2008[87]	Lower fat	20 cross.	LDL-CE	1	0.24
	Itsiopoulos, 2011[88]	Habitual	27 cross.	PL	3	0.07
C22:0	Muzsik, 2019[108]	Healthy	59/58	Erythr.	4	0.17
C22:1n-11	Hagfors, 2005[100]	Habitual	26/25	PL	3	0.68
C22:1n-9	Muzsik, 2019[108]	Healthy	59/58	Erythr.	4	0.01
C22:2n-6						0.13
C22:4n-6	Bihuniak, 2016[92]	N/A	16/-	Plasma	6	-0.96‡
	Muzsik, 2019[108]	Healthy	59/58	Erythr.	4	0.07
C22:5n-6						0.11
C24:0	Hagfors, 2005[100]	Habitual	26/25	PL	3	0.35
C24:1n-9	Muzsik, 2019[108]	Healthy	59/58	Erythr.	4	0.18
(PUFA+MUFA):SFA						0.16
n-3 exc. C18:3n-3	Lyon Diet Heart Study[62]	Habitual	219/213	Plasma	2	0.33‡
n-6 exc. C18:2n-6						-0.47‡
C18:2n-6:C18:3n-6	Muzsik, 2019[108]	Healthy	59/58	Erythr.	4	0.26
	Bemelmans, 2002[98]	Habitual	97/154	CE	12	-0.38
C20:4n-6:C20:5n-3	NU-AGE[110]	Habitual	70/70	Plasma	12	0.14
	Tutino, 2018[126]	Healthy	18/15	Erythr.	3	-0.08
		Habitual	46/29			-0.34
C20:5n-3:C22:4n-6	Muzsik, 2019[108]	Healthy	59/58		4	0.27
C22:6n-3:C22:4n-6						0.18
(EPA+DHA):C22:4n-6						0.19
PUFA:MUFA						0.14
PUFA:SFA	Michalsen, 2006[53]	Habitual	48/53	Plasma	12	0.04
(n-6+SFA):n-3	Bihuniak, 2016[92]	N/A	16/-	Plasma	6	-1.08†

Abbreviations: CE – cholesteryl esters; Erythr. – erythrocytes; EPA – eicosapentaenoic acid (C20:5n-3); DHA – docosahexaenoic acid (C22:6n-3); LDL – low-density lipoprotein; mo. – months; MUFA – monounsaturated fatty acids; N/A – not applicable; Nt/Nc – numbers of participants in the intervention/control group; PL – phospholipids; SFA – saturated fatty acids; SMD – standardised mean difference

Blue colour indicates negative values, and red colour indicates positive values with intensity proportional to effect size. The SMD for Bihuniak, 2016 was estimated relative to baseline values. Study names were used for studies reporting across multiple publications in the current review.

* p difference < 0.05; † p difference < 0.01; ‡ p difference < 0.001

2.4.9 Amino acids and amino-acid related analytes

Four parallel RCTs compared the effects of Mediterranean and control diets on serum or plasma amino acids and related analytes, including arginine, branch-chained amino acids, cysteine, trimethylamine N-oxide (TMAO) and the dietary precursors of its gut microbial formation: choline, carnitine and betaine (**Table 2.9**). Among 29 results, only the oxidised form of cysteine (SMD = -0.84) and valine (SMD = -0.17) were found to be affected by Mediterranean diet, both of which were reported on by single studies. Null results for betaine, choline and TMAO were reported by two studies each (**Table 2.9**).

Table 2.9 Standardised mean differences in biomarkers of amino acids and amino acid-related analytes after Mediterranean diet interventions relative to control diets

Biomarker	Study (first author, year or study name)	Control diet	Nt/Nc	Time, mo.	SMD
Arginine	Thomazella, 2011[125]	Lower fat	20/19	3	
ADMA					-0.21
L-arginine					0.34
L-arginine/ADMA					0.41
Alphaglycerophosphocholine	PREDIMED[118]	Lower fat	677/303	12	0.00
Betaine					0.05
	HECCPS[101]	Healthy	45/45	6	0.05
Carnitine					0.30
Choline					0.15
	PREDIMED[118]	Lower fat	677/303	12	0.07
Cysteine	Jaacks, 2018[106]	Habitual	11/9	2	-1.46
Cystine					-0.84*
Isoleucine	PREDIMED[120]	Lower fat	632/267	12	-0.02
Leucine					-0.01
Phosphocholine	PREDIMED[118]	Lower fat	677/303	12	0.05
TMAO					0.00
	HECCPS[101]	Healthy	45/45	6	0.00
TMAO:betaine					0.00
TMAO:carnitine					-0.17
TMAO:choline					0.00
TMAO:γ-butyrobetaine					-0.05
Valine	PREDIMED[120]	Lower fat	632/267	12	-0.17*
γ-Butyrobetaine	HECCPS[101]	Healthy	45/45	6	0.25

Abbreviations: ADMA – asymmetric dimethylarginine; HECCPS – Healthy Eating for Colon Cancer Prevention Study; N/A – not applicable; Nt/Nc – numbers of participants in the intervention/control group; PREDIMED – Prevención con Dieta Mediterránea; SMD – standardised mean difference; TMAO – trimethylamine N-oxide

Blue colour indicates negative values, and red colour indicates positive values with intensity proportional to effect size. Study names were used for studies reporting across multiple publications in the current review.

* p difference < 0.05; †p difference < 0.01; ‡ p difference < 0.001

2.4.10 Publication bias of biomarkers eligible for meta-analysis

The extent of publication bias was heterogeneous across the 25 biomarkers eligible for meta-analysis (**Figures 2.2-2.4**). Major asymmetry of Doi plots with a right skew was detected for β-carotene (LFK index = 2.17), β-cryptoxanthin (LFK index = 2.81), vitamin C (LFK index = 2.83), and total n-3 fatty acids (LFK index = 2.41). Major left-skewed asymmetry was present for total PUFA (LFK index = -2.53) and the ratio of n-6:n-3 fatty acids (LFK index = -2.41). There was evidence of minor asymmetry with a right skew for total n-3 (LFK index = 1.79) and the fatty acids C18:0 (LFK index = 1.27), C22:6n-3 (LFK index = 1.37) and C20:4n-6 (LFK index = 1.95). Minor left-skewed asymmetry was detected for lycopene (LFK index = -1.99), C14:0 (LFK index = -1.94), C16:1n-7c (LFK index = -1.60) and C18:3n-3 (LFK index = -1.35). There was no evidence of Doi plot asymmetry for α-carotene, retinol, SFA, MUFA, C16:0. C18:1n-9c, C20:5n-3, total n-6 fatty acids, C18:2n-6, C20:3n-6 and C20:4n-6 (LFK index range: -0.99, 0.94).

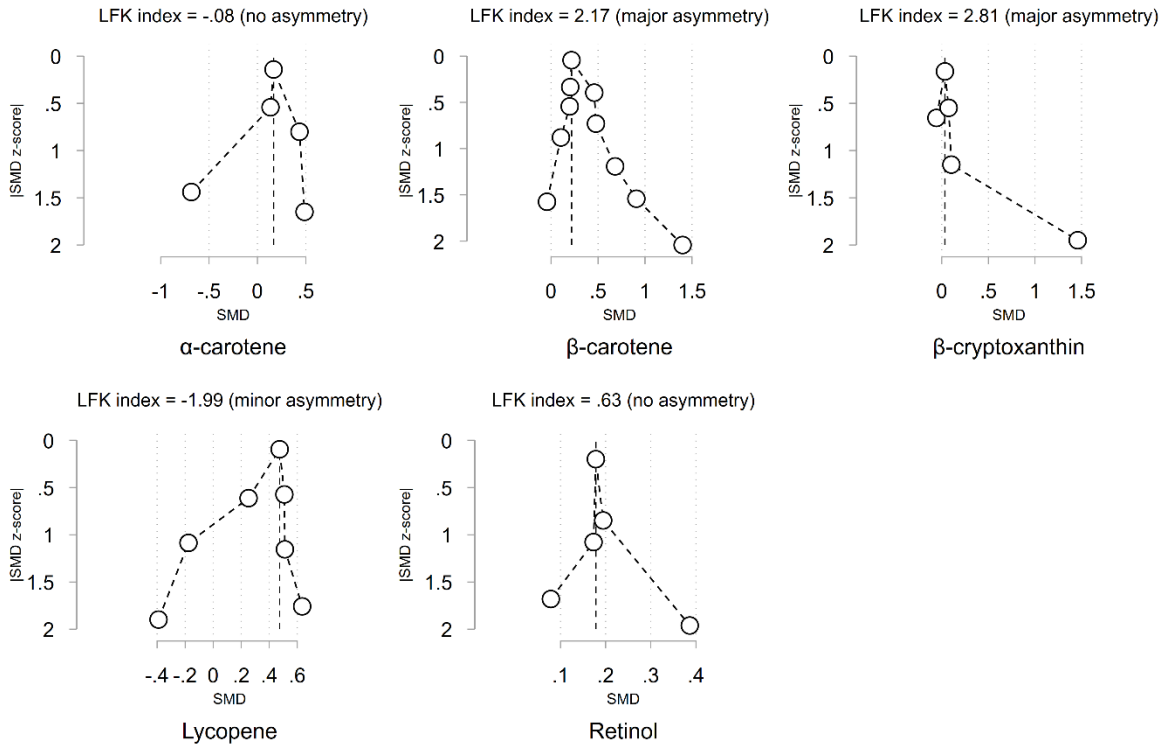


Figure 2.2 Doi plots and LFK indices of publication bias of the effects of the Mediterranean diet on circulating carotenoids and retinol

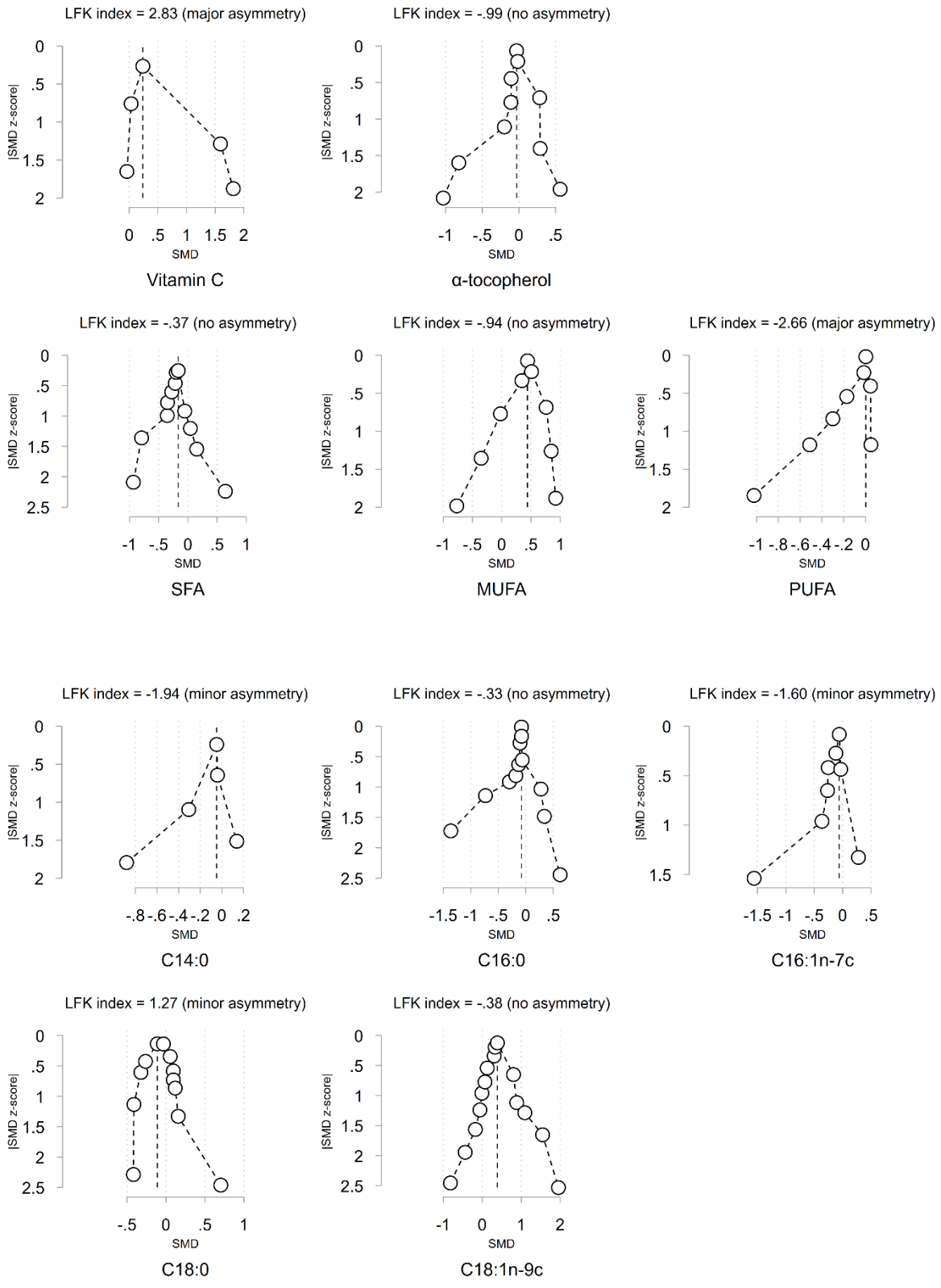


Figure 2.3 Doi plots and LFK indices of publication bias of the effects of the Mediterranean diet on circulating vitamin C, α-tocopherol, saturated, monounsaturated, and total polyunsaturated fatty acids

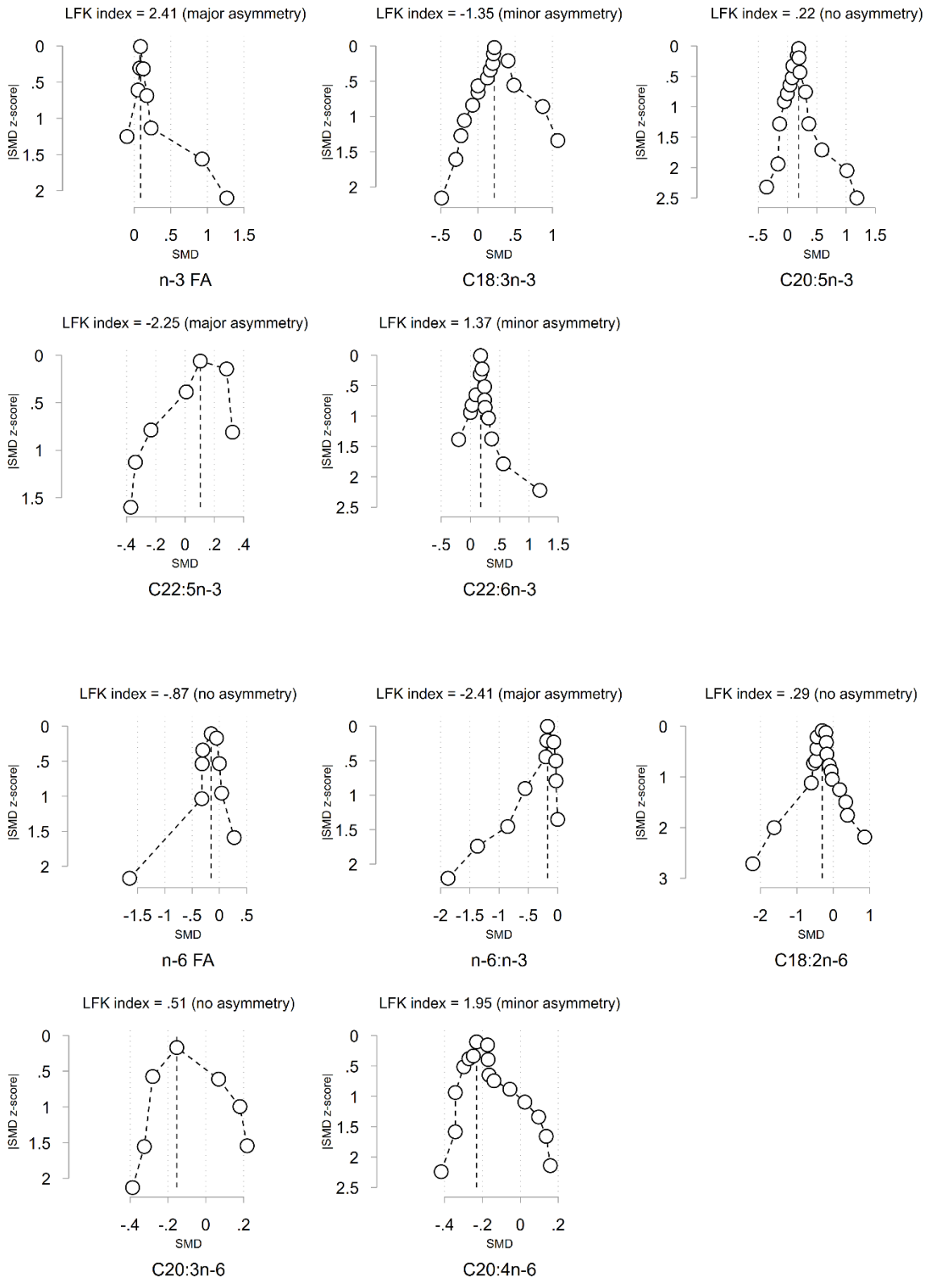


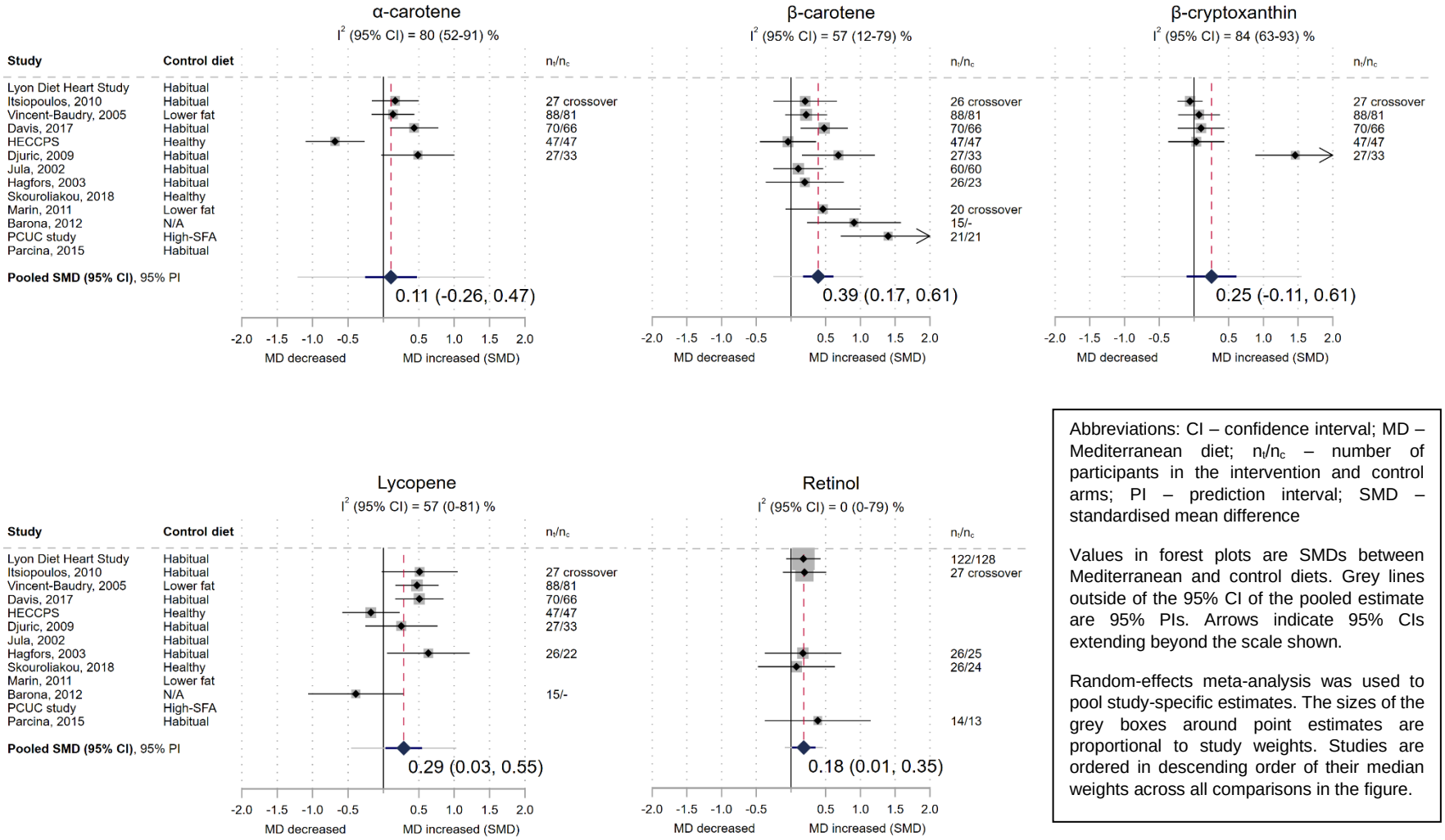
Figure 2.4 Doi plots and LFK indices of publication bias of the effects of the Mediterranean diet on circulating n-3 and n-6 fatty acids, and n-6:n-3 ratio

2.4.11 Meta-analysis

Thirteen nutritional biomarkers out of the 25 included in the meta-analysis statistically significantly differed between the Mediterranean and control diets post-intervention (**Figures 2.5-2.9**). Mediterranean diet increased circulating β-carotene, lycopene, retinol (**Figure 2.5**), vitamin C, total MUFA (**Figure 2.6**), C18:1n-9c (**Figure 2.7**), total n-3 fatty acids, and individual C20:5n-3 and C22:6n-3 fatty acids (**Figure 2.8**). The effect sizes ranged from SMD (95% CI) of 0.18 (0.00, 0.36) for retinol to 0.67 (0.06, 1.27) for vitamin C. Mediterranean diet decreased circulating total SFA (**Figure 2.6**), n-6:n-3 ratio, and C18:2n-6 and C20:4n-6 fatty acids (**Figure 2.9**), with effect sizes ranging from SMD (95% CI) of -0.18 (-0.26, -0.09) for C20:4n-6 to -0.41 (-0.67, -0.15) for the n-6:n-3 ratio. Among the above biomarkers only C20:4n-6 had a 95% prediction interval which did not include the null (-0.30, -0.06), suggesting a high degree of reproducibility.

Pooled estimates were compatible with both increased and decreased circulating α-carotene, β-cryptoxanthin (**Figure 2.5**), α-tocopherol, total PUFA (**Figure 2.6**), C14:0, C16:0, C16:1n-7c. C18:0 (**Figure 2.7**), C18:3n-3, C22:5n-3 (**Figure 2.8**), and total n-6 and C20:3n-6 fatty acids (**Figure 2.9**) following Mediterranean diet interventions. The magnitude of point estimates was similar as for the biomarkers for which statistically significant results were detected in the cases of β-cryptoxanthin (SMD = 0.28, 95% CI: -0.13), total PUFA (SMD = -0.19, 95% CI: -0.42, 0.03), C14:0 (SMD = -0.24, 95% CI: -0.73, 0.25), C16:0 (SMD = -0.15, 95% CI: -0.41, 0.11), C16:1n-7c (SMD = -0.29, 95% CI: -0.69, 0.10), C18:3n-3 (SMD = 0.18 (-0.08, 0.44) and total n-6 fatty acids (SMD = -0.20, 95% CI: -0.45, 0.04).

The I^2 values were highly variable across the meta-analysed biomarkers and indicated moderate or high relative heterogeneity for most comparisons (**Figures 2.5-2.9**). Low heterogeneity was detected only for retinol (I^2, 95% CI: 0, 0-79 %) and C20:4n-6 (I^2, 95% CI: 7, 0-43 %). However, high and moderate heterogeneity, respectively, could not be ruled for these biomarkers as indicated by the associated 95% CIs.

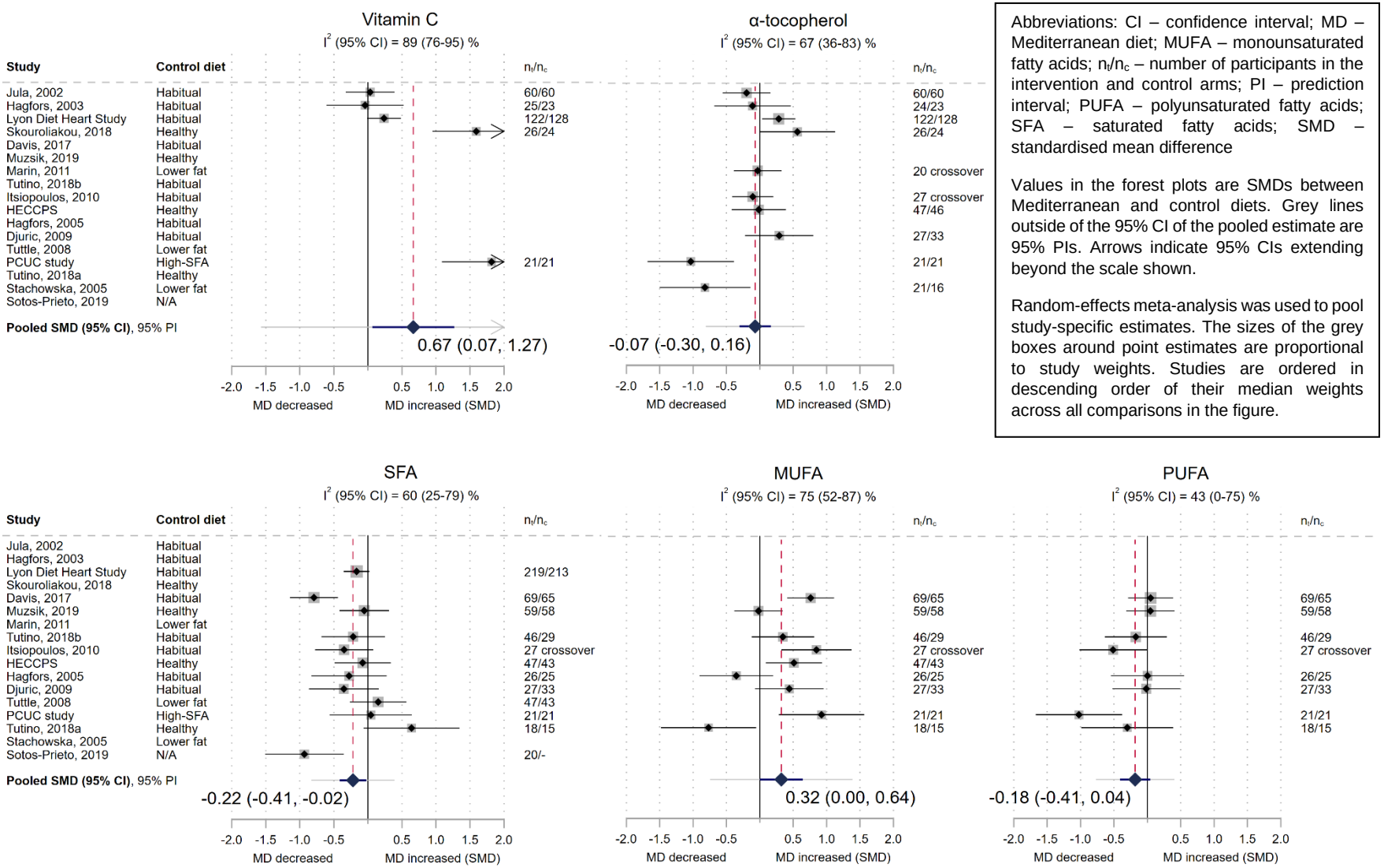


Abbreviations: CI – confidence interval; MD – Mediterranean diet; n_t/n_c – number of participants in the intervention and control arms; PI – prediction interval; SMD – standardised mean difference

Values in forest plots are SMDs between Mediterranean and control diets. Grey lines outside of the 95% CI of the pooled estimate are 95% PIs. Arrows indicate 95% CIs extending beyond the scale shown.

Random-effects meta-analysis was used to pool study-specific estimates. The sizes of the grey boxes around point estimates are proportional to study weights. Studies are ordered in descending order of their median weights across all comparisons in the figure.

Figure 2.5 Effects of the Mediterranean diet on circulating carotenoids and retinol: standardised mean differences

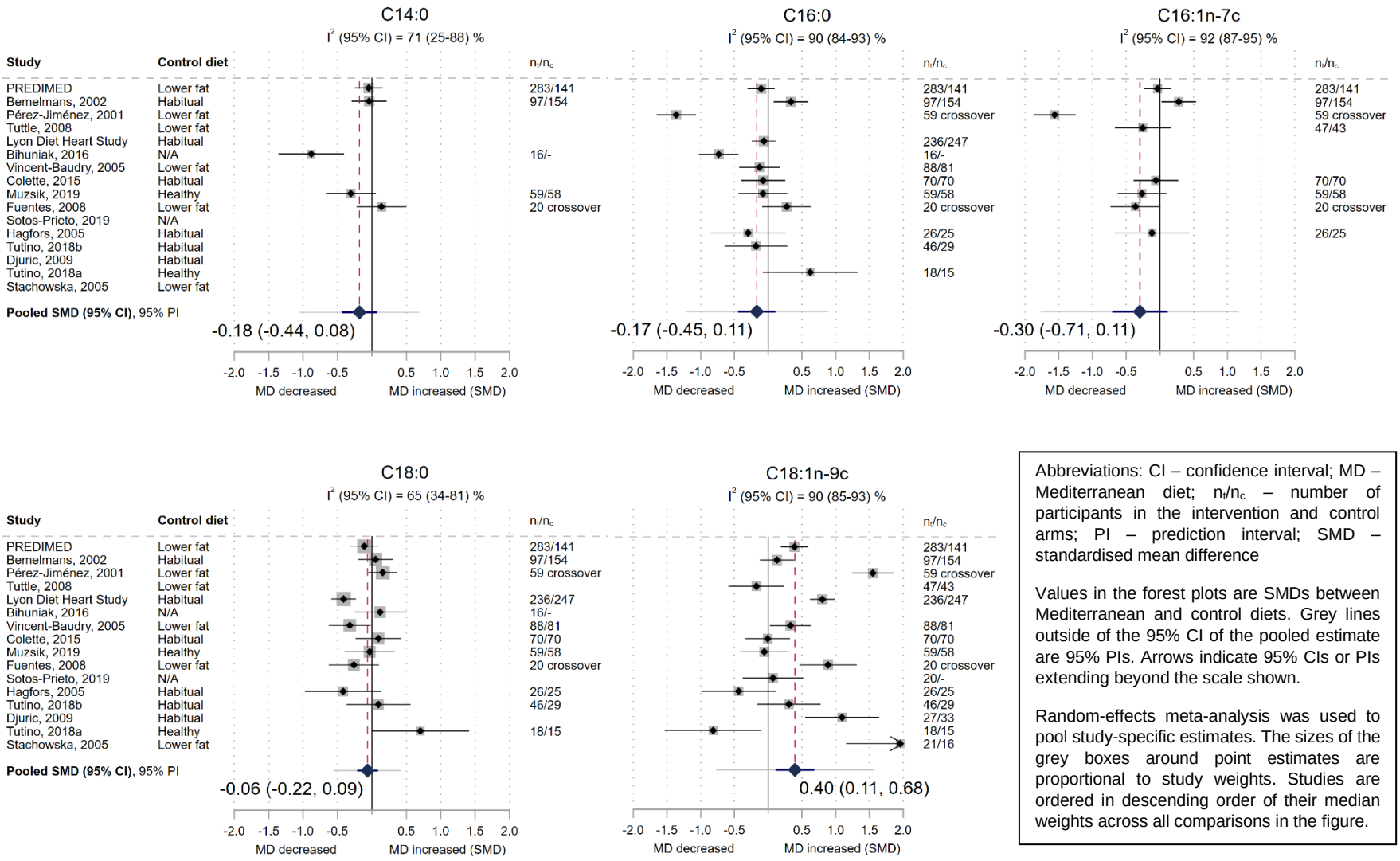


Abbreviations: CI – confidence interval; MD – Mediterranean diet; MUFA – monounsaturated fatty acids; n_t/n_c – number of participants in the intervention and control arms; PI – prediction interval; PUFA – polyunsaturated fatty acids; SFA – saturated fatty acids; SMD – standardised mean difference

Values in the forest plots are SMDs between Mediterranean and control diets. Grey lines outside of the 95% CI of the pooled estimate are 95% PIs. Arrows indicate 95% CIs extending beyond the scale shown.

Random-effects meta-analysis was used to pool study-specific estimates. The sizes of the grey boxes around point estimates are proportional to study weights. Studies are ordered in descending order of their median weights across all comparisons in the figure.

Figure 2.6 Effects of the Mediterranean diet on circulating vitamin C, α-tocopherol, saturated, mono- and polyunsaturated fatty acids: standardised mean differences

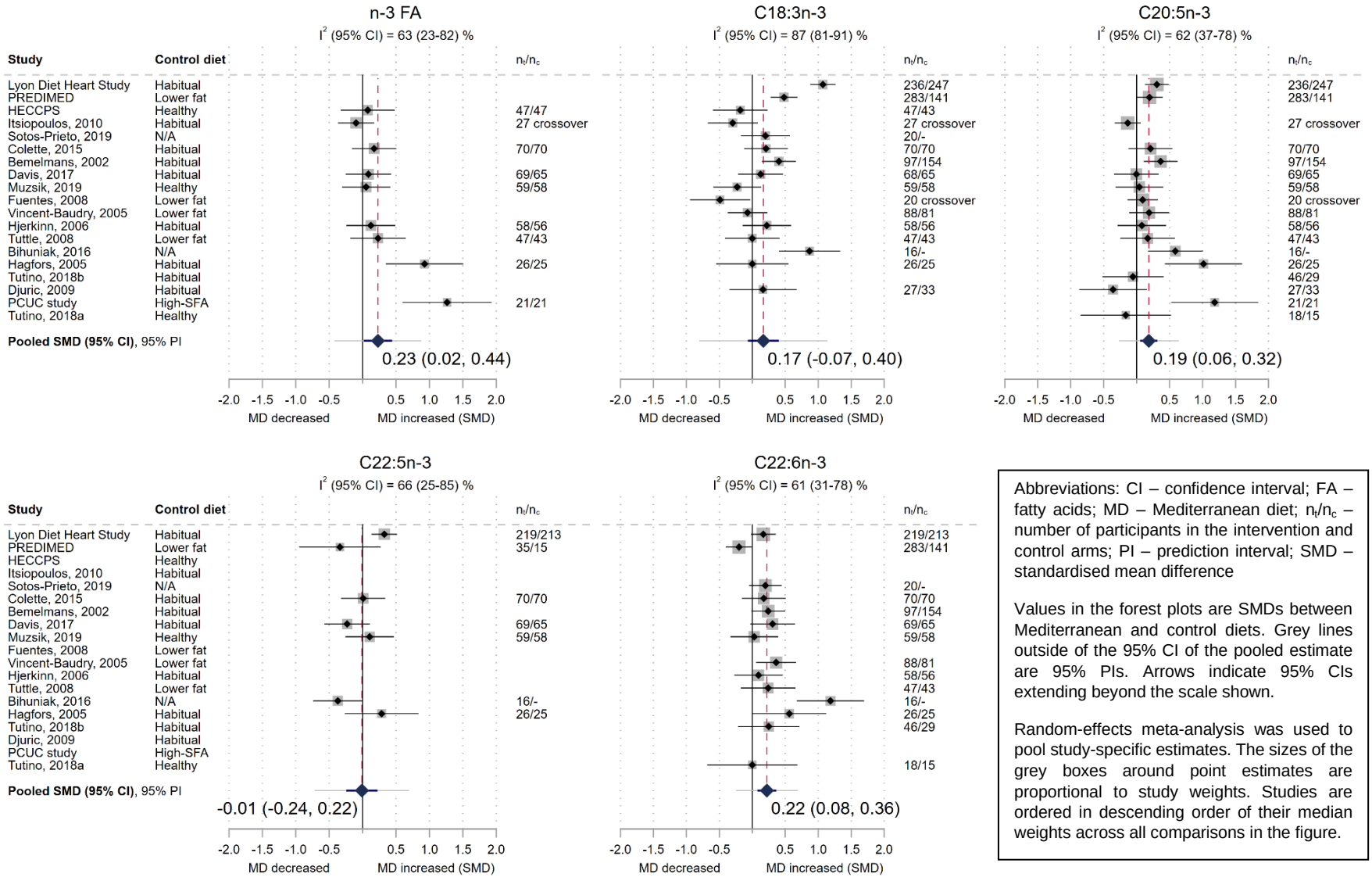


Abbreviations: CI – confidence interval; MD – Mediterranean diet; n_t/n_c – number of participants in the intervention and control arms; PI – prediction interval; SMD – standardised mean difference

Values in the forest plots are SMDs between Mediterranean and control diets. Grey lines outside of the 95% CI of the pooled estimate are 95% PIs. Arrows indicate 95% CIs or PIs extending beyond the scale shown.

Random-effects meta-analysis was used to pool study-specific estimates. The sizes of the grey boxes around point estimates are proportional to study weights. Studies are ordered in descending order of their median weights across all comparisons in the figure.

Figure 2.7 Effects of the Mediterranean diet on circulating saturated and monounsaturated fatty acids: standardised mean differences

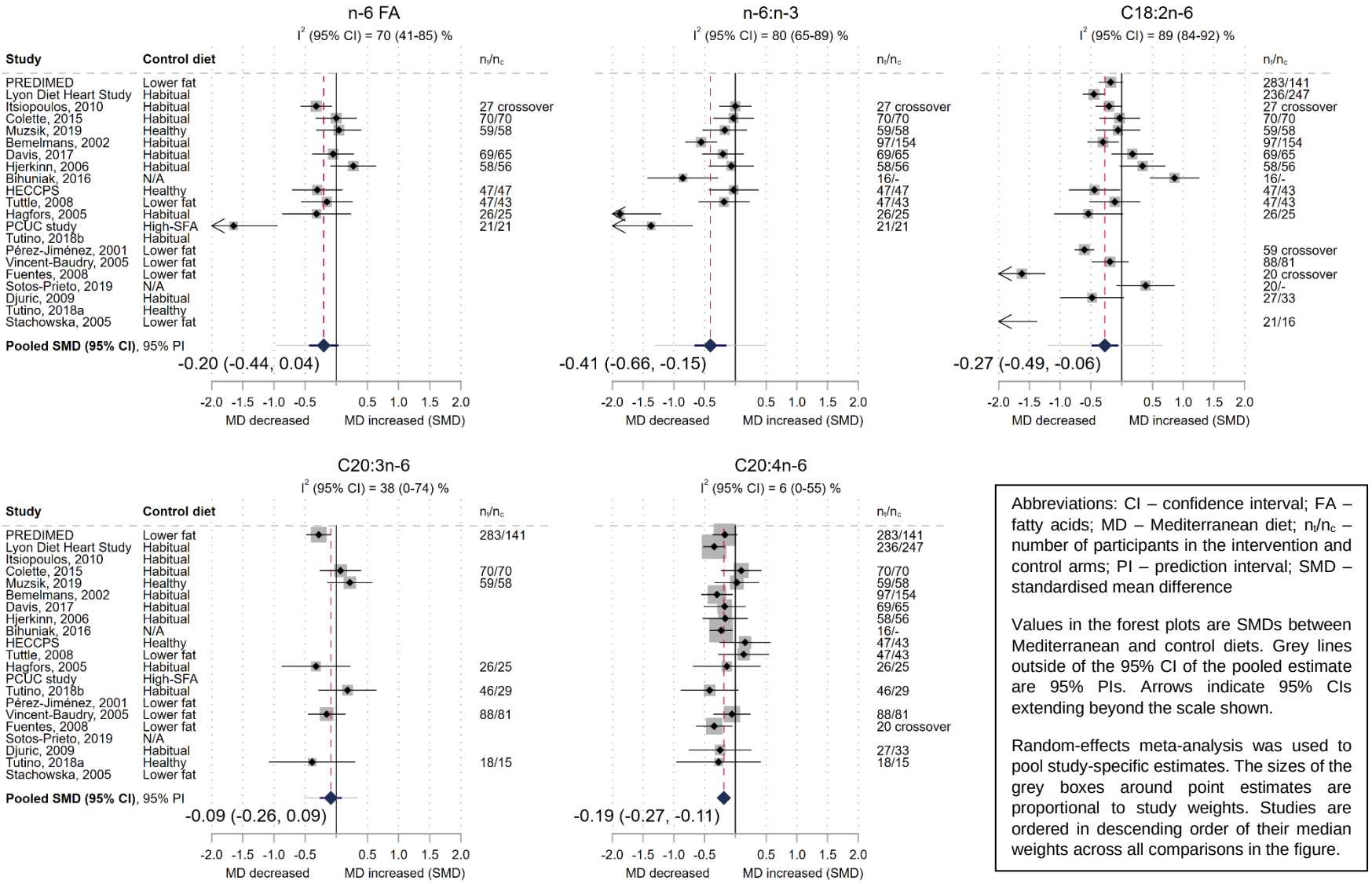


Abbreviations: CI – confidence interval; FA – fatty acids; MD – Mediterranean diet; n_t/n_c – number of participants in the intervention and control arms; PI – prediction interval; SMD – standardised mean difference

Values in the forest plots are SMDs between Mediterranean and control diets. Grey lines outside of the 95% CI of the pooled estimate are 95% PIs. Arrows indicate 95% CIs extending beyond the scale shown.

Random-effects meta-analysis was used to pool study-specific estimates. The sizes of the grey boxes around point estimates are proportional to study weights. Studies are ordered in descending order of their median weights across all comparisons in the figure.

Figure 2.8 Effects of the Mediterranean diet on circulating total and individual n-3 fatty acids: standardised mean differences



Abbreviations: CI – confidence interval; FA – fatty acids; MD – Mediterranean diet; n_t/n_c – number of participants in the intervention and control arms; PI – prediction interval; SMD – standardised mean difference

Values in the forest plots are SMDs between Mediterranean and control diets. Grey lines outside of the 95% CI of the pooled estimate are 95% PIs. Arrows indicate 95% CIs extending beyond the scale shown.

Random-effects meta-analysis was used to pool study-specific estimates. The sizes of the grey boxes around point estimates are proportional to study weights. Studies are ordered in descending order of their median weights across all comparisons in the figure.

Figure 2.9 Effects of the Mediterranean diet on circulating total and individual n-6 fatty acids: standardised mean differences

2.4.12 Meta-regression and subgroup analysis

I identified several continuous covariates as potential effect modifiers of the effects of Mediterranean diet on nutritional biomarkers in meta-regression analysis (**Table 2.10**). Baseline biomarker concentrations were inversely associated with post-intervention SMD in plasma total n-6 fatty acids (-0.87 per % of total fatty acids, 95% CI: -1.48, -0.26). Following Winsorisation of trial duration at 6 months, each additional month of the intervention (95% CI) was associated with increased SMD in C16:1n-7c by 0.18 (0.01, 0.35). Taking into account the directionality of the pooled SMD in the main analysis (**Figure 2.7**), this coefficient would translate to a smaller response with longer duration. Each year of mean baseline age was associated with higher pooled SMDs in total n-6 (0.03, 95% CI: 0.01, 0.05) and C18:2n-6 (0.03, 95% CI: 0.01, 0.04), and lower SMDs in total n-3 (-0.02, 95% CI: -0.03, -0.00), C18:1n-9c (-0.03, 95% CI: -0.05, -0.01) and C22:5n-3 (-0.03, 95% CI:-0.04, -0.02) fatty acids. For C18:1n-9c, a negative association was additionally detected between baseline mean BMI and post-intervention SMD where each kg/m^2 unit was associated with a -0.15 (95% CI: -0.23, -0.06) difference.

Meta-regression coefficients with p values between 0.05-0.10 were detected for the relationship between mean baseline age and SMDs in β-carotene, α-tocopherol, and total PUFA; percentage of women and SMDs in β-cryptoxanthin, α-tocopherol, C18:0, total PUFA and C22:5n-3; baseline mean BMI and SMDs in C16:0 and C16:1n-7c; trial duration and SMDs in total PUFA, total n-6 fatty acids and n-6:n-3 ratio (**Table 2.10**). All the above biomarker-covariate combinations were taken forward to the subgroup analysis and dichotomised at covariate medians for each analysis, pending availability of at least five studies in each grouping.

Subgroup analysis suggested presence of effect modification for several biomarker-covariate combinations and modestly lower relative heterogeneity in some of the subgroups compared to the main results (**Table 2.11**). For C18:1n9-c (main analysis I^2 = 90%), an effect of Mediterranean diet was observed in studies with mean baseline age ≤53.8 years (SMD = 0.92, 95% CI: 0.52, 1.32; I^2 = 87%) but not in studies with age >53.8 years (SMD = 0.02, 95% CI: -0.20, 0.23; I^2 = 69%; p heterogeneity <0.001). A similar pattern emerged for BMI, where the pooled SMD in studies with mean baseline BMI ≤28.7 kg/m^2 was 0.81 (95% CI: 0.31, 1.31; I^2 = 92%) and in studies with mean BMI >28.7 kg/m^2 the pooled SMD was 0.11 (95% CI: -0.10, 0.31; I^2 = 61%; p heterogeneity = 0.010). For C20:5n-3 (main analysis I^2 = 91%), an effect of the intervention was detected in studies which utilised a modified Mediterranean diet (SMD =

0.34; 95% CI: 0.14, 0.55; I^2 = 65%), however, it was not apparent in studies that applied a traditional Mediterranean diet (SMD = 0.05, 95% CI: -0.05, 0.15); I^2 = 0%; p heterogeneity = 0.005). There was evidence to suggest effect modification of the results for C18:2n-6 (main analysis I^2 = 80%) by baseline age (p heterogeneity = 0.001) and blood fraction (p heterogeneity = 0.014). In studies with mean baseline age ≤57.5 years, the pooled SMD was -0.59 (95% CI: -0.89, -0.28; I^2 = 89%), and in studies with age >57.5 years the pooled SMD was 0.03 (95% CI: -0.19, 0.25; I^2 = 75%). Studies which performed C18:2n-6 assays in total plasma did not detect an effect of Mediterranean diet interventions (SMD = -0.14; 95% CI: -0.33, 0.05; I^2 = 82%), as opposed to studies which used any other blood fraction (SMD = -0.59; 95% CI: -0.96, -0.22; I^2 = 91%). The blood fractions in the later subgroup included erythrocytes, plasma phospholipids, cholesteryl esters, LDL cholesterol esters and triglycerides (**Table 2.7**).

Subgroup analysis of C18:3n-3 by blood fraction revealed biologically important differential effects despite the lack of formal evidence to suggest presence of effect modification (p heterogeneity = 0.084). The main analysis indicated an average treatment effect compatible with both increased and decreased concentrations (SMD = 0.17; 95% CI: -0.07, 0.40; I^2 = 87%). Total plasma assays yielded a pooled positive result (SMD = 0.31; 95% CI: 0.01, 0.60; I^2 = 88%) while a combination of the remaining blood fractions did not (SMD = -0.06; 95% CI: -0.35, 0.23; I^2 = 72%; **Table 2.11**).

Table 2.10 Meta-regression analysis of the effects of Mediterranean versus control diet interventions on 25 nutritional biomarkers: results with p values for meta-regression coefficients < 0.10

Biomarker and covariate	SMD difference (95% CI)	n of studies	I^2	p
β-carotene				
Age (per year)	-0.02 (-0.04, 0.00)	9	59%	0.098
β-cryptoxanthin				
Percent of women (per %)	0.02 (-0.00, 0.05)	5	63%	0.065
Lycopene				
Percent of women (per %)	-0.01 (-0.02, 0.00)	7	32%	0.080
α-tocopherol				
Age (per year)	0.03 (-0.01, 0.06)	9	67%	0.090
C16:0				
BMI (per kg/m^2)	0.08 (-0.01, 0.17)	12	87%	0.066
C16:1n-7c				
Duration (per month)	0.08 (-0.01, 0.16)	8	82%	0.062
Winsorised at 6 months	0.18 (0.01, 0.35)	8	85%	0.040
Age (per year)	0.02 (-0.00, 0.04)	8	86%	0.050
BMI (per kg/m^2)	0.10 (-0.02, 0.23)	8	87%	0.090
C18:0				
Percent of women (per %)	0.005 (-0.000, 0.010)	12	48%	0.066
C18:1n-9c				
Age (per year)	-0.03 (-0.05, -0.01)	15	84%	0.010
BMI (per kg/m^2)	-0.15 (-0.23, -0.06)	15	76%	0.002
PUFA				
Duration (per month)	0.14 (-0.02, 0.30)	8	14%	0.077
Winsorised at 6 months	0.14 (-0.02, 0.30)	8	14%	0.077
Age (per year)	0.02 (-0.00, 0.03)	8	11%	0.070
Percent of women (per %)	0.01 (-0.00, 0.01)	8	17%	0.082
n-3				
Age (per year)	-0.02 (-0.03, -0.00)	10	50%	0.049
C22:5n-3				
Age (per year)	-0.03 (-0.05, -0.01)	7	0%	0.011
Percent of women (per %)	-0.01 (-0.01, 0.00)	7	33%	0.090
n-6				
Duration (per month)	0.07 (-0.02, 0.15)	9	60%	0.094
winsorised at 6 months	0.18 (-0.02, 0.38)	9	60%	0.068
Age (per year)	0.03 (0.01, 0.05)	9	3%	0.003
Baseline biomarker (per %fatty acids)	-0.87 (-1.48, -0.26)	5	13%	0.020
n-6/n-3				
Duration (per month)	0.06 (-0.04, 0.16)	11	83%	0.195
Winsorised at 6 months	0.22 (-0.00, 0.44)	11	81%	0.051
C18:2n-6				
Age (per year)	0.03 (0.01, 0.04)	18	77%	0.001
C20:4n-6				
Duration (per month)	0.01 (-0.01, 0.03)	16	3%	0.329
Winsorised at 6 months	0.04 (-0.00, 0.09)	16	0%	0.052

Abbreviations: BMI – body mass index; CI – confidence interval; SMD – standardised mean difference; PUFA – polyunsaturated fatty acids

The following study characteristics were tested for all meta-analysed biomarkers in the current review pending availability of the data for ≥5 studies: baseline age, BMI, baseline biomarker concentrations, percent of women, trial duration with and without Winsorisation at 6 months.

Table 2.11 Subgroup analysis of the effects of Mediterranean versus control diet interventions on nutritional biomarkers: results with p values for heterogeneity < 0.10 and ≥5 studies per subgroup

Biomarker and subgroup	SMD (95% CI)	n of studies	I^2	p heterogeneity
C18:1n-9c				
Main result	0.40 (0.11, 0.68)	15	90%	
Age, years				
<53.8	0.92 (0.52, 1.32)	7	87%	
≥53.8	0.02 (-0.20, 0.23)	8	69%	<0.001
BMI, kg/m^2				
<28.7	0.81 (0.31, 1.31)	7	92%	
≥28.7	0.11 (-0.10, 0.31)	8	61%	0.010
C18:3n-3				
Main result	0.17 (-0.07, 0.40)	16	87%	
Blood fraction				
Plasma	0.31 (0.01, 0.60)	10	88%	
Other	-0.06 (-0.35, 0.23)	6	72%	0.084
C20:5n-3				
Main result	0.19 (0.06, 0.32)	17	62%	
Mediterranean diet				
Traditional	0.05 (-0.05, 0.15)	8	0%	
Modified	0.34 (0.14, 0.55)	9	65%	0.005
C22:6n-3				
Main result	0.22 (0.08, 0.36)	13	62%	
Feeding				
No or little feeding	0.10 (-0.06, 0.26)	7	45%	
Partial or full-feeding	0.35 (0.13, 0.57)	7	62%	0.066
n-6:n-3				
Main result	-0.41 (-0.66, -0.15)	11	80%	
Duration, months				
2-4	-0.80 (-1.41, -0.18)	5	90%	
≥6	-0.20 (-0.39, -0.01)	6	46%	0.080
C18:2n-6				
Main result	-0.27 (-0.49, -0.06)	18	89%	
Age, y				
≤57.5	-0.59 (-0.89, -0.28)	9	89%	
>57.5	0.03 (-0.19, 0.25)	9	75%	0.001
Blood fraction				
Plasma	-0.04 (-0.28, 0.20)	10	82%	
Other	-0.59 (-0.96, -0.22)	8	91%	0.014

Abbreviations: BMI – body mass index; CI – confidence interval; SMD – standardised mean difference; PUFA – polyunsaturated fatty acids

The following study characteristics were tested for all meta-analysed biomarkers in the current review pending availability of the data for ≥5 studies in each subgroup: baseline age, BMI, baseline biomarker concentrations, percent of women, trial duration, feeding design, Mediterranean diet type, control diet type, geographical location, blood fraction of fatty acids.

2.4.13 Sensitivity analyses

Results of the sensitivity analyses comparing the main results on the SMD scale with the unweighted mean difference scale and investigating the influence of within-person trials and non-randomised designs are summarised in **Tables 2.12-2.15**.

Results from the analysis using unweighted mean differences were fully consistent with the main results in terms of the statistical significance apart from C18:3n-3 (SMD = 0.18, 95% CI: -0.04, 0.40; mean difference = 0.05 %total fatty acids, 95% CI: 0.00, 0.09). They also yielded mostly similar (+/- 10%) I^2 values with the exceptions of C14:0 (**Table 2.13**), n-3 PUFA, C22:5n-3, C22:6n-3 (**Table 2.14**), C20:3n-6 and C20:4n-6 (**Table 2.15**). There was no apparent pattern in either of the SMD or unweighted mean difference scales yielding consistently lower I^2 values for these biomarkers.

The influence on the main results of the imputed correlation coefficients used for estimating standard errors in within-person trials was minor, and inference was affected for two biomarkers in the sensitivity analyses using minimum or maximum plausible correlation coefficients. The main result (pooled SMD; 95% CI) for lycopene based on seven studies (0.29; 0.03, 0.55) was no longer statistically significant when using the maximum correlation coefficient (0.25; -0.09, 0.58). However, the effect size and precision increased when restricting the analysis to five studies which did not require imputed coefficients (0.33; 0.06, 0.61). The main result for retinol based on five studies (0.18; 0.01, 0.35) was no longer statistically significant after applying the minimum correlation (0.18; -0.01, 0.37) or exclusion of the single crossover trial (0.18; -0.02, 0.38) (**Table 2.12**). Exclusion of within-person trials with imputed values of the correlations also attenuated the results to the null for total SFA (-0.23; -0.42, -0.04 to -0.17; -0.37, 0.03), MUFA (0.32; 0.00, 0.64 to 0.26; -0.08, 0.60) and C18:1n-9c (0.40; 0.11, 0.68 to 0.29; -0.01, 0.59). For MUFA, this analysis was equivalent to excluding cross-over trials (**Table 2.13**).

Exclusion of within-person trials from the analysis materially affected the results (pooled SMD; 95% CI) for several other biomarkers. For lycopene, excluding one pre-post study without a control group increased the effect size and precision (0.36; 0.12, 0.60), whereas exclusion of one crossover RCT decreased the effect size and precision below the threshold of statistical significance (0.23; -0.08, 0.54) (**Table 2.12**). Exclusion of crossover RCTs had the same influence on C18:2n-6 (-0.15; -0.36, 0.06 compared to the main result of -0.27; -0.49, -0.06) (**Table 2.14**). By contrast, the main result for C18:3n-3 (0.17; 95% CI:

-0.07, 0.40) was no longer null following this sensitivity analysis (0.25; 0.01, 0.48) (**Table 2.13**). Apart from these results for C18:3n-3, all null results from the main analyses were robust to the above sensitivity analyses. Among biomarkers for which pooled SMDs statistically significantly differed between the Mediterranean and control diets in the main analyses, only β-carotene and C20:4n-6 were robust to all sensitivity analyses.

Table 2.12 Sensitivity analyses of meta-analyses of the effects of Mediterranean versus control diet interventions on carotenoids and vitamins

Biomarker	Main result (SMD)	Mean difference (ng/dL)	Imputed correlations in within-person trials (SMD)			Excluding non-RCT designs (SMD)	Excluding crossover RCTs (SMD)
			Minimum	Maximum	Excluding imputed		
α-carotene							
n of studies	5	5	5	5	4	N/A	4
n of participants	567	567	567	567	459		459
SMD/MD (95% CI)	0.11 (-0.26, 0.47)	0.16 (-0.23, 0.56)	0.10 (-0.29, 0.50)	0.11 (-0.21, 0.43)	0.09 (-0.40, 0.58)		0.09 (-0.40, 0.58)
I^2 (95% CI)	80 (52-91) %	82 (59-92) %	80 (52-91) %	80 (52-91) %	85 (62-94) %		85 (62-94) %
β-carotene							
n of studies	10	10	10	10	7	9	8
n of participants	869	869	869	869	670	854	685
SMD/MD (95% CI)	0.39 (0.17, 0.61)	0.78 (0.39, 1.16)	0.39 (0.17, 0.61)	0.42 (0.21, 0.63)	0.37 (0.10, 0.64)	0.35 (0.13, 0.57)	0.42 (0.15, 0.69)
I^2 (95% CI)	57 (12-79) %	53 (3-77) %	55 (8-78) %	75 (54-87) %	65 (22-85) %	55 (5-79) %	65 (26-84) %
β-cryptoxanthin							
n of studies	5	5	5	5	4	N/A	4
n of participants	567	567	567	567	459		459
SMD/MD (95% CI)	0.25 (-0.11, 0.61)	0.17 (-0.12, 0.47)	0.28 (-0.13, 0.70)	0.25 (-0.11, 0.61)	0.37 (-0.13, 0.87)		0.37 (-0.13, 0.87)
I^2 (95% CI)	84 (63-93) %	86 (70-94) %	81 (55-92) %	84 (63-93) %	85 (63-94) %		85 (63-94) %
Lycopene							
n of studies	7	7	7	7	5	6	6
n of participants	630	630	630	630	507	615	522
SMD/MD (95% CI)	0.29 (0.03, 0.55)	0.49 (0.05, 0.94)	0.29 (0.03, 0.55)	0.25 (-0.09, 0.58)	0.33 (0.06, 0.61)	0.36 (0.12, 0.60)	0.25 (-0.04, 0.55)
I^2 (95% CI)	57 (0-81) %	69 (31-86) %	57 (0-81) %	86 (74-93) %	55 (0-84) %	46 (0-79) %	62 (9-85) %
Retinol							
n of studies	5	5	5	5	4	N/A	4
n of participants	486	486	486	486	378		378
SMD/MD (95% CI)	0.18 (0.01, 0.35)	0.31 (0.05, 0.57)	0.18 (-0.01, 0.37)	0.19 (0.05, 0.32)	0.18 (-0.02, 0.38)		0.18 (-0.02, 0.38)
I^2 (95% CI)	0 (0-79) %	0 (0-79) %	0 (0-79) %	0 (0-79) %	0 (0-85) %		0 (0-85) %
Vitamin C							
n of studies	5	5	5	5	N/A	N/A	N/A
n of participants	510	510	510	510			
SMD/MD (95% CI)	0.67 (0.07, 1.27)	1.62 (0.13, 3.11)	0.67 (0.07, 1.27)	0.67 (0.07, 1.27)			
I^2 (95% CI)	89 (76-95) %	91 (81-95) %	89 (76-95) %	89 (76-95) %			
α-tocopherol							
n of studies	10	10	10	10	8	N/A	8
n of participants	887	887	887	887	699		699
SMD/MD (95% CI)	-0.07 (-0.30, 0.16)	-0.26 (-1.02, 0.49)	-0.08 (-0.33, 0.18)	-0.06 (-0.26, 0.14)	-0.09 (-0.40, 0.23)		-0.09 (-0.40, 0.23)
I^2 (95% CI)	67 (36-83) %	71 (44-85) %	66 (35-83) %	68 (37-83) %	74 (46-87) %		74 (46-87) %

Abbreviations: CI – confidence interval; MD – mean difference; RCT – randomised controlled trial; SMD – standardised mean difference.

Numbers of participants in crossover trials were double-counted for comparability with parallel trials.

Table 2.13 Sensitivity analyses of meta-analyses of the effects of Mediterranean versus control diet interventions on saturated and monounsaturated fatty acids

Biomarker	Main result (SMD)	Mean difference (%total fatty acids)	Imputed correlations in within-person trials (SMD)			Excluding non-RCT designs (SMD)	Excluding crossover RCTs (SMD)
			Minimum	Maximum	Excluding imputed		
SFA							
n of studies	12	12	12	12	10	11	11
n of participants	1,199	1,199	1,199	1,199	1,071	1,179	1,091
SMD/MD (95% CI)	-0.23 (-0.42, -0.04)	-0.55 (-0.92, -0.18)	-0.23 (-0.42, -0.03)	-0.24 (-0.42, -0.06)	-0.17 (-0.37, 0.03)	-0.19 (-0.37, -0.00)	-0.22 (-0.43, -0.01)
I^2 (95% CI)	60 (24-79) %	47 (0-73) %	60 (24-79) %	61 (27-79) %	57 (13-79) %	53 (7-76) %	63 (29-81) %
C14:0							
n of studies	5	5	5	5	2	3	4
n of participants	643	643	643	643	399	479	563
SMD/MD (95% CI)	-0.18 (-0.44, 0.08)	-0.03 (-0.08, 0.02)	-0.21 (-0.47, 0.06)	-0.16 (-0.42, 0.09)	-0.13 (-0.37, 0.11)	-0.06 (-0.27, 0.14)	-0.26 (-0.55, 0.04)
I^2 (95% CI)	71 (25-88) %	70 (23-88) %	68 (16-87) %	76 (41-90) %	32 %	32 (0-93) %	74 (26-91) %
C16:0							
n of studies	12	12	12	12	8	10	10
n of participants	1,797	1,797	1,797	1,797	1,317	1,633	1,481
SMD/MD (95% CI)	-0.17 (-0.45, 0.11)	-0.53 (-1.30, 0.25)	-0.16 (-0.39, 0.06)	-0.16 (-0.52, 0.20)	-0.08 (-0.19, 0.02)	-0.16 (-0.46, 0.13)	-0.10 (-0.29, 0.10)
I^2 (95% CI)	90 (84-93) %	93 (90-95) %	82 (70-89) %	95 (93-97) %	0 (0-68) %	88 (81-93) %	73 (50-86) %
C18:0							
n of studies	12	12	12	12	8	10	10
n of participants	1,797	1,797	1,797	1,797	1,317	1,633	1,481
SMD/MD (95% CI)	-0.06 (-0.22, 0.09)	-0.11 (-0.30, 0.09)	-0.07 (-0.23, 0.08)	-0.06 (-0.21, 0.09)	-0.12 (-0.31, 0.07)	-0.09 (-0.27, 0.08)	-0.08 (-0.24, 0.09)
I^2 (95% CI)	65 (34-81) %	70 (46-83) %	57 (18-77) %	75 (55-86) %	63 (20-83) %	68 (37-83) %	62 (25-81) %
MUFA							
n of studies	9	9	9	9	8	N/A	8
n of participants	677	677	677	677	569		569
SMD/MD (95% CI)	0.32 (0.00, 0.64)	0.91 (0.14, 1.69)	0.32 (0.00, 0.64)	0.34 (0.01, 0.67)	0.26 (-0.08, 0.60)		0.26 (-0.08, 0.60)
I^2 (95% CI)	75 (52-87) %	72 (45-86) %	75 (51-87) %	82 (67-90) %	75 (51-88) %		75 (51-88) %
C16:1n-7c							
n of studies	8	8	8	8	5	7	6
n of participants	1,144	1,144	1,144	1,144	680	996	828
SMD/MD (95% CI)	-0.30 (-0.71, 0.11)	-0.07 (-0.22, 0.09)	-0.26 (-0.57, 0.05)	-0.30 (-0.81, 0.21)	-0.10 (-0.25, 0.04)	-0.38 (-0.82, 0.06)	-0.04 (-0.21, 0.13)
I^2 (95% CI)	92 (87-95) %	96 (95-98) %	84 (69-91) %	96 (95-98) %	0 (0-79) %	92 (85-95) %	41 (0-76) %
C18:1n-9c							
n of studies	15	14	15	15	11	13	13
n of participants	1,988	1,951	1,988	1,988	1,504	1,820	1,672
SMD/MD (95% CI)	0.40 (0.11, 0.68)	1.56 (0.18, 2.94)	0.38 (0.11, 0.64)	0.40 (0.08, 0.73)	0.29 (-0.01, 0.59)	0.44 (0.12, 0.77)	0.26 (0.00, 0.52)
I^2 (95% CI)	90 (85-93) %	96 (95-97) %	87 (80-91) %	94 (91-96) %	87 (79-92) %	91 (86-94) %	86 (77-91) %

Abbreviations: CI – confidence interval; MD – mean difference; MUFA – monounsaturated fatty acids; RCT – randomised controlled trial; SFA – saturated fatty acids; SMD – standardised mean difference.

Numbers of participants in crossover trials were double-counted for comparability with parallel trials. The 95% CIs of I^2 could not be estimated with <3 trials.

Table 2.14 Sensitivity analyses of meta-analyses of the effects of Mediterranean versus control diet interventions on total polyunsaturated and n-3 fatty acids

Biomarker	Main result (SMD)	Mean difference (%total fatty acids)	Imputed correlations in within-person trials (SMD)			Excluding non-RCT designs (SMD)	Excluding crossover RCTs (SMD)
			Minimum	Maximum	Excluding imputed		
PUFA							
n of studies	8	8	8	8	7	N/A	7
n of participants	587	587	587	587	479		479
SMD/MD (95% CI)	-0.18 (-0.41, 0.04)	-0.83 (-1.68, 0.02)	-0.18 (-0.40, 0.05)	-0.22 (-0.47, 0.03)	-0.13 (-0.37, 0.10)		-0.13 (-0.37, 0.10)
I^2 (95% CI)	43 (0-75) %	56 (4-80) %	41 (0-74) %	64 (24-83) %	39 (0-74) %		39 (0-74) %
n-3 PUFA							
n of studies	9	9	9	9	8	N/A	8
n of participants	890	890	890	890	782		782
SMD/MD (95% CI)	0.23 (0.02, 0.44)	0.58 (0.05, 1.10)	0.24 (0.03, 0.46)	0.22 (0.01, 0.44)	0.28 (0.05, 0.51)		0.28 (0.05, 0.51)
I^2 (95% CI)	63 (23-82) %	73 (47-86) %	57 (11-80) %	68 (35-84) %	60 (12-81) %		60 (12-81) %
C18:3n-3							
n of studies	16	16	16	16	11	13	14
n of participants	2,101	2,101	2,101	2,101	1,729	1,917	1,913
SMD/MD (95% CI)	0.17 (-0.07, 0.40)	0.05 (0.00, 0.09)	0.18 (-0.05, 0.41)	0.16 (-0.10, 0.41)	0.18 (-0.11, 0.47)	0.09 (-0.19, 0.37)	0.25 (0.01, 0.48)
I^2 (95% CI)	87 (81-91) %	92 (89-94) %	86 (79-91) %	91 (88-94) %	89 (82-93) %	89 (83-93) %	86 (78-91) %
C20:5n-3							
n of studies	17	17	17	17	13	15	15
n of participants	2,109	2,109	2,109	2,109	1,757	1,945	1,921
SMD/MD (95% CI)	0.19 (0.06, 0.32)	0.15 (0.04, 0.25)	0.21 (0.08, 0.34)	0.18 (0.05, 0.31)	0.19 (0.04, 0.34)	0.15 (0.01, 0.28)	0.23 (0.09, 0.36)
I^2 (95% CI)	62 (37-78) %	68 (47-81) %	52 (16-72) %	64 (39-79) %	55 (16-76) %	61 (31-78) %	55 (19-75) %
C22:5n-3							
n of studies	7	7	7	7	6	6	N/A
n of participants	898	898	898	898	882	882	
SMD/MD (95% CI)	-0.01 (-0.24, 0.22)	-0.02 (-0.07, 0.04)	-0.01 (-0.24, 0.22)	-0.01 (-0.24, 0.22)	0.06 (-0.16, 0.28)	0.06 (-0.16, 0.28)	
I^2 (95% CI)	66 (25-85) %	67 (27-85) %	66 (25-85) %	66 (25-85) %	57 (0-83) %	57 (0-83) %	
C22:6n-3							
n of studies	14	14	14	14	11	11	N/A
n of participants	1,764	1,764	1,764	1,764	1,580	1,580	
SMD/MD (95% CI)	0.22 (0.08, 0.36)	0.17 (0.05, 0.30)	0.22 (0.08, 0.36)	0.22 (0.08, 0.36)	0.15 (0.02, 0.28)	0.15 (0.02, 0.28)	
I^2 (95% CI)	61 (31-78) %	69 (45-82) %	61 (31-78) %	61 (31-78) %	41 (0-71) %	41 (0-71) %	

Abbreviations: CI – confidence interval; MD – mean difference; PUFA – polyunsaturated fatty acids; RCT – randomised controlled trial; SMD – standardised mean difference.

Numbers of participants in crossover trials were double-counted for comparability with parallel trials.

Table 2.15 Sensitivity analyses of meta-analyses of the effects of Mediterranean versus control diet interventions on n-6 fatty acids

Biomarker	Main result (SMD)	Mean difference (%total fatty acids)	Imputed correlations in within-person trials (SMD)			Excluding non-RCT designs (SMD)	Excluding crossover RCTs (SMD)
			Minimum	Maximum	Excluding imputed		
n-6 PUFA							
n of studies	9	9	9	9	8	N/A	8
n of participants	890	890	890	890	782		782
SMD/MD (95% CI)	-0.20 (-0.44, 0.04)	-0.99 (-2.04, 0.06)	-0.20 (-0.45, 0.05)	-0.20 (-0.43, 0.03)	-0.19 (-0.46, 0.08)		-0.19 (-0.46, 0.08)
I^2 (95% CI)	70 (41-85) %	75 (52-87) %	69 (37-84) %	72 (44-86) %	72 (42-86) %		72 (42-86) %
n-6:n-3							
n of studies	11	11	11	11	8	9	10
n of participants	1,054	1,054	1,054	1,054	782	890	946
SMD/MD (95% CI)	-0.41 (-0.66, -0.15)	-0.99 (-1.53, -0.45)	-0.42 (-0.68, -0.15)	-0.40 (-0.66, -0.15)	-0.41 (-0.75, -0.07)	-0.35 (-0.64, -0.05)	-0.46 (-0.74, -0.17)
I^2 (95% CI)	80 (65-89) %	88 (81-93) %	78 (62-88) %	82 (68-89) %	81 (64-90) %	80 (63-89) %	80 (64-89) %
C18:2n-6							
n of studies	18	17	18	18	12	15	15
n of participants	2,375	2,338	2,375	2,375	1,767	2,191	1,951
SMD/MD (95% CI)	-0.27 (-0.49, -0.06)	-1.03 (-1.99, -0.08)	-0.23 (-0.43, -0.03)	-0.28 (-0.50, -0.06)	-0.25 (-0.46, -0.04)	-0.38 (-0.60, -0.16)	-0.15 (-0.36, 0.06)
I^2 (95% CI)	89 (84-92) %	93 (91-95) %	82 (73-88) %	90 (86-93) %	77 (60-87) %	87 (80-92) %	83 (72-89) %
C20:3n-6							
n of studies	7	7	7	7	N/A	N/A	N/A
n of participants	834	834	834	834			
SMD/MD (95% CI)	-0.09 (-0.26, 0.09)	-0.06 (-0.14, 0.02)	-0.09 (-0.26, 0.09)	-0.09 (-0.26, 0.09)			
I^2 (95% CI)	38 (0-74) %	11 (0-74) %	38 (0-74) %	38 (0-74) %			
C20:4n-6							
n of studies	16	16	16	16	13	14	15
n of participants	2,049	2,049	2,049	2,049	1,805	1,885	1,969
SMD/MD (95% CI)	-0.19 (-0.27, -0.11)	-0.26 (-0.43, -0.09)	-0.18 (-0.26, -0.11)	-0.19 (-0.27, -0.11)	-0.14 (-0.24, -0.05)	-0.16 (-0.25, -0.06)	-0.18 (-0.26, -0.10)
I^2 (95% CI)	6 (0-55) %	51 (14-73) %	1 (0-53) %	10 (0-48) %	10 (0-49) %	12 (0-51) %	6 (0-56) %

Abbreviations: CI – confidence interval; MD – mean difference; PUFA – polyunsaturated fatty acids; RCT – randomised controlled trial; SMD – standardised mean difference.

Numbers of participants in crossover trials were double-counted for comparability with parallel trials.

2.5 Discussion

The current systematic review consisted in an outcome-wide appraisal of the effects of the Mediterranean diet on nutritional biomarkers, and meta-analytical pooling if at least five study-specific estimates were available. It identified 45 publications from 29 trials that cumulatively reported on 127 biomarkers or biomarker ratios, 25 of which were eligible for meta-analysis. Nineteen studies had an overall poor rating in the risk of bias assessment, which was primarily driven by RCTs not providing information on randomisation procedures in sufficient detail.

The key findings for the meta-analysed biomarkers were that relative to control diets the Mediterranean diet increased circulating β-carotene, lycopene, retinol, vitamin C, total MUFA, C18:1n-9c (oleic), total n-3, and C20:5n-3 (eicosapentaenoic) and C22:6n-3 (docosahexaenoic) fatty acids, and decreased circulating total SFA, n-6:n-3 ratio, C18:2n-6 (linoleic) and C20:4n-6 (arachidonic) fatty acids. The pooled effect sizes were small to moderate with absolute SMDs ranging from 0.18 to 0.67. Arachidonic acid was identified as the only biomarker for which the 95% prediction interval did not include the null, suggesting that this finding would be expected to be reliably replicated in future trials of similar designs as those included in the current review. Furthermore, arachidonic acid and β-carotene were the only biomarkers which were robust to all sensitivity analyses that explored the influence of randomised cross-over trials and pre-post studies without a control group. There were no statistically significant differences between the Mediterranean and control diets in circulating α-carotene, β-cryptoxanthin, α-tocopherol, total PUFA, individual even-chained SFAs, C16:1n-7c (palmitoleic), C18:3n-3 (α-linolenic), C22:5n-3 (docosapentaenoic) and total n-6 and C20:3n-6 (dihomo-γ-linolenic) fatty acids. There was evidence of a major positive publication bias for β-carotene and vitamin C, C14:0 and C22:5n-3 fatty acids, and a major negative publication bias for total PUFA and the ratio of n-6:n-3 fatty acids. The relative heterogeneity was substantial (I^2 >50%) for 21 out of the 25 biomarkers. The narrative synthesis of the evidence on biomarkers not eligible for meta-analysis identified decreased γ-tocopherol as an additional potential biomarker of the Mediterranean diet.

2.5.1 Strengths and limitations

This systematic review is the first appraisal to date of the evidence on nutritional biomarkers as compliance measures of Mediterranean diet interventions. Its strengths include a

comprehensive search strategy, large numbers of included studies and outcomes, and a detailed meta-analysis appropriately incorporating within-person trials.[85,86] The risk of publication bias was assessed using a novel method, the LFK index, which consists in evaluation of the symmetry of the distribution of study-specific estimates and does not rely on frequentist testing. It eliminates the subjectivity inherent in the visual assessment of funnel plots, and it has been shown to outperform the Egger's test in terms of the overall predictive performance and sensitivity in detecting publication bias, particularly with small numbers of studies.[84]

This systematic review had several weaknesses. Seven out of the 45 included publications were missed in the primary search of three databases. The search strategy was designed for a broader search than the topic of the current review, covering a wider range of dietary patterns, study designs and types of biomarkers. This may have decreased its sensitivity to identify experimental designs on the effects of the Mediterranean diet on nutritional biomarkers. The included trials were heterogeneous in terms of the definition of the Mediterranean diet, the type of control diet, the methods of implementation of the dietary interventions and their duration, characteristics of study participants, and, for fatty acids, the blood fraction of measurement. This reduced the comparability of the results in the narrative synthesis and likely increased the statistical heterogeneity in the meta-analysis. The assessment of consistency of the results between trials was limited by the low numbers of studies reporting on most of the biomarkers. Selective reporting on fatty acids was evident as most of the trials measured a full fatty acid profile, however, all measured fatty acids were reported on only by one study.[108]

The risk of bias assessment was hampered by the inadequate reporting of most of the RCTs on the randomisation procedures and concealment to allocation, leading to an overall "poor" rating in the NHLBI Study Quality Assessment Tool for 18 out of the 25 RCTs. It should be noted that these aspects of the study design may have been appropriately implemented in some of the trials, and the true quality of the studies included in the review may have in fact been higher. Of note, deviations from the randomisation protocol in recruitment of the participants have been confirmed by the PREDIMED trial[112–115] and in the study of the Pontifical Catholic University of Chile.[81,82,116–120]

There were several analytical limitations. The number of data points for some of the meta-analytical comparisons was low (n = 5 for five outcomes) which limited the assessment of heterogeneity, sensitivity analyses and the statistical power. I used SMDs as the unit of the effects of the Mediterranean diet in the narrative synthesis and meta-analysis to facilitate the

comparison between different biomarkers.[135,136] However, the SMD depends on both the average treatment effect and its standard deviation. Thus, between-trial variation in the spread of treatment effects decreases the comparability of the effect sizes on the SMD scale.[137] Moreover, while the meta-analysis on the SMD scale technically allowed for pooling of the results for fatty acids measured in different blood fractions and expressed in different units, it could not fully account for the variation introduced by the heterogeneity of biological specimens and units.[80,138,139] For example, the largest effect size for linoleic acid (C18:2n-6; pooled SMD = -0.31) was reported by a trial which assayed this biomarker in triglycerides in absolute concentrations (SMD = -2.21).[124] This would be expected to result in a larger magnitude of effect compared to expressing the concentration as the proportion of total fatty acids,[138] which was the predominant metric in the meta-analysed trials. The studies reporting the second[87] (SMD = -1.62) and the third[89] (SMD = -0.61) largest effect size for C18:2n-6 assayed the fatty acid profile in LDL cholesteryl esters. Relative concentrations of this fatty acid in cholesteryl esters are nearly two-fold higher than in plasma whereas the population spread is similar in both compartments.[80] Thus, larger SMDs were more likely to be detected in the former compartment. Consistent with these observations, I found evidence of heterogeneity of the pooled estimates for C18:2n-6 when stratified by blood fraction. There was no evidence of an effect of the Mediterranean diet in studies which assayed total plasma fatty acids and there was evidence of a negative effect in all other blood compartments combined. Of note, the risk of bias assessment on the SMD scale may have increased the rate of false-positive results[140], however, the influence of the SMD metric compared to the weighted mean difference on the LFK index has not been evaluated.[84]

A further analytical limitation was the combined use of post-intervention differences in means between the intervention and control diets and the differences in change from baseline. In principle, these two measures should not be combined while using standardised effect sizes because their standard deviations, and hence the SMD values, depend on different sources of variability. The latter is influenced by within-person variation which is ignored by the former.[141] However, previous empirical evidence suggests that combining standardised follow-up and change data in meta-analyses of continuous outcomes yields valid results.[79] There are no contraindications to include the two measures in meta-analysis of weighted mean differences[141] which I used as sensitivity analysis. It largely confirmed validity of the results on the SMD scale, showing no impact on inference for all biomarkers except for alpha-linolenic acid (C18:3n-3).

Estimations of SMDs in most of the within-person trials required additional assumptions. Out of the ten included trials, only three pre-post studies without a control group reported coefficients of change readily convertible into the SMDs[92,121] or the summary statistics required for estimation of within-person correlations of biomarkers.[65,92] The remaining within-person trials required the use of imputed correlation coefficients, however, the impact of correlation values on the main meta-analytical results was negligible for most outcomes. Moreover, the assumption of lack of carry-over effects in cross-over RCTs may have been violated.[85]

2.5.2 Comparison with previous research

Biomarkers measured by studies included in the current review were either aimed at assessing adherence to different aspects of the Mediterranean diet or the purpose of their measurement was not specified. Increased intake of fruit and vegetables was a stated goal of all interventions which provided details on the Mediterranean diets used. Most circulating carotenoids (α-carotene, β-carotene, β-cryptoxanthin and lutein but not lycopene) and vitamin C were previously reported to increase in response to increased intake of these foods.[135] By contrast, in the current meta-analysis only levels of β-carotene, lycopene and vitamin C were found to be higher in the Mediterranean diet interventions compared to control diets. The results for the remaining carotenoids may have been null due to limited numbers of studies (n = 5), heterogeneity of comparator groups and heterogeneity of dietary sources of these compounds. The positive result for lycopene could be explained by the emphasis of the Mediterranean diet on lycopene-rich foods,[68] increased absorption of this carotenoid with concurrent consumption of added fat[142,143] and increased conversion to more bioavailable isomers by cooking with olive oil, onion and garlic.[144]

The result of higher serum retinol following Mediterranean diet interventions does not have a clear interpretation with regards to potential mechanisms. Changes in dietary intakes of retinol or carotenoids would not be expected to meaningfully impact this biomarker in the context of adequate status of vitamin A in the study populations.[145] Moreover, the Mediterranean diets were likely to decrease intakes of retinol given the reduced consumption of animal-based products, as supported by estimates from dietary self-report in one of the included trials.[99] Beyond the possibility of a false-positive finding in the current review, evidence from non-human animal models suggests that conversion from provitamin A carotenoids to retinol is upregulated by high dietary content of MUFAs and n-3 PUFAs.[146] These characteristics are

consistent with the Mediterranean diet, and they may have potentially contributed to the comparatively higher serum retinol. The interpretation of the results for tocopherols is likewise challenging. There was a null pooled effect on circulating α-tocopherol and there were null or negative effects on γ-tocopherol in individual studies. This contrasts with higher estimated intakes of vitamin E in Mediterranean diet intervention arms compared to control diets.[64,99] The evidence on the relationship between estimated habitual intakes of tocopherols and their circulating concentrations is conflicting, with both null associations and up to moderate positive correlations reported previously in the literature.[44,147–150] Several factors could contribute to the lack or negative effects on circulating tocopherols despite increased dietary vitamin E in the context of Mediterranean diet interventions: decreased systemic oxidative stress and inflammation,[151] changes in plasma lipoproteins,[99] and higher PUFA intake requiring utilisation of tocopherols for prevention of oxidation of fatty acids.[152]

Interpretation of the results for SFA requires consideration of dietary sources and metabolic processes. The pool of circulating SFAs is predominantly made up of C16:0 (palmitic) and C18:0 (stearic) fatty acids, which account for ~20-45% of total fatty acids depending on the blood fraction.[80] The meta-analytical result of lower total circulating SFA after Mediterranean diet interventions was likely driven by changes in these major even-chain fatty acids which are influenced by de-novo lipogenesis and associated dietary factors (e.g. high-glycaemic load), and dietary intakes of these compounds.[153] However, the pooled results for individual even-chain SFAs and palmitoleic acid (MUFA) were only directionally but not statistically significantly negative, albeit based on an incomplete overlap with the studies reporting on total SFA. This apparent discrepancy could potentially be attributed to higher heterogeneity in responses of individual even-chain SFAs than that of total SFA or the compositional nature of the data. All but one trial[108] contributing to these results reported on fatty acids as proportions of total fatty acids, and thus the decrease in total SFAs might have been partly driven by increases in MUFAs and n-3 PUFAs. Speculatively, influence of the Mediterranean diet on the low concentration odd-chain[154] and very-long chain[155] saturated fatty acids may have collectively meaningfully contributed to the negative result for total SFA. Effects of Mediterranean diet interventions on these compounds remain unclear.[108]

For MUFA, higher circulating C18:1n-9c (oleic acid) and total MUFA can be attributed to consumption of olive oil.[156] External validity of this finding for assessment of adherence to the Mediterranean diet is likely limited to dietary interventions or settings with high habitual consumption of olive oil. Circulating C18:1n-9c can be synthesised endogenously via

desaturation of C18:0 (stearic acid) or derived from diet.[80] Main sources and dietary correlates of biomarkers are heterogeneous across populations, and can include olive oil in the Mediterranean region of Europe and meat in central Europe and Scandinavian countries.[157]

For PUFA, higher n-3 PUFA status was detected following Mediterranean diet interventions compared to control diets. This included total n-3 PUFA, and C20:5n-3 (eicosapentaenoic) and C22:6n-3 (docosahexaenoic) fatty acids which likely reflected the increased intake of fish and seafood.[66] Concurrently, there was a decrease in n-6:n-3 ratio, and C18:2n-6 (linoleic) and C20:4n-6 (arachidonic) fatty acids. Decreased consumption of dietary C18:2n-6 was a plausible driver of the effect on its circulating levels,[158,159] particularly given the lack of endogenous synthesis.[80] However, intakes alone may account for a minority of variation in blood concentrations,[52] suggesting a sizeable contribution of metabolic control of the concentrations of C18:2n-6 independent from dietary content.[160–162] The evidence on determinants of circulating C20:4n-6 is limited[163] though conversion from C18:2n-6 has been confirmed to be negligible.[164,165] Evidence from supplementation trials suggests presence of a dose-response relationship between ingested and circulating C20:4n-6.[163,166] Thus, the low intakes of animal-based foods[163] in the Mediterranean diet are consistent with the current meta-analytical result. Of note, a cross-over RCT found that olive oil (25 mL/day) decreased LDL content of C20:4n-6 with a concurrent increase in C18:1n-9c compared to washout periods of habitual diet.[167] This result raises the possibility that olive oil as part of Mediterranean diet interventions may have contributed to the consistent negative result for C20:4n-6 identified in the current review.

The n-6:n-3 ratio has been a longstanding proposition[168] to quantify the balance between dietary n-6 and n-3 PUFAs as a risk factor for a wide range of chronic non-communicable diseases.[169] It was based on the notions of purported prothrombotic, proconstrictive, and proinflammatory effects of high intakes of n-6 PUFAs and opposing actions of n-3 PUFAs.[169] The n-6:n-3 ratio has been criticised on several grounds leading to calls for abandoning its use, including assigning equal weights to the 18-carbon essential fatty acids and long-chain PUFAs, the oversimplified or incorrect assumptions that high n-6 intakes are deleterious and high n-3 intakes are beneficial for disease prevention, and not accounting for absolute intakes of n-3 and n-6 PUFAs.[67,169–171] Utility of the circulating n-6:n-3 ratio as a dietary biomarker has not been evaluated beyond reports of correlations with estimates of this metric from FFQs which were of similar magnitudes as for other biomarkers of PUFA.[172–174] In the current review, the negative effect of interventions on the n-6:n-3 ratio likely reflected the above described changes

in individual PUFAs, and in particular the most abundant C18:2n-6.[80] However, the ratio would not be expected to be a robust dietary biomarker given that (i) it does not account for absolute intakes or tissue levels of fatty acids from either group, (ii) a virtually infinite number of combinations of individual PUFA molecules can yield the same value of this metric, and (iii) manipulating the ratio can be achieved through altering intakes of either or both components simultaneously.

One previous systematic review without meta-analysis summarised the relationship between the Mediterranean diet and polyunsaturated fatty acids.[133] The main finding with regards to the interventional evidence was that approximately two thirds of the RCTs reported increases in tissue concentrations of any n-3 fatty acids or a decrease in the n-6:n-3 ratio, and a high level of variability between studies in responses of the n-6 fatty acids. These patterns are largely consistent with the findings of the current work which improves on the earlier systematic review[133] by evaluating each fatty acid or fatty acid grouping separately and conducting a formal quantitative synthesis.

2.5.3 What this study adds and implications of this research

All trials included in this systematic review considered nutritional biomarkers as univariate measures of compliance. The Mediterranean diet is, however, a multimodal dietary intervention, and as such single biomarkers lack face, content, and construct validity for objective assessment of adherence to this dietary pattern. This notion is supported by the results of the current meta-analysis which yielded mostly small effect sizes for individual biomarkers. Application of nutritional biomarkers in trials of dietary patterns at large should move towards combining multiple analytes to develop statistical models that can robustly classify participants into their intervention arms. The current review identified a number of candidate biomarkers, primarily circulating carotenoids and fatty acids, that could be considered for inclusion in such models. These findings can help guide selection of biomarkers of compliance to Mediterranean diet interventions in future trials. Additionally, they can serve as a reference for evaluating in observational settings the validity of self-reported adherence to the Mediterranean[175] diet against nutritional biomarkers.

Research on dietary patterns other than the Mediterranean diet or analytes other than nutritional biomarkers provides insights into the application of the multi-biomarker approach. For example, the SYSDIET study was a partial-feeding RCT which compared the effects on

cardiometabolic risk factors of a healthy Nordic diet with a diet based on mean nutrient intakes in Nordic countries.[176] The trial measured several nutritional biomarkers post-intervention as pre-specified biomarkers of selected components of the intervention: plasma phospholipid fatty acids C22:5n-3 and C22:5n-3 (eicosapentaenoic and docosahexaenoic) for fish, C18:3n-3 (α-linolenic) for canola oil, serum C15:0 (pentadecanoic) fatty acid for dairy fat, plasma β-carotene for vegetable, and alkylresorcinols for whole grains. Composite biomarkers of adherence were derived in 154 participants using principal component analysis and weighted sum of ranks of the above biomarkers, yielding highly significant ($p < 0.001$) SMDs of approximately 1.1 and 1.7, respectively (estimated from values read from box-plots[74]).[176] Metabolomic profiling, which consists in building multivariable models based on tens or hundreds of metabolites, has been the mainstay of metabolomic research with examples of application to the Mediterranean diet, including in the context of trials.[60] In a subgroup analysis in one of the centres of the PREDIMED trial, urinary metabolomic profiles at 1- or 3 years post-randomisation were able to correctly classify 93%, 85% and 68% of participants to their respective intervention arms of the Mediterranean diet with either olive oil or nuts and the control lower-fat diet.[177] The potential of urinary metabolomics to develop robust biomarkers of adherence to dietary patterns has been demonstrated in an inpatient feeding crossover RCT of four levels of concordance with the World Health Organization healthy eating guidelines.[178] The same analytical approaches of combining multiple analytes in metabolomics research could be extended to nutritional biomarkers measured in a targeted manner.

2.5.4 Conclusions

This systematic review identified circulating β-carotene, lycopene, retinol, vitamin C, and several mono- and polyunsaturated fatty acids or fatty acid groupings as candidate biomarkers of adherence to the Mediterranean diet in interventional study designs. Lower concentrations of arachidonic acid and higher levels of β-carotene relative to control diets were the most consistent findings. Individual nutritional biomarkers were unlikely to capture the complexity of whole-diet interventions, but rather they may have reflected their individual components. Neither of the identified trials considered combining separate analytes into composite biomarkers which should be considered by future research.

Chapter 3

Nutritional biomarker scores of dietary patterns: derivation and associations with incident type 2 diabetes

Abstract

Background: Mediterranean diet scores (MDS), alternative Healthy Eating Index-2010 (aHEI-2010), and Dietary Approaches to Stop Hypertension (DASH) have been inversely associated with incidence of type 2 diabetes (T2D), but the well-recognised measurement error of dietary self-report limits the quality of evidence. The largest nutritional study of incidence of T2D, the InterAct case-cohort study nested within the European Prospective Investigation into Cancer and Nutrition (EPIC-InterAct), has previously reported null results on aHEI and DASH, which contrasts with the overall body of evidence, and a modestly inverse association for MDS.

Methods: I derived biomarker scores of MDS, aHEI-2010 and DASH estimated from food frequency questionnaires and diet histories in the EPIC-InterAct subcohort using bootstrap-selection enhanced elastic net regression based on 49 circulating nutritional biomarkers: 37 plasma phospholipid fatty acids, six carotenoids, vitamins C and 25(OH)D, iron status biomarkers and serum cations. The biomarker scores were calculated as linear predictions from the regression models. Participants in the derivation samples had individual biomarker score equations re-derived using the leave-one-out principle to increase independence from self-reported diet. I then evaluated associations of the biomarker scores and self-reported dietary patterns with incident T2D using country-specific Prentice-weighted Cox regression with pooling via random-effects meta-analysis. Up to 21,549 participants, including 9,181 incident T2D cases, were available for these analyses out of 27,779 participants of the original case-cohort sampled from a cohort of 340,234 individuals in eight countries.

Findings The biomarker scores were modestly positively correlated with their respective dietary patterns in the subcohort (r range: 0.31-0.34) and the non-subcohort incident T2D cases (r range: 0.27-0.33) Adjusting for potential confounders, including measures of adiposity, the scores were inversely associated with incident T2D: the hazard ratios (95% confidence interval) per standard deviation of the scores were 0.85 (0.79-0.92) for MDS, 0.75 (0.68-0.82)

for aHEI-2010, and 0.83 (0.77-0.90) for DASH. Corresponding results based on dietary self-report were 0.90 (0.86-0.95), 0.96 (0.91-1.02) and 0.96 (0.91, 1.02), respectively.

Conclusions: These results suggest utility of combining nutritional biomarkers into composite measures to objectively assess adherence to dietary patterns. Using such biomarker scores as exposure variables in investigations of T2D incidence may yield inverse associations in instances when results based on dietary self-report are null. However, validity of the biomarker scores as measures of adherence to dietary patterns is unclear, and further work is required to develop and validate such biomarkers in trials of dietary pattern interventions.

3.1 Background

Previous research has indicated that healthy pre-defined dietary patterns are inversely associated with incidence of T2D.[26] Adjusted summary hazard ratios (95% CI) from prospective cohort studies comparing extreme categories of adherence were 0.79 (0.74-0.85) for any healthy dietary pattern, 0.85 (0.76-0.95) for the Mediterranean diet scores (MDS), 0.79 (0.73-0.85) for the Alternative Healthy Eating Index (aHEI), and 0.80 (0.73-0.88) for indices of the Dietary Approaches to Stop Hypertension (DASH).[6] The quality of evidence was low-to-moderate.[6] Subjective assessment of habitual diet and modest effect sizes were some of the key factors which decreased the rating of evidence.[6]

The European Prospective Investigation into Cancer and Nutrition (EPIC)-InterAct case-cohort study is the world's largest study of incidence of T2D which has accumulated nearly 4 million person-years of follow-up.[179] It has been an important contributor to quantitative synthesis of the evidence on healthy pre-defined dietary patterns and risk of new-onset T2D, accounting for ~15-25% of weight in random-effects meta-analyses.[180,181] For the MDS, EPIC-InterAct has reported an inverse association[182] with a modest effect size similar to the above pooled meta-analytical estimate.[6] However, there is uncertainty whether the association was driven by the overall dietary pattern as opposed to a small proportion of its individual components. In a secondary analysis, the MDS was recalculated by leaving out a single food or nutrient component. The inverse association was attenuated to the null when excluding in turn meat and meat products or alcohol, and it was substantially weakened after exclusion of olive oil.[182]

For aHEI and DASH, EPIC-InterAct has reported null results[183] which were at variance with the overall body of evidence suggesting inverse associations with incident T2D.[180,181] Furthermore, short-to-midterm randomised interventions have shown that the DASH diet decreases fasting insulin (but not fasting glucose and HOMA-IR),[184] and this effect may be independent from changes in body weight.[185] Evidence for the DASH diet was tested in an RCT which demonstrated its blood pressure-lowering effects,[186] and systolic blood pressure is a robust, potentially causal risk factor for new-onset T2D.[187] All studies included in the published meta-analysis on aHEI and DASH and incidence of T2D other than EPIC-InterAct have been conducted in the USA, and they have mostly reported inverse associations.[180,181] The aHEI score has been originally designed for assessment of preventative potential of diet against chronic diseases in the USA,[188] and thus its generalisability as a predictor of T2D in other countries may be limited. However, recent investigations from Singapore[189] and Australia[190]

suggest that it may also be inversely associated with incidence of T2D in other geographical regions. It remains unclear whether the null associations previously detected in EPIC-Interact for aHEI and DASH[183] represent true causal effects in European populations.

Measurement error of self-reported diet[22] should be considered as a potential factor which may have decreased the validity of associations between dietary patterns and incidence of T2D.[26] Assessment of adherence to dietary patterns could be improved by developing objective measures of exposure, such as nutritional biomarkers. Dietary patterns, however, lack biologically plausible single biomarkers and, therefore, necessitate combining multiple analytes into composite biomarker scores.[50,51] Such scores could potentially complement or replace traditional dietary self-report in evaluating diet-disease associations. The EPIC-InterAct study measured a broad range of circulating nutritional biomarkers in ~80% of its participants, including plasma carotenoids, phospholipid fatty acids, vitamin C, vitamin D metabolites, and serum magnesium, calcium, and iron status markers. I hypothesised that combinations of some of these biomarkers could be used to jointly characterise adherence to dietary patterns, and that such composite biomarkers would be inversely associated with incidence of T2D.

3.2 Aim

The aim of this work was to re-evaluate the associations between dietary patterns and incident T2D in the EPIC-InterAct study by using biomarker-based exposure assessment in lieu of dietary self-report. My objectives were to derive nutritional biomarker scores of adherence to MDS, aHEI-2010 and DASH indices, and to test the associations of the biomarker scores with incident T2D.

3.3 Methods: the EPIC-InterAct study

EPIC-InterAct is a case-cohort study of T2D embedded within eight of the ten countries in the EPIC study: Denmark, France, Germany, Italy, the Netherlands, Spain, Sweden, and the United Kingdom.[179] Recruitment and baseline data collection took place during 1991-1998 in two to six centres per country. There were 12,403 individuals with ascertained and verified incident T2D over 3.99 million-person years of follow-up from a cohort of 340,234 participants with stored blood samples in the EPIC study. Incident T2D cases were ascertained from several

sources, including self-report, primary and secondary care registers, drug registers, hospital admissions, and mortality data, as described in detail previously.[179] All self-reported cases were confirmed using at least one additional method of verification. Linkage with diabetes and drug registers was used for adjudication in Denmark and Sweden without reliance on self-report. Participants were followed-up until 31st of December 2007 and cases were censored at the date of diagnosis or death. Information on the vital status was collected via linkage with regional or national mortality registries.

From the cohort of 340,234 participants, a centre-stratified subcohort was assembled by randomly selecting 16,835 individuals. A total of 16,154 participants remained in the subcohort after exclusions (n = 548 with prevalent diabetes; n = 133 with uncertain diabetes status). **Figure 4.1** in the subsequent chapter depicts the process of assembly of the case-cohort.

3.3.1 Nutritional and metabolic biomarkers

Venous blood samples were collected at varying times of the day from fasted and non-fasted participants. Plasma and serum samples were stored in liquid nitrogen (up to −196°C), except for Umeå where freezers were used (−80°C). The following nutritional biomarkers were measured: six plasma carotenoids (α-carotene, β-carotene, lycopene, lutein, zeaxanthin, β-cryptoxanthin), 37 phospholipid fatty acids, vitamins C and 25-OH D (25-OH D_2, 25-OH D_3 and epimers), and serum transferrin, iron, magnesium and calcium. Vitamin C measurements were not undertaken in Swedish participants due to unavailability of samples stabilised with meta-phosphoric acid (n = 5,401), and carotenoids and vitamin D metabolites were not assayed in participants from the Malmo recruitment centre (n = 3,556). In the Umeå centre, iron, transferrin, magnesium and calcium were not measured due to lack of availability of serum samples (n = 1,845).

High performance liquid chromatography with ultraviolet light detection was used to measure plasma vitamin C and carotenoids (α-carotene, β-carotene, lycopene, lutein, zeaxanthin, β-cryptoxanthin).[49,191,192] Aliquots were stabilised with meta-phosphoric acid for the vitamin C assay and butylated hydroxytoluene for measurement of carotenoids. Coefficients of variation were between 4.2-4.5% for vitamin C and ranged from 2.7 to 6.7% for carotenoids. High-to-moderate reproducibility after long term storage was observed in EPIC-Norfolk.[193,194] Individual fatty acids were measured as molar percentages of total fatty acids using automated,

high throughput gas chromatography at the Medical Research Council Human Nutrition Research (Cambridge, United Kingdom).[195,196] Plasma 25(OH)D metabolites were measured using liquid chromatography–tandem mass spectrometry at a Vitamin D External Quality Assessment Scheme certified laboratory (Vitas AS, Oslo, Norway).[197]

The following serum nutritional and metabolic biomarkers were measured at the Stichting Ingenhousz Laboratory (Etten-Leur, Netherlands) using Roche Hitachi Modular P: serum ferritin, transferrin, iron, calcium and magnesium were measured, total cholesterol (TC), high-density lipoprotein cholesterol (HDL-C), triglycerides (TG), high-sensitivity C-reactive protein (hs-CRP), alanine aminotransferase (ALT), aspartate aminotransferase (AST) and gamma-glutamyl transferase (GGT). Plasma was used for assays of lipids, hs-CRP and liver enzymes in the Umeå centre. Haemoglobin A1c (HbA1c) was measured in the erythrocyte fraction using the Tosoh-G8 analyser (Tosoh Bioscience, Japan). Laboratory staff were blinded to case status of participants, and samples were processed in a random order.

I excluded from the analysis 10 fatty acids with mean concentrations <0.05 mol% in the subcohort, leaving 27 fatty acids available for analysis, and all vitamin D forms other than non-epimeric 25-OH D_3 due to large proportions of participants with concentrations below the detection limit.[197]

3.3.2 Dietary patterns

The MDS, aHEI-2010 and DASH scores were constructed with minor modifications to accommodate unavailability of data or between-country heterogeneity in assessment of some of their components in EPIC-InterAct, as described previously.[182,183] The earlier work in EPIC-InterAct used the initial aHEI score[198] and I applied the updated aHEI-2010 score.[188] I reconstructed the EPIC-InterAct MDS and DASH indices[182,183] without any changes. Calculation of the dietary pattern indices was restricted to participants with plausible estimated energy intakes of 800-4,000 kcal/day in men and 500-3,500 kcal/day in women. Scoring of the dietary patterns is shown in **Table 3.1**.

Briefly, the MDS was a variation of the relative MDS indices by Trichopoulou et al.[199,200] The key changes were using an amended list of components and tertiles rather than medians to establish the cut-offs for scoring.[201] The MDS included assessment of intake of nine components: (positively scored) vegetables, legumes, fruits and nuts, cereal, fish and seafood,

olive oil, moderate alcohol use, and (negatively scored) meat and meat products, and dairy products. The scoring cut-offs for food groups were based on tertiles of the EPIC-InterAct subcohort distributions of energy-standardised estimated intakes. Thus, each food group was assigned into a tertile category of 0, 1 or 2 for the adherence to each component and summed, producing a range between 0 and 18. The olive oil intake was scored as 0 in non-consumers, 1 for estimated energy-standardised intake below the median of subcohort consumers, and 2 for estimated intake at or above the median. For alcohol, the estimated intakes within sex-specific ranges of moderate consumption were scored as 2 points and the estimated intakes outside of these ranges were assigned 0 points. All components were summed, producing an integer range between 0 and 18.[201]

Compared to the original aHEI-2010 score,[188] I excluded the components on intakes of trans-saturated fat and sodium due to unavailability of data in EPIC-InterAct. I replaced the assessment of intake of long-chain n-3 PUFA with fish and shellfish, as previously used in the initial version of the aHEI score.[198] The intake of whole grains was assessed solely based on estimated intakes of non-white bread. The modified aHEI-2010 included assessment of nine components: (positively scored) vegetable, fruits, whole grains, nuts and legumes, fish and shellfish, % of energy from PUFA, and (negatively scored) soft drinks and juice, red and processed meat, and alcohol use. Each component was scored between 0-10 points using pre-defined cut-offs for minimum and maximum points. For intermediate values of estimated intakes, points were assigned proportionally to intakes. All nine components were summed, producing a range between 0 and 90.

The DASH scoring system was based on the index developed by Günther et al.[202] Individual components were scored between 0-10 in the same manner as those from the aHEI-2010 score with the exceptions of the grains and dairy components. Total intakes of both food groups were scored positively between 0-5. Participants received additional 5 points if their estimated intake of cereal fibre was in the top fifth of subcohort distribution and if their estimated intake of dairy fat was in the bottom fifth. Remaining components included: (positively scored) vegetable, fruits, sources of plant proteins, and (negatively scored) sources of animal proteins, added fats and oils, and sweets.

Table 3.1 Scoring of dietary pattern indices in the EPIC-InterAct study

Component	Range of points	Minimum points	Maximum points
MDS (g/1,000 kcal)*			
Vegetable	0-2	<57.6	>100.3
Legumes	0-2	<0.49	>6.37
Fruits and nuts	0-2	<66.0	>133.8
Cereals	0-2	<81.3	>113.5
Fish and seafood	0-2	<9.48	>20.45
Meat and meat products	0-2	>59.8	<40.8
Dairy	0-2	>194	<102
Olive oil	0-2	Non-consumers	>6.85
Ethanol (g/day)	0 or 2	Intake outside of ranges for maximum points	Men: 10-50 Women: 5-25
aHEI-2010 (servings)‡			
Vegetable	0–10	0/day	≥5/day
Fruits	0–10	0/day	≥4/day
Whole grains	0–10	0/day	Men: 90 g/day Women: 75 g/day
Soft drinks or juice	0–10	≥1/day	0/day
Nuts, seeds, and legumes	0–10	0/day	≥1/day
Red & processed meat	0–10	≥1.5/day	0/day
Fish and shellfish	0–10	0/day	≥2/day
PUFA, % of energy	0–10	≤2	≥10
Alcohol	0–10	Men: ≥3.5 /day Women: ≥2.5 /day	Men: 0.5-2/day Women: 0.5-1.5/day
DASH (servings) †			
Cereals	0–5	0/day	≥6/day
Cereal fibre	0–5	No grain intake	Subcohort Q5 (≥12.1 g/day)
Vegetable	0–10	0/day	≥4/day
Fruits	0–10	0/day	≥4/day
Dairy	0–5	0/day	≥2/day
Dairy fat	0–5	No dairy intake	Subcohort Q1 (≤7.4 g/day)
Meat, poultry, fish, eggs	0–10	≥4/day	≤1/day
Nuts, seeds, legumes	0–10	0/day	≥4/day
Fats and oils	0–10	≥6/day	≤3/day
Sweets	0–10	≥10/week	≤5/week

MDS – Mediterranean Diet Score; DASH – Dietary Approaches to Stop Hypertension; Q – quintile; aHEI-2010 – alternative Healthy Eating Index - 2010

*Integer points were used. One point was assigned for estimated intakes between cut-offs for minimum and maximum points, except for ethanol for which either 0 or 2 points were assigned.

†Continuous points proportional to intake were used for estimated intakes between cut-offs for minimum and maximum points. Serving sizes were as follows: vegetable and fruit - 125 g; nuts and seeds - 30 g; grains - 50 g; dairy products: milk and yoghurt - 150 g, cheese - 45 g; meat, poultry, fish and eggs - 30 g; legumes - 100 g; fats and oils - 10 g; sweets: chocolate - 20 g, ice cream - 50 g.

‡Continuous points proportional to intake were used for estimated intakes between cut-offs for minimum and maximum points. Sodium and trans-fat components were excluded, and long-chain n-3 fatty acids component of the original score was replaced with fish and shellfish due to unavailability of data in EPIC-InterAct. Serving sizes were as follows: vegetable and fruit - 125 g; soft drinks or juice - 250 g; nuts and seeds - 30 g; legumes - 100 g; red meat – 120g; processed meat – 30 g.

3.3.3 Covariates

Questionnaires and physical examination were used at baseline to collect standardised information on covariates, including sociodemographic, medical and health behavioural factors and anthropometry. Weight, height and waist circumference were measured at baseline in all EPIC centres, except for Umeå, Sweden, where waist circumference was not measured (n = 1,845).[179] Subgroups of participants from France and the Oxford (UK) centre had self-reported anthropometry. Physical activity was assessed using a validated questionnaire.[203] Self-reported diet was measured using country-specific, validated food frequency questionnaires or diet histories. Estimation of intake of foods, energy and nutrients was harmonised across the EPIC cohorts.[204,205] Information on family history of T2D was not collected in Italy, Spain, and Oxford and Heidelberg (Germany) centres. Information on current use of vitamin or mineral supplements was not collected in Heidelberg. Covariates had <30% missing data within countries except for Germany with 45% of missing data for vitamin/mineral supplement use and family history of T2D. The prevalence of missing information for all countries combined was the highest for family history of T2D (22%), and otherwise up to 9% (dietary supplements), with several covariates having complete data for all participants.

3.3.4 Analytical samples and exclusions

I excluded participants from the Swedish Malmo recruitment centre (n = 3,556) from all analyses due to unavailability of measurements of six plasma carotenoids as objective markers of intake of fruits and vegetables.[49] Further participants were excluded because of incomplete data required for calculation of biomarker scores: 2,674 for aHEI-2010 and 2,930 for MDS and DASH. Between 21,293 and 21,549 participants were available for the analyses of prospective associations between the biomarker scores and T2D, including 12,768-12,920 subcohort participants. There were 552-549 incident cases in the subcohort as a design feature of the case-cohort study.

For derivation of the biomarker scores, I restricted the analytical sample to the subcohort (n = 16,154) and excluded participants from Malmo (n = 1,929). Additional exclusions were due to lack of self-reported diet or implausible estimated energy intakes (further 340 participants), missing data on nutritional biomarkers (further 1,456 participants) and country-specific outliers in biomarker concentrations (further 768 participants). Serum ferritin concentrations higher than 1,000 μg/L was a further pre-specified exclusion criterion, however, no participants had

such levels after removing the outliers from the dataset. Seven participants from Spain had missing data required for calculation of the DASH score.

For derivation of multi-country biomarker scores, I excluded participants from Umeå due to an incomplete set of nutritional biomarkers (n = 899) and the health-conscious arm of the UK Oxford centre because of maximising recruitment of vegetarians (n = 200),[206] leaving between 10,562 and 10,569 participants. These subsamples were included in derivation of country-specific biomarker scores, for which a total of 11,641-11,648 participants were available: 453 in France, 1,537 in Italy, up to 3,083 in Spain, 865 and 187 in the UK in the general population and the health-conscious recruitment arms, respectively; 1,235 in the Netherlands, 1,686 in Germany, 892 in Sweden and 1,710 in Denmark.

3.4 Methods: statistical analysis

Stata 16.1 was used for all analyses except for quantile regression imputation which was done in R, version 4.0.2 (package imputeLCMD). For analyses involving statistical significance testing, two-sided $\alpha = 0.05$ was used. For descriptive statistics, two-sided $\alpha = 0.10$ was used to test for trend across quintiles to capture both statistically significant and marginal associations.

3.4.1 Pre-treatment of nutritional biomarkers and dietary pattern variables

Zero values in fatty acids (12% for C17:1 and otherwise <6% for 15 of the 26 remaining fatty acids) were assumed to be below the limit of detection and they were imputed using quantile regression imputation.[207,208] Fatty acids were then re-scaled to sum up to 100% within the set of fatty acids used in the analysis. All nutritional biomarkers and total cholesterol were natural logarithm-transformed to stabilise variance in subsequent analyses. Carotenoids were adjusted for total cholesterol using the residual method in order to account for their correlations with dietary carotenoids.[44,209] Vitamin 25-OH D_3 was residual-adjusted for seasonality via the cosinor model,[210] whereby the concentrations were regressed on sine and cosine functions of the day of the year with interaction by latitude of recruitment centre. Biomarkers not measured in the Umeå centre (vitamin C, iron, transferrin, magnesium and calcium) had missing values replaced with subcohort medians.

Dietary patterns were residual-adjusted for potential confounders of the associations between dietary patterns and nutritional biomarkers. Adjustments were made separately within each country for the following covariates: age at blood draw (years, continuous), sex, recruitment centre, prevalent cancer, cardiovascular disease, hypertension and hyperlipidaemia; familial history of T2D, smoking status (never, former, current smoker), physical activity index (inactive, moderately inactive, moderately active, active), day of the year of the blood draw (sine and cosine function of the day of the year of blood draw), fasting status (<3, 3-6, >6 hours), current use of vitamin or mineral supplements, menopausal status (pre-, peri-, postmenopausal, bilateral oophorectomy), current hormone replacement therapy use, marital status (single, married or cohabiting, divorced or separated, widowed), educational attainment (none, primary school, technical or professional school, secondary school, post-secondary school education), current employment, estimated energy intake, body mass index (BMI; kg/m^2, continuous) and waist circumference (cm, continuous). Continuous covariates were Winsorised at 4 standard deviations (SD) below and above their mean in the subcohort. Missing covariate data were imputed by multiple imputation using chained equations in 10 datasets.[211] The imputation model included all the above covariates, MDS, aHEI-2010, DASH, nutritional biomarkers, total serum cholesterol, and the Nelson–Aalen cumulative hazard estimate for T2D taking into account country-specific sampling fractions as auxiliary variables. Upon confirming low variability between the imputed datasets in residual-adjusted dietary pattern scores (intraclass correlation coefficients >0.99), the means of the residual-adjusted values from the imputed datasets were used for derivation of the biomarker scores. These residual-adjusted dietary patterns were further residual-adjusted for country prior to derivation of multi-country biomarker scores.

3.4.2 Derivation of biomarker scores

I derived multi-country biomarker scores as the primary analysis and country-specific biomarker scores as a secondary approach. Unless a specific reference to the country-specific scores is made, biomarker scores refer to the multi-country ones throughout this chapter. Participants with at least one log-transformed biomarker value outside of the country-specific 25^{th} percentile minus 3 times the interquartile range (IQR) or the 75^{th} percentile plus 3 times the IQR were excluded from derivation of the biomarker scores.

I applied elastic net regression with bootstrap selection stability to select nutritional biomarkers jointly predictive of self-reported adherence to dietary patterns.[212–214] One hundred bootstrap samples were used. Within each sample, ten-fold cross-validation was used to select the λ and α penalties based on minimising cross-validated prediction errors. The α values were tested in 0.1 increments between 0.1 and 0.9. The boundaries of 0 and 1 were omitted to prevent, respectively, lack of variable selection and unstable selection of predictors in presence of multi-collinearity.[212] A grid of 100 λ values was tested per each α value. Biomarkers which were selected in ≥90 % of the bootstrap samples were included in the biomarker scores. Residual-adjusted dietary pattern scores were then regressed in the derivation samples on the bootstrap-selected predictors by means of ridge regression (elastic net regression with α penalty = 0) to estimate penalised coefficients for the biomarker terms.[215] The λ penalties were selected based on minimising cross-validated prediction errors in the empirical samples. Standardised coefficients from ordinary least squares regression were used for comparative presentation of the biomarkers scores to avoid the impact of differential tuning parameters on coefficient shrinkage between scores.

Linear predictions from the ridge regression models were used to calculate the biomarker scores in participants who were not included in the derivation samples, i.e., the non-subcohort incident T2D cases and subcohort participants excluded from derivation. For participants from the derivation samples, I applied the leave-one-out approach to reduce model overfitting. Each participant had their individual prediction equation re-estimated by excluding them from the derivation sample and re-running the ridge regression with the previously selected λ value, followed by calculation of linear prediction.

I assessed the performance of the biomarker scores based on the strength of their correlations with the corresponding dietary pattern indices without prior residual-adjustment. Country-specific Pearson correlation coefficients were estimated and pooled using random-effects meta-analysis.[78] Fisher's Z-transformation was used to estimate standard errors within countries. Correlations between the biomarker scores and components of dietary patterns were also evaluated.

3.4.3 Associations of biomarker scores with incident type 2 diabetes

Individual nutritional biomarkers were Winsorised at 4 standard deviations (SD) below or above the subcohort means and were then used to calculate the biomarker scores with the scoring algorithms developed as described above. The multi-country biomarker scores were standardised using the means and SD and categorised into quintiles based on the distributions in the subcohort. I performed Prentice-weighted Cox regression analysis with a robust variance estimator to estimate hazard ratios (HRs) for associations between the biomarker scores and incident T2D.[216] Country-specific HRs were estimated and pooled using random-effects meta-analysis, followed by calculation of the 95% confidence and prediction intervals.[78] Within the UK, analyses were stratified by recruitment arms from the general population and health conscious participants and combined meta-analytically with random-effects prior to pooling of country-specific results. The generalised least squares method for trend estimation was used to calculate the p trend values across quintile medians of biomarker scores.[217] Restricted cubic splines with five knots were used to assess potential non-linearity of the associations between biomarker scores and T2D incidence. Multivariate random-effects meta-analysis was used to pool the country-specific estimates of the spline parameters.[218]

The multivariable-adjusted model included the following covariates: age (as the underlying timescale), sex, recruitment centre, prevalent comorbidity (cancer, cardiovascular disease, hypertension, hyperlipidaemia), family history of T2D, smoking status (never, former, current smoker), physical activity index (inactive, moderately inactive, moderately active, active), seasonality (sine and cosine function of the day of the year of blood draw), fasting status (<3, 3-6, >6 hours), current use of vitamin or mineral supplements, marital status (single, married or cohabiting, divorced or separated, widowed), educational attainment (none, primary school, technical or professional school, secondary school, post-secondary school education), current employment, and in women, menopausal status (pre-, peri-, postmenopausal, bilateral oophorectomy) and current hormone replacement therapy use. A further model was fitted with adjustment for adiposity, i.e., BMI and waist circumference (main analytical model). Waist circumference was not available in Sweden and only BMI was adjusted for in this country. Continuous covariates were Winsorised at 4 SDs below or above the subcohort means. For comparison with the biomarker-based assessment, I estimated the HR for the association with incident T2D of standardised indices of dietary patterns estimated from self-report using the main analytical model.

Additional models explored the effects of adjustment for biomarkers constituting the biomarker scores and metabolic factors: blood lipids (total cholesterol, high density lipoprotein cholesterol, triglycerides), haemoglobin A1c, high-sensitivity C-reactive protein and circulating liver enzymes (alanine transaminase, aspartate transaminase, gamma-glutamyl transferase). Triglycerides were log-transformed and continuous covariates were otherwise entered into the models as linear terms in the original units. In the analysis with further adjustment for nutritional biomarkers constituting the scores, individual biomarkers were entered in turn into the models on the normal (non-log) scale as linear effects and quadratic fractional polynomial terms. Simultaneous adjustment for all such biomarkers was also undertaken. Only the quadratic terms which had statistically significant pooled effects in the analyses adjusting for individual biomarkers were included in these models to decrease the probability of model non-convergence.

Missing covariate data were imputed by country-specific multiple imputation using chained equations in 10 datasets.[211] I selected this number of datasets a priori based on computational efficiency considerations and confirmed its suitability based on the Monte Carlo errors for the main exposures.[211] The imputation model included all the covariates specified above for the Cox model, the biomarker scores, individual nutritional biomarkers, case status, subcohort membership, the Nelson–Aalen cumulative hazard estimate taking into account country-specific sampling fractions, and interaction terms as separate variables for the above pre-specified effect modifiers using predictive mean matching.[219,220] Female-specific covariates were imputed using models restricted to women. The p values for trend and non-linearity were calculated based on averaging the Z- and chi^2-statistic, respectively, from values obtained within each imputed dataset. In the analyses with restricted cubic splines, multivariate meta-analyses were performed separately in each imputed dataset. For graphing purposes, predicted values and the standard errors of the fitted values were calculated within each imputed dataset. Predicted values were averaged and the predicted standard errors were pooled by combining the within- and between imputation variances using Rubin's rules. I also performed complete-case analyses.

I examined multiplicative interactions of the biomarker scores with the following covariates: baseline age, sex, BMI, seasonality, fasting status, use of dietary supplements, physical activity and smoking using the adiposity-adjusted model (excluding waist circumference for interaction by BMI) and variable specifications as outlined above. The underlying time variable in the Cox model was switched from age to duration of follow-up when testing for effect modification by

age. Interactions by continuous and binary covariates were tested by meta-analytical pooling of the interaction coefficients. The hypotheses of joint equality of interaction coefficients to zero for non-binary categorical variables and seasonality modelled via the cosinor approach were tested with the Wald test using a pooled analysis without meta-analysis while removing waist circumference (unavailable in Sweden) from the main analytical model and adjusting for country and recruitment centre. In the multiply imputed analyses, this test for interaction was performed under the assumption of proportionality of between- and within-imputation variance.[221] Stratified estimates were calculated meta-analytically and presented if the p value for interaction was <0.05 in either the multiply imputed or complete-case analysis.

3.4.4 Sensitivity analyses

I performed several sensitivity analyses to assess robustness of the main findings. I repeated the derivation of biomarker scores with alternative analytical decisions to evaluate how they impact on the HR estimates, including: use of a single elastic net regression, calculation of biomarker scores with unpenalised ordinary least squares coefficients, application of a higher λ penalty according to the "one-standard-error rule" ($\lambda_{+1\ SE}$)[222] and lack of residual-adjustment of dietary patterns for potential confounders. The $\lambda_{+1\ SE}$ rule selects the λ penalty as the largest λ value that is within 1 SE of the minimum of the cross-validation function, as opposed to the optimal λ selected via cross-validation.[222] To evaluate the impact of exclusion of Swedish participants from derivation of the primary multi-country biomarker scores and substitution of unmeasured biomarkers in Sweden with subcohort medians, I assessed the associations of biomarker scores with incidence of T2D using country-specific scores and multi-country scores derived using only carotenoids and fatty acids (without exclusion of Swedish participants). Independence of the prospective associations from alcohol intake and dietary factors previously reported in EPIC-InterAct as drivers of the inverse association between MDS and incident T2D[182] was evaluated by adjusting the main analytical model for ethanol (g/day, restricted cubic splines with 4 knots), and ethanol, meat (g/day, continuous), olive oil (g/day) with adjustment for energy intake via the nutrient residual model.[223]

Time-varying effects were assessed by splitting follow-up time at 7 years and performing stratified analysis. Potential reverse causation bias was evaluated by separately (i) excluding the first 2 years of follow-up, (ii) excluding participants with prevalent major disease

conditions (cancer, myocardial infarction, or stroke), and (iii) excluding participants with baseline HbA1c concentrations ≥ 6.5% (48 mmol/mol).

3.5 Results

3.5.1 Background characteristics by country

Country-specific background characteristics and median biomarker concentrations of the analytical samples for derivation of biomarker scores are shown in **Table 3.2**. Approximately 63% of the participants were women, ranging from 47% in Denmark to 100% in France. MDS varied substantially between countries, with a median of 5 points in Sweden and 11 points in Spain, Italy and the UK health-conscious recruitment arm on a scale up to 18 points. The aHEI-2010 and DASH scores exhibited little variation between countries except for the UK health-conscious participants who consistently had notably higher median points. Use of unspecified vitamin and mineral supplements was low in Italy and Germany (12% and 16% respectively) and over 40% in Spain, the UK and Sweden. There was marked variation in the fasting status at blood draw, with participants from the UK, the Netherlands and Denmark being almost exclusively non-fasted, and participants from Sweden being exclusively fasted, and mixed fasting status in the remaining countries. Sizeable differences were observed in median concentrations of some biomarkers across countries. For serum carotenoids and ferritin, 2-3-fold differences were observed between countries with the lowest and the highest concentrations. Omega-3 PUFAs varied 1.5-2-fold between countries with the lowest and the highest median concentrations; whereas omega-6 PUFAs where largely similar, in particular the most abundant linoleic acid (C18:2). For elaidic trans-fatty acid (C18:1n9t), a multimodal distribution of country medians was observed, with concentrations in the UK, the Netherlands and Sweden in the range of 0.35-0.40 mol% and 0.13-0.20 mol% in the other countries. With minor exceptions, medians of the remaining fatty acids were similar across the EPIC-InterAct countries.

Table 3.2 Background characteristics and median concentrations of biomarkers in biomarker derivation samples: the EPIC-InterAct subcohort*

Variable	France	Italy	Spain	UK-GP	UK-HC	Netherlands	Germany	Sweden	Denmark
Numbers of participants	453	1,537	3,083	865	187	1,235	1,686	892	1,710
Background characteristics									
Age, years	55	50	48	58	48	53	50	50	56
Women, %	100	66	63	58	81	83	60	52	47
MDS, points (0-18 scale)	10.0	11.0	11.0	8.0	11.0	7.0	7.0	5.0	7.0
aHEI-2010, points (0-90 scale)	43	39	44	43	50	41	38	36	42
DASH diet, points (0-80 scale)	43	41	43	38	47	38	38	42	37
Moderately active or active, %	45	33	29	34	48	65	47	47	60
Current smokers, %	10	26	27	15	5	27	19	23	34
Tertiary education, %	41	14	12	16	58	21	36	20	21
Dietary supplement use, %	23	12	41	47	59	36	16	46	74
>6 hours of fasting, %	48	80	66	3	11	2	11	100	1
Carotenoids and vitamin C									
α-carotene (ng/mL)	101	43	27	53	84	33	49	50	43
ß-carotene (ng/mL)	428	256	148	240	299	200	253	228	172
ß-cryptoxanthin (ng/mL)	146	164	186	84	107	114	97	72	48
Lycopene (ng/mL)	229	347	198	237	333	195	219	210	179
Lutein (ng/mL)	208	247	147	124	151	130	138	128	115
Zeaxanthin (ng/mL)	24	22	44	15	19	19	24	16	11
Vitamin C (µmol/L)	45	39	42	43	49	48	48	-†	40
Phospholipid fatty acids									
C14:0 (%mol)	0.410	0.360	0.282	0.381	0.373	0.401	0.390	0.400	0.389
C16:0 (%mol)	30	30	29	30	31	30	30	31	31
C18:0 (%mol)	14	14	15	14	14	14	14	14	14
C15:0 (%mol)	0.270	0.210	0.172	0.240	0.223	0.242	0.220	0.220	0.206
C17:0 (%mol)	0.492	0.411	0.444	0.421	0.411	0.402	0.390	0.391	0.389
C20:0 (%mol)	0.130	0.130	0.124	0.150	0.157	0.140	0.120	0.140	0.121
C22:0 (%mol)	0.250	0.210	0.225	0.270	0.230	0.280	0.200	0.230	0.217
C23:0 (%mol)	0.130	0.110	0.103	0.120	0.100	0.120	0.090	0.110	0.095

Variable	France	Italy	Spain	UK-GP	UK-HC	Netherlands	Germany	Sweden	Denmark
C24:0 (%mol)	0.260	0.220	0.220	0.230	0.218	0.261	0.200	0.211	0.206
C18:3n-3 (%mol)	0.280	0.280	0.190	0.330	0.319	0.262	0.310	0.391	0.285
C20:5n-3 (%mol)	1.071	0.710	0.863	1.152	0.845	0.863	0.971	1.291	1.587
C22:5n-3 (%mol)	0.989	0.750	0.659	1.023	0.910	0.961	0.920	1.061	1.029
C22:6n-3 (%mol)	4.79	3.45	4.61	4.17	3.57	3.35	3.64	4.05	4.69
C18:2n-6c (%mol)	21	22	23	23	24	24	23	22	22
C18:3n-6 (%mol)	0.050	0.090	0.072	0.070	0.059	0.080	0.080	0.060	0.077
C20:2 (%mol)	0.380	0.360	0.368	0.391	0.392	0.400	0.380	0.371	0.346
C20:3n-6 (%mol)	3.07	3.63	3.04	3.16	3.04	3.28	3.09	3.10	2.76
C20:4n-6 (%mol)	9.6	10.3	9.7	8.3	8.1	9.3	9.6	8.4	8.5
C22:4 (%mol)	0.300	0.330	0.257	0.261	0.285	0.310	0.300	0.280	0.241
C22:5n-6 (%mol)	0.220	0.276	0.180	0.157	0.173	0.210	0.210	0.160	0.124
C18:1n-9t (%mol)	0.170	0.130	0.130	0.401	0.375	0.370	0.180	0.350	0.197
C18:2n-6t (%mol)	0.070	0.060	0.060	0.070	0.075	0.080	0.060	0.080	0.060
C16:1 (%mol)	0.440	0.470	0.317	0.501	0.422	0.490	0.530	0.510	0.568
C17:1 (%mol)	0.030	0.070	0.052	0.040	0.045	0.040	0.070	0.060	0.067
C18:1n-9c (%mol)	9.0	10.9	9.8	9.4	9.2	8.8	9.3	10.4	9.7
C20:1 (%mol)	0.260	0.230	0.202	0.312	0.315	0.252	0.241	0.280	0.247
C24:1 (%mol)	0.370	0.350	0.328	0.341	0.291	0.320	0.320	0.360	0.357
Iron status and other biomarkers									
Ferritin (µg/L)	85	69	50	74	36	97	111	79	113
Iron (µmol/L)	18	16	16	17	19	18	16	-	16
Transferrin (mg/dL)	67	70	71	70	74	68	67	-	64
Vitamin 25(OH)D_3 (nmol/L)	37	36	37	41	38	42	38	56	39
Magnesium (mmol/L)	0.860	0.840	0.850	0.860	0.900	0.850	0.850	-	0.820
Calcium (mmol/L)	2.39	2.46	2.49	2.45	2.47	2.46	2.44	-	2.34

Abbreviations: aHEI – alternative Healthy Eating Index; DASH – Dietary Approaches to Stop Hypertension; GP – general practice recruitment arm; HC – health-conscious recruitment arm; %mol – molar percentage of all fatty acids measured; MDS – Mediterranean diet score

*Values or medians or percentages

†Biomarkers not measured

3.5.2 Background characteristics by adherence to dietary patterns

Most baseline characteristics of the subcohort participants differed by quintiles of dietary patterns as assessed either by the self-report or the multi-country biomarker scores (**Table 3.3**). The patterns of distribution were largely directionally consistent between the dietary patterns and the two exposure assessment methods. However, for several comparisons the differences between the values in extreme quintiles (Q5 vs Q1) were more pronounced with biomarker-based assessment than when using dietary-self-report, particularly for clinical markers of cardiometabolic risk and the aHEI-2010 and DASH dietary patterns. There was some discordance between dietary-self report and biomarker-based assessment of MDS. For example, age decreased across the fifths of self-reported MDS (Q5 vs Q1: mean 50.5 vs 51.1 years) whereas it increased with higher levels of the biomarker score of MDS (Q5 vs Q1: 52.9 vs 51.3). This change in the confounding structure of the biomarker score compared to self-report was likely introduced only for MDS due to its sizeable variation by country and adjustment for country in the process of derivation of the biomarker scores.

Table 3.3 Baseline characteristics of the EPIC-InterAct subcohort participants by quintiles of dietary patterns assessed by self-report and nutritional biomarker scores

Covariate and dietary pattern	Q1	Q2	Q3	Q4	Q5	p
Age, years (n=12,920)						
MDS: self-report	51.1 (10.4)	52.4 (9.3)	52.3 (9.0)	51.6 (8.7)	50.5 (8.3)	*
MDS: biomarker score	51.3 (9.8)	50.8 (9.2)	51.1 (9.3)	51.8 (8.6)	52.9 (8.4)	*
aHEI-2010: self-report	50.9 (9.6)	50.9 (9.2)	51.7 (8.7)	52.0 (8.9)	52.5 (9.0)	*
aHEI-2010: biomarker score	51.4 (9.0)	51.0 (9.3)	51.4 (9.3)	51.6 (9.2)	52.3 (8.8)	*
DASH: self-report	51.5 (10.0)	51.0 (9.1)	51.8 (8.9)	51.5 (8.7)	52.3 (8.9)	*
DASH: biomarker score	51.5 (9.5)	50.9 (9.3)	51.2 (9.2)	51.5 (8.9)	52.7 (8.6)	*
Women, % (n=12,920)						
MDS: self-report	51	61	66	68	66	*
MDS: biomarker score	59	61	65	65	64	*
aHEI-2010: self-report	53	59	64	68	72	*
aHEI-2010: biomarker score	49	58	66	69	71	*
DASH: self-report	47	60	65	70	74	*
DASH: biomarker score	46	57	62	71	78	*
Postmenopausal, % (n=8,415)						
MDS: self-report	33	44	47	48	43	*
MDS: biomarker score	39	39	43	45	47	*
aHEI-2010: self-report	32	37	44	50	54	*
aHEI-2010: biomarker score	30	37	45	49	54	*
DASH: self-report	28	38	45	50	57	*
DASH: biomarker score	27	35	40	52	64	*
HRT use, % (n=5,921)						
MDS: self-report	13	22	21	20	16	
MDS: biomarker score	14	16	20	22	21	*
aHEI-2010: self-report	12	15	20	24	26	*
aHEI-2010: biomarker score	12	15	20	22	26	*
DASH: self-report	13	18	20	23	25	*
DASH: biomarker score	10	14	17	26	35	*
Dietary supplement use, % (n=11,975)						
MDS: self-report	43	40	42	39	35	*
MDS: biomarker score	36	36	38	41	44	*
aHEI-2010: self-report	33	37	39	43	46	*
aHEI-2010: biomarker score	34	35	37	41	48	*
DASH: self-report	40	36	38	39	44	*
DASH: biomarker score	38	38	38	40	42	*
Current smokers, % (n=12,761)						
MDS: self-report	35	27	24	22	23	*
MDS: biomarker score	35	27	24	22	19	*
aHEI-2010: self-report	32	29	25	23	18	*
aHEI-2010: biomarker score	40	29	24	19	16	*
DASH: self-report	32	30	25	23	17	*
DASH: biomarker score	44	29	23	18	14	*
Moderately active or active, % (n=12,722)						
MDS: self-report	51	48	44	37	36	*
MDS: biomarker score	44	43	43	41	44	
aHEI-2010: self-report	42	43	43	42	43	
aHEI-2010: biomarker score	43	43	41	42	46	
DASH: self-report	48	43	41	41	39	*
DASH: biomarker score	43	40	42	43	46	*

Covariate and dietary pattern	Q1	Q2	Q3	Q4	Q5	p
Tertiary education, % (n=12,647)						
MDS: self-report	19	21	22	21	19	
MDS: biomarker score	15	19	21	22	25	*
aHEI-2010: self-report	20	21	20	21	21	
aHEI-2010: biomarker score	17	18	20	22	25	*
DASH: self-report	21	22	21	18	21	*
DASH: biomarker score	15	18	21	22	26	*
Currently employed, % (n=9,432)						
MDS: self-report	69	68	67	66	62	*
MDS: biomarker score	61	66	68	70	69	*
aHEI-2010: self-report	69	68	67	66	64	*
aHEI-2010: biomarker score	66	68	66	67	68	
DASH: self-report	70	69	66	65	62	*
DASH: biomarker score	66	67	67	68	66	
Family history of T2D, % (n=6,097)						
MDS: self-report	18	18	18	18	14	
MDS: biomarker score	20	18	18	17	15	*
aHEI-2010: self-report	17	18	17	19	18	
aHEI-2010: biomarker score	19	18	18	17	17	
DASH: self-report	17	17	19	19	17	
DASH: biomarker score	16	20	18	17	18	
Prevalent hypertension, % (n=12,559)						
MDS: self-report	20	21	19	17	19	*
MDS: biomarker score	21	18	18	18	19	
aHEI-2010: self-report	21	19	20	19	17	*
aHEI-2010: biomarker score	22	21	19	18	16	*
DASH: self-report	20	19	19	20	18	
DASH: biomarker score	20	20	19	18	18	*
Prevalent hyperlipidaemia, % (n=11,875)						
MDS: self-report	11	15	17	18	20	*
MDS: biomarker score	14	14	16	16	23	*
aHEI-2010: self-report	16	15	18	18	17	
aHEI-2010: biomarker score	18	17	16	16	17	
DASH: self-report	14	16	18	17	19	*
DASH: biomarker score	14	16	16	18	19	*
Prevalent CVD, % (n=11,883)						
MDS: self-report	3.2	2.4	2.0	1.4	1.2	*
MDS: biomarker score	2.0	2.0	1.9	1.8	1.7	
aHEI-2010: self-report	2.2	1.7	2.1	1.6	2.0	
aHEI-2010: biomarker score	2.4	1.8	1.8	2.0	1.5	*
DASH: self-report	2.5	1.9	1.6	1.7	2.0	
DASH: biomarker score	2.5	2.0	2.1	1.8	1.1	*
Prevalent cancer, % (n=12,920)						
MDS: self-report	2.4	3.0	3.0	2.4	1.2	*
MDS: biomarker score	2.6	2.5	2.5	2.3	2.2	
aHEI-2010: self-report	2.4	2.1	1.8	2.9	2.8	*
aHEI-2010: biomarker score	2.9	2.5	2.2	2.2	2.3	
DASH: self-report	2.7	2.5	2.1	2.2	2.5	
DASH: biomarker score	2.4	2.2	2.0	2.5	3.0	

Covariate and dietary pattern	Q1	Q2	Q3	Q4	Q5	p
BMI, kg/m^2 (n=12,818)						
MDS: self-report	25.7 (4.1)	25.9 (4.3)	25.9 (4.1)	26.3 (4.2)	26.6 (4.3)	*
MDS: biomarker score	26.7 (4.6)	26.3 (4.2)	26.1 (4.1)	26.0 (4.2)	25.6 (4.0)	*
aHEI-2010: self-report	25.8 (4.1)	26.0 (4.1)	26.1 (4.1)	26.3 (4.4)	26.2 (4.3)	*
aHEI-2010: biomarker score	26.9 (4.5)	26.4 (4.3)	26.2 (4.2)	25.8 (4.1)	25.3 (4.0)	*
DASH: self-report	25.9 (4.0)	26.0 (4.2)	26.2 (4.1)	26.4 (4.4)	26.1 (4.4)	*
DASH: biomarker score	27.0 (4.4)	26.6 (4.2)	26.2 (4.1)	25.8 (4.1)	25.1 (4.0)	*
HbA1c, mmol/mol (n=12,779)						
MDS: self-report	36.1 (5.3)	36.1 (5.2)	35.8 (4.8)	35.8 (4.5)	35.7 (4.7)	*
MDS: biomarker score	36.3 (6.2)	35.9 (5.2)	35.6 (4.6)	35.9 (4.4)	35.7 (3.9)	*
aHEI-2010: self-report	35.8 (5.2)	35.8 (4.6)	35.9 (4.8)	36.0 (5.6)	35.9 (4.1)	
aHEI-2010: biomarker score	36.2 (5.7)	36.0 (5.6)	35.8 (4.9)	35.6 (4.2)	35.8 (3.9)	*
DASH: self-report	35.9 (5.0)	35.8 (4.4)	35.8 (4.8)	35.8 (4.8)	36.0 (5.4)	
DASH: biomarker score	36.4 (5.7)	35.9 (4.9)	35.7 (4.6)	35.8 (5.0)	35.6 (4.3)	*
Triglycerides, mmol/L (n=12,920)						
MDS: self-report	1.53 (1.01)	1.44 (0.92)	1.30 (0.93)	1.23 (0.80)	1.19 (0.78)	*
MDS: biomarker score	1.53 (1.15)	1.35 (0.89)	1.28 (0.83)	1.25 (0.78)	1.22 (0.76)	*
aHEI-2010: self-report	1.42 (0.90)	1.38 (0.95)	1.31 (0.85)	1.30 (0.96)	1.22 (0.79)	*
aHEI-2010: biomarker score	1.64 (1.20)	1.41 (0.93)	1.27 (0.80)	1.19 (0.70)	1.12 (0.66)	*
DASH: self-report	1.53 (1.01)	1.34 (0.89)	1.30 (0.95)	1.23 (0.76)	1.23 (0.81)	*
DASH: biomarker score	1.52 (1.13)	1.37 (0.91)	1.32 (0.85)	1.23 (0.78)	1.19 (0.72)	*
CRP, μmol/L (n=12,912)						
MDS: self-report	2.57 (4.31)	2.32 (4.74)	2.10 (4.02)	2.14 (3.80)	1.99 (3.73)	*
MDS: biomarker score	2.78 (4.78)	2.25 (4.09)	2.08 (3.61)	2.08 (4.44)	1.80 (3.33)	*
aHEI-2010: self-report	2.28 (3.92)	2.19 (3.85)	2.31 (4.36)	2.13 (3.76)	2.09 (4.63)	*
aHEI-2010: biomarker score	2.85 (4.65)	2.33 (4.55)	2.13 (4.08)	1.94 (3.46)	1.76 (3.63)	*
DASH: self-report	2.44 (4.32)	2.23 (3.93)	2.12 (3.35)	2.10 (4.47)	2.12 (4.41)	*
DASH: biomarker score	2.87 (4.73)	2.32 (4.11)	2.12 (4.51)	1.94 (3.65)	1.73 (3.20)	*
HDL-C, mmol/L (n=12,920)						
MDS: self-report	1.41 (0.42)	1.47 (0.42)	1.53 (0.43)	1.53 (0.42)	1.53 (0.41)	*
MDS: biomarker score	1.40 (0.40)	1.46 (0.41)	1.52 (0.41)	1.52 (0.41)	1.59 (0.44)	*
aHEI-2010: self-report	1.45 (0.41)	1.48 (0.42)	1.51 (0.42)	1.53 (0.43)	1.53 (0.40)	*
aHEI-2010: biomarker score	1.42 (0.42)	1.46 (0.41)	1.50 (0.41)	1.53 (0.43)	1.57 (0.42)	*
DASH: self-report	1.42 (0.42)	1.49 (0.41)	1.52 (0.42)	1.53 (0.42)	1.53 (0.41)	*
DASH: biomarker score	1.37 (0.40)	1.45 (0.40)	1.50 (0.41)	1.55 (0.42)	1.62 (0.44)	*
GGT, U/L (n=12,905)						
MDS: self-report	38 (76)	32 (70)	29 (59)	26 (26)	26 (30)	*
MDS: biomarker score	37 (84)	30 (40)	27 (30)	28 (66)	27 (28)	*
aHEI-2010: self-report	38 (83)	32 (71)	28 (31)	28 (36)	24 (26)	*
aHEI-2010: biomarker score	46 (90)	31 (67)	26 (28)	24 (21)	22 (23)	*
DASH: self-report	36 (62)	34 (91)	28 (33)	26 (26)	25 (32)	*
DASH: biomarker score	40 (75)	30 (53)	29 (65)	26 (35)	24 (24)	*

Abbreviations: aHEI – alternative Healthy Eating Index; BMI – body mass index, CRP – C-reactive protein; CVD – cardiovascular disease; DASH – Dietary Approaches to Stop Hypertension; GGT – gamma glutamyl transferase; HbA1c – Haemoglobin A1c; HDL-C – high-density lipoprotein cholesterol; HRT – hormone replacement therapy; MDS – Mediterranean diet score; T2D – type 2 diabetes

Ranges of participants per fifths were 1,971-2,584 (Q1), 2,488-2,584 (Q2), 2,520-2,809 (Q3), 2,519-2,691 (Q4), 2,521-2,649 (Q5). Values are means (standard deviation) or percentages in participants with non-missing data for a given covariate. Tests for trend were calculated by regressing the covariates on fifths of dietary patterns entered into linear or logistic regression models as continuous variables with values equal to quantile numbers.

*p trend < 0.10

3.5.3 Nutritional biomarker scores

Forty of the 49 nutritional biomarkers considered as predictors of adherence to dietary patterns were selected into any of the biomarker scores (**Table 3.4**). There was a large overlap between the biomarker scores of the three dietary patterns in terms of the biomarkers selected and the coefficients of standardised effect sizes. Consistently positively scored biomarkers included α-carotene, β-cryptoxanthin, lutein, vitamin C and lignoceric acid (C24:0). By contrast, behenic (C22:0) and tricosanoic (C23:0) acids were the only biomarkers scored consistently negatively across the three dietary patterns. Among the remaining carotenoids, coefficients for β-carotene and lycopene were directionally variable and small (absolute standardised $\beta \leq 0.03$) and zeaxanthin was scored consistently negatively in all three biomarker scores.

Even-chain SFAs were frequently selected into the scores of aHEI-2010 and DASH with negative weighting, whereas only palmitic acid (C16:0) was included in the score of MDS with positive weighting. Consistent with the negative scoring of dairy products in MDS, both pentadecanoic (C15:0) and heptadecanoic (C17:0) acids had negative coefficients for this dietary pattern as candidate biomarkers of dairy fat intake,[224] while the selection and scoring of these even-chain SFAs was variable for aHEI-2010 and DASH. Among n-3 PUFAs, docosahexaenoic acid (C22:6n-3) was the only biomarker selected into all three scores. It had the largest effect sizes of all fatty acids from this group with positive scoring for MDS and aHEI-2010 and negative scoring for DASH. For n-6 PUFA, γ-linolenic (C18:3n-6) and dihomo-γ-linolenic acid (C20:3n-6) were scored positively with small effect sizes for all three dietary patterns. The remaining n-6 fatty acids were mostly negatively weighted and robustly selected into the biomarker score of DASH but more sparsely for MDS and aHEI-2010. Linoleic acid (C18:2n-6c) was notably included only in the DASH biomarker score with the largest absolute coefficient ($\beta = -0.16$) across all biomarkers and dietary patterns. Arachidonic (C20:4n-6) acid osbond (C22:5n-6) acids were selected as predictors of aHEI-2010 and DASH dietary patterns, whereas adrenic acid (C22:4) was a predictor of MDS and DASH. Cis-MUFAs were frequently selected into the biomarker scores and had dietary index-specific patterns. Mostly positive coefficients were estimated for these fatty acids, including oleic acid (C18:1n-9c), for prediction of the MDS. By contrast, the effect sizes for oleic acid were negative and relatively large in case of the aHEI-2010 and DASH indices. The remaining cis-MUFAs in the biomarker scores of these two dietary patterns had coefficients close to zero. Trans-fatty acids, vitamin 25(OH)D_3, and iron status biomarkers were frequently selected into the biomarker

scores of MDS and DASH but only serum ferritin and iron were included in the aHEI-2010 biomarker score. Coefficients for these biomarkers were mostly close to zero.

Country-specific biomarker scores were overall more parsimonious than the above-described multi-country scores (**Appendices 3.1-3.3**). The Spanish biomarker scores of MDS and aHEI-2010 and the Danish biomarker score of the DASH diet were qualitatively similar to their respective multi-country scores. Application of the $\lambda_{+1\ SE}$ penalty led to selection of more parsimonious multi-country models for the MDS (α-carotene×0.05 + lutein×0.09 + vitamin C×0.07 - C22:4×0.10 + C22:6n-3×0.08) and aHEI-2010 (α-carotene×0.14 - C16:1×0.07 + C17:0×0.04 - C22:4×0.09). No predictors were discovered for the DASH diet under this increased penalisation.

Table 3.4 Nutritional biomarker scores of dietary patterns derived in the EPIC-InterAct subcohort: standardised coefficients*

	MDS	aHEI-2010	DASH
α-carotene	0.07	0.10	0.12
β-carotene	-†	0.02	-0.03
β-cryptoxanthin	0.07	0.06	0.15
Lycopene	0.03	-	0.00
Lutein	0.07	0.05	0.03
Zeaxanthin	-0.02	-0.10	-0.04
Vitamin C	0.07	0.02	0.05
C14:0	-	-	-0.01
C16:0	0.04	-0.09	-0.08
C18:0	-	-0.09	-0.10
C15:0	-0.08	-	-0.05
C17:0	-0.02	0.04	0.06
C20:0	-0.01	0.01	-0.01
C22:0	-0.05	-0.05	-0.08
C23:0	-0.01	-0.03	-0.02
C24:0	0.05	0.09	0.03
C18:3n-3	0.02	-	0.00
C20:5n-3	0.00	-	-0.06
C22:5n-3	-	-0.07	0.03
C22:6n-3	0.06	0.09	-0.08
C18:2n-6c	-	-	-0.16
C18:3n-6	0.03	0.05	0.04
C20:2	-	-	0.01
C20:3n-6	0.02	0.05	-
C20:4n-6	-	-0.06	-0.08
C22:4	-0.14	-	-0.04
C22:5n-6	-	-0.08	-0.06
C16:1	-	-0.07	0.03
C17:1	-0.01	0.01	0.00
C18:1n-9c	0.03	-0.09	-0.13
C20:1	0.01	0.00	0.01
C24:1	0.07	-0.03	0.02
C18:1n-9t	0.02	0.02	-0.01
C18:2n-6t	-	-0.02	0.00
Magnesium	-0.01	-0.01	-0.01
Calcium	0.00	-	0.00
Vitamin 25(OH) D_3	-0.02	-	-0.01
Ferritin	0.04	0.01	0.01
Iron	0.00	0.01	0.01
Transferrin	0.04	-	0.02

Abbreviations: aHEI – alternative Healthy Eating Index; DASH – Dietary Approaches to Stop Hypertension; Mediterranean diet score

*Derived in 10,562-10,569 subcohort participants. Red colour highlights positive coefficients, and blue colour highlights negative coefficients with intensity proportional to value.

†Biomarker not selected for a given score

3.5.4 Associations of biomarker scores with self-reported diet

The biomarker scores were modestly correlated with their respective dietary patterns (**Table 3.5**). In the subcohort, the pooled correlation coefficients were 0.31, 0.34 and 0.32 for MDS, aHEI-2010 and the DASH score, respectively. The ranges of correlation coefficients by country were wide, spanning from 0.12-0.42 for MDS, 0.17-0.48 for aHEI-2010, and 0.16-0.41 for the DASH diet. In the test samples of non-subcohort T2D cases, the performance of the scores was lower by r_{Δ} of -0.03, -0.01 and -0.05, respectively. The country-specific biomarker scores yielded materially similar results in the subcohort as the multi-country scores. The biomarker scores of the three dietary patterns were moderately correlated with one another, with a pooled r range of the multi-country scores in the subcohort of 0.55-0.67. These correlations were stronger than those between pairs of dietary patterns as assessed by dietary self-report (r range: 0.40-0.45).

Among individual foods and nutrients (**Table 3.6**), estimated fruit and vegetable intakes were positively correlated with the three biomarker scores ($r > 0.2$). Cereal fibre was positively correlated with the biomarker scores of aHEI-2010 and DASH ($r = 0.13$ each) and to a lesser degree with the biomarker score of MDS ($r = 0.07$). Conversely, the correlation with fish and shellfish was stronger for MDS ($r = 0.23$) than for the two remaining biomarker scores ($r = 0.14$ and $r = 0.06$, respectively). Estimated red and processed meat intake was inversely correlated with the biomarker scores of MDS ($r = -0.06$), aHEI-2010 ($r = -0.12$) and DASH ($r = -0.15$), and more strongly so than for total meat and meat products (**Table 3.6**). Divergent relationships with ethanol emerged, with correlations coefficients of 0.10, -0.22 and -0.07, respectively. The percentage of energy from PUFA was only meaningfully correlated with the biomarker score of aHEI-2010 ($r = 0.18$). The remaining foods included in derivation of the dietary patterns had correlations with the biomarker scores between -0.07 and 0.07, i.e., legumes, nuts and seeds, dairy products, cereals, olive oil, sweets, sugar sweetened beverages and fruit juice. Directionality was largely consistent with the underlying scoring algorithms of the respective dietary patterns.

Table 3.5 Correlations of biomarker scores of dietary patterns with dietary patterns assessed by self-report in the EPIC-InterAct study (minimum, maximum country-specific values)*

Dietary pattern or biomarker score	MDS	aHEI-2010	DASH	MDS biomarker	aHEI-2010 biomarker	DASH biomarker
Subcohort, multi-country biomarker scores (n = 12,455)						
MDS	-	(0.30, 0.50)	(0.35, 0.51)	(0.12, 0.42)	(0.05, 0.39)	(0.03, 0.44)
aHEI-2010	0.40	-	(0.27, 0.52)	(0.10, 0.29)	(0.17, 0.48)	(0.15, 0.39)
DASH	0.45	0.44	-	(0.07, 0.33)	(-0.02, 0.29)	(0.16, 0.41)
MDS biomarker	0.31	0.22	0.21	-	(0.43, 0.70)	(0.61, 0.83)
aHEI-2010 biomarker	0.22	0.34	0.19	0.55	-	(0.49, 0.72)
DASH biomarker	0.26	0.25	0.32	0.67	0.60	-
Non-subcohort T2D cases, multi-country biomarker scores (n = 8,239)						
MDS	-	(0.28, 0.51)	(0.31, 0.53)	(0.09, 0.43)	(0.10, 0.37)	(0.02, 0.42)
aHEI-2010	0.41	-	(0.29, 0.52)	(0.03, 0.35)	(0.25, 0.42)	(0.06, 0.33)
DASH	0.41	0.43	-	(-0.03, 0.29)	(0.01, 0.26)	(0.12, 0.37)
MDS biomarker	0.28	0.20	0.16	-	(0.36, 0.65)	(0.59, 0.82)
aHEI-2010 biomarker	0.21	0.33	0.19	0.52	-	(0.43, 0.66)
DASH biomarker	0.22	0.22	0.27	0.69	0.58	-
Subcohort, country-specific biomarker scores (n = 12,573)						
MDS	-	(0.30, 0.50)	(0.35, 0.51)	(0.11, 0.45)	(0.11, 0.33)	(0.11, 0.40)
aHEI-2010	0.40	-	(0.27, 0.52)	(0.13, 0.35)	(0.19, 0.48)	(0.08, 0.31)
DASH	0.45	0.44	-	(0.04, 0.36)	(0.15, 0.32)	(0.19, 0.46)
MDS biomarker	0.33	0.23	0.23	-	(0.31, 0.87)	(0.19, 0.76)
aHEI-2010 biomarker	0.22	0.34	0.20	0.53	-	(0.39, 0.65)
DASH biomarker	0.26	0.26	0.33	0.59	0.54	-

Abbreviations: aHEI – alternative Healthy Eating Index; DASH – Dietary Approaches to Stop Hypertension; MDS – Mediterranean diet score; min, max – minimum, maximum country-specific correlation; T2D – type 2 diabetes

*Highlighted values are pooled country-specific correlations using random-effects meta-analysis. Intensity of the red colour is proportional to correlation coefficients. Values in brackets are ranges of country-specific correlations. The subcohort includes 10,562 participants who constituted derivation samples for the multi-country biomarker scores and 11,648 participants for the country-specific scores. These individuals had their biomarker score equations re-estimated using the leave-one-out approach.

Table 3.6 Correlations of biomarker scores of dietary patterns with food and nutrients components of dietary patterns assessed by self-report (minimum, maximum country-specific values) in the EPIC-InterAct subcohort (n = 12,455)*

Food or nutrient†	Mean daily intake (SD)	MDS		aHEI-2010		DASH	
		r	(min, max)	r	(min, max)	r	(min, max)
Fruits, g	240 (190)	0.22	(0.08, 0.30)	0.20	(0.09, 0.27)	0.34	(0.18, 0.43)
Vegetable, g	188 (122)	0.24	(0.12, 0.36)	0.21	(0.08, 0.34)	0.23	(0.09, 0.41)
Legumes, g	19 (29)	-0.03	(-0.20, 0.08)	-0.03	(-0.21, 0.08)	-0.02	(-0.13, 0.08)
Nuts and seeds, g	3.3 (8.2)	0.03	(0.01, 0.06)	0.07	(-0.01, 0.16)	0.03	(-0.06, 0.16)
Dairy products, g	328 (235)	-0.06	(-0.17, 0.01)	0.03	(-0.13, 0.09)	0.03	(-0.17, 0.12)
Cereals, g	219 (111)	0.03	(-0.09, 0.14)	0.08	(-0.05, 0.26)	0.07	(-0.09, 0.18)
Cereal fibre, g	8.6 (5.0)	0.07	(-0.05, 0.18)	0.13	(-0.01, 0.33)	0.13	(-0.05, 0.28)
Meat and meat products, g	110 (60)	-0.02	(-0.17, 0.25)	-0.08	(-0.24, 0.20)	-0.11	(-0.24, -0.02)
Red and processed meat, g	81 (51)	-0.06	(-0.24, 0.19)	-0.12	(-0.29, 0.13)	-0.15	(-0.31, -0.03)
Fish and shellfish, g	37 (33)	0.23	(0.15, 0.38)	0.14	(0.10, 0.19)	0.06	(-0.04, 0.15)
Olive oil, g	9.1 (13.8)	0.06	(-0.15, 0.28)	0.01	(-0.08, 0.10)	0.01	(-0.11, 0.09)
PUFA, E%	5.6 (2.0)	-0.03	(-0.19, 0.10)	0.18	(0.01, 0.25)	0.00	(-0.07, 0.04)
Sweets excluding SSBs, g	19 (41)	-0.07	(-0.12, -0.02)	-0.02	(-0.13, 0.04)	-0.04	(-0.18, 0.04)
SSBs and fruit juice, g	86 (155)	0.02	(-0.06, 0.07)	-0.02	(-0.09, 0.03)	0.03	(-0.04, 0.08)
Ethanol, g	13.4 (18.6)	0.10	(0.00, 0.23)	-0.22	(-0.35, -0.07)	-0.07	(-0.17, 0.05)

Abbreviations: aHEI – alternative Healthy Eating Index; DASH – Dietary Approaches to Stop Hypertension; MDS – Mediterranean diet score; min, max – minimum, maximum country-specific correlation; PUFA – polyunsaturated fatty acids; SSB – sugar-sweetened beverages

*Highlighted values are pooled country-specific correlations using random-effects meta-analysis. Red colour highlights positive coefficients, and blue colour highlights negative coefficients with intensity proportional to value.

†Adjusted for estimated energy intake using the residual method.

3.5.5 Associations of biomarker scores and self-reported dietary patterns with incident T2D

Biomarker scores of all three dietary patterns were inversely associated with incident T2D (**Table 3.7**). In the adiposity-adjusted multivariable model, the HRs (95% CI) for the top quintiles of the biomarker scores compared to the bottom quintiles were 0.61 (0.47-0.80) for MDS, 0.47 (0.37-0.61) for aHEI-2010, and 0.68 (0.54-0.86) for the DASH diet. There was evidence of a trend across the fifths of biomarker scores of MDS (p trend = 0.012) and aHEI-2010 (p trend <0.001), but not DASH (p trend = 0.186). The HRs per 1 SD were 0.85 (0.79-0.92), 0.75 (0.68-0.82) and 0.83 (0.77-0.90), respectively, with high heterogeneity between country-specific estimates (I^2 values: 66-79). The patterns of country-specific estimates were similar between the biomarker scores in terms of the relative magnitude of effect sizes and precision (**Figure 3.1**). The biomarker score of MDS was statistically significantly inversely associated with incident T2D in France, Italy, Spain and the UK, but not the Netherlands, Germany, Sweden and Denmark. The biomarker score of aHEI-2010 had inverse associations in all countries, while the biomarker score of DASH had non-significant associations in Sweden and Denmark. The 95% prediction intervals included the null for all three dietary patterns. There was no evidence of departure from linearity (**Figure 3.2**; all p-values > 0.52).

The crude model adjusted for age, sex and recruitment centre, and the multivariable model without adjustment for adiposity yielded materially similar results (**Table 3.7**). The adiposity-adjusted HRs were attenuated towards the null by approximately 0.1 per 1 SD of the biomarker scores compared to the multivariable model which did not include BMI and waist circumference. Further adjustment of the adiposity-adjusted model for biomarkers of major cardiometabolic pathways led to additional attenuation (**Table 3.8**). All estimates per 1 SD remained statistically significantly inverse, however, the Q5 vs Q1 comparisons were attenuated to the null after adjustment for the associations of MDS and DASH biomarker scores for HbA1c.

The HRs (95% CI) per 1 SD of adherence to dietary patterns estimated from dietary self-report were 0.90 (0.86-0.95) for MDS, 0.96 (0.91-1.02) for aHEI-2010 and 0.96 (0.91, 1.02) for the DASH diet (main analytical model with residual adjustment for estimated energy intake).

Table 3.7 Associations between biomarker scores of dietary patterns and incidence of type 2 diabetes in the EPIC-InterAct case-cohort study

Biomarker score and model*	Quintiles					p_{trend}†	Per 1 SD	I^2, % (95% CI)
	Q1	Q2	Q3	Q4	Q5			
MDS (n = 21, 293)								
Number of cases	2,467	1,898	1,665	1,572	1,472			
IR per 100,000 person-years	505	390	337	284	297			
Pooled HRs (95% CIs)								
Age, sex, and centre adjusted	1.0 (Ref.)	0.73 (0.65-0.82)	0.59 (0.50-0.70)	0.54 (0.49-0.60)	0.44 (0.37-0.53)	<0.001	0.75 (0.70-0.79)	67 (34-84)
Multivariable adjusted	1.0 (Ref.)	0.77 (0.65-0.90)	0.62 (0.49-0.79)	0.58 (0.50-0.66)	0.46 (0.36-0.60)	<0.001	0.75 (0.69-0.82)	77 (57-88)
+adiposity	1.0 (Ref.)	0.78 (0.64-0.95)	0.68 (0.52-0.88)	0.64 (0.52-0.79)	0.61 (0.47-0.80)	0.012	0.85 (0.79-0.92)	66 (28-84)
aHEI-2010 (n = 21,549)								
Number of cases	3,179	2,041	1,626	1,290	1,045			
IR per 100,000 person-years	576	399	354	273	208			
Pooled HRs (95% CIs)								
Age, sex, and centre adjusted	1.0 (Ref.)	0.62 (0.53-0.72)	0.48 (0.40-0.58)	0.37 (0.31-0.44)	0.28 (0.22-0.36)	<0.001	0.63 (0.58-0.69)	84 (71-91)
Multivariable adjusted	1.0 (Ref.)	0.61 (0.50-0.75)	0.49 (0.39-0.61)	0.39 (0.32-0.49)	0.29 (0.21-0.40)	<0.001	0.64 (0.58-0.71)	85 (73-92)
+adiposity	1.0 (Ref.)	0.70 (0.58-0.84)	0.59 (0.46-0.74)	0.52 (0.42-0.65)	0.47 (0.37-0.61)	<0.001	0.75 (0.68-0.82)	79 (59-89)
DASH (n = 21,293)								
Number of cases	2,780	1,948	1,626	1,484	1,236			
IR per 100,000 person-years	556	420	339	292	206			
Pooled HRs (95% CIs)								
Age, sex, and centre adjusted	1.0 (Ref.)	0.71 (0.64-0.79)	0.59 (0.53-0.64)	0.53 (0.46-0.61)	0.41 (0.34-0.50)	<0.001	0.70 (0.65-0.76)	80 (63-89)
Multivariable adjusted	1.0 (Ref.)	0.74 (0.66-0.83)	0.61 (0.54-0.69)	0.56 (0.45-0.68)	0.44 (0.35-0.57)	0.005	0.70 (0.63-0.78)	85 (73-92)
+adiposity	1.0 (Ref.)	0.81 (0.68-0.96)	0.74 (0.66-0.84)	0.73 (0.60-0.90)	0.68 (0.54-0.86)	0.186	0.83 (0.77-0.90)	61 (16-82)

Abbreviations: aHEI-2010 – alternative Healthy Eating Index 2010; CI – confidence interval; DASH – Dietary Approaches to Stop Hypertension; HR – hazard ratio; IR – incidence rate; MDS – Mediterranean diet score; SD – standard deviation

*Hazard ratios were pooled from country-specific estimates. Multivariable adjusted model included the following covariates: age (as timescale), sex, recruitment centre, prevalent cancer, cardiovascular disease, hypertension and hyperlipidaemia; familial history of T2D, smoking status (never, former, current smoker), physical activity index (inactive, moderately inactive, moderately active, active), seasonality (sine and cosine function of the day of the year), fasting status (<3, 3-6, >6 hours), current use of vitamin or mineral supplements, marital status (single, married or cohabiting, divorced or separated, widowed), educational attainment (none, primary school, technical or professional school, secondary school, post-secondary school education), current employment, and in women, menopausal status (pre-, peri-, postmenopausal, bilateral oophorectomy), current hormone replacement therapy use. Adjustment for adiposity included body mass index and waist circumference.

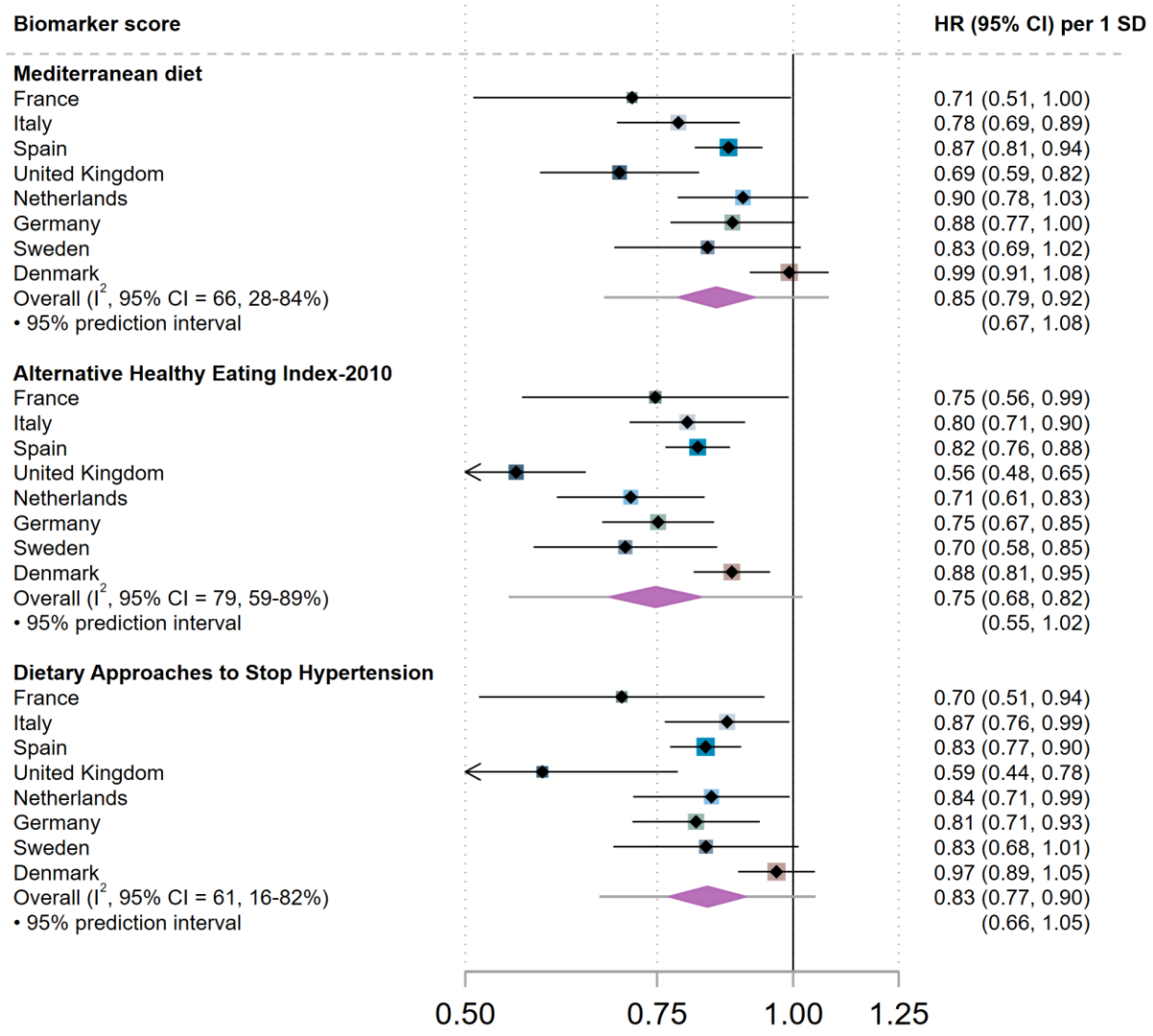


Figure 3.1 Associations between nutritional biomarker scores of dietary patterns and incidence of type 2 diabetes in the EPIC-InterAct case-cohort study

Abbreviations: aHEI-2010 – alternative Healthy Eating Index 2010; CI – confidence interval; DASH – Dietary Approaches to Stop Hypertension; HR – hazard ratio; IR – incidence rate; MDS – Mediterranean diet score; SD – standard deviation

Associations were adjusted for: age (as timescale), sex, recruitment centre, prevalent cancer, cardiovascular disease, hypertension and hyperlipidaemia; familial history of type 2 diabetes, smoking status (never, former, current smoker), physical activity index (inactive, moderately inactive, moderately active, active), seasonality (sine and cosine function of the day of the year), fasting status (<3, 3-6, >6 hours), current use of vitamin or mineral supplements, marital status (single, married or cohabiting, divorced or separated, widowed), educational attainment (none, primary school, technical or professional school, secondary school, post-secondary school education), current employment, body mass index and waist circumference, and in women, menopausal status (pre-, peri-, postmenopausal, bilateral oophorectomy), current hormone replacement therapy use. 21,293-21,549 participants were included in the analysis.

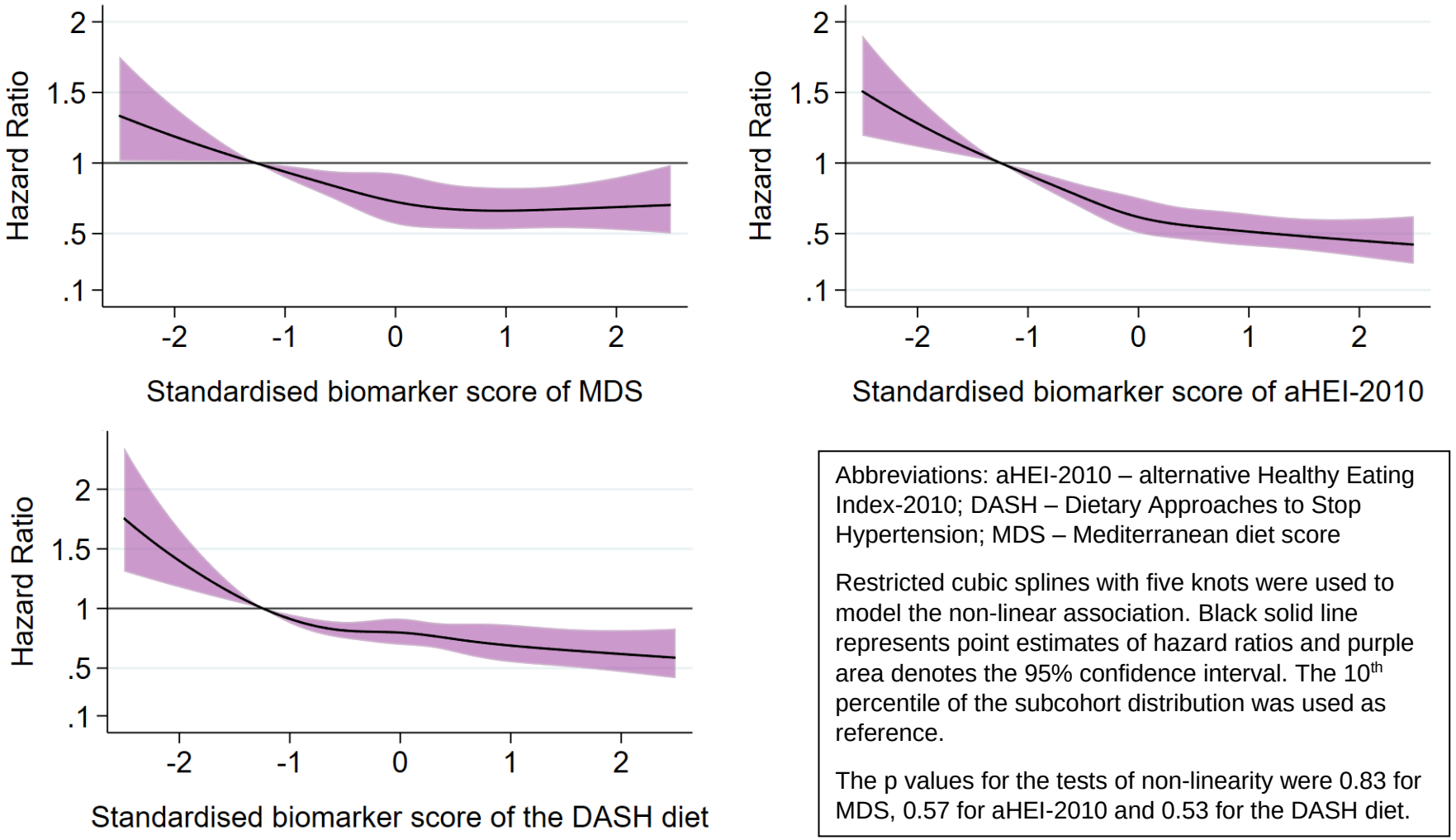


Figure 3.2 Non-linear associations between nutritional biomarker scores of dietary patterns and incidence of type 2 diabetes in EPIC-InterAct

Table 3.8 Associations between nutritional biomarker scores of dietary patterns and incidence of type 2 diabetes in the EPIC-InterAct case-cohort study: pooled hazard ratios (95% CI) after additional adjustment for circulating metabolic factors

Biomarker score and model*	Quintiles					p_{trend}†	Per 1 SD	I^2, % (95% CI)
	Q1	Q2	Q3	Q4	Q5			
MDS (n = 21, 293)								
Main result	1.0 (Ref.)	0.78 (0.64-0.95)	0.68 (0.52-0.88)	0.64 (0.52-0.79)	0.61 (0.47-0.80)	0.012	0.85 (0.79-0.92)	66 (28-84)
+blood lipids†	1.0 (Ref.)	0.80 (0.64-1.00)	0.77 (0.60-0.98)	0.72 (0.59-0.88)	0.72 (0.55-0.94)	0.104	0.90 (0.83-0.97)	61 (19-81)
+liver enzymes‡	1.0 (Ref.)	0.75 (0.61-0.93)	0.66 (0.50-0.88)	0.62 (0.49-0.79)	0.63 (0.48-0.83)	0.056	0.85 (0.78-0.92)	62 (23-82)
+haemoglobin A1c	1.0 (Ref.)	0.78 (0.64-0.95)	0.81 (0.66-0.99)	0.73 (0.58-0.91)	0.79 (0.59-1.05)	0.865	0.90 (0.86-0.95)	4 (0-66)
+hsCRP	1.0 (Ref.)	0.79 (0.65-0.95)	0.69 (0.53-0.90)	0.66 (0.54-0.81)	0.63 (0.49-0.82)	0.024	0.86 (0.80-0.93)	61 (19-81)
aHEI-2010 (n = 21,549)								
Main result	1.0 (Ref.)	0.70 (0.58-0.84)	0.59 (0.46-0.74)	0.52 (0.42-0.65)	0.47 (0.37-0.61)	<0.001	0.75 (0.68-0.82)	79 (59-89)
+blood lipids†	1.0 (Ref.)	0.76 (0.64-0.89)	0.64 (0.51-0.81)	0.59 (0.49-0.72)	0.55 (0.44-0.70)	0.002	0.77 (0.71-0.84)	70 (41-85)
+liver enzymes‡	1.0 (Ref.)	0.74 (0.61-0.89)	0.65 (0.52-0.83)	0.59 (0.47-0.75)	0.55 (0.43-0.71)	0.007	0.79 (0.72-0.85)	65 (29-83)
+haemoglobin A1c	1.0 (Ref.)	0.70 (0.60-0.82)	0.68 (0.57-0.80)	0.56 (0.43-0.73)	0.52 (0.39-0.68)	0.237	0.75 (0.70-0.82)	54 (3-78)
+hsCRP	1.0 (Ref.)	0.71 (0.60-0.85)	0.60 (0.48-0.75)	0.54 (0.43-0.67)	0.49 (0.38-0.63)	<0.001	0.75 (0.68-0.83)	76 (55-88)
DASH (n = 21,293)								
Main result	1.0 (Ref.)	0.81 (0.68-0.96)	0.74 (0.66-0.84)	0.73 (0.60-0.90)	0.68 (0.54-0.86)	0.186	0.83 (0.77-0.90)	61 (16-82)
+blood lipids†	1.0 (Ref.)	0.81 (0.70-0.93)	0.75 (0.66-0.86)	0.75 (0.59-0.97)	0.73 (0.56-0.95)	0.821	0.82 (0.74-0.91)	75 (52-87)
+liver enzymes‡	1.0 (Ref.)	0.84 (0.69-1.04)	0.79 (0.69-0.90)	0.80 (0.63-1.00)	0.76 (0.59-0.98)	0.440	0.83 (0.76-0.91)	72 (45-86)
+haemoglobin A1c	1.0 (Ref.)	0.77 (0.59-1.01)	0.80 (0.69-0.92)	0.83 (0.67-1.03)	0.85 (0.65-1.09)	0.711	0.89 (0.82-0.96)	46 (0-75)
+hsCRP	1.0 (Ref.)	0.83 (0.70-0.99)	0.77 (0.68-0.86)	0.77 (0.64-0.93)	0.71 (0.56-0.91)	0.276	0.83 (0.76-0.91)	71 (42-85)

Abbreviations: aHEI-2010 – alternative Healthy Eating Index 2010; CI – confidence interval; DASH – Dietary Approaches to Stop Hypertension; HR – hazard ratio; hsCRP – high sensitivity C-reactive protein; IR – incidence rate; MDS – Mediterranean diet score; n – number of participants; SD – standard deviation

*The main results were adjusted for: age (as timescale), sex, recruitment centre, prevalent cancer, cardiovascular disease, hypertension and hyperlipidaemia; familial history of T2D, smoking status (never, former, current smoker), physical activity index (inactive, moderately inactive, moderately active, active), seasonality (sine and cosine function of the day of the year), fasting status (<3, 3-6, >6 hours), current use of vitamin or mineral supplements, marital status (single, married or cohabiting, divorced or separated, widowed), educational attainment (none, primary school, technical or professional school, secondary school, post-secondary school education), current employment, and in women, menopausal status (pre-, peri-, postmenopausal, bilateral oophorectomy), current hormone replacement therapy use.

†High-density lipoprotein cholesterol and triglycerides

‡Alanine aminotransferase, aspartate transaminase and gamma-glutamyl transferase

3.5.6 Associations of biomarker scores with incident T2D: sensitivity and secondary analyses

The main results from the adiposity-adjusted multivariable model were robust to multiple sensitivity analyses which explored the influence of time of follow-up, the impact of exclusion of the Swedish participants from derivation of the biomarker scores, potential reverse causality, alternative analytical decisions in deriving the biomarker scores and adjustment for dietary factors (**Table 3.9**). The main results were materially similar between the multiply-imputed and complete-case analysis, however, there was evidence of effect modification by covariates only in the former (**Table 3.10**). For example, the HR per 1 SD of the biomarker score of MDS was 0.84 (95% CI: 0.77-0.91) and 0.82 (95% CI: 0.76-0.90), respectively. In the multiply-imputed analysis non-users of dietary supplements had a HR of 0.81 (95% CI: 0.75-0.88) and users 0.94 (95% CI: 0.87-1.01; $p_{interaction} = 0.006$). The corresponding stratum-specific estimates in the complete-case analysis were 0.81 (95% CI: 0.73-0.90) and 0.87 (95% CI: 0.81-0.94; $p_{interaction} = 0.18$). Beyond the above non-significant result in supplement users, all stratum-specific estimates in the interaction analysis were statistically significantly inverse. The biomarker scores of aHEI-2010 and DASH interacted with baseline age whereby the HRs were higher with increasing age, and the biomarker score of DASH interacted with use of dietary supplements with higher HR in users (**Table 3.10**). There was no evidence of effect modification by sex, BMI, seasonality, fasting status, physical activity, and smoking status ($p_{interaction}$ values > 0.05).

Additional adjustment of the main results for individual nutritional biomarkers suggested that the inverse associations of the biomarker scores of aHEI-2010 and DASH, but not MDS, were not driven by any single biomarker comprising the scores (**Table 3.11**). For the biomarker score of MDS, attenuation to the null occurred after adjustment for (HR; 95% CI): α-carotene (0.93; 0.84-1.02), lutein (0.92; 0.84-1.01) and C24:1 (0.93; 0.84-1.03). These results were not confirmed when using the more parsimonious biomarker score with $\lambda_{+1\ SE}$ penalisation (5 versus 31 biomarkers in the primary score). Mutual adjustment for all components of the biomarker scores attenuated the result to the null for MDS (HR 0.91; 95% CI: 0.46-1.80) and DASH (HR 0.76; 95% CI: 0.52-1.12) but not aHEI-2010 (HR 0.48; 95% CI: 0.26-0.89). Moreover, the more the biomarker scores of MDS and aHEI-2010 with $\lambda_{+1\ SE}$ penalisation remained inversely associated with incident T2D following the mutual adjustment (**Table 3.11**).

Table 3.9 Nutritional biomarker scores of dietary patterns and incidence of type 2 diabetes in the EPIC-InterAct case-cohort study: sensitivity analyses

Model	HR (95% CI) per 1 SD		
	MDS	aHEI-2010	DASH
Main result*	0.85 (0.79-0.92)	0.75 (0.68-0.82)	0.83 (0.77-0.90)
Stratification by follow-up time			
first 7 years	0.80 (0.74-0.87)	0.73 (0.65-0.82)	0.81 (0.72-0.90)
> 7 years	0.87 (0.80-0.94)	0.77 (0.68-0.86)	0.83 (0.74-0.92)
Biomarker scores including Sweden in derivation			
derived using carotenoids and fatty acids only	0.83 (0.77-0.90)	0.75 (0.67-0.84)	0.84 (0.78-0.89)
country-specific scores	0.77 (0.69-0.87)	0.81 (0.76-0.87)	0.79 (0.73-0.85)
Exclusions			
first 2 years of follow-up	0.83 (0.77-0.89)	0.74 (0.67-0.81)	0.81 (0.74-0.89)
participants with HbA1c > 48mmol/mol	0.83 (0.77-0.90)	0.76 (0.69-0.83)	0.82 (0.74-0.90)
participants with prevalent cancer, MI or stroke	0.83 (0.77-0.89)	0.74 (0.67-0.81)	0.80 (0.72-0.88)
outliers in nutritional biomarkers†	0.82 (0.76-0.88)	0.71 (0.65-0.79)	0.79 (0.73-0.86)
Alternative biomarker scores			
single elastic net regression	0.82 (0.76-0.88)	0.77 (0.70-0.84)	0.84 (0.76-0.92)
unpenalised coefficients	0.83 (0.77-0.89)	0.74 (0.68-0.81)	0.81 (0.74-0.89)
increased penalisation ($\lambda_{+1\ SE}$)‡	0.74 (0.67-0.82)	0.70 (0.63-0.79)	-
unadjusted at derivation stage	0.75 (0.67-0.84)	0.68 (0.59-0.78)	0.55 (0.45-0.66)
Additional adjustments¶			
alcohol	0.83 (0.77-0.90)	0.71 (0.65-0.77)	0.80 (0.72-0.87)
meat, olive oil and alcohol	0.83 (0.76-0.90)	0.70 (0.64-0.77)	0.80 (0.72-0.88)

Abbreviations: aHEI-2010 – alternative Healthy Eating Index 2010; CI – confidence interval; DASH – Dietary Approaches to Stop Hypertension; HR – hazard ratio; MDS – Mediterranean diet score; MI – myocardial infarction; SD – standard deviation

*Hazard ratios were pooled from country-specific estimates. Multivariable adjusted model included the following covariates: age (as timescale), sex, recruitment centre, prevalent cancer, cardiovascular disease, hypertension and hyperlipidaemia; familial history of T2D, smoking status (never, former, current smoker), physical activity index (inactive, moderately inactive, moderately active, active), seasonality (sine and cosine function of the day of the year), fasting status (<3, 3-6, >6 hours), current use of vitamin or mineral supplements, marital status (single, married or cohabiting, divorced or separated, widowed), educational attainment (none, primary school, technical or professional school, secondary school, post-secondary school education), current employment, body mass index and waist circumference, and in women, menopausal status (pre-, peri-, postmenopausal, bilateral oophorectomy), current hormone replacement therapy use.

†Excluding participants with ≥1 biomarker required for calculation of a given score outside of the 25th percentile minus 3 times the interquartile range (IQR) or the 75th percentile plus 3 times the IQ of log-transformed, country-specific distributions.

‡The $\lambda_{+1\ SE}$ penalty selects the λ penalty as the largest λ value that is within 1 standard of the minimum of the cross-validation function, as opposed to the optimal λ selected via cross-validation, leading to selection of more parsimonious biomarker score equations. No predictors were selected for the DASH diet under this increased penalisation.

¶Adjusted for estimated energy intake using the nutrient residual method. Alcohol was modelled using restricted cubic splines (4 knots) and meat and olive oil were entered into the models as continuous terms.

Table 3.10 Nutritional biomarker score of dietary patterns and incidence of type 2 diabetes in EPIC-InterAct: associations per 1 standard deviation by categories of covariates*

Biomarker score and covariate category	Multiply imputed analysis			Complete-case analysis		
	n†	HR (95% CI)	$p_{interaction}$‡	n	HR (95% CI)	$p_{interaction}$‡
MDS						
Main result	21,293	0.85 (0.79-0.92)	-	14,857	0.82 (0.76-0.90)	-
Dietary supplements						
Non-users	13,220	0.81 (0.75-0.88)			0.81 (0.73-0.90)	
Users	8,073	0.94 (0.87-1.01)	0.006		0.87 (0.81-0.94)	0.18
aHEI-2010						
Main result	21,549	0.75 (0.68-0.82)	-	14,998	0.71 (0.64-0.79)	-
Age at baseline, years						
<45	4,108	0.67 (0.57-0.79)			0.73 (0.66-0.82)	
45-60	12,516	0.73 (0.66-0.81)			0.69 (0.61-0.78)	
>60	4,925	0.78 (0.70-0.88)	0.035		0.75 (0.66-0.85)	0.07
DASH						
Main result	21,293	0.83 (0.77-0.90)	-	14,857	0.78 (0.70-0.87)	-
Dietary supplements						
Non-users	13,220	0.78 (0.72-0.85)			0.78 (0.70-0.88)	
Users	8,073	0.88 (0.78-0.99)	0.019		0.82 (0.70-0.95)	0.47
Age at baseline, years						
<45	4,055	0.76 (0.63-0.92)			0.82 (0.62-1.08)	
45-60	12,374	0.82 (0.76-0.88)			0.78 (0.71-0.86)	
>60	4,864	0.85 (0.76-0.95)	0.039		0.81 (0.70-0.94)	0.31

Abbreviations: aHEI-2010 – alternative Healthy Eating Index 2010; CI – confidence interval; DASH – Dietary Approaches to Stop Hypertension; HR – hazard ratio; MDS – Mediterranean diet score; MI – myocardial infarction; SD – standard deviation

*Hazard ratios were pooled from country-specific estimates. Multivariable adjusted model included the following covariates: age (as timescale), sex, recruitment centre, prevalent cancer, cardiovascular disease, hypertension and hyperlipidaemia; familial history of T2D, smoking status (never, former, current smoker), physical activity index (inactive, moderately inactive, moderately active, active), seasonality (sine and cosine function of the day of the year), fasting status (<3, 3-6, >6 hours), current use of vitamin or mineral supplements, marital status (single, married or cohabiting, divorced or separated, widowed), educational attainment (none, primary school, technical or professional school, secondary school, post-secondary school education), current employment, body mass index and waist circumference, and in women, menopausal status (pre-, peri-, postmenopausal, bilateral oophorectomy), current hormone replacement therapy use. Presence of interaction was also evaluated for sex, BMI, seasonality, fasting status, physical activity and smoking status ($p_{interaction}$ values for biomarker score-covariate pairs not reported in table > 0.05).

†Numbers of participants by use of dietary supplements in multiply imputed analysis are mid-point values between the smallest and the largest values in the imputation datasets.

‡Interaction p values for age are based on continuous-by-continuous interaction terms between age and biomarker score.

Table 3.11 Nutritional biomarker scores of dietary patterns and incidence of type 2 diabetes in the EPIC-InterAct case-cohort study: adjustment for component biomarkers*

Model†	MDS		aHEI-2010		DASH
	λ_{CV}‡	$\lambda_{+1\ SE}$ ¶	λ_{CV}	$\lambda_{+1\ SE}$	λ_{CV}
Main result	0.85 (0.78-0.93)	0.73 (0.65-0.83)	0.75 (0.68-0.82)	0.73 (0.65-0.83)	0.83 (0.77-0.90)
+α-carotene	0.93 (0.84-1.02)	0.77 (0.68-0.89)	0.80 (0.72-0.89)	0.77 (0.67-0.89)	0.91 (0.81-1.03)
+ß-carotene	-§	-	0.84 (0.75-0.93)	-	0.92 (0.83-1.02)
+ß-crypt.	0.89 (0.81-0.97)	-	0.77 (0.70-0.85)	-	0.85 (0.76-0.94)
+Lycopene	0.88 (0.81-0.97)	-	-	-	0.84 (0.77-0.92)
+Lutein	0.92 (0.84-1.01)	0.76 (0.66-0.88)	0.79 (0.71-0.87)	-	0.88 (0.80-0.96)
+Zeaxanthin	0.86 (0.79-0.94)	-	0.75 (0.67-0.84)	-	0.82 (0.75-0.90)
+Vitamin C	0.89 (0.82-0.96)	0.73 (0.64-0.84)	0.77 (0.70-0.84)	-	0.86 (0.79-0.93)
+C14:0	-	-	-	-	0.81 (0.73-0.89)
+C15:0	0.82 (0.75-0.90)	-	-	-	0.82 (0.74-0.91)
+C16:0	0.83 (0.75-0.91)	-	0.79 (0.72-0.86)	-	0.81 (0.73-0.90)
+C16:1	-	-	0.80 (0.72-0.89)	0.75 (0.63-0.89)	0.80 (0.72-0.89)
+C17:0	0.84 (0.76-0.91)	-	0.82 (0.74-0.91)	0.81 (0.70-0.93)	0.86 (0.78-0.95)
+C17:1	0.85 (0.78-0.92)	-	0.75 (0.67-0.82)	-	0.82 (0.74-0.90)
+C18:0	-	-	0.75 (0.68-0.82)	-	0.81 (0.73-0.90)
+C18:1n-9c	0.85 (0.78-0.93)	-	0.69 (0.61-0.77)	0.62 (0.52-0.73)	0.81 (0.74-0.90)
+C18:1n-9t	0.83 (0.76-0.92)	-	0.76 (0.69-0.84)	-	0.81 (0.73-0.90)
+C18:2n-6	-	-	-	-	0.82 (0.74-0.91)
+C18:2n-6t	-	-	0.75 (0.67-0.85)	-	0.82 (0.74-0.91)
+C18:3n-3	0.85 (0.78-0.93)	-	-	-	0.82 (0.74-0.90)
+C18:3n-6	0.87 (0.80-0.95)	-	0.77 (0.70-0.85)	-	0.80 (0.72-0.89)
+C20:0	0.85 (0.77-0.93)	-	0.79 (0.69-0.89)	-	0.78 (0.70-0.88)
+C20:1	0.85 (0.78-0.93)	-	0.74 (0.67-0.83)	-	0.82 (0.74-0.90)
+C20:2	-	-	-	-	0.82 (0.75-0.91)
+C20:3n-6	0.89 (0.82-0.97)	-	0.77 (0.69-0.85)	-	-
+C20:4n-6	-	-	0.74 (0.67-0.82)	-	0.82 (0.74-0.90)
+C20:5n-3	0.82 (0.75-0.90)	-	-	-	0.82 (0.74-0.90)
+C22:0	0.85 (0.78-0.92)	-	0.78 (0.70-0.87)	-	0.78 (0.70-0.87)
+C22:4n-6	0.86 (0.79-0.93)	0.66 (0.57-0.76)	-	0.69 (0.60-0.79)	0.82 (0.75-0.90)
+C22:5n-3	-	-	0.73 (0.66-0.82)	-	0.82 (0.74-0.91)
+C22:5n-6	-	-	0.73 (0.67-0.81)	-	0.82 (0.75-0.90)
+C22:6n-3	0.84 (0.77-0.92)	0.68 (0.58-0.78)	0.75 (0.67-0.83)	-	0.81 (0.74-0.90)
+C23:0	0.84 (0.76-0.93)	-	0.77 (0.67-0.87)	-	0.80 (0.72-0.90)
+C24:0	0.87 (0.79-0.96)	-	0.79 (0.69-0.91)	-	0.82 (0.74-0.91)
+C24:1	0.93 (0.84-1.03)	-	0.78 (0.71-0.87)	-	0.82 (0.73-0.92)
+Ferritin	0.84 (0.76-0.93)	-	0.78 (0.71-0.87)	-	0.83 (0.74-0.92)
+Transferrin	0.85 (0.78-0.92)	-	-	-	0.81 (0.73-0.90)
+Iron	0.85 (0.78-0.93)	-	0.75 (0.68-0.83)	-	0.82 (0.74-0.90)
+25(OH)D_3	0.86 (0.78-0.94)	-	-	-	0.82 (0.74-0.91)
+Magnesium	0.85 (0.78-0.93)	-	0.75 (0.67-0.84)	-	0.81 (0.73-0.90)
+Calcium	0.84 (0.77-0.92)	-	-	-	0.81 (0.74-0.90)
+All the above	0.91 (0.46-1.80)	0.48 (0.26-0.89)	0.46 (0.29-0.72)	0.42 (0.22-0.79)	0.76 (0.52-1.12)

Abbreviations: aHEI-2010 – alternative Healthy Eating Index 2010; CI – confidence interval; DASH – Dietary Approaches to Stop Hypertension; HR – hazard ratio; MDS – Mediterranean diet score; MI – myocardial infarction; SD – standard deviation

*Hazard ratios were pooled from country-specific estimates. Participants from Sweden were excluded from this analysis due to lack of measurement of vitamin C, transferrin and serum cations. Between 19,719-20,619 participants were included.

†The main result was adjusted for: age (as timescale), sex, recruitment centre, prevalent cancer, cardiovascular disease, hypertension and hyperlipidaemia; familial history of T2D, smoking status (never, former, current smoker), physical activity index (inactive, moderately inactive, moderately active, active), seasonality (sine and cosine function of the day of the year), fasting status (<3, 3-6, >6 hours), current use of vitamin or mineral supplements, marital status (single, married or cohabiting, divorced or separated, widowed), educational attainment (none, primary school, technical or professional school, secondary school, post-secondary school education), current employment, body mass index and waist circumference, and in women, menopausal status (pre-, peri-, postmenopausal, bilateral oophorectomy), current hormone replacement therapy use. Nutritional biomarkers were entered into the models as linear and quadratic terms on the normal (non-log) scale. For the analyses adjusting for all components of a given biomarker score, only the quadratic terms were retained which were statistically significant in analyses of individual biomarkers.

‡Primary biomarker scores derived using elastic net regression with λ penalty selected based on minimising cross-validated prediction error.

¶The $\lambda_{+1\ SE}$ penalty selects the λ penalty as the largest λ value that is within 1 standard of the minimum of the cross-validation function, as opposed to the optimal λ selected via cross-validation, leading to selection of more parsimonious biomarker score equations. No predictors were selected for the DASH diet under this increased penalisation.

§Biomarker not included in a given score.

3.6 Discussion

In this investigation, I evaluated combinations of nutritional biomarkers as biomarkers of pre-defined dietary patterns and predictors of incident T2D. The key findings were that biomarker scores of MDS, aHEI-2010 and DASH were modestly correlated with adherence to their respective dietary patterns estimated from self-report, and that the biomarker scores were inversely associated with incidence of T2D in the pan-European EPIC-InterAct study. The inverse association of the biomarker score of MDS corroborated the earlier report from this study on MDS estimated from dietary self-report.[182] Null associations were previously reported in EPIC-InterAct for self-reported aHEI and DASH indices, which I confirmed for aHEI-2010 and replicated for DASH.[183] By contrast, I found strong inverse associations when using the objectively measured markers of these dietary patterns. The biomarker score of aHEI-2010 had the most robust positive relationship with dietary self-report and the strongest inverse association with incident T2D. There was evidence to suggest that this association was independent from the biomarkers constituting the score. This independence was also observed for a secondary parsimonious biomarker score of the MDS, but not for the primary one or the biomarker score of the DASH diet index.

3.6.1 Strengths

The major strength of the current research was the use of a novel approach to derivation of biomarker scores of dietary patterns and its application to evaluating diet-disease associations. The main results were robust to several sensitivity analyses and alternative modelling of biomarker scores. The analysis was based on the largest study to date of incident T2D which measured at scale a comprehensive set of nutritional biomarkers. Data from over 10,000 subcohort participants were used to derive the biomarker scores, with the large sample size favourably affecting their prediction accuracy,[225] and approximately 9,000 incident cases were available to evaluate their associations with T2D, thus providing ample statistical power for longitudinal investigations. Both stages of analysis were adjusted for a wide range of potential confounding factors.

For derivation of the biomarker scores, this adjustment increased the likelihood of independence of the measurement errors of biomarker scores from confounders of diet-disease associations, which is an important criterion for validity of dietary biomarkers.[226,227] Moreover,

it minimised the selection of biomarkers into the scores due to confounding structures specific to the EPIC-InterAct population, potentially resulting in improved external generalisability. I stabilised the variable selection process by utilising bootstrap-enhanced elastic net regression[212–214] and applying a stringent criterion of at least 90% selection rate across the bootstrap samples for inclusion into the biomarker scores. This process was not only likely to reduce type I error, but also to decrease model overfitting. I applied a leave-one-out cross-validation framework in the derivation samples to further minimise overfitting. Independent datasets of the oversampled non-subcohort incident T2D cases confirmed a similar magnitude of correlations between the biomarker scores and their respective dietary patterns as in the subcohort. The correlation coefficients were only marginally lower which suggested reasonable external performance and low spectrum bias,[228] particularly given the differential population characteristics of the incident cases which may have negatively impacted on the models' performance.

3.6.2 Limitations: derivation of the biomarker scores

Major limitations of derivation of the biomarker scores for application to associations with incident T2D were the internal development and lack of an objective validation criterion of the scores as dietary biomarkers. Correlations between the biomarker scores and their respective dietary patterns were modest, but in the range typically observed between established biomarkers of intake and their corresponding measures from self-report.[229,230]

The biomarker scores were derived based on prediction of dietary self-report. Effectively, the outcome variables were combinations of the true levels of adherence to dietary patterns, and random and systematic measurement errors. Under the assumption of uncorrelated measurement errors between the dietary patterns and the biomarker predictor variables, the outcome measurement error would be expected to result in a downward bias in predictive accuracy of the true values of dietary pattern indices, and, potentially, attenuation of the associations between the biomarker scores and incident T2D.[231,232] However, an alternative scenario of correlated measurement errors must be considered.[233] Suppose that a variable known to be a common source of systematic dietary misreporting, such as BMI,[234] could also follow a non-classical measurement error in relation to biomarker concentrations due to biological phenomena. For example, adiposity could be positively related to preferential storage in adipose tissue of fat-soluble biomarkers, e.g., carotenoids, and in consequence a

systematic relative decrease of circulating carotenoids as a proportion of the total body pool. Implications of such correlated measurement errors could lead to misspecification of the biomarker score algorithms in terms of both variable selection and bias in the coefficients. Their impact on performance of the biomarker scores as biomarkers of dietary patterns would be unclear,[233] and the effect on diet-disease associations difficult to predict. Key criteria for validity of biomarkers of intake for studying diet-disease associations are lack of systematic error in relation to the dietary variable of interest and confounding factors,[226] both of which may have not been met. Overall, use of the biomarker scores in the current research was likely to reduce measurement error, however, a partial 'carryover effect' of the systematic error of dietary self-report into the biomarker scores could not be ruled out.

Further limitations of the biomarker scores included the biological properties of individual biomarkers. They are candidate biomarkers of intake, concentration and function which are not only affected by dietary intakes, but also bioavailability, endogenous synthesis, genetic variation, homeostatic control, and nutrient metabolism.[43,80,235] Some of the biomarkers had heterogeneous relationships with dietary exposures with important differences by country. For example, plasma phospholipid oleic acid was previously found in the EPIC study to be correlated with estimated intakes of olive oil in countries of the Mediterranean region, whereas meat intake was its strongest correlate in central European and Scandinavian countries.[157] Such limitations of the biomarkers raise concerns about specificity of the biomarker scores to dietary quality overall, as well as specificity to individual dietary patterns. An additional limitation was the lack standardisation of blood draw procedures, resulting in a varying fasting status of the study participants. Given the lack of specificity of individual biomarkers to dietary patterns or most of their components, the biomarker scores had low content validity and they were inherently predictive, rather than being direct biological measures of adherence to dietary patterns. However, there was evidence to support construct and criterion validity based on the positive associations with dietary patterns and replication of an inverse association between self-reported MDS and incident T2D with use of the biomarker-based assessment.

Limitations in the prospective associations between the biomarker scores and incidence of T2D included the likelihood of several potential sources of bias, i.e., reverse causality early in the follow-up (though unlikely, given the results of my sensitivity analyses), residual confounding, differential misclassification in ascertainment of the outcome, and within-person variation and measurement error of nutritional biomarkers. There may have been degradation of nutritional biomarkers despite storage in liquid nitrogen, but I would not expect bias in relative risk

estimates assuming degradation rates were non-differential by incident case status or biomarker level. Nutritional biomarkers and covariates were measured only once at baseline, and thus I was unable to account for their changes during follow-up. The random measurement error in the nutritional biomarkers would be expected to bias the association towards the null when considered in isolation from other sources of error, but the context of multivariable statistical modelling precludes any inference on the direction of the potential bias.[22] The potential differential misclassification in the ascertainment of the outcome also may have biased the results in an unknown direction. I can speculate that higher adherence to dietary patterns may have been associated with greater health-consciousness and healthcare-seeking behaviours, higher likelihood of T2D diagnosis, and an underestimation of the inverse association. I standardised measures of the biomarker scores and adherence dietary patterns estimated from self-report to allow for a comparison between their strength of association. This approach conditioned the effect sizes on the underlying distributions of the exposure variables in EPIC-InterAct which may have limited the quantitative and comparative interpretation.[236] The use of dietary supplements was available as a binary variable which allowed only for a crude assessment of effect modification.

3.6.3 Comparison with previous research: biomarkers of dietary patterns

Validity of combinations of nutritional biomarker scores as biomarkers of dietary patterns has recently been demonstrated in a novel feeding design of individualised habitual diets in an ancillary study of the Women's Health Initiatie (WHI) in 153 postmenopausal women.[50] The study quantified the level of adherence to the alternative Mediterranean diet (aMED) index, aHEI-2010, Healthy Eating Index-2010 (HEI-2010) and the DASH diet (index by Fung et al.[237]) based on foods and beverages provided to participants, thereby eliminating measurement error due to self-report. The aMED index was conceptually similar to the relative MDS used in the current study, and it differed primarily in the application of scoring cut-offs based on median intakes, rather than tertiles.[37] The set of candidate nutritional biomarker predictors had a substantial overlap with the biomarker available in EPIC-InterAct, and it included serum phospholipid fatty acids, carotenoids, tocopherols, retinol, folate, vitamin B_{12}, and 24-hour urinary nitrogen, sodium, potassium and energy expenditure estimated from the doubly labelled water technique. The study used an internal validation criterion of ≥0.36 (R^2) variance in the dietary patterns explained by nutritional biomarkers.[50] This cut-off was established based

on the correlation of 0.60 observed between feeding amounts of protein and 24-hour urinary nitrogen excretion.[44] Following variable selection by means of cross-validated least absolute shrinkage and selection operator (lasso; elastic net regression with α penalty = 1), the internal validation criterion was met for biomarkers of aMED and HEI-2010, but not for aHEI-2010 and DASH (R^2 = 0.30 each). Primarily serum carotenoids and phospholipid fatty acids were included in the scores. This was similar to the current research, thus providing some evidence in support of construct validity of the EPIC-InterAct biomarker scores. The WHI scores were, however, more parsimonious, consistent with the use of lasso which favours sparser regression solutions than elastic net regression.[212] Of note, SFAs were generally not selected into the WHI scores except for C15:0 for HEI-2010 and C24:0 for aHEI-2010.

Among reports from observational research, Gerber developed a biomarker score based on plasma ß-carotene and vitamin E, and erythrocyte eicosapentaenoic and docosahexaenoic fatty acids, which had a Spearman correlation of 0.52 with the Mediterranean Diet Quality Index (DQI) in the general population of a French Mediterranean region.[238] Neuhouser et al. identified a multivariable model predicting (a non-Mediterranean) DQI which consisted of plasma vitamin C, α-tocopherol, ß-carotene, ß-cryptoxanthin, and oleic (C18:1-n9c) and stearic (C18:0) fatty acids. It explained 36% of the variance in DQI in postmenopausal women, equivalent to a correlation coefficient of 0.60.[239] These correlation coefficients were higher than observed in the current study, which may have been due to use of more homogeneous study samples, not accounting for confounding by non-dietary factors, and using statistical methods which do not prevent overfitting in derivation samples. Alternatively, the nutrient adequacy-focused DQI may have been better suited as a construct for prediction from biomarker concentrations than the chronic disease-related dietary pattern indices used in the current investigation.

3.6.4 Comparison with previous research: diet-disease associations using biomarker scores

To my knowledge, the only example of derivation and application of nutritional biomarker scores of dietary patterns to diet-disease associations was a recent investigation in 642 participants of the InCHIANTI study on the Mediterranean diet and mortality (435 deaths).[240] Using the EPIC MDS[201] as a conceptual reference, the authors applied its tertile-based scoring algorithm to concentrations of nutritional biomarkers in lieu of estimated dietary intakes. Each

MDS component had a pre-specified biomarker assigned to it based on prior subject matter knowledge. The biomarker score was positively correlated with MDS (rho = 0.26) and inversely associated with mortality ($HR_{T3vs.T1}$ = 0.72; 95% CI: 0.56-0.91) whereas the self-reported MDS was not ($HR_{T3vs.T1}$ = 0.90; 95% CI: 0.69-1.19). The approach taken to derivation of the biomarker score was fundamentally flawed by applying a hypothesis-free method which did not aim to maximise the predictive power of the biomarker for the Mediterranean diet, thereby ignoring the non-dietary regulation of concentrations of nutritional biomarkers.[39] Furthermore, the single biomarkers assigned to each MDS component were unlikely to be valid and specific biomarkers of intake or exposure to their respective foods or nutrients. This led to debatable choices of biomarkers for some exposures; for example, the sum of long[153] and very-long even-chain SFA for meat,[241] plasma selenium for cereals[242] or serum vitamin B_{12} for dairy.[243,244]

Metabolomic profiling has been the primary method of development of biomarkers of dietary patterns,[51,69] and applications to incidence of T2D have been reported in the literature. Shi et al. identified metabolite scores of Baltic Sea Diet Score and Healthy Nordic Food Index using a two-stage approach.[245] First, random forest modelling was applied to identify predictors of the diet indices estimated from self-report, followed by principal component analysis performed on these pre-selected metabolites. The first principal component was modestly correlated with the indices (partial Spearman correlation: 0.25-0.27) but not with incidence of T2D in a nested case-control study (odds ratio per 1 SD ≈ 1.1, 95% CI 0.95-1.20; values read from graph). Similar to the above described InCHIANTI study, the hypothesis-free approach taken to derivation of the metabolomic profiles was likely to limit the ability of the score to capture the variation in metabolite levels attributable to adherence to the dietary patterns under investigation. An analysis from the Nurses Health Studies and the Health Professionals Follow-up Study has employed elastic net regression to derive metabolite profiles of plant-based diet indices.[246] The analytical approach and the key findings were similar to the ones for MDS in the current investigation: the metabolite scores were positively correlated with their respective dietary pattern indices (r range: 0.33-0.45) and metabolite scores of healthy dietary patterns were inversely associated with incidence of T2D (HR per 1 SD ~0.8), whereas associations for adherence estimated from self-report were moderately weaker.[246]

3.6.5 What this study adds and implications of this research

Previous investigations in the EPIC-InterAct study suggested the aHEI and DASH indices were not associated with incidence of T2D[183] and that MDS had an inverse association, though it did not remain statistically significant in several secondary analyses.[182] In the current study, I have found inverse associations of all three dietary patterns using objectively measured biomarkers which were largely robust to multiple sensitivity analyses. These results strengthen the evidence for benefits of the Mediterranean dietary pattern in prevention of T2D. They also raise the possibility that aHEI and DASH may be inversely related to new-onset T2D in European populations, and they increase the consistency of findings from the EPIC-InterAct study with the overall body of evidence for the two dietary patterns.[181] The inverse associations of the biomarker scores with incidence of T2D were largely independent from circulating biomarkers of cardiometabolic health or individual nutritional biomarkers. This suggests that biomarker scores of dietary quality may be novel risk factors for incident T2D with potential utility for enhancing risk prediction models which could be explored in future research.

The above interpretation of the current findings rests on the assumption of validity of the biomarker scores as biomarkers of dietary patterns which was untestable within the framework of the current research. Further work on derivation of biomarkers of dietary patterns is required using feeding and interventional designs to exclude or minimise the influence of subjective measurement of diet. Additionally, future research should test a broader range of nutritional biomarkers for development of biomarkers of dietary patterns than the set available in the current study and compare the utility of targeted measurements of classical nutritional biomarkers with omics approaches.

Dietary patterns represent an amalgamate of dietary goals which may be fulfilled or unmet through multiple combinations of dietary choices. This unique characteristic among dietary exposures has implications for biomarker development. It indicates that there may be heterogeneous clusters of dietary behaviours within participants with the same level of quantified adherence to a given dietary pattern. For example, certain individuals may maximise intakes of recommended foods or nutrients while consuming considerable amounts of unhealthy components with negative weighting or vice versa. A single biomarker score derived from coefficients of linear biomarker terms is unlikely to adequately capture the level of adherence at such extremes. Derivation of the biomarker scores could be improved by considering interactions between nutritional biomarkers in the modelling approach to account

for potential heterogeneity in biomarker concentrations due to the above-outlined variability in clustering of dietary behaviours.

3.6.6 Conclusions

Adherence to the pre-defined healthy dietary patterns may potentially be assessed by combining information from multiple circulating nutritional biomarkers into biomarker scores. Biomarker scores of the Mediterranean Diet Score, Alternative Healthy Eating Index-2010 and the DASH diet were inversely associated with incidence of T2D. Only the Mediterranean Diet Score had an inverse association when examining these relationships using dietary self-report. This raises the possibility that objective assessment of adherence to dietary patterns using biomarkers can enable the unmasking of associations attenuated to the null in analyses based on subjective quantification of dietary intakes.

Chapter 4

A nutritional biomarker score of the Mediterranean diet and incident type 2 diabetes: experimental and observational analysis from the MedLey randomised controlled trial and the EPIC-InterAct case-cohort study

Abstract

Background: Self-reported adherence to the Mediterranean diet has been modestly inversely associated with incidence of type 2 diabetes (T2D) in cohort studies. There is uncertainty about the validity and magnitude of this association due to subjective reporting of diet. The association has not been evaluated using an objectively measured biomarker of the Mediterranean diet.

Methods: I derived a biomarker score based on five circulating carotenoids and 24 fatty acids that discriminated between the Mediterranean or habitual diet arms of a parallel design, partial-feeding randomised controlled trial, the MedLey trial (128 participants out of 166 randomised). I then applied this biomarker score in an observational study, the European Prospective Investigation into Cancer and Nutrition (EPIC)-InterAct case-cohort study, to assess the association of the score with T2D incidence. It included 22,202 participants, of whom 9,453 were T2D cases, with relevant biomarkers from an original case-cohort of 27,779 participants sampled from a cohort of 340,234 people.

Findings Within the trial, the biomarker score discriminated well between the two arms; the cross-validated C-statistic was 0.88 (95% confidence interval (CI) 0.82 to 0.94). The score was inversely associated with incident T2D in EPIC-InterAct: the hazard ratio (HR) per standard deviation of the score was 0.71 (95% CI: 0.65 to 0.77). In comparison, the HR per standard deviation of the self-reported Mediterranean diet was 0.91 (95% CI: 0.86 to 0.95). Assuming the score was causally associated with T2D, higher adherence to the Mediterranean diet in Western European adults by 10 percentiles of the score was estimated to reduce the incidence of T2D by 11% (95% CI: 7 to 14%).

Conclusions: These findings suggest that objectively assessed adherence to the Mediterranean diet is associated with lower risk of T2D and that even modestly higher adherence may have the potential to reduce the population burden of T2D meaningfully.

4.1 Background

Assessment of adherence to dietary patterns using biomarkers has been based almost exclusively on observational study designs.[51,55,60,69] Typically, development of biomarkers of complex dietary exposures in such settings consists in building multivariable models whereby self-reported intakes, the dependent variables, are predicted from concentrations of multiple circulating or urinary compounds. It has been proposed to improve dietary assessment compared to food frequency questionnaires alone,[51] however, it inevitably incorporates the dietary misreporting into the analytical framework. The extent to which the measurement error of dietary data impacts on performance of the resultant composite biomarkers has not been evaluated, and thus their validity remains unknown. An additional concern specific to dietary patterns is the construct validity of the indices used to assess adherence to a given pattern.[247] For example, indices of the Mediterranean diet may not adequately capture the adherence to this diet when applied in non-Mediterranean settings.[175] These limitations could be addressed by using experimental, preferably feeding designs for derivation of biomarkers or validation of biomarkers developed in observational settings. Thus, I sought to integrate the EPIC-InterAct data with an intervention with the Mediterranean diet which measured a similar set of nutritional biomarkers.

4.1.1 External derivation of biomarkers: study design considerations

Using the systematic review (Chapter 2) and background reading of the literature, I have identified and invited for collaboration studies which could be used for external derivation and validation of biomarkers of the Mediterranean diet which had a large overlap with EPIC-InterAct in terms of the nutritional biomarkers measured. Candidate studies included a feeding study within the Women's Health Initiative (WHI),[44] the PREDIMED trial[117] and the MedLey Study (referred from here onwards as the MedLey trial).[64]

I initially approached the Nutrition and Physical Activity Assessment Study Feeding Study (NPAAS-FS), an ancillary study of the WHI.[44] The collaboration was declined on the basis of already ongoing work on biomarkers of dietary patterns (later published and applied in Chapter 5).[50] This study introduced a novel feeding design in which 153 participants were provided for two weeks with personalised total diets that aimed to recreate their baseline habitual diets. Following the feeding period, a range of nutritional biomarkers were measured, including

serum carotenoids[44] and plasma phospholipid fatty acids.[52] Prediction equations for multiple dietary exposures were then developed based on nutritional biomarker concentrations and personal characteristics.[44,50,52] Though technically a feeding intervention, this study aimed to minimise the effect of the intervention on the baseline equilibrium of the biomarkers, and to preserve the baseline inter-individual variation in dietary intakes. These two features are the opposites of typical objectives of dietary interventions. As a result, the study was conceptually akin to a cross-sectional design, and it used a similar analytical framework as I applied in Chapter 3.[50] Strengths of this novel feeding design were minimisation of the measurement error, preservation of habitual distributions of intake, minimal concerns about equilibrium of biomarker concentrations, and the possibility of deriving biomarker scores for a wide range of nutrients,[44,52,248–250] foods[248] and dietary patterns.[50] A major limitation was the potential for reliance on associations between dietary intakes and nutritional biomarkers arising due to confounding to maximise the explanatory power of the prediction equations. This raises concerns about external generalisability and predictive power of the equations outside of the population of the WHI. Further exacerbating this issue, the prediction equations may require incorporating into them participants' personal characteristics, such as BMI, to meet the study's internal validity criterion ($R^2 \geq 0.36$)[50] which has implications for applying the equations to studying diet-disease associations.[251]

Conventional randomised controlled trials (RCT) of dietary interventions were a second class of study design which I considered. A major strength of such designs is the possibility to incorporate the randomised allocation into derivation of biomarker scores.[252] This entails developing classifiers which aim to predict the allocation based on end-of-intervention biomarker concentrations, thereby using the randomised design to minimise confounding. For example, King et al compared the effects of six weeks of a low-fat (17% energy) and moderate-fat (34% energy) diet on circulating fatty acids.[252] They derived a logistic regression model which almost perfectly discriminated between the two interventions based on post-intervention biomarker levels (C-statistic = 0.99). The choice of control diets in such designs is of major importance for derivation of a biomarker score. Some comparisons for interventions with healthy dietary patterns, e.g., a different healthy dietary pattern, may yield composite biomarkers of limited utility for epidemiological applications.

I briefly considered approaching the PREDIMED trial for collaboration, however, given the above consideration, I judged the study design as inappropriate for derivation of a biomarker score of the Mediterranean diet, particularly for application in a pan-European study.

Specifically, PREDIMED intervened on Spanish participants with advice on following the Mediterranean diet and provision of extra-virgin olive oil or mixed nuts compared with advice to follow a healthy diet with reduced fat intake.[117] The per protocol assignment was randomised, however, deviations from the randomisation sequence emerged in 21% of participants with no material influence on the primary outcome.[117] The differences in estimated dietary intakes between the trial arms pertained primarily to the supplemental foods and were otherwise modest, raising concerns whether PREDIMED actually tested the effects of the Mediterranean diet per se.[19] Of note, the difference in the estimated proportion of energy from fat between the Mediterranean diet and the control lower-fat diet was only 4%.[117] Other concerns over suitability of the PREDIMED trial for joint work with the EPIC-InterAct study included an overlap only in circulating fatty acids,[81] lack of clarity on the extent of deviations from randomisation protocol in the subsample with fatty acids, and use of a population at high cardiovascular risk,[117] which could potentially limit external generalisability of the biomarker score.

Finally, I identified the MedLey partial-feeding RCT which tested the effects of the Mediterranean diet against continuations of a habitual diet on circulating carotenoids and fatty acids.[64] I judged it to be well-suited for the proposed work. Moreover, the trial deposited some of the nutritional biomarker data as an online supplement,[64] which allowed me to test the feasibility of the project prior to arranging the data transfer.

4.1.2 Overview of previous research

The Mediterranean diet has been reported to be associated with decreased incidence of multiple non-communicable diseases including type 2 diabetes (T2D).[26,181] However, the evidence for T2D stems predominantly from observational research using self-reported dietary assessment, with modest effect sizes.[6] Dietary self-report is subject to systematic and random measurement error which may bias the associations with disease risk in an unknown direction.[22] Thus, the relationship between the Mediterranean diet and the incidence of T2D may have been misquantified. Uncertainty about the validity and size of this association limits robustness of the evidence for this dietary pattern and T2D incidence.[6]

Research on the derivation of objective biomarkers of adherence to the Mediterranean diet has had a number of limitations. Previous studies used mostly cross-sectional designs without external validation,[60] and though evidence from feeding[50] or experimental[177] designs is

promising, it has been hampered by challenges such as the different definitions of the Mediterranean diet or interventions used,[19,50,175] the targeting of specific populations,[50,177] and the use of sub-group analyses or not fully randomised enrolment of participants.[50,117] Overall, the validity and external generalisability of biomarkers of this dietary pattern reported in the literature remain largely unknown. Given these limitations, biomarker-based assessment of the Mediterranean diet has rarely been applied to the associations with disease outcomes,[60] and never previously for T2D (beyond the work presented in Chapter 3).

4.2 Aim

The objectives were to derive a nutritional biomarker score that could discriminate between a Mediterranean diet intervention and a habitual diet arm of a RCT, and to test the association of the biomarker score with incident T2D in a population based observational study. Additionally, I aimed to estimate the potential population impact of greater adherence to the Mediterranean diet, as assessed by the biomarker score, and future risk of T2D. As a secondary objective, I sought to evaluate the effects of the Mediterranean diet on individual nutritional biomarkers, relative to the habitual diet of the RCT.

4.3 Methods

The overall study design and participant flows for the main analysis are displayed in **Figure 4.1**. Briefly, I derived a nutritional biomarker score in the MedLey trial as an objective measure of adherence to the Mediterranean diet, and I applied the score in EPIC-InterAct to test its association with incident T2D. Additionally, I used data from non-EPIC-InterAct participants of one of the cohorts participating in EPIC-InterAct, EPIC-Norfolk,[253] for secondary analyses of properties of the biomarker score.

4.3.1 The MedLey trial

The MedLey trial is an RCT that randomised 166 Australians aged ≥65 years from metropolitan Adelaide, Australia, to one of two groups: either the Mediterranean diet with maintenance of baseline body weight or continuation of habitual diet.[64,254] It measured circulating

carotenoids[83,88,90,99] and fatty acids[63,81,83,100,128] as pre-specified secondary outcomes and compliance measures at the end of a 6-month intervention conducted in years 2013 and 2014. The Mediterranean diet intervention consisted of fortnightly sessions with the study dietitian and provision of key shelf-stable foods amounting to 30–35% of estimated energy requirements: virgin olive oil, low-fat Greek yoghurt, unsalted nuts, canned legumes, and canned tuna ($n_{randomised}$ = 85; $n_{commenced}$ = 80). Participants in the usual diet control group received gift vouchers to local food stores ($n_{randomised}$ = 81; $n_{commenced}$ = 72). I used the MedLey trial to derive a biomarker score of discrimination between the Mediterranean and habitual diet arms. End-of-study nutritional biomarkers were available in 68 and 65 participants, respectively. Participants with at least one biomarker value outside of the 25th percentile minus 3 times the interquartile range (IQR) or the 75th percentile plus 3 times the IQR were excluded, leaving 67 and 61 participants in the Mediterranean and habitual diet groups for derivation of the biomarker score.

All participants provided written informed consent and the study was approved by the University of South Australia Human Research Ethics Committee. The MedLey trial was prospectively registered in the Australian New Zealand Clinical Trials Registry (ACTRN12613000602729).

4.3.2 The EPIC-InterAct study

Description of this case-cohort study of T2D and data collection procedures is provided in Chapter 3. Only additional information specific to the current chapter are reported herein.

From the original case-cohort of 27,779 individuals, I excluded participants from the recruitment centre in Malmö, Sweden, because of unavailability of data on plasma carotenoids (n = 3,556) and the remaining participants without measurements of nutritional biomarkers overlapping with those available in the MedLey trial (n = 2,021), leaving 22,202 participants available for analysis, with 9,453 participants who developed incident T2D and 13,313 subcohort participants. As a design feature of the case-cohort study, the subcohort included 564 cases of T2D (**Figure 4.1**).[179]

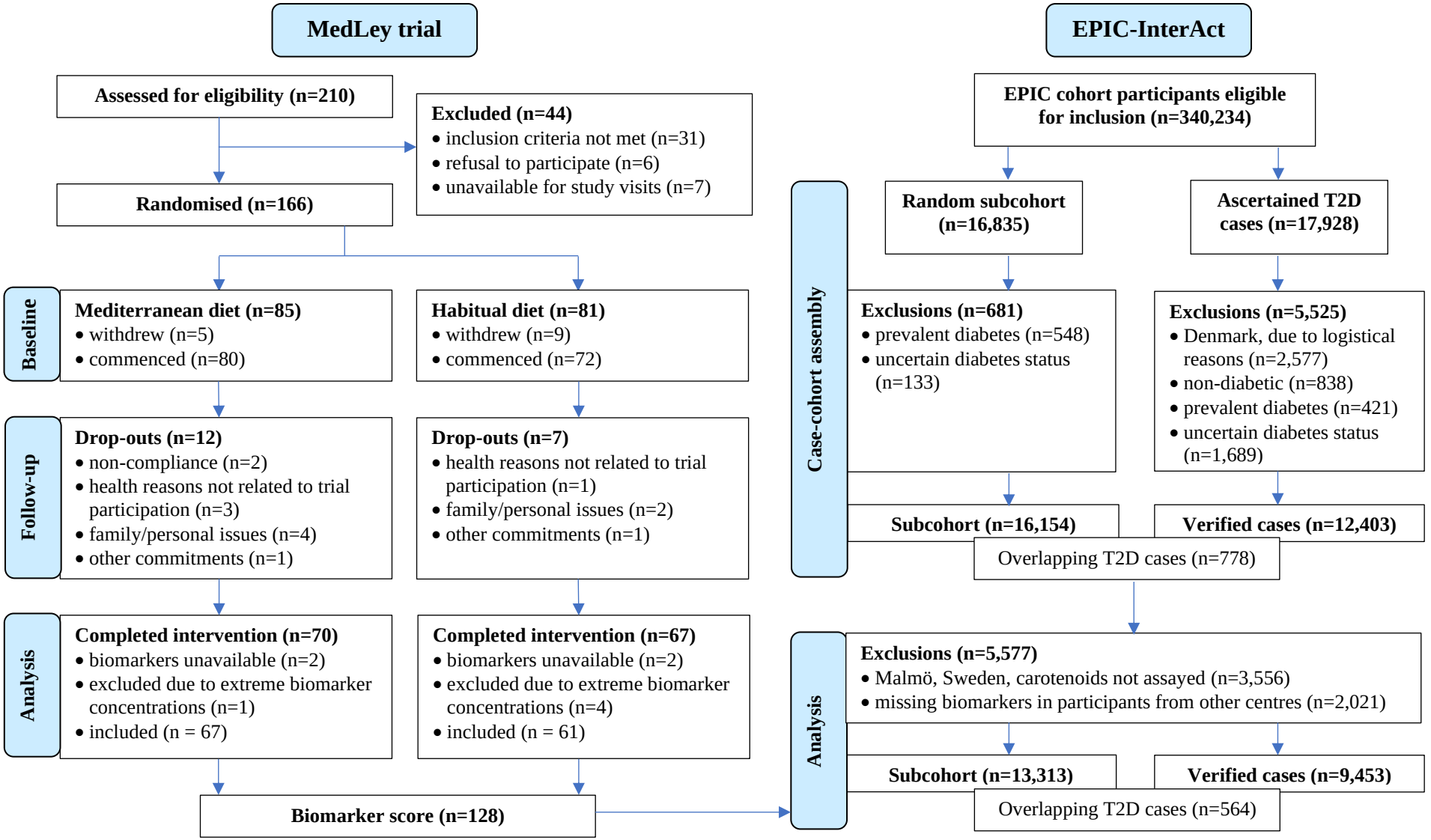


Figure 4.1 Designs of the MedLey trial and the EPIC-InterAct case-cohort study and numbers of participants included in the analysis

4.3.3 Biomarkers

In the MedLey trial, venous blood samples and 24-hour urine were taken at baseline, 3- and 6 months post-randomisation after 8 hours of fasting, centrifuged and stored at -80°C. High performance liquid chromatography with photo diode array detection was used to assay carotenoids.[255] Individual erythrocyte fatty acids were assayed via direct transesterification as weight% of all the fatty acids measured, followed by gas chromatographic analysis.[256] Plasma total cholesterol was measured using a Siemens ADVIA chemistry analyser at SA Pathology (Adelaide, Australia). Urinary 24-h excretion of sodium, potassium, magnesium and calcium was assayed at a laboratory accredited by the Australian National Association of Testing Authorities. Urine was collected starting from the second void until the first void of the following day. Laboratory staff were blinded to randomised intervention allocation.

Five carotenoid variables were measured in both the MedLey trial and EPIC-InterAct study: α-carotene, β-carotene, β-cryptoxanthin, lutein, and sum of lutein and zeaxanthin. For fatty acids measured in the MedLey trial, units were converted from weight% to mol% to harmonise the data with assays of fatty acids in the EPIC-InterAct study. After exclusion of fatty acids with mean concentrations <0.05 mol% of total fatty acids measured, 24 fatty acids overlapped between the MedLey trial and EPIC-InterAct.

4.3.4 Measurement of covariates and self-reported diet

In the MedLey trial, questionnaires and physical examination were used to collect data on baseline characteristics. Self-reported diet was measured with three-day weighed food diaries at 0, 2- and 4-months from either Thursday to Saturday or Sunday to Tuesday. Records were entered and analysed using FoodWorks Professional Version 7.0.3016 (Xyris Software Spring Hill, QLD, Australia) to estimate mean daily dietary intakes. A local nutrient database was used to estimate the intakes of energy and ethanol.[64]

The EPIC-InterAct relative Mediterranean Diet Score was used to quantify the self-reported adherence to this dietary pattern (Chapter 3, **Table 3.1**). The scoring cut-offs established based on tertiles of estimated intakes of foods and nutrients in the EPIC-InterAct subcohort were used when applying the score to the MedLey trial data.

4.3.5 Methods for the EPIC-Norfolk study: impact of assays on the biomarker score

I used EPIC-Norfolk data to test reliability of the biomarker score and the stability of its association with incident T2D when applied to different biomarker assays. Briefly, the Norfolk arm of the EPIC study is a population-based cohort study of 25,639 middle-aged adults in East of England.[253] Detailed information on this cohort is reported in Chapters 6 and 7. Subsamples between 140 to 732 EPIC-Norfolk participants were available for the current analyses.

Venous blood samples were collected at varying times of the day from non-fasted participants and stored in liquid nitrogen. Baseline circulating biomarkers were measured in a subset of non-EPIC-InterAct participants, from EPIC-Norfolk, with the same assays and procedures as used in EPIC-InterAct and described previously (Chapter 3). Several additional assays of the same biomarkers were performed. Plasma carotenoids were measured at the Nutrition and Hormones Laboratory at the International Agency for Research on Cancer (IARC; Lyon, France). All 5 carotenoid variables overlapping between the MedLey Trial and EPIC-InterAct were available. Serum total cholesterol was assayed with the RA 1000 (Bayer Diagnostics, Basingstoke, UK).

Circulating fatty acids were measured in phospholipids and erythrocytes. Phospholipid fatty acids were assayed at IARC using a HP-5980 gas chromatograph (Agilent, Palo Alto, CA) in absolute concentrations and converted to molar percentages of total fatty acids used in the analysis.[257] Composition of erythrocyte fatty acids (mol%) was determined using a GC-3900 gas chromatograph (Varian Associates, Palo Alto, CA) at the Rijksinstituut voor Volksgezondheid en Milieu (RIVM) Laboratory at the National Institute for Public Health and the Environment in Bilthoven, Netherlands. Out of the 24 fatty acids overlapping between EPIC-InterAct and the MedLey trial, C22:0 was not measured in phospholipids and C18:1-n9t was not used due to concerns over accuracy stated in the study's data dictionary, and C17:1, C22:0 and C22:5n-6 were not measured in erythrocytes. I re-derived the biomarker score in the MedLey trial with exclusion of these fatty acids for subsequent analyses in EPIC-Norfolk.

The following subsets of participants were available to test reliability of the biomarker score and the stability of its association with incident T2D:

- Reliability and stability subset 1: From a nested case-control study of incident T2D of 397 participants,[257] I excluded 12 participants with prevalent T2D or unverified incident T2D status according to the EPIC-InterAct case ascertainment criteria[179] and three

participants with missing biomarker data. There were 197 incident T2D cases and 185 non-cases with the IARC assays (plasma carotenoids and phospholipid fatty acids) and erythrocytes fatty acids.

- Stability subset 2: All participants from subset #1 and additional 247 incident cases and 103 non-cases had the IARC assays available, giving a total of 444 incident cases and 288 non-cases.
- Reliability subset 2: 140 participants had data from the IARC assays and the EPIC-InterAct assays.

The biomarker score and individual biomarkers were assessed for reliability by calculating Pearson and intraclass correlation coefficients between measures obtained from different laboratory assays. Stability of the association of the biomarker score with incident T2D was evaluated by estimating odds ratios per 1 SD of the biomarker score in complete-case analysis. The multivariable-adjusted model included the same set of covariates as the main analytical model from the analysis in EPIC-InterAct. These results were compared to the HR from Prentice-weighted Cox regression estimated in EPIC-InterAct for EPIC-Norfolk participants (688 incident cases and 886 non-cases).

4.4 Statistical analysis

Stata 16.1 was used for all analyses except for quantile regression imputation which was done in R, version 4.0.2 (package imputeLCMD). For analyses involving statistical significance testing, two-sided $\alpha = 0.05$ was used.

4.4.1 Derivation of the biomarker score

I imputed values of fatty acids below the limit of detection using quantile regression imputation.[208] In the MedLey trial, values were imputed for five out of the 24 fatty acids and the proportion of missingness was the highest for C17:1 at 21%, and otherwise <13%. The imputation was stratified by trial arm and study visit. In EPIC-InterAct, values were imputed for 16 fatty acids, and the corresponding proportions of missingness were 12% (also for C17:1) and < 6%, respectively. Fatty acids were re-scaled to sum up to 100% within the sets of fatty acids overlapping between the MedLey trial and EPIC-InterAct. Concentrations of circulating

carotenoids (ng/mL) were adjusted for TC using the residual method in order to account for their correlations with dietary carotenoids.[209]

I used logistic elastic net regression to derive the biomarker score in MedLey trial.[212,258] Natural-log-transformed nutritional biomarkers measured at 6 months were used to predict the binary randomised assignment to the Mediterranean or habitual diet groups. Multivariable fractional polynomial logistic regression was used to evaluate the validity of fitting the log-transformed linear biomarker terms under the log-linearity assumption in logistic regression.[259] In the elastic net regression, the λ penalties were selected from a grid of 100 values per each α value.[212] The α values were tested in 0.1 increments between 0.5 and 0.9. This α range favouring the lasso penalty was applied to enable selection of sufficiently parsimonious models for convergence of the post-selection unpenalised regressions given the sample size of the MedLey trial. I repeated the elastic net regression with random 10-fold cross-validation 1,000 times to stabilise the variable selection and included in the final model the predictors with selection rate ≥90%.[258] Ridge regression was used to calculate penalised coefficients of the final model.[215] The biomarker score was calculated as a linear prediction from the ridge regression model.

Predictors were selected from five carotenoid and 24 fatty acid variables, and 406 pairwise interaction terms between them. The rationale for considering interactions between biomarkers was three-fold. First, at any given value of the outcome of the logistic regression model (log odds of being under the intervention), participants may have differentially adhered to individual components of the Mediterranean diet. Thus, a model with a higher variance introduced by the interaction terms may have been better suited to accommodate differential patterns in the uptake of the intervention. Second, there was prior evidence that interactions between fatty acids, such as ratios representing enzymatic activity regulating the metabolism of this group of compounds, could positively contribute to prediction of dietary exposures.[260] Third, the interactions could plausibly increase independence of the biomarker score from the individual biomarkers. In turn, this would be expected to minimise undue influence of the component biomarkers as potential drivers of the association between the biomarker score and disease outcomes.

Discriminatory performance of the biomarker score was assessed by calculating 5-fold cross-validated C-statistic. Cross-validation was implemented by randomly splitting the sample into 5 folds with sampling stratified by the randomised group assignment. Sequentially, 4 folds

were used to re-estimate the coefficients for the biomarker score, followed by calculation of the C-statistic in the left-out fold. The cross-validated C-statistic was the average of the C-statistic values calculated in this manner. The associated 95% confidence interval (CI) was estimated based on the Hanley and McNeil variance formula.[261] I derived several alternative biomarker scores for sensitivity analyses of the impact of analytical decisions on variable selection, discriminatory performance, and stability of associations with incident T2D. This included: changing the selection rate cut-off for inclusion of the predictors from 90% to 95% and 99% across repeated elastic net regression models, changing the unit of fatty acids from mol% to weight%, using all nutritional biomarkers available in the MedLey trial for variable selection (including those non-overlapping with EPIC-InterAct).

As a secondary approach, I derived the biomarker score using only linear biomarker terms by means of the above-described elastic net regression procedure and (separately) backward selection of predictors based on minimising the Bayesian Information Criterion (BIC) in unpenalised logistic regression.[262]

4.4.2 Effects of the Mediterranean diet intervention on biomarkers

Mixed linear modelling with unstructured covariance under the intention-to-treat analysis[263] was the per-protocol analytical approach for estimation of differences between the Mediterranean and habitual diet groups in continuous outcomes.[64,254] I assessed the end-of-trial differences in individual biomarkers and their C-statistic values to facilitate interpretation of the effects of intervention, and to compare the magnitude of effect with the biomarker score. The outcomes were standardised using the baseline means and standard deviations (log-transformed for individual biomarkers; mean = 0, standard deviation = 1). All participants with data available for a given biomarker in at least one timepoint contributed to these analyses. Between 131-136 participants had non-missing data at 6 months. No exclusions were made based on biomarker concentrations.

4.4.3 Associations of the biomarker score with incident T2D

Methods for the country-specific Cox regression and multiple imputation in the EPIC-InterAct study are described in Chapter 3. For comparison with the primary interactions-based

biomarker score, I estimated the HR for the association with incident T2D of biomarker scores derived using only linear biomarker terms.

To assess the potential public health impact of a feasible shift in the average population adherence to the Mediterranean diet, I modelled the population attributable fraction (PAF)[264] and the absolute difference in disease-free time[265] if the value of the biomarker score for each individual in the study population were to increase by 10 percentiles, assuming a causal relationship. The PAF was estimated in the EPIC-InterAct subcohort separately in each country in the adiposity-adjusted multivariable model, and then combined across countries using random effects meta-analysis.[78] I used country-specific inverse-probability weights to account for the case-cohort design.[266,267] The absolute time difference was calculated for the survival percentile equivalent to the cumulative incidence of T2D in the subcohort (4.24 %). To facilitate interpretation of the PAF and disease-free time by comparison with an established risk factor, we estimated this measure for a 10 percentile lower BMI after removing from the multivariable model the biomarker score, waist circumference and physical activity.

4.4.4 Sensitivity analyses

I performed several sensitivity analyses to assess the robustness of the main findings. I repeated the derivation and assessment of discriminatory performance of the biomarker score in the MedLey trial with several alternative analytical decisions in the elastic net regression, followed by longitudinal analysis in EPIC-InterAct for each alternative score. Using the primary biomarker score in EPIC-InterAct, time-varying effects were assessed by splitting follow-up time at 7 years and performing stratified analysis. Potential reverse causation bias was evaluated by separately (i) censoring T2D cases observed during the first 2 years of follow-up, (ii) excluding participants with prevalent major disease conditions (cancer, myocardial infarction, or stroke), and (iii) excluding participants with baseline HbA1c concentrations ≥ 6.5% (48 mmol/mol).

4.4.5 Associations of the biomarker score with self-reported diet

I estimated Pearson correlation coefficients between the MDS and its components, and the biomarker score in the MedLey trial and in the EPIC-InterAct subcohort. In the MedLey trial, I undertook two types of analysis using mixed linear models to evaluate longitudinally the

dose-response relationship between change in the MDS and biomarker score values.

First, I estimated the differences in end-of-trial biomarker score values between the control habitual diet group and each tertile of the MDS in the intervention group, followed by a test for trend across the tertiles. I derived the tertiles of the MDS using best unbiased linear predictions (BLUP) in participants who completed at least 2 out of 3 food diaries or completed the last food diary. In order to test for trend across the tertiles, I ran a mixed linear model restricted to the intervention group, treating the grouping of thirds of adherence as a continuous variable with values equal to BLUP group medians of the MDS. The test for trend was calculated based on a linear combination of the coefficients for (i.) the grouping of thirds of the MDS and (ii.) the interaction between the grouping of thirds of the MDS and the end-of-trial visit indicator.

Second, I estimated standardised coefficients for change in the biomarker score values relative to change in the MDS in each trial arm. Mixed linear models were fitted including an interaction term between the MDS and trial arm indicator. Standardised coefficients of change were estimated in each arm as linear combinations of the main effect for the MDS and the interaction term. The interaction term was also used to test the null hypothesis of equality of the coefficients.

The above analyses included 68 participants from the Mediterranean diet group and 63 individuals from the habitual diet group who had non-missing biomarker score and food diary data at the last assessment. To enable mixed linear modelling, self-reported dietary data collected at months 2 and 4 were fitted as collected at the same timepoints as blood draws for biomarker assessment at months 3 and 6, respectively.

4.5 Results

Compared to the MedLey trial participants, the EPIC-InterAct subcohort members were younger (mean age 52 years, EPIC-InterAct, versus 71 years, MedLey trial), had lower tertiary educational attainment (20% versus 53%), were less likely to have a family history of T2D (18% versus 30%), and had higher prevalence of hypertension (36% versus 19%) and hyperlipidaemia (38% versus 17%). The study populations were similar with regards to dietary and cardiometabolic phenotypes (**Table 4.1**). Blood levels of some of the nutritional biomarkers were materially different between the MedLey trial and EPIC-InterAct, with

substantial variation between the EPIC-InterAct countries. Comparing median concentrations in the MedLey trial habitual diet group and ranges of medians by country in the EPIC-InterAct subcohort, there were differences for ß-carotene (726 and 144-419 ng/mL), lycopene (110 and 172-348 ng/mL), C24:1 (1.22 and 0.32-0.37 %mol) and C24:0 (1.07 and 0.20-0.26 %mol) fatty acids (**Table 4.2**).

4.5.1 Effects of the MedLey trial intervention on nutritional biomarkers

After six months of the intervention, the Mediterranean diet group differed from the habitual diet arm in concentrations of 13 out of the 29 biomarkers overlapping with the Medley trial. The Mediterranean diet intervention resulted in an increase in the concentrations of ß-carotene, lycopene, C22:6-n3 (docosahexaenoic), C20:2-n6 (eicosadienoic), and long- and very-long chain monounsaturated fatty acids. Concurrently, there was a decrease in C22:5-n6 (osbond) and C17:1 (heptadecenoic) and several saturated fatty acids (**Figure 4.2**). The effect sizes (95% CI) relative to the habitual diet control group ranged from -0.67 (-1.00, -0.34) SD for C14:0 (myristic) to 0.76 (0.44, 1.09) SD for C20:1 (gondoic acid). The highest discriminatory performance was detected for C18:1n-9c (oleic acid; C-statistic = 0.70).

Among biomarkers not overlapping with the EPIC-InterAct study, there was a decrease in C18:0 dimethyl acetal and trans-vaccenic acid (C18:1-n7t), an increase in cis-vaccenic acid (C18:1-n7c) and urinary excretion of magnesium, and a marginal increase in potassium excretion (**Appendix 4.1**). Concurrently, there was no effect on C16:0 dimethyl acetal, and urinary sodium and calcium. The highest discriminatory performance was detected for C18:0 dimethyl acetal (C-statistic = 0.72).

Table 4.1 Baseline characteristics of participants of the MedLey randomised partial-feeding controlled trial of the Mediterranean diet and the EPIC-InterAct case-cohort study*

	MedLey trial	EPIC-InterAct	
		Subcohort	Cases of T2D
Number of participants	128	13,313	9,453
Age, years	71 (5)	52 (9)	55 (8)
Women, %	54	63	50
Postmenopausal, %†	100	44	60
Hormone therapy use, %†	5	14	14
Current smokers, %	0	26	28
Moderately active or active, %	-‡	43	38
Tertiary education, %	56	20	13
Currently employed, %	21	67	58
Family history of T2D, %	31	18	34
Disease history, %			
Hypertension	36	19	39
Hyperlipidaemia	38	17	27
Cardiovascular disease	5.0	1.9	4.0
Cancer	-	2.4	2.5
Mediterranean diet score, 0-18 points	9.7 (2.6)	8.8 (3.1)	8.5 (3.2)
Score components, g/1,000 kcal			
Vegetable	83 (67)	94 (64)	92 (66)
Legumes	8.2 (17.4)	9.2 (13.2)	9.4 (14.3)
Fruits and nuts	136 (75)	123 (99)	119 (102)
Grains and grain products	72 (39)	104 (41)	103 (43)
Fish and shellfish	24 (26)	18 (16)	20 (18)
Meat and meat products	35 (32)	52 (24)	58 (25)
Milk and milk products	130 (95)	163 (114)	160 (122)
Olive oil	2.0 (3.4)	4.3 (6.3)	4.0 (6.3)
Ethanol	4.6 (5.0)	6.0 (7.7)	6.5 (8.7)
Dietary supplement use, %	66	39	41
Body mass index, kg/m^2	26.5 (3.5)	26.1 (4.2)	29.9 (4.7)
Waist circumference, cm	90 (13)	87 (13)	98 (12)
Haemoglobin A1c, mmol/mol	-	35.9 (4.9)	43.1 (11.1)
≥ 6.5% (48 mmol/mol), %	-	1	17
Total cholesterol, mmol/L	5.2 (0.9)	5.9 (1.1)	6.1 (1.2)
HDL cholesterol, mmol/L	1.65 (0.44)	1.50 (0.42)	1.25 (0.37)
Triglycerides, mmol/L	1.13 (0.45)	1.33 (0.91)	1.98 (1.41)

Abbreviations: EPIC – European Prospective Investigation into Cancer and Nutrition; HDL – high density lipoprotein; T2D – type 2 diabetes

*Values are means (standard deviations) or percentages. The EPIC-InterAct subcohort participants and incident cases of T2D were independently sampled from the underlying EPIC cohort. As a feature of the case-cohort design, 564 incident cases included in the analysis were simultaneously subcohort participants.

†Calculated in women.

‡Data unavailable in the MedLey trial

Table 4.2 Medians (25th, 75th percentile) of circulating carotenoids and fatty acids: the MedLey trial post-intervention and the EPIC-InterAct subcohort at baseline

	MedLey trial		EPIC-InterAct							
	Hab-diet	Med-diet	France	Italy	Spain	UK	Netherlands	Germany	Sweden	Denmark
	n=65	n=68	n=529	n=1,910	n=3,423	n=1,230	n=1,426	n=1,890	n=924	n=1,981
Carotenoids, ng/mL*										
α-carotene	20	35	98	42	26	57	33	47	50	41
	(9, 82)	(7, 187)	(65, 157)	(26, 74)	(16, 44)	(34, 92)	(19, 53)	(27, 85)	(28, 88)	(22, 77)
β-carotene	726	998	419	251	144	246	197	248	227	163
	(440, 1029)	(637, 1588)	(278, 578)	(173, 365)	(94, 213)	(161, 350)	(126, 278)	(158, 395)	(150, 335)	(98, 266)
β-cryptoxanthin	31	45	140	163	182	84	110	94	72	44
	(14, 80)	(14, 80)	(92, 225)	(90, 289)	(110, 302)	(48, 133)	(69, 182)	(57, 154)	(40, 124)	(24, 85)
Lycopene	110	153	225	348	195	252	192	217	212	172
	(86, 150)	(102, 237)	(145, 314)	(261, 460)	(118, 285)	(157, 357)	(116, 303)	(143, 308)	(133, 301)	(104, 258)
Lutein & zeaxanthin	513	426	234	273	194	143	151	162	146	123
	(370, 656)	(301, 664)	(175, 314)	(213, 353)	(148, 250)	(111, 192)	(111, 202)	(120, 212)	(113, 191)	(89, 168)
Fatty acids, mol%†										
C14:0	0.46	0.39	0.41	0.36	0.28	0.38	0.40	0.39	0.40	0.39
	(0.39, 0.51)	(0.33, 0.45)	(0.34, 0.50)	(0.31, 0.42)	(0.23, 0.34)	(0.32, 0.45)	(0.33, 0.47)	(0.33, 0.47)	(0.34, 0.48)	(0.33, 0.46)
C16:0	28	27	30	30	29	30	30	30	31	31
	(27, 28)	(27, 28)	(29, 32)	(29, 31)	(28, 30)	(29, 31)	(29, 31)	(29, 31)	(30, 32)	(30, 32)
C18:0	12	12	14	14	15	14	14	14	14	14
	(11, 13)	(11, 12)	(14, 15)	(13, 15)	(14, 16)	(13, 15)	(13, 15)	(13, 15)	(13, 14)	(13, 14)
C15:0	0.29	0.28	0.27	0.21	0.17	0.23	0.24	0.22	0.22	0.20
	(0.25, 0.35)	(0.22, 0.32)	(0.23, 0.31)	(0.18, 0.25)	(0.14, 0.21)	(0.20, 0.27)	(0.21, 0.28)	(0.18, 0.27)	(0.18, 0.25)	(0.17, 0.24)
C17:0	0.41	0.41	0.49	0.42	0.44	0.42	0.40	0.39	0.39	0.39
	(0.38, 0.46)	(0.36, 0.44)	(0.44, 0.55)	(0.36, 0.46)	(0.38, 0.50)	(0.37, 0.47)	(0.36, 0.44)	(0.34, 0.43)	(0.36, 0.43)	(0.33, 0.43)
C20:0	0.15	0.15	0.13	0.13	0.12	0.15	0.14	0.12	0.14	0.12
	(0.13, 0.16)	(0.13, 0.16)	(0.12, 0.15)	(0.11, 0.14)	(0.11, 0.14)	(0.13, 0.18)	(0.12, 0.16)	(0.11, 0.14)	(0.12, 0.15)	(0.11, 0.14)
C22:0	0.39	0.36	0.24	0.21	0.22	0.26	0.28	0.20	0.23	0.22
	(0.33, 0.44)	(0.32, 0.41)	(0.21, 0.28)	(0.18, 0.24)	(0.19, 0.26)	(0.22, 0.30)	(0.24, 0.32)	(0.17, 0.23)	(0.21, 0.25)	(0.19, 0.25)
C24:0	1.07	1.02	0.25	0.22	0.22	0.23	0.26	0.20	0.21	0.21
	(0.94, 1.27)	(0.89, 1.15)	(0.22, 0.29)	(0.20, 0.25)	(0.19, 0.25)	(0.20, 0.26)	(0.23, 0.30)	(0.18, 0.23)	(0.19, 0.24)	(0.18, 0.23)
C18:3n-3	0.17	0.17	0.28	0.28	0.19	0.33	0.26	0.31	0.39	0.29
	(0.15, 0.19)	(0.15, 0.20)	(0.21, 0.39)	(0.20, 0.37)	(0.13, 0.28)	(0.23, 0.47)	(0.18, 0.38)	(0.24, 0.42)	(0.32, 0.51)	(0.23, 0.36)
C20:5n-3	1.46	1.79	1.07	0.71	0.86	1.10	0.85	0.98	1.30	1.59
	(0.96, 2.62)	(1.16, 2.29)	(0.79, 1.50)	(0.54, 0.90)	(0.60, 1.26)	(0.80, 1.56)	(0.62, 1.17)	(0.74, 1.32)	(0.99, 1.65)	(1.20, 2.26)
C22:5n-3	2.78	2.66	0.99	0.75	0.66	1.01	0.97	0.92	1.06	1.03
	(2.49, 3.10)	(2.37, 2.92)	(0.85, 1.14)	(0.66, 0.86)	(0.58, 0.75)	(0.87, 1.16)	(0.82, 1.09)	(0.80, 1.04)	(0.94, 1.20)	(0.91, 1.17)
C22:6n-3	5.4	5.8	4.8	3.5	4.6	4.1	3.4	3.6	4.1	4.7
	(4.4, 6.3)	(5.1, 6.3)	(4.0, 5.5)	(2.9, 4.1)	(3.9, 5.3)	(3.3, 5.0)	(2.8, 4.1)	(3.0, 4.3)	(3.4, 4.8)	(3.9, 5.5)
C18:2n-6c	12	12	21	22	23	23	24	23	22	22
	(10, 12)	(11, 13)	(19, 23)	(20, 23)	(21, 25)	(21, 25)	(22, 26)	(21, 25)	(21, 24)	(20, 24)

	MedLey trial		EPIC-InterAct							
	Hab-diet	Med-diet	France	Italy	Spain	UK	Netherlands	Germany	Sweden	Denmark
C20:2	0.21	0.22	0.38	0.36	0.37	0.39	0.40	0.38	0.37	0.35
	(0.19, 0.24)	(0.21, 0.26)	(0.34, 0.42)	(0.33, 0.40)	(0.33, 0.41)	(0.35, 0.44)	(0.36, 0.45)	(0.34, 0.43)	(0.34, 0.42)	(0.32, 0.39)
C20:3n-6	1.4	1.5	3.1	3.7	3.0	3.2	3.3	3.1	3.1	2.8
	(1.3, 1.7)	(1.3, 1.7)	(2.5, 3.6)	(3.2, 4.2)	(2.5, 3.5)	(2.7, 3.7)	(2.8, 3.7)	(2.7, 3.6)	(2.7, 3.6)	(2.4, 3.2)
C20:4n-6	12.5	11.8	9.6	10.3	9.7	8.3	9.3	9.6	8.4	8.5
	(11.2, 13.4)	(11.3, 12.7)	(8.6, 10.9)	(9.1, 11.5)	(8.6, 11.0)	(7.3, 9.5)	(8.2, 10.4)	(8.5, 10.7)	(7.5, 9.4)	(7.5, 9.4)
C22:4	2.27	2.06	0.30	0.33	0.26	0.27	0.31	0.30	0.28	0.24
	(1.63, 2.88)	(1.69, 2.44)	(0.26, 0.35)	(0.29, 0.39)	(0.22, 0.31)	(0.23, 0.32)	(0.27, 0.36)	(0.26, 0.35)	(0.25, 0.31)	(0.21, 0.29)
C22:5n-6	0.28	0.25	0.22	0.28	0.18	0.16	0.21	0.21	0.16	0.13
	(0.21, 0.36)	(0.21, 0.28)	(0.18, 0.27)	(0.23, 0.33)	(0.15, 0.22)	(0.13, 0.20)	(0.16, 0.26)	(0.17, 0.25)	(0.13, 0.19)	(0.10, 0.16)
C16:1	0.41	0.39	0.44	0.46	0.32	0.49	0.48	0.53	0.51	0.58
	(0.34, 0.53)	(0.30, 0.50)	(0.35, 0.55)	(0.38, 0.60)	(0.26, 0.42)	(0.38, 0.63)	(0.38, 0.64)	(0.42, 0.70)	(0.42, 0.64)	(0.45, 0.76)
C17:1	0.11	0.09	0.03	0.07	0.05	0.04	0.04	0.07	0.06	0.07
	(0.08, 0.13)	(0.07, 0.11)	(0.02, 0.09)	(0.05, 0.10)	(0.04, 0.07)	(0.02, 0.08)	(0.02, 0.08)	(0.03, 0.11)	(0.04, 0.08)	(0.05, 0.10)
C18:1n-9c	16.8	17.8	9.0	10.8	9.8	9.4	8.7	9.3	10.5	9.8
	(16.4, 17.4)	(16.8, 18.1)	(8.2, 10.0)	(9.7, 12.0)	(8.3, 11.2)	(8.4, 10.4)	(7.9, 9.7)	(8.6, 10.2)	(9.8, 11.3)	(9.0, 10.8)
C20:1	0.23	0.25	0.26	0.23	0.20	0.31	0.25	0.25	0.28	0.25
	(0.21, 0.25)	(0.23, 0.29)	(0.23, 0.29)	(0.20, 0.26)	(0.18, 0.23)	(0.27, 0.37)	(0.22, 0.29)	(0.21, 0.29)	(0.26, 0.32)	(0.20, 0.29)
C24:1	1.22	1.31	0.37	0.35	0.33	0.33	0.32	0.32	0.36	0.36
	(1.02, 1.35)	(1.18, 1.43)	(0.32, 0.42)	(0.30, 0.40)	(0.28, 0.38)	(0.28, 0.40)	(0.27, 0.38)	(0.28, 0.37)	(0.31, 0.40)	(0.31, 0.41)
C18:1n-9t	0.10	0.10	0.17	0.13	0.13	0.40	0.37	0.18	0.35	0.20
	(0.08, 0.11)	(0.07, 0.11)	(0.12, 0.29)	(0.10, 0.18)	(0.10, 0.18)	(0.29, 0.56)	(0.26, 0.54)	(0.13, 0.24)	(0.25, 0.49)	(0.15, 0.27)

Abbreviations: Hab – habitual; Med – Mediterranean

*Carotenoids were measured in serum in the MedLey trial and in plasma in EPIC-InterAct.

†Fatty acids were measured in erythrocytes in the MedLey trial and in plasma phospholipids in EPIC-InterAct. The denominator for the mol% unit was the sum of all fatty acids presented in this table.

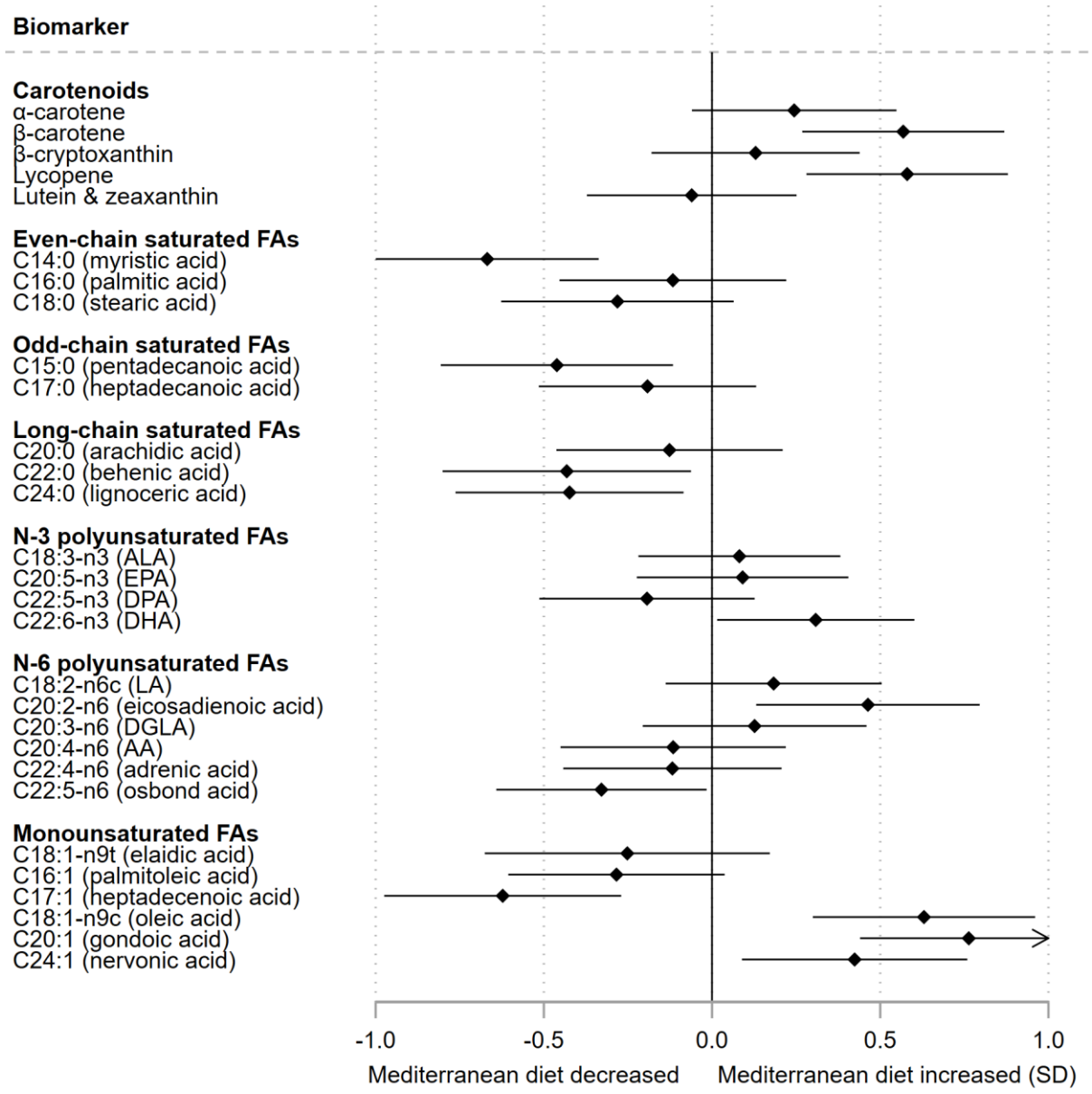


Figure 4.2 Differences in standardised means of nutritional biomarkers between the Mediterranean and habitual diet groups in the MedLey trial at 6 months

Abbreviations: AA – arachidonic acid; ALA – α-linolenic acid; DGLA – dihomo-γ-linolenic acid; DHA – docosahexaenoic acid; DPA – docosapentaenoic acid; EPA – eicosapentaenoic acid; FAs – fatty acids; LA – linoleic acid

Mixed linear models were used to estimate standardised differences after 6 months of partial-feeding intervention. Standardised values were calculated using baseline means and standard deviations of natural logarithm-transformed values of biomarkers. Horizontal error bars indicate 95% confidence intervals. Between 131-136 participants had non-missing biomarker data at 6 months.

4.5.2 Derivation of the biomarker score

The biomarker score consisted of a linear combination of 23 biomarkers in total. The score distinguished two arms of the Mediterranean and habitual diet groups in the MedLey trial with a cross-validated C-statistic = 0.88 (95% CI: 0.82-0.94). The discriminatory performance was not influenced by modelling assumptions and overall, it was not strongly driven by individual biomarkers. However, exclusion of C20:1 (gondoic acid) from calculation of the score resulted in a sizeable decrease of the C-statistic to 0.78 (**Table 4.3**). Reproducibility of the biomarker score in the control group between baseline and month 3 was high (r = 0.75).

After standardisation of the score in the MedLey trial, mean score values were higher by 1.81 (95% CI: 1.45-2.18) points in the Mediterranean than the habitual diet group. The score remained stable in the habitual diet group throughout the duration of the study, while progressively increasing in the Mediterranean diet intervention arm (**Figure 4.3**). In EPIC-InterAct, baseline medians of the biomarker score standardised using the overall subcohort distribution ranged between countries from -0.35 in Spain to 0.71 SD in Sweden (**Figure 4.4**). Non-standardised biomarker score values were substantially lower in EPIC-InterAct than in the MedLey trial habitual diet group, potentially indicating lower adherence to the Mediterranean diet in the EPIC population or miscalibration of the biomarker score model when applied to the InterAct data for estimation of absolute predicted values (**Table 4.4**).

Table 4.3 Biomarker scores of discrimination between the Mediterranean and habitual diet in the MedLey trial*

Biomarker†	Natural log mean (SD)	C-statistic after exclusion from the score with selection rate ≥ 90% and mol% fatty acids‡	Interacting biomarker	Scoring coefficients by selection rate cut-off and the unit of fatty acids§			
				≥ 90%, mol%	≥ 95%, mol%	≥ 99%, mol%	≥ 90%, wt%
ß-carotene	6.78 (0.81)	0.86	-				
			C18:1n-9c	0.054	0.066	0.100	0.117
			C22:0	-0.054			
β-crypt.	3.55 (1.06)	0.84	C24:1	0.244	0.245		
Lycopene	4.93 (0.69)	0.86	-				
			C22:6n-3	0.047	0.066	0.071	
			C18:1n-9c	0.045			
Lutein and zeaxanthin	6.17 (0.50)	0.82	C20:1	0.166	0.174	0.276	0.250
C15:0	-1.28 (0.25)	0.84	ß-carotene	-0.043	-0.048	-0.112	
C17:0	-0.91 (0.15)	0.84	ß-carotene	-0.089	-0.104		
C18:0	2.47 (0.06)	0.84	C20:1	0.309	0.318	0.522	0.550
C22:0	-1.00 (0.19)	0.84	C24:1	-1.423	-1.478	-2.855	-3.762
C24:0	0.04 (0.18)	0.84	-				
			C24:1	-3.915	-4.246	-5.655	1.193
C18:3n-3	-1.76 (0.24)	0.84	C20:1	-0.317	-0.333		
C20:5n-3	0.49 (0.52)	0.84	C24:0	-1.467	-1.609	-2.018	
C22:5n-3	1.00 (0.17)	0.84	C22:4n-6	-0.495	-0.497		
C22:6n-3	1.71 (0.19)	0.85	C17:1	-0.128	-0.122		
C18:2n-6	2.46 (0.14)	0.84	C20:1	0.292	0.297		
C20:4n-6	2.49 (0.13)	0.84	C20:1	0.228	0.219	0.445	0.395
C22:5n-6	-1.36 (0.28)	0.85	-				
			ß-carotene	-0.051	0.054		-0.072
			Lycopene	-0.065	-0.077	-0.119	-0.101
			C15:0	0.209	0.202		0.372
			C17:0	0.431	0.424	1.046	0.742
			C22:0	0.333	0.364	0.717	0.515
			C17:1	0.159	0.155		
C16:1	-0.90 (0.36)	0.84	C24:0	0.813	0.918		
C17:1	-2.33 (0.32)	0.85	-				
			ß-carotene	-0.045	-0.050	-0.082	-0.085
C18:1n-9c	2.84 (0.06)	0.86	-	3.247	3.589	4.218	4.052
C20:1	-1.41 (0.19)	0.78	-				
			C24:0	0.429	0.491		
C24:1	0.21 (0.19)	0.84	-				
			C22:5n-6	-0.765	-0.781		
C18:1n-9t	-2.40 (0.29)	0.84	C20:1	-0.280	-0.296		
Base odds				-10.463	-10.933	-14.276	-13.893

Abbreviations: mol% - molar percent; RCT – randomised controlled trial; SD – standard deviation; wt% - weight percent; β-crypt. – β-cryptoxanthin

*n = 67 in the Mediterranean diet group and n = 61 in the continuation of habitual diet group

†Serum carotenoids were adjusted for total cholesterol using the residual method, re-scaled to the unadjusted mean, and expressed in ng/mL. Erythrocyte fatty acids were proportions of total fatty acids used in the analysis.

‡Values were estimated by calculating the score with omission of a given biomarker and its interaction terms and testing the discriminatory performance. The full score without omissions had a C-statistic of 0.91.

§Values are unstandardised coefficients of natural logarithm-transformed biomarkers from elastic net logistic regression models (on the log scale). Cross-validated logistic elastic net regression was repeated 1,000 times and predictors were included at pre-specified selection rate cut-offs. The primary pre-specified cut-off was ≥ 90%. The wt% fatty acids model was the same for the ≥ 90% and ≥ 95% cut-offs. C-statistic values ranged from 0.89-0.91, and 0.85-0.87 with 5-fold cross-validation.

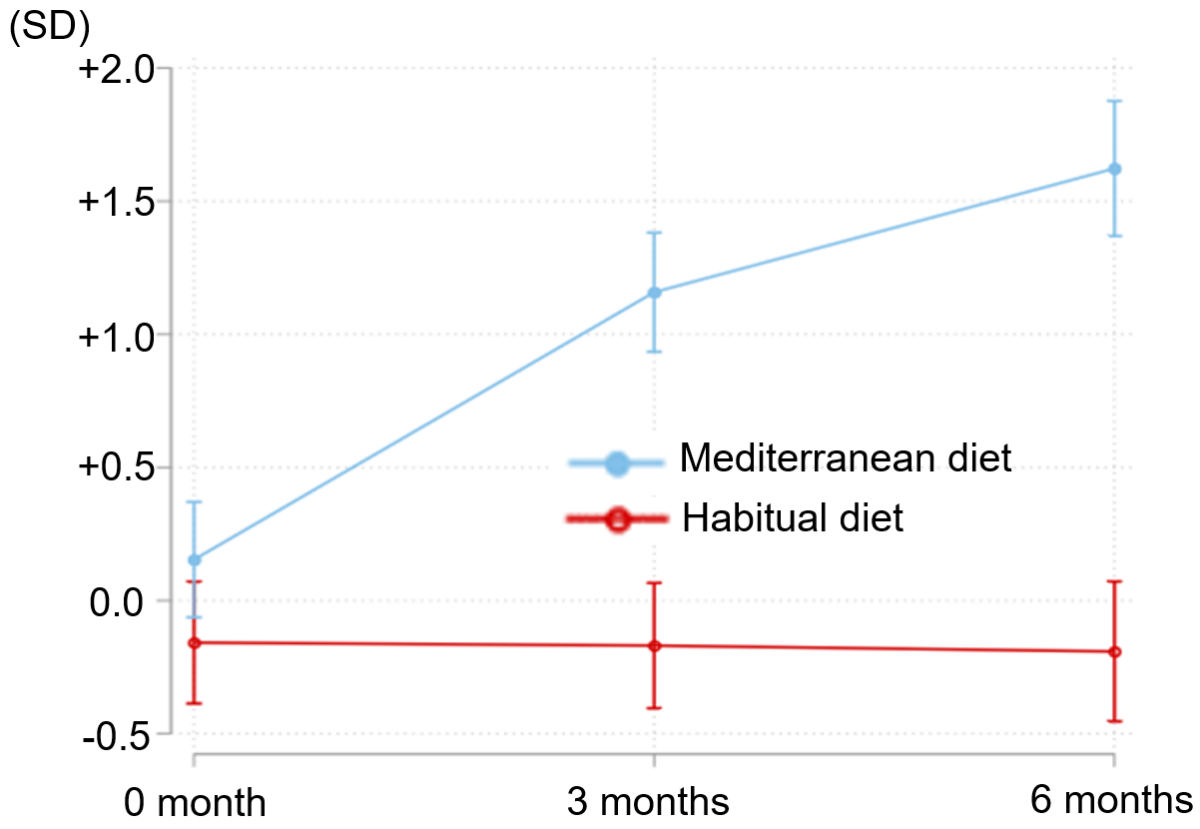


Figure 4.3 Standardised differences in the biomarker score of the Mediterranean diet between the arms of the MedLey trial.

The biomarker score was derived as a discriminatory model between the Mediterranean and habitual diet in the MedLey randomised partial-feeding controlled trial. Circulating carotenoids and fatty acids were used to calculate the score as linear predictions from the discriminatory model. Vertical bars represent 95% confidence intervals.

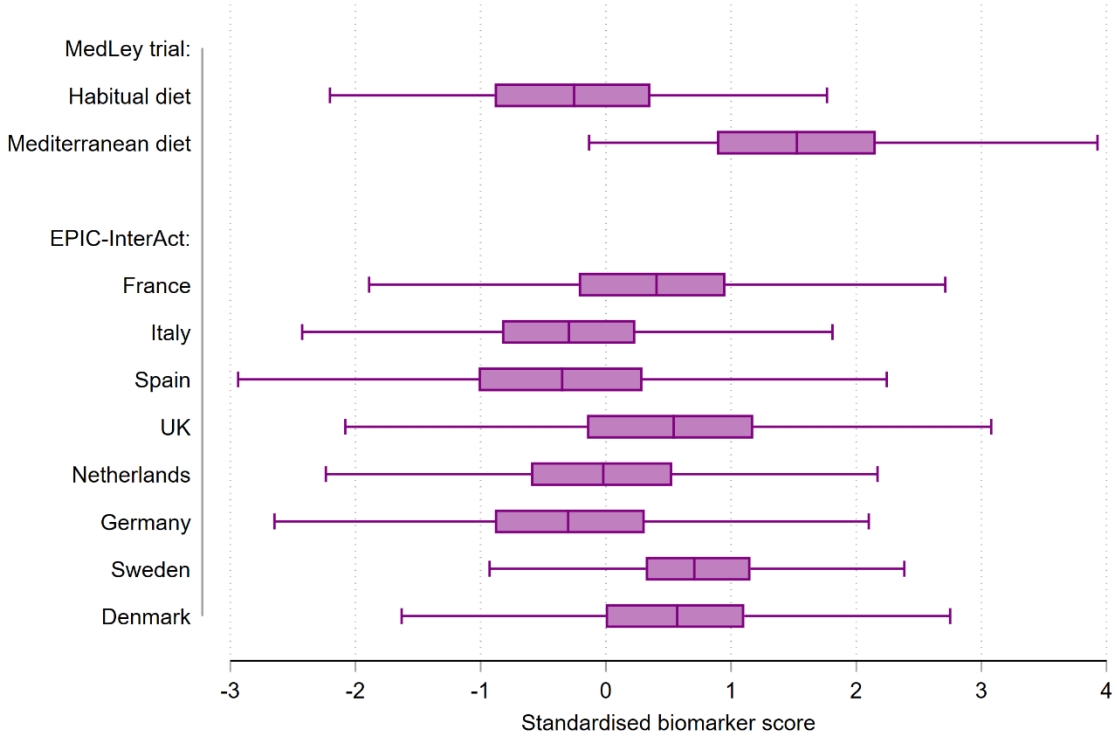


Figure 4.4 Nutritional biomarker score of the Mediterranean diet: study-specific distribution by the MedLey trial arms post-intervention and the EPIC-InterAct subcohort by country at baseline.

Abbreviations: EPIC – European Prospective Investigation into Cancer and Nutrition; UK – United Kingdom

The biomarker score was derived as a discriminatory model between the Mediterranean and habitual diet in the MedLey randomised partial-feeding controlled trial. Circulating carotenoids and fatty acids were used to calculate the score as linear predictions from the discriminatory model. The score was standardised separately within the MedLey trial and the EPIC-InterAct subcohort. Boxes denote the interquartile ranges (IQR) and medians inside; and whiskers, values up to 1.5 IQR outside of these percentiles.

Table 4.4 Medians (25th, 75th percentile) of the nutritional biomarker score of the Mediterranean diet in the MedLey trial post-intervention and in the EPIC-InterAct subcohort

Biomarker score*	MedLey trial		EPIC-InterAct							
	Hab-diet	Med-diet	France	Italy	Spain	UK	Netherlands	Germany	Sweden	Denmark
	n=65	n=68	n=529	n=1,910	n=3,423	n=1,230	n=1,426	n=1,890	n=924	n=1,981
Standardised values	-0.25 (-0.89, 0.35)	1.53 (0.89, 2.16)	0.40 (-0.22, 0.96)	-0.30 (-0.83, 0.24)	-0.35 (-1.02, 0.29)	0.54 (-0.15, 1.18)	-0.02 (-0.60, 0.53)	-0.30 (-0.89, 0.31)	0.71 (0.32, 1.15)	0.57 (-0.00, 1.11)
Linear predictions	-1.1 (-2.0, -0.4)	1.2 (0.3, 2.0)	-9.0 (-11.5, -6.7)	-11.9 (-14.0, -9.7)	-12.1 (-14.8, -9.4)	-8.4 (-11.3, -5.8)	-10.7 (-13.1, -8.5)	-11.9 (-14.3, -9.4)	-7.7 (-9.3, -5.9)	-8.3 (-10.6, -6.1)
Med-/hab-diet probability	0.24 (0.12, 0.41)	0.76 (0.58, 0.88)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)

Abbreviations: Hab – habitual; Med – Mediterranean; n – number of participants

*The biomarker score was derived as a discriminatory model between the Mediterranean and habitual diet in the MedLey randomised partial-feeding controlled trial using circulating carotenoids and fatty acids. The standardised values are study-specific. The linear predictions are the log odds and the Med-/Hab-diet probabilities are the predicted probabilities of the assignment to the Mediterranean or habitual diet arms of the MedLey trial.

4.5.3 Association of the biomarker score with incident T2D

The biomarker score of the Mediterranean diet was inversely associated with incident T2D (**Table 4.5**). In the adiposity-adjusted multivariable model, the HR (95% CI) for the top fifth of the biomarker score compared to the bottom fifth was 0.38 (0.30-0.50) (p trend = 0.012). The HR (95% CI) per 1 SD was 0.71 (0.65-0.77) with inverse associations in all countries, moderate heterogeneity between country-specific estimates (I^2 = 67%), and a 95% prediction interval of 0.55-0.91 (**Figure 4.5**). I found weak evidence of departure from linearity (p non-linearity = 0.044, **Figure 4.5**) where the inverse association levelled off in approximately the upper fifth of the distributions of the score. The association remained statistically significant after additional adjustments for individual and all biomarkers or interaction terms included in the score (**Table 4.6**).

The estimated PAF (95% CI) was 11% (7-14%), indicating that the incidence of T2D could be reduced by 11% if the biomarker score were increased by 10 percentiles, assuming a causal relationship. For comparison, the estimated PAF (95% CI) for a 10 percentile lower BMI was 28% (20-35%). The corresponding estimates for the delay in incidence were 117 (87-148) days of disease-free time and an acceleration by 357 (303-410) days, respectively.

The main result was robust to multiple sensitivity analyses which explored the effects of time of follow-up, reverse causation, and alternative analytical decisions at the stage of derivation of the biomarker score (**Table 4.6**). Among the potential effect modifiers, I found evidence of interaction of the biomarker score with age and the use of dietary supplements (**Table 4.7**). The HR (95% CI) per 1 SD of the biomarker score was 0.76 (0.69-0.84) in supplement users and 0.69 (0.62-0.75) in non-users. The stratum-specific estimates by age at baseline <45, 45-60 and >60 years were 0.54 (0.42-0.69), 0.74 (0.69-0.80) and 0.74 (0.67-0.82), respectively. The results from complete-case analysis were similar to the multiple imputation estimates (**Table 4.7**).

Table 4.5 Associations between the nutritional biomarker score of the Mediterranean diet* with incidence of type 2 diabetes in the European Prospective Investigation into Cancer and Nutrition-InterAct (n = 22,202)

Model	Quintiles					p_{trend}‡	Per 1 SD	I^2, % (95% CI)
	Q1	Q2	Q3	Q4	Q5			
Number of cases	2,779	1,954	1,698	1,508	1,514			
IR per 100,000 person-years	572	407	286	252	268			
Pooled HRs (95% CIs)†								
Age, sex, and centre adjusted	1.0 (Ref.)	0.64 (0.58-0.69)	0.49 (0.44-0.54)	0.40 (0.34-0.47)	0.34 (0.28-0.41)	<0.001	0.65 (0.61-0.69)	69 (39-85)
Multivariable adjusted†	1.0 (Ref.)	0.66 (0.60-0.73)	0.53 (0.46-0.60)	0.43 (0.37-0.49)	0.36 (0.29-0.45)	<0.001	0.67 (0.62-0.72)	67 (33-84)
+adiposity	1.0 (Ref.)	0.69 (0.62-0.77)	0.59 (0.52-0.67)	0.48 (0.40-0.58)	0.38 (0.30-0.50)	0.012	0.71 (0.65-0.77)	67 (30-84)

Abbreviations: CI – confidence interval; HR – hazard ratio; IR – incidence rate; SD – standard deviation

*The biomarker score was derived as a discriminatory model between the Mediterranean and habitual diet in the MedLey randomised partial-feeding controlled trial. Circulating carotenoids and fatty acids were used to calculate the score as linear predictions from the discriminatory model.

†Hazard ratios were pooled from country-specific estimates. Multivariable adjusted model included the following covariates: age (as timescale), sex, recruitment centre, prevalent cancer, cardiovascular disease, hypertension and hyperlipidaemia; familial history of T2D, smoking status (never, former, current smoker), physical activity index (inactive, moderately inactive, moderately active, active), seasonality (sine and cosine function of the day of the year), fasting status (<3, 3-6, >6 hours), current use of vitamin or mineral supplements, marital status (single, married or cohabiting, divorced or separated, widowed), educational attainment (none, primary school, technical or professional school, secondary school, post-secondary school education), current employment, and in women, menopausal status (pre-, peri-, postmenopausal, bilateral oophorectomy), current hormone replacement therapy use. Adjustment for adiposity included body mass index and waist circumference.

‡ Generalised least-squares trend estimation method was used to calculate p-values for a linear trend over an ordinal variable of median biomarker scores of the five quintile groups.

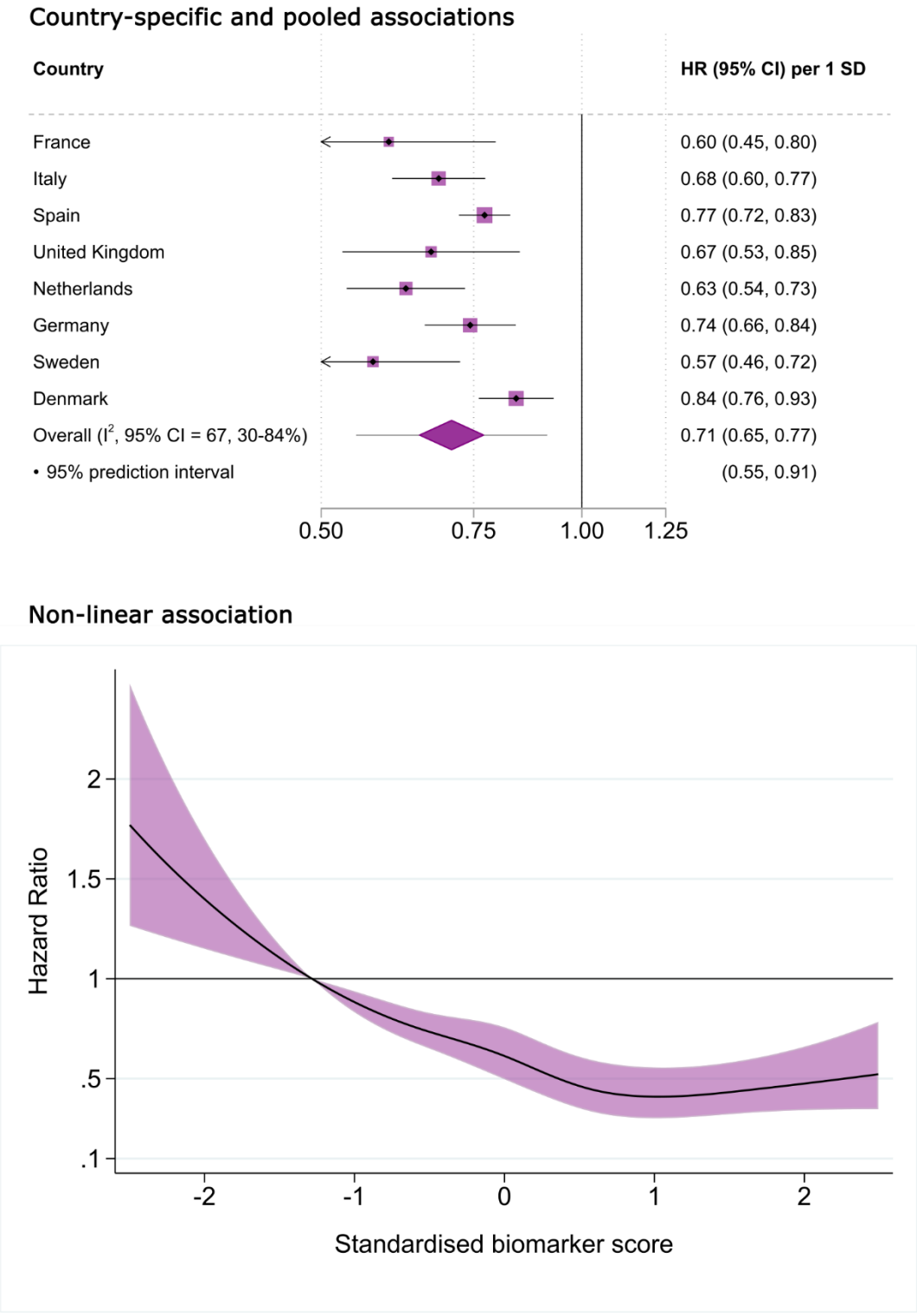


Figure 4.5 Association between the nutritional biomarker score of the Mediterranean diet and incidence of type 2 diabetes in the EPIC-InterAct case-cohort study (n = 22,202).

Abbreviations: CI – confidence interval; HR – hazard ratio; SD – standard deviation

The biomarker score was derived as a discriminatory model between the Mediterranean and habitual diet in the MedLey randomised partial-feeding controlled trial. Circulating carotenoids and fatty acids were used to calculate the score as linear predictions from the discriminatory model. Associations were assessed with the Prentice-weighted Cox regression and pooled by random-effects meta-analysis.

Top: The diamond and error bars of the pooled estimate represent the 95% confidence and prediction intervals.

Bottom: Restricted cubic splines with five knots were used to model the non-linear association. The p value for the test of non-linearity was 0.044. Black solid line represents point estimates of hazard ratios and purple area denotes the 95% confidence interval. The 10^{th} percentile of the subcohort distribution was used as reference.

Associations were adjusted for: age (as timescale), sex, recruitment centre, prevalent cancer, cardiovascular disease, hypertension and hyperlipidaemia; familial history of type 2 diabetes, smoking status (never, former, current smoker), physical activity index (inactive, moderately inactive, moderately active, active), seasonality (sine and cosine function of the day of the year), fasting status (<3, 3-6, >6 hours), current use of vitamin or mineral supplements, marital status (single, married or cohabiting, divorced or separated, widowed), educational attainment (none, primary school, technical or professional school, secondary school, post-secondary school education), current employment, body mass index and waist circumference, and in women, menopausal status (pre-, peri-, postmenopausal, bilateral oophorectomy), current hormone replacement therapy use.

Table 4.6 Nutritional biomarker score of the Mediterranean diet derived in the MedLey trial and incidence of type 2 diabetes in EPIC-InterAct*: sensitivity analyses

Model	HR (95% CI) per 1 SD
Main result	0.71 (0.65-0.77)
First 7 years of follow-up	0.68 (0.60-0.77)
> 7 years of follow-up	0.73 (0.65-0.81)
Excluding the first 2 years of follow-up	0.71 (0.66-0.78)
Excluding participants with HbA1c > 48mmol/mol	0.73 (0.68-0.80)
Excluding participants with prevalent cancer, MI or stroke	0.71 (0.65-0.77)
Excluding outliers in nutritional biomarkers†	0.69 (0.64-0.75)
Biomarker score from a single elastic net regression	0.71 (0.66-0.77)
Biomarker score calculated using unpenalised coefficients	0.77 (0.73-0.82)
Biomarker score derived with predictors' selection rate cut-off 95%	0.71 (0.65-0.77)
Biomarker score derived with predictors' selection rate cut-off 99%	0.74 (0.69-0.79)
Biomarker score derived with fatty acids as %weight instead of %mol	0.73 (0.65-0.82)
Additional adjustment for components of the biomarker score:‡	
C18:1n-9c	0.70 (0.65-0.76)
ß-carotene # C15:0	0.76 (0.71-0.81)
ß-carotene # C17:0	0.77 (0.72-0.83)
ß-carotene # C22:0	0.77 (0.72-0.83)
ß-carotene # C22:5n-6	0.72 (0.66-0.77)
ß-carotene # C17:1	0.71 (0.65-0.77)
ß-carotene # C18:1n-9c	0.76 (0.70-0.81)
ß-cryptoxanthin # C24:1	0.75 (0.70-0.81)
Lycopene # C22:6n-3	0.72 (0.67-0.78)
Lycopene # C22:5n-6	0.70 (0.65-0.76)
Lycopene # C18:1n-9c	0.71 (0.66-0.77)
Lutein & zeaxanthin # C20:1	0.73 (0.67-0.78)
C18:0 # C20:1	0.68 (0.62-0.74)
C15:0 # C22:5n-6	0.70 (0.64-0.75)
C17:0 # C22:5n-6	0.70 (0.65-0.75)
C22:0 # C22:5n-6	0.70 (0.65-0.75)
C22:0 # C24:1	0.78 (0.65-0.92)
C24:0 # C20:5n-3	0.59 (0.51-0.68)
C24:0 # C16:1	0.69 (0.63-0.75)
C24:0 # C20:1	0.72 (0.66-0.79)
C24:0 # C24:1	0.80 (0.67-0.97)
C18:3n-3 # C20:1	0.70 (0.65-0.76)
C22:5n-3 # C22:4	0.70 (0.65-0.76)
C22:6n-3 # C17:1	0.70 (0.64-0.76)
C18:2n-6c # C20:1	0.71 (0.65-0.77)
C20:4n-6 # C20:1	0.70 (0.64-0.76)
C22:5n-6 # C17:1	0.69 (0.64-0.75)
C22:5n-6 # C24:1	0.71 (0.65-0.77)
C18:1n-9t # C20:1	0.71 (0.66-0.77)
Simultaneous adjustment for all components of the biomarker score	0.46 (0.30-0.72)

Abbreviations: CI – confidence interval; EPIC – European Prospective Investigation into Cancer and Nutrition; HR – hazard ratio; MI – myocardial infarction; RCT – randomised controlled trial; SD – standard deviation

*The biomarker score was derived as a discriminatory model between the Mediterranean and habitual diet in the MedLey randomised partial-feeding controlled trial. Repeated elastic net regression models were used for variable selection, and predictors with selection rate ≥ 90% were included in the biomarker score. Circulating carotenoids and fatty acids were used to calculate the score as linear predictions from the discriminatory model. The multivariable adjusted model included the following covariates: age (as timescale), sex, recruitment centre, prevalent cancer, cardiovascular disease, hypertension and hyperlipidaemia; familial history of type 2 diabetes, smoking status (never, former, current smoker), physical activity index (inactive, moderately inactive, moderately active, active), seasonality (sine and cosine function of the day of the year), fasting status (<3, 3-6, >6 hours), current use of vitamin or mineral supplements, marital status (single, married or cohabiting, divorced or separated, widowed), educational attainment (none, primary school, technical or professional school, secondary school, post-secondary school education), current employment, body mass index and waist circumference, and in women, menopausal status (pre-, peri-, postmenopausal, bilateral oophorectomy), current hormone replacement therapy use. Hazard ratios were pooled from country-specific estimates. 22,202 participants with non-missing biomarker score data were included in the analysis.

†Participants with nutritional biomarker values 4SDs above or below subcohort means in any of the component biomarkers of the biomarker score.

‡A hash mark denotes interaction. Biomarkers or biomarker-biomarker interactions were entered into the models as linear and squared terms. Only squared terms which were statistically significant when adjusting for individual components of the biomarker score were used in the analysis simultaneously adjusting for all components.

Table 4.7 Nutritional biomarker score of the Mediterranean diet derived in the MedLey trial and incidence of type 2 diabetes in EPIC-InterAct: associations per 1 standard deviation by categories of covariates*

Covariate category	Multiply imputed analysis			Complete-case analysis		
	n†	HR (95% CI)	$p_{interaction}$‡	n	HR (95% CI)	$p_{interaction}$‡
Main result	22,202	0.71 (0.65-0.77)		15,430	0.70 (0.64-0.76)	-
Dietary supplements						
Non-users	13,819	0.69 (0.62-0.75)		9,816	0.67 (0.61-0.74)	
Users	8,383	0.76 (0.69-0.84)	0.21	5,614	0.75 (0.66-0.85)	0.03
Age at baseline, years						
<45	4,234	0.54 (0.42-0.69)		3,174	0.52 (0.38-0.70)	
45-60	12,892	0.74 (0.69-0.80)		8,884	0.72 (0.65-0.80)	
>60	5,076	0.74 (0.67-0.82)	0.02	3,372	0.74 (0.65-0.84)	0.01

Abbreviations: CI – confidence interval; EPIC – European Prospective Investigation into Cancer and Nutrition; HR – hazard ratio; RCT – randomised controlled trial

*The biomarker score was derived as a discriminatory model between the Mediterranean and habitual diet in the MedLey randomised partial-feeding controlled trial. Circulating carotenoids and fatty acids were used to calculate the score as linear predictions from the discriminatory model. Hazard ratios were pooled from country-specific estimates. Models were adjusted for: age (as timescale for effect modification by supplement use), sex, recruitment centre, prevalent cancer, cardiovascular disease, hypertension and hyperlipidaemia; familial history of type 2 diabetes, smoking status (never, former, current smoker), physical activity index (inactive, moderately inactive, moderately active, active), seasonality (sine and cosine function of the day of the year), fasting status (<3, 3-6, >6 hours), current use of vitamin or mineral supplements, marital status (single, married or cohabiting, divorced or separated, widowed), educational attainment (none, primary school, technical or professional school, secondary school, post-secondary school education), current employment, body mass index and waist circumference, and in women, menopausal status (pre-, peri-, postmenopausal, bilateral oophorectomy), current hormone replacement therapy use. Presence of interaction was also evaluated for sex, BMI, seasonality, fasting status, physical activity and smoking status ($p_{interaction}$ values > 0.05).

†Numbers of participants by use of dietary supplements in multiply imputed analysis are mid-point values between the smallest and the largest values in the imputation datasets.

‡Interaction p values for age are based on continuous-by-continuous interaction terms between age and biomarker score.

4.5.4 Impact of biomarker assays on reliability of the biomarker score and stability of its associations with incident T2D in EPIC-Norfolk

Reliability of the biomarker score was poor in two subsamples of the EPIC-Norfolk study in which I investigated the impact of applying the score to measurements from different biomarker assays. The correlation was -0.06 when comparing the IARC assays (non-InterAct) for plasma phospholipid fatty acids and carotenoids with erythrocyte fatty acids and IARC carotenoids (n = 382), and 0.05 for the comparison of the IARC assays and the assays used in the EPIC-InterAct study (n = 140). The reliability of measurements of individual biomarkers was high for carotenoids (intraclass correlation coefficients ≥ 0.87) and variable for fatty acids with poor reliability of several low concentration fatty acids (**Table 4.8**). The overall reliability of quantification of fatty acids was similar between the measurements of erythrocyte and plasma phospholipid fatty acids and two assays of plasma phospholipid fatty acids.

Associations of the biomarker score with incident T2D in EPIC-Norfolk were inverse when using different assays of plasma carotenoids and phospholipid fatty acids (**Table 4.9**). However, the association was null when the biomarker score was calculated using erythrocyte fatty acids, though this was based on a 2-3-fold lower number of cases than the other comparisons.

Table 4.8 Medians (25^{th}, 75^{th} percentile) and intraclass correlation coefficients of biomarkers by laboratory assays in subsamples of the EPIC-Norfolk study

	Sample 1 (n=384)*			Sample 2 (n=140)†		
	Set 1: IARC carotenoids & RIVM EFAs	Set 2: IARC carotenoids & IARC PPFAs	ICC	Set 2: IARC carotenoids & IARC PPFAs	Set 3:Vitas carotenoids & HNR PPFAs	ICC
Carotenoids (ng/mL)						
α-carotene	62 (40, 103)		-	90 (57, 131)	73 (46, 114)	0.95
β-carotene	184 (128, 278)		-	231 (151, 314)	298 (188, 449)	0.87
β-crypt.	70 (43, 106)		-	83 (50, 131)	102 (62, 159)	0.92
Lycopene	265 (180, 350)		-	293 (194, 438)	265 (173, 387)	0.94
Lutein and zeaxanthin	181 (139, 240)		-	200 (155, 267)	177 (131, 236)	0.94
Fatty acids (mol%)‡						
C14:0	0.43 (0.37, 0.50)	0.35 (0.28, 0.44)	0.48	0.36 (0.26, 0.50)	0.42 (0.36, 0.48)	0.39
C16:0	23 (23, 24)	27 (25, 29)	0.19	27 (24, 29)	31 (30, 32)	0.20
C18:0	15 (14, 16)	13 (12, 14)	0.33	13 (12, 14)	15 (14, 16)	0.50
C15:0	0.28 (0.25, 0.31)	0.18 (0.14, 0.21)	0.54	0.18 (0.15, 0.21)	0.23 (0.20, 0.26)	0.60
C17:0	0.36 (0.33, 0.40)	0.38 (0.33, 0.44)	0.62	0.41 (0.35, 0.47)	0.42 (0.38, 0.48)	0.79
C20:0	0.37 (0.32, 0.42)	0.03 (0.02, 0.05)	0.05	0.03 (0.02, 0.04)	0.13 (0.12, 0.15)	0.20
C24:0	3.82 (3.25, 4.25)	0.05 (0.03, 0.08)	0.00	0.03 (0.02, 0.05)	0.22 (0.19, 0.24)	0.00
C18:3n-3	0.15 (0.12, 0.17)	0.22 (0.17, 0.27)	0.60	0.23 (0.18, 0.29)	0.29 (0.22, 0.38)	0.56
C20:5n-3	0.93 (0.75, 1.25)	1.04 (0.73, 1.38)	0.79	1.19 (0.89, 1.55)	1.26 (0.92, 1.64)	0.98
C22:5n-3	2.65 (2.43, 2.90)	1.37 (1.16, 1.56)	0.47	1.37 (1.18, 1.63)	0.98 (0.85, 1.10)	0.50
C22:6n-3	5.78 (5.03, 6.34)	4.69 (3.97, 5.88)	0.70	5.20 (4.27, 6.16)	4.08 (3.55, 4.93)	0.80
C18:2n-6c	12 (11, 13)	24 (22, 27)	0.59	25 (23, 27)	23 (21, 25)	0.92
C20:2	0.26 (0.23, 0.29)	0.46 (0.41, 0.52)	0.42	0.46 (0.41, 0.52)	0.38 (0.35, 0.43)	0.61
C20:3n-6	1.73 (1.54, 1.93)	4.31 (3.75, 4.91)	0.41	3.92 (3.40, 4.95)	2.89 (2.43, 3.49)	0.80
C20:4n-6	13.8 (12.8, 14.5)	9.6 (8.6, 10.8)	0.59	9.4 (8.5, 10.5)	7.9 (7.1, 9.1)	0.77
C22:4	2.50 (2.24, 2.84)	0.32 (0.26, 0.38)	0.18	0.30 (0.25, 0.37)	0.24 (0.21, 0.29)	0.55
C16:1	0.53 (0.45, 0.61)	0.76 (0.63, 0.95)	0.67	0.74 (0.62, 0.93)	0.54 (0.39, 0.66)	0.71
C18:1n-9c	13.0 (12.3, 13.7)	10.3 (9.1, 11.3)	0.51	10.2 (8.9, 11.2)	9.2 (8.4, 10.3)	0.74
C20:1	0.26 (0.23, 0.30)	0.13 (0.10, 0.17)	0.44	0.14 (0.11, 0.17)	0.30 (0.25, 0.34)	0.16
C24:1	3.46 (3.03, 3.95)	0.03 (0.01, 0.10)	0.01	0.01 (0.00, 0.03)	0.31 (0.27, 0.36)	0.03

Abbreviations: crypt – cryptoxanthin; EFAs – erythrocyte fatty acids; HNR – Human Nutrition Research (Medical Research Council HNR Unit, Cambridge, United Kingdom); IARC – International Agency for Research on Cancer; ICC – intraclass correlation coefficient; PPFAs – plasma phospholipid fatty acids; RIVM – Rijksinstituut voor Volksgezondheid en Milieu (Laboratory at the National Institute for Public Health and the Environment, Bilthoven, Netherlands)

*Nested case-control study of incident type 2 diabetes (197 cases, 187 controls). Carotenoids were not measured in 2 control participants. Non-cases were weighted by the inverse of cumulative incidence of type 2 diabetes in EPIC-Norfolk for calculation of medians and interquartile ranges.

†Convenience sample of participants with nutritional biomarkers measured by two methods.

‡The denominator was the sum of all fatty acids presented in this table.

Table 4.9 Associations between the nutritional biomarker score of the Mediterranean diet* with incidence of type 2 diabetes in EPIC-Norfolk: comparison of results from different sets† of biomarkers per 1 standard deviation higher score values

Biomarker score and model	Set 1: EFAs & IARC carotenoids	Set 2: IARC PPFAs & IARC carotenoids		Set 3: HNR PPFAs & Vitas carotenoids
	OR (95% CI)	OR (95% CI)	OR (95% CI)	HR (95% CI)
Cases/non-cases (n)	197/185	197/185	444/288	688/886
Age & sex adjusted	1.00 (0.82-1.24)	0.77 (0.59-1.00)	0.78 (0.66-0.90)	0.69 (0.61-0.78)
Multivariable	1.03 (0.81-1.31)	0.85 (0.63-1.16)	0.82 (0.67-0.98)	0.72 (0.61-0.85)

Abbreviations: CI – confidence interval; EFAs – erythrocyte fatty acids; HNR - Human Nutrition Research; HR – hazard ratio; IARC – International Agency for Research on Cancer; n – number of participants; OR – odds ratio; PPFAs – plasma phospholipid fatty acids

*The biomarker score was derived as a discriminatory model between the Mediterranean and habitual diet in the MedLey randomised partial-feeding controlled trial. Circulating carotenoids and fatty acids were used to calculate the score as linear predictions from the discriminatory model.

†Set 1 and Set 2 were case-control studies nested in the EPIC-Norfolk case-cohort study participating in EPIC-InterAct. Set 3 includes the same biomarker assays as were used in the main analysis in EPIC-InterAct. Set 3 participants constituted the case-cohort study. Results for Set 2 are presented for participants overlapping with Set 1, and separately for all available Set 2 participants. Set 1 and Set 2 were analysed using complete-case logistic regression with the original EPIC-Norfolk data. The proportion of participants with missing data in the multivariable model was 2%. Set 3 was analysed using harmonised EPIC-InterAct data with Prentice-weighted Cox regression on multiply imputed datasets. The proportion of participants with missing data was 20%.

The multivariable model for was adjusted for the following covariates: age (as timescale in Cox regression; age and age-squared in logistic regression), sex, prevalent cancer, cardiovascular disease, hypertension and hyperlipidaemia; familial history of T2D, smoking status (never, former, current smoker), physical activity index (inactive, moderately inactive, moderately active, active), seasonality (sine and cosine function of the day of the year), fasting status (<3, 3-6, >6 hours), current use of vitamin or mineral supplements, menopausal status (pre-, peri-, postmenopausal, bilateral ovariectomy), current hormone replacement therapy use, marital status (single, married or cohabiting, divorced or separated, widowed), educational attainment (none, primary school, technical or professional school, secondary school, post-secondary school education), current employment, body mass index and waist circumference. In Set 1 and Set 2, variables or categories with cell counts < 20 participants were merged with similar variables or adjacent categories.

4.5.5 Associations of the biomarker score with self-reported diet

There was evidence of a positive relationship between higher self-reported adherence to the intervention and the biomarker score in the MedLey trial. Stratified by thirds of self-reported adherence to the intervention in the Mediterranean group (as quantified by the MDS), mean differences (95% CI) in the biomarker score relative to the habitual diet were 1.45 (0.93, 1.97), 1.83 (1.32, 2.34) and 2.18 (1.67, 2.69) SD (p trend = 0.027). The standardised coefficient for change in the biomarker score relative to change in the MDS was 0.29 (0.17, 0.40) in the Mediterranean and 0.13 (-0.00, 0.26) in the habitual diet group (p interaction = 0.075). Thus, one SD increase in the self-reported MDS was longitudinally associated, respectively, with a 0.29 SD increase and a marginal 0.13 SD increase in the biomarker score.

Cross-sectionally, the biomarker score was modestly positively correlated with the MDS (r = 0.27) in the Mediterranean diet arm of the MedLey trial post-intervention, and weakly positively correlated in the baseline sample of the MedLey trial (r = 0.18) and in the InterAct subcohort (r = 0.11) (**Table 4.10**). Correlations in the subcohort ranged from 0.01 in the largely vegetarian British participants from the 'health-conscious' recruitment arm to 0.24 in Spain, and otherwise 0.06-0.15. The correlations of the biomarker score with components of the MDS were heterogeneous across the three samples. In the MedLey trial intervention group, positive correlations were strongest for the self-reported intakes of vegetables (r = 0.27) and legumes (r = 0.24), and a weak inverse correlation was observed for meat and meat products (r = -0.05) and olive oil (r = -0.08). In the MedLey trial baseline sample, olive oil was the most robust correlate (r = 0.23), exceeding the strength of correlation for the overall MDS (r = 0.11), followed by meat and meat products (r = 0.17) and fish and seafood (r = 0.14). In EPIC-InterAct, fish and seafood had the strongest pooled correlation with the biomarker score (r = 0.17), which was higher than that of the MDS. It was also the only MDS component which had consistently positive correlations across all countries.

Table 4.10 Pearson correlation coefficients (minimum, maximum country-specific values in EPIC-InterAct) between a biomarker score of the Mediterranean diet derived in the Medley trial and the Mediterranean Diet Score estimated from self-report

MDS or MDS component, g/1,000 kcal*	Mean intake (SD)	r†	(min, max)
MedLey trial, Mediterranean diet group (n = 68)			
MDS (points, max. 18)	13.8 (1.9)	0.27	
Vegetable	113 (56)	0.27	
Legumes	23 (24)	0.24	
Fruits and nuts	211 (100)	0.10	
Cereals	72 (26)	0.09	
Fish and seafood	40 (26)	0.14	
Meat and meat products	23 (24)	-0.05	
Dairy	161 (65)	0.14	
Olive oil	34 (18)	-0.08	
Alcohol (g/day)	8.1 (8.1)	0.06	
MedLey trial, baseline sample (n = 133)			
MDS (points, max. 18)	9.6 (2.6)	0.18	
Vegetable	81 (64)	0.10	
Legumes	8.4 (17.5)	0.05	
Fruits and nuts	138 (75)	0.03	
Cereals	71 (39)	-0.06	
Fish and seafood	22 (26)	0.14	
Meat and meat products	34 (31)	0.17	
Dairy	130 (94)	-0.10	
Olive oil	4.1 (7.7)	0.23	
Alcohol (g/day)	9.2 (10.6)	0.12	
EPIC-InterAct subcohort (n = 12,625)			
MDS (points, max. 18)	8.8 (3.1)	0.11	(-0.01, 0.24)
Vegetable	94 (65)	0.08	(0.04, 0.17)
Legumes	9.1 (13.2)	-0.01	(-0.12, 0.08)
Fruits and nuts	122 (99)	0.04	(-0.02, 0.10)
Cereals	104 (41)	-0.01	(-0.12, 0.04)
Fish and seafood	18 (16)	0.17	(0.11, 0.31)
Meat and meat products	52 (24)	0.03	(-0.06, 0.33)
Dairy	163 (114)	0.00	(-0.06, 0.06)
Olive oil	9.1 (13.8)	0.06	(-0.05, 0.30)
Alcohol (g/day)	13.4 (18.7)	-0.01	(-0.08, 0.09)

Abbreviations: MDS – Mediterranean Diet Score, min – minimum; max – maximum; SD – standard deviation

*Adjusted for estimated energy intake using the residual method.

†Values in the EPIC-InterAct sample are pooled estimates of country-specific correlations. Fisher z-transformation was used to obtain standard errors of correlation coefficients for pooling via random-effects meta-analysis. Red colour highlights positive coefficients, and blue colour highlights negative coefficients with intensity proportional to value.

4.5.6 Secondary biomarker scores

Secondary biomarker scores derived using only linear biomarker terms (without pairwise interactions) had statistically significantly lower discriminatory performance between the Mediterranean and habitual diet arms of the MedLey trial than the primary interactions-based model (5-fold cross-validated C-statistic values 0.83 for both, p difference ≤0.033). Association with incident T2D was compatible with both decreased and increased risk for the score based on elastic net regression (HR per 1 SD = 0.87; 95% CI: 0.74-1.02). It was inverse for the BIC-based score, albeit the effect size and precision were lower than for the primary biomarker score (HR per 1 SD = 0.82; 95% CI: 0.70-0.96). The null or weaker associations were driven by positive associations detected in Denmark with HRs (95% CI) of 1.38 (1.24-1.54) and 1.26 (1.15-1.38), respectively (**Appendix 4.2**). Comparisons of top versus bottom fifths of the biomarker scores in the overall EPIC-InterAct study were statistically significantly inverse for both scores at 0.65 (0.42-0.99) and 0.59 (0.38-0.93), respectively.

I conducted post-hoc exploratory analyses to investigate the drivers of the unexpected change of direction of estimates in Denmark compared to the primary biomarker score (**Appendix 4.3**). Adjustments or exclusions from calculations of the biomarker scores of lignoceric (C24:0) or gondoic (C20:1) acids attenuated the positive associations to the null. No other individual component biomarkers appeared to influence the results, and neither of the adjustments or exclusions allowed to replicate the inverse association observed with the primary biomarker score. Both C24:0 and C20:1 had associations with incident T2D which were distinctly different in Denmark than in the whole EPIC-InterAct study. HRs (95% CI) per 1 SD higher C24:0 and C20:1 in Denmark were, respectively, 0.52 (0.45-0.60) compared to the overall 0.76 (0.66-0.86) in EPIC-InterAct and 1.29 (1.19-1.40) compared to the overall 0.76 (0.66-0.86) in EPIC-InterAct. C24:0 was negatively weighted and C20:1 was positively weighted in both secondary biomarker scores (**Appendix 4.3**). Overall, these results suggested that lignoceric (C24:0) and gondoic (C20:1) acids were driving the positive associations of the secondary biomarker scores with incident T2D in Denmark.

4.6 Discussion

In the current research I combined information from experimental and observational studies to investigate the association between a composite biomarker score of adherence to the Mediterranean diet and incident T2D. The key findings were that a biomarker score derived within the MedLey RCT had a high discriminatory performance between the Mediterranean and habitual diet arms, and that when this biomarker score was applied to the pan-European EPIC-InterAct Study, there was an inverse association with incident T2D. The 95% prediction interval for this association did not include the null, suggesting that the finding would be expected to be replicated in similar populations.[268] Higher adherence to the Mediterranean diet (as reflected by a 10-percentile higher value of the biomarker score) could reduce the incidence of T2D by 11%. The magnitude of association was approximately 3-fold stronger when using the biomarker score compared to a score derived from dietary self-report (Chapter 3).

4.6.1 Effects of the Mediterranean diet on nutritional biomarkers

Differences in individual nutritional biomarkers between the Mediterranean and habitual diet groups of the MedLey trial, such as higher concentrations of circulating carotenoids, and docosahexaenoic and oleic fatty acids, were consistent with the overall body of interventional evidence (Chapter 2). Several effects reported in the current work are novel or otherwise merit discussion. Higher concentrations of ß-carotene, lycopene and the marginally higher urinary output of potassium were indicative of a successful implementation of dietary counselling for increased consumption of fruits and vegetables, which were not a part of the feeding intervention.[269]

For fatty acids, the Mediterranean diet resulted in a decrease of several SFAs: myristic (C14:0), pentadecanoic (C15:0), behenic (C22:0) and lignoceric acid (C24:0). The intervention did not have statistically significant effects on the major products of de-novo lipogenesis (DNL) – palmitic (C16:0), palmitoleic (C16:1) and stearic acids (C18:0) – though there was a decrease in C14:0.[270] Beyond the DNL activity, the decrease in C14:0 may have been driven by its lower intakes, with the major dietary sources being dairy fat and other animal fats.[80,154] Likewise, the decrease in C15:0 is compatible with counselling for a moderate intake of dairy and provision of low-fat yogurt in the intervention arm.[271] Other potential causes include decreased endogenous synthesis of this fatty acid[271] (possibly via modulation of liver fat content[271,272]) or

a response to increased proportion of energy derived from dietary fat,[269] particularly given no effect of the MedLey intervention on C17:0.[273] Concurrently, there was a decrease in heptadecenoic acid (C17:1) which is another candidate biomarker of dairy fat intake.[274] Beyond dairy fat, it is also present in ruminant adipose tissue[275] and thus the result may have been additionally driven by lowering intakes of ruminant meat and adipose tissue.

The current investigation is the first to report evidence from a RCT that the Mediterranean diet decreases relative concentrations of circulating long-chain SFAs and increases concentrations of long-chain MUFAs (≥C20). Notably, interventions with supplementation of olive oil alone have found a similar pattern of effect on the profile of fatty acids.[276,277] Long-chain SFAs have been an area of active research in recent years in observational epidemiology as factors inversely related to risk of T2D and several cardiovascular diseases.[278] However, their biology is poorly understood, and the relative contributions of diet and metabolic control to regulation of their concentrations remain unknown. Small amounts of these fatty acids are present in some fatty plant foods, such as peanuts, macadamia nuts, and vegetable oils, e.g., canola and sunflower oils (but not olive oil).[278] Short-term trials have shown that circulating long-chain SFAa increase in response to intervention with peanuts[279] and macadamia nuts.[280] Observational evidence suggests that they are weakly correlated with a broad range of dietary exposures.[196,241] The magnitude of reported positive correlations was similar between nuts and seeds and other fatty foods and dietary fats, including dairy fat.[196,241] Inverse correlates include alcohol,[196,241] grain products,[196] carbohydrate in men (but not in women),[241] fish,[196] red and processed meat,[196] unprocessed meat,[241] and olive oil and other vegetable oils[196] (though opposite directionality has also been observed for the latter[241]). The mechanisms behind the concentration-lowering effect of the Mediterranean diet on long-chain SFAs in the MedLey trial are unclear. They may have been driven directly by a combination of changes in dietary intakes of foods and nutrients, or via the effects of the intervention on lowering the rate of elongation of C18:0.[278] The increase in long chain MUFAs, gondoic (C20:1) and nervonic acid (C24:1), could be attributed to upregulation of elongation of oleic acid (C18:1-n9c) driven by abundance of this substrate[281] or higher dietary intakes of these fatty acids from marine fish and vegetable oils.[282,283]

For n-6 PUFAs, eicosadienoic acid (C20:2-n6) increased and osbond acid (C22:5-n6) decreased in response to the intervention. Little is known about the dietary determinants of circulating C20:2-n6. It is unlikely to be a biomarker specific to the Mediterranean diet or its components. Beyond the results on this dietary pattern,[108] it has also been reported to increase

in response to a high-carbohydrate, low-fat diet (57% and 23% of energy, respectively) in comparison to moderate- or low-carbohydrate diets.[284] Osbond acid has been identified as a candidate biomarker of intake of eggs in metabolomic profiling of habitual diet.[285]

Overall, nutritional biomarkers included in the biomarker score were likely affected by changes in intakes of fruit, vegetables, fish, meat, eggs, and dietary fat (in particular from olive oil and dairy). Effects of the Mediterranean diet per se on individual biomarkers were also plausible via its impact on metabolic control. The intervention goals of higher consumption of whole-grain products and legumes did not have matching biomarkers with direct links to dietary intakes.

Among biomarkers not used for derivation of the biomarker score, there was a decrease in C18:0 dimethyl acetal and trans-vaccenic acid (C18:1-n7t) and an increase in cis-vaccenic acid (C18:1-n7c). Additionally, urinary excretion of magnesium was higher in the Mediterranean than the habitual diet group, likely reflecting across changes in dietary intakes but also homeostatic regulation of magnesium status.[286,287] C18:0 dimethyl acetal is a plasmalogen endogenously synthesised by humans and it is present in considerable amounts in meat and fish.[288] It has been shown to be unaffected by interventions with plant or dairy fats,[288] and it was found to be lower in habitual vegans than meat-eaters.[289] Lower trans-vaccenic acid in the Mediterranean diet group was likely yet another indicator of lower intake of ruminant fat.[290] In turn, higher cis-vaccenic acid is a surprising finding given lack of effect of the intervention on major DNL products, including the precursor palmitoleic acid (C16:1). Also, this result was not previously identified in RCTs.[81,82,100,108] A common features of these trials was a concurrent lack of effect on oleic acid (also inconsistent with the MedLey trial), possibly as a result of conducting the investigations in populations with high baseline intake of olive oil[81,82] or use of modified Mediterranean diets with alternative sources of dietary fats.[100,108] Cis-vaccenic acid has recently been identified in olive oil in considerable amounts[291] which provides a plausible explanation for higher concentrations of this fatty acids in the MedLey trial intervention group.

Research on the effects of the Mediterranean diet on circulating fatty acids has been limited by selective reporting focusing on PUFAs and MUFAs, and omission of results on low-concentration fatty acids in journal articles (Chapter 2). Only one previous trial reported on a full profile of 35 fatty acids measured in erythrocytes.[108] It compared the effects of hypo- and isocaloric Mediterranean and Central European diets in Poland. This RCT found the same effects of the Mediterranean diet as did the MedLey trial on gondoic (C20:1) and eicosadienoic

acid (C20:2n-6) but an opposite response (increase) of pentadecanoic acid (C15:0), consistent with use of full-fat dairy products in the Mediterranean diet.[108] In contrast to the MedLey trial, all remaining fatty acids did not differ between the arms. The small number of significant results was likely driven by using absolute rather than relative concentrations in the analysis[138] in the context of a weight-loss intervention. There was an approximately 2-fold decrease of total fatty acids in erythrocyte membranes in both trial arms[108] which may have obscured changes in the composition in the profile of fatty acids when analysed on the absolute scale.

4.6.2 Strengths and limitations

The major strength of the current research was the use of a novel analytical approach that combined the derivation of an objective measure of the Mediterranean diet in a partial-feeding study (the MedLey RCT) and its application in a large observational study (the EPIC-InterAct Study). The RCT compared the effects of this dietary pattern (without weight loss) with continuation of habitual diet on nutritional biomarkers. Such an experimental design allowed me to derive a biomarker score of the Mediterranean diet in a manner that was free from influences of other dietary and non-dietary factors, while using a control group suitable for application of the score to study participants in an observational setting in Western countries. I used a set of biomarkers that reflect dietary exposures over the past weeks or months[292] which is a desirable timeframe for assessment of habitual diet for epidemiological applications. Beyond the novelty of deriving a biomarker score, this work identified several novel or rarely reported on effects of the Mediterranean diet on individual biomarkers. My analysis was based on the largest study to-date of nutritional biomarkers and T2D, including over 9,000 incident cases. Among other strengths, the observational analyses adjusted for a comprehensive range of potential confounding factors and included several sensitivity analyses and modelled the population impact of greater adherence to the Mediterranean diet on future T2D risk.

This research had several limitations. The intervention in the MedLey trial was administered at one level of intensity. This allowed for modelling of the Mediterranean diet only as a binary variable and precluded objective evaluation of the dose-response relationship between the adherence to this dietary pattern and the biomarker score.[226] However, I found evidence of a dose-response relationship in secondary analyses which tested the relationship between self-reported level of adherence and the biomarker score in the intervention group. Second, the MedLey trial was a partial-feeding RCT which may have resulted in lower adherence to the

dietary components not provided (e.g., fruits and vegetables) than the food items provided as part of the Mediterranean diet intervention. Third, the study used a combination of candidate biomarkers of intake, concentration and function which are not only affected by dietary intakes, but also bioavailability, endogenous synthesis, genetic variation, homeostatic control, and nutrient metabolism.[43,80,235] Thus, changes in nutritional biomarkers in the MedLey trial may have represented a metabolic response to a healthy diet, rather than a specific biomarker profile indicative of adherence to the Mediterranean diet. Fourth, I was unable to assess whether participants in the intervention group reached equilibrium in concentrations of nutritional biomarkers and values of the biomarker score. Fifth, the trial experienced a moderate degree of drop-outs between the randomisation and the end-of-trial assessment (20% in each arm, including missing biomarker data). Both the unknown equilibrium status and the loss to follow-up[293] may have biased the derivation of the RCT score and evaluation of its performance as a classifier. Sensitivity and specificity of the biomarker score, as well as external validity, remain unknown at present and require evaluation of the score in external trials of interventions with the Mediterranean diet and other dietary patterns.

Limitations in the observational research in EPIC-InterAct included the likelihood of several potential sources of bias (Chapter 3). Briefly, this included reverse causality early in the follow-up, differential misclassification in ascertainment of the outcome, and within-person variation in nutritional biomarkers. Given the results of the sensitivity analyses and practical considerations, these sources of bias would be expected to underestimate the association or not have a meaningful impact on the results (Chapter 3). The random measurement error in the nutritional biomarkers could bias the association both towards and away from the null in the context of multivariable models.[22] Residual confounding was unlikely to fully account for the inverse association given its large effect size and the effect sizes of risk factors for T2D previously reported in the literature.[187,294] However, the above mentioned potentially limited specificity may have contributed to positive residual confounding.

Finally, reliability of the externally derived biomarker score applied to biomarker assays used in EPIC-InterAct remains unknown. Importantly, circulating fatty acids were measured in MedLey Study in erythrocytes, and in plasma phospholipids in EPIC-InterAct. Profiles of fatty acids measured in both fractions are overall similar, but have sizeable differences in several fatty acids.[80] I was unable to objectively evaluate the reliability due to unavailability of samples of participants with biomarker data measured using both the MedLey and InterAct methods. I have, however, compared it in subsamples of British participants of the EPIC study who had

circulating carotenoids and fatty acids measured with multiple methods. The key findings were that overall reliability of quantification of fatty acids was similar when comparing phospholipid fatty acids from two assays and phospholipid fatty acids with erythrocyte fatty acids. The biomarker score calculated from these assays and two assays of carotenoids was not reproducible across the different combinations of biomarker measurements, yielding null correlations, with unclear implications for diet-disease associations. The inverse association of the biomarker score with incident T2D from the main analysis in EPIC-InterAct was replicated in one of the reproducibility subsamples and there was a null association in a second one. Despite these inconsistencies, the association in the EPIC-InterAct study was remarkably robust and stable across sensitivity analyses with biomarker scores re-developed with several alternative analytical decisions leading to selection of different sets of biomarker predictors and biomarker-biomarker interactions. This decreases concerns over the potentially low reproducibility giving rise to a chance finding.

Of note, within the EPIC-InterAct subcohort I observed the highest biomarker score values in Scandinavian countries, and the lowest in the countries of the Mediterranean region. The reasons for this are unclear but may include the MedLey trial design whereby the Mediterranean diet intervention was adapted to a non-Mediterranean setting. Moreover, there is also evidence to suggest that adherence to the Mediterranean diet has been moderate and declining in European Mediterranean countries.[295,296] Though not a limitation per se, the biomarker score had a low positive correlation with the self-reported MDS in the EPIC-InterAct subcohort. Mediterranean diet scores applied to dietary intakes estimated from food frequency questionnaires and diet histories in Western countries are unlikely to capture adherence to an actual Mediterranean diet, hence the biomarker and the self-reported score are likely two different constructs and low correlations can be reasonably expected.

4.6.3 Comparison with previous studies

Attempts to derive biomarkers of the Mediterranean diet and other dietary patterns have previously been largely confined to metabolomic profiling using cross-sectional designs.[60,69] In a subgroup analysis in one of the centres of the PREDIMED trial, urinary metabolomic profiles at 1- or 3 years post-randomisation were able to correctly classify 93%, 85% and 68% of participants to their respective intervention arms of the Mediterranean diet with either olive oil or nuts and the control lower-fat diet.[177] Only one previous study considered a score

comprised of nutritional biomarkers.[5] Using a novel feeding design of individualised habitual diets and performing data-driven variable selection from a wider range of nutritional biomarkers than available in our investigation, the study derived and internally validated a biomarker score which included two circulating carotenoids and seven fatty acids, urinary potassium and doubly labelled water (objective measurement of total energy expenditure).[50] This finding corroborates our results and the hypothesis that combinations of nutritional biomarkers, and in particular circulating carotenoids and fatty acids, can be used to objectively assess habitual adherence to the Mediterranean diet. Research on other dietary patterns likewise suggests utility of combinations of carotenoids and fatty acids for their objective assessment.[50,176,238,239]

My work on derivation of the nutritional biomarker score of the Mediterranean diet adds to the previous literature by incorporating a randomised controlled intervention. Owing to the RCT design of the MedLey trial, I was able to minimise confounding by non-nutritional factors, and thus derive a biomarker score with| potentially improved generalisability to external populations. A further novelty of this work is the application of the experimentally-derived score to prospective associations between dietary patterns at large and disease outcomes. To my knowledge, the only other example of derivation and application of a putative biomarker of the Mediterranean diet to associations with disease endpoints was a metabolomics-based analysis in the PREDIMED study and prospective cohorts in the USA. Similar to my analysis, it observed inverse associations for cardiovascular disease (CVD) using an observationally derived metabolite score (HR range per 1 SD: 0.71-0.86) and weaker or null associations when using the self-reported Mediterranean diet.[297] Of note, the study used baseline data from a case-cohort design nested within PREDIMED for derivation of the metabolite score without exclusion of the oversampled incident cases of CVD. This may have decreased the specificity of the score to the Mediterranean diet while increasing the likelihood of observing an inverse association with incidence of CVD. The PREDIMED trial has shown that the Mediterranean diet reduced the risk of CVD.[117] Thus future cases would be expected to have metabolomic profiles compatible with lower adherence to this dietary pattern than non-cases while also exhibiting differences in metabolomic profiles unrelated to diet. Regressing adherence to the Mediterranean diet on metabolite concentrations (i.e., deriving the metabolite score) in the case-cohort sample would be expected to be conditioned more strongly on cardiometabolic health and non-dietary exposures compared to restricting the derivation sample to the subcohort. Indeed, the composition of the metabolite score reported in this study supports these

concerns. For example, cotinine – a metabolite of nicotine and a biologically implausible dietary biomarker – was selected into the score as a negative correlate of the Mediterranean diet.[297]

For incident T2D, the PREDIMED trial of primary CVD prevention reported a 30% risk reduction in the Mediterranean diet intervention arms relative to the control lower fat diet group (273 incident cases in 3,541 participants).[17] The CORDIOPREV trial of secondary CVD prevention reported a statistically non-significantly higher incidence of T2D in the Mediterranean diet arm compared to a low-fat control group (HR = 1.35; 95% CI: 0.91-2.01; 107 cases in 462 participants).[298] This finding from a predominantly pre-diabetic sample (85% prevalence) of a secondary CVD prevention trial is of lower relevance to prevention of T2D in the general population than that of the PREDIMED trial. Beyond moderating effects of population characteristics, the result of the CORDIOPREV trial may have been driven towards favouring the lower-fat diet by a small mean weight loss over 5 years (-1.14 kg) compared to a small average weight gain in the Mediterranean diet arm (+0.78 kg), thus potentially introducing an indirect mediating effect via body weight. Five-year effects of the PREDIMED interventions on body weight were decisively null.[131] Inverse associations between the self-reported Mediterranean diet and incident T2D in middle-aged adults have previously been reported in EPIC-InterAct[182] and other prospective cohorts.[181] The largest reduction in the incidence between extreme categories of the self-reported adherence to the Mediterranean diet was 25% (top versus bottom fifth),[299] which is a substantially smaller effect size than the 62% observed in the current study using the biomarker score.

4.6.4 What this study adds and implications of this research

The analysis of dietary patterns aims to evaluate the cumulative impact of dietary exposures on disease risk to inform development of dietary guidelines.[26] High quality evidence on dietary patterns and the primary prevention of T2D is lacking, partly because of the limitations of dietary assessment methods.[6] Self-reported tools have been used almost exclusively in the published studies on this topic.[6,245] I have developed a method of objective assessment of the Mediterranean diet via nutritional biomarkers in order to investigate the association of this dietary pattern and T2D in a more robust manner. This approach yielded substantially greater magnitude and stronger inverse associations than for the Mediterranean diet assessed with subjective methods (Chapter 3), thus also addressing the limitation of small effect sizes often

observed in nutritional epidemiology.[300] My modelling indicates that even a modest, 10-percentile increase in the objectively assessed Mediterranean diet could potentially avert 11% of new T2D cases. In comparative analysis, this effect size was approximately 40% of the estimated effect of decreasing mean BMI by 10 percentiles. The BMI is a well-established, causal risk-factor for T2D with a large effect size,[187] and this comparison illustrates the important potential of improving dietary quality for prevention of this disease.

Of note, the inverse associations of the biomarker score with T2D in the current study were independent from measured adiposity, with no evidence of effect modification by BMI. This suggests that a sizeable decrease in the incidence of T2D could potentially be achieved through improved adherence to the Mediterranean diet even in the absence of modification of body weight, and across the spectrum of adiposity in the population. Investigations of adherence to dietary patterns estimated from self-report and disease outcomes typically aim to emulate an isocaloric interventions with different levels of dietary quality by statistical adjustment for estimated energy intake[223] and covariates, including measures of adiposity to test independence of the association from body fat or potential mediation. This approach has limitations which preclude inference whether observed associations are truly independent from energy balance. Energy intakes estimated from self-report perform poorly as measures of true energy intakes,[301] and the impact of adjustment for estimated energy intake may be limited to decreasing measurement error of dietary exposures by 'cancelling out' correlated errors,[302] and not necessarily simulating the isocaloric conditions in practice. Excess adiposity is one of the key correlates of dietary misreporting,[234] and thus adjustment of a diet-disease association for indicators of body fat may provide a biased estimate for the dietary exposure. In the current study, I was able to remove the need for energy adjustment by deriving the biomarker score in a trial under isocaloric conditions and ensure independence of errors in the biomarker score from adiposity.

Comparison of biomarker scores derived with and without pairwise interactions between individual biomarkers was another methodological contribution of this investigation. The standard approach to model building with multiple analytes involves considering only linear biomarker terms as predictors despite the recognised need to account for effect modification and non-linear relationships.[297] The primary interactions-based model had consistent, inverse associations with incident T2D in EPIC-InterAct as a whole and within individual countries, and across a range of sensitivity and subgroup analyses. By contrast, two approaches to deriving the biomarker score using only linear biomarker terms yielded a positive association

in Denmark and less consistently inverse or null associations in the remaining countries. This led to overall pooled associations that were null or substantially weaker than that for the primary biomarker score. It is plausible that the interaction terms in the biomarker score acted as a guarding mechanism against undue influence of component biomarkers on the association with incident T2D.

Successful application of the nutritional biomarker score in the current study opens up avenues for a broader use of this method for the Mediterranean diet and other dietary exposures in observational research and in monitoring of compliance in dietary trials and potentially of preventive interventions in the real world. My primary finding of an inverse association between the Mediterranean diet and incident T2D in pan-European populations is consistent with interventional evidence from a Mediterranean (Spanish) population,[17] and it lends itself to consideration for developing dietary guidelines, public health policy and personalised dietary advice.

4.6.5 Conclusions

The findings of the current study have demonstrated the utility of combining circulating carotenoids and fatty acids as a composite biomarker of the Mediterranean diet. The inverse association between a biomarker score of this dietary pattern and T2D were approximately 3-fold larger than for adherence to the Mediterranean diet estimated from dietary self-report, thus raising the possibility that previous prospective studies may have substantially underestimated the strength of the association. These results add to the evidence in favour of adopting a Mediterranean-type diet in European adults for prevention of T2D.

Chapter 5

Generalisability of nutritional biomarker scores of the Mediterranean diet

Abstract

Background: Composite measures of biomarkers have been used in nutritional epidemiology for assessment of dietary patterns based on prediction of dietary self-report. Validity and generalisability of such composite biomarkers is unclear and has not been previously evaluated.

Methods: I derived biomarker scores based on prediction of Mediterranean diet scores (MDS) from five circulating carotenoids and 24 fatty acids in multiple cross-sectional samples in the European Prospective Investigation into Cancer and Nutrition (EPIC) study (up to 12,495 participants) and the baseline sample of the MedLey trial (n = 144). I then tested their validity by examining whether they differed after 6 months of intervention in the MedLey trial between participants under a Mediterranean diet intervention or continuation of habitual diet (n = 133 out of 166 randomised). Associations of the scores with incident T2D were evaluated in the EPIC-InterAct case-cohort study. It included 22,202 participants, of whom 9,453 were T2D cases, with relevant biomarkers from an original case-cohort of 27,779 participants.

Findings Controlling for multiple testing, values of 8/13 biomarker scores were higher in the Mediterranean diet intervention than the control habitual diet group of the MedLey trial, and 10/13 scores, including 2 out of 4 externally derived, were inversely associated with incidence of T2D in EPIC-InterAct. For example, the biomarker score derived in the EPIC-InterAct subcohort differed by 0.49 standard deviation between the MedLey trial arms (95% confidence interval: 0.16, 0.82; Q-value (false discovery rate-corrected p value) = 0.020) and the hazard ratio of incident T2D was 0.82 (95% confidence interval: 0.76-0.89; Q-value <0.001) per standard deviation. The magnitude of the differences and associations were similar for biomarker scores which met the validity criterion in the MedLey trial or had an inverse relationship with new-onset T2D in EPIC-Interact.

Conclusions: These findings suggest that biomarker scores of the Mediterranean diet derived based on prediction of dietary self-report may be valid measures of adherence to the Mediterranean dietary pattern and largely generalisable inverse correlates of T2D risk. Further

research is needed to confirm validity of biomarker scores derived in cross-sectional designs based on dietary self-report as biomarkers of the Mediterranean diet.

5.1 Background

The advent of large-scale measurements of nutritional biomarkers and omics platforms in prospective cohort studies presents an opportunity to improve dietary assessment by leveraging data from dietary self-report and objective biomarkers.[51] This enables investigators to derive multi-analyte scores of complex dietary exposures such as dietary patterns[245,246] and foods[303] which can be used to test associations of the underlying exposure variables with incidence of T2D.[245,246,303] The primary motivation for application of this analytical framework is to decrease the influence of measurement error introduced by subjective dietary reporting, and to increase the precision of dietary assessment.[51]

Evaluation of adherence to dietary patterns using nutritional biomarkers is a novel approach to objectively evaluate the overall diet quality.[50] It consists in combining information from multiple analytes to explain a sizeable proportion of variation in adherence to dietary patterns in observational settings (Chapter 3)[50] or to allow for robust discrimination between participants under interventions with different dietary patterns (Chapter 4). Variable selection and derivation of such composite biomarkers is driven by whole diets; however, the process depends on the structure of dietary intakes and strength of the relationships between components of dietary patterns and individual biomarkers in derivation samples. The ensuant statistical models are inherently predictive, and they may not be generalisable to external datasets with different compositions of foods and nutrients or otherwise divergent population characteristics which may affect the levels of nutritional biomarkers. Therefore, generalisability of the biomarker scores as correlates of risk of disease outcomes may also be limited. Construct validity of biomarker scores is a further concern with potential implications for specificity to a given dietary pattern. For example, a biomarker score which predicts self-reported Mediterranean dietary pattern in a non-Mediterranean setting may not adequately capture adherence to the traditional Mediterranean diet.

Displacement of memory-based dietary assessment with an objective measurement in the analysis is a conceptually appealing prospect, however, the validity of such analysis rests on the assumption that a given biomarker score is a valid measure of the dietary exposure of interest. This assumption is unverifiable by evaluating the strength of the relationship between the score and the corresponding measure estimated from dietary-self report, which is typically the only benchmark available within contemporary prospective cohorts. Integrating data from observational studies with a dietary intervention can provide an objective criterion for testing

validity of such composite biomarker measures. Thus, using the case of the Mediterranean diet, I assessed whether nutritional biomarker scores derived based on prediction of dietary self-report in multiple heterogeneous samples differ between study participants under a Mediterranean diet intervention and individuals following a habitual Western diet. I then tested the biomarker scores for generalisability to associations with incident T2D.

5.2 Aim

The aim of this chapter was to assess generalisability of observationally derived biomarker scores of the Mediterranean diet as biomarkers of this dietary pattern and predictors of incident T2D. The specific objectives were to:

1) derive nutritional biomarker scores of self-reported Mediterranean diet in multiple cross-sectional samples of the EPIC-InterAct study, non-InterAct participants of the EPIC-Norfolk study, and the baseline sample of the MedLey trial
2) test for post-intervention differences in the biomarker scores between the Mediterranean and control diet arms of the MedLey trial as a validation criterion
3) evaluate associations of the biomarker scores with incidence of T2D in EPIC-InterAct
4) evaluate the association of a literature-based biomarker score of the Mediterranean diet with incidence of T2D in EPIC-InterAct

Correlations between the biomarker scores and self-reported Mediterranean diet score (MDS) in the InterAct-subcohort were used as a secondary measure of performance of the biomarker scores. I additionally repeated objectives #1 and #2 for aHEI-2010 and DASH as an indirect test of specificity of the biomarker scores of MDS to the Mediterranean diet. I hypothesised that the biomarker scores of aHEI-2010 and DASH would not differ between the MedLey trial arms, or that they would have a smaller magnitude of differences than the biomarker scores of the Mediterranean diet.

5.3 Methods

This investigation was conducted in the EPIC-InterAct study (Chapters 3 and 4), the MedLey trial following harmonisation of biomarker data with EPIC-InterAct (Chapter 4). The EPIC-Norfolk study (Chapter 4), one of the cohorts participating in EPIC-InterAct, was additionally

used to externally derive biomarker scores of the InterAct MDS index and a MDS for adherence to pyramid-based guidelines (MDS-pyramid; **Table 5.1**) in non-InterAct participants. The MDS-pyramid was previously shown to have the strongest associations with incident cardiovascular disease among several scores of self-reported Mediterranean diet tested in EPIC-Norfolk .[304] Five carotenoid variables and 24 fatty acids overlapping between the EPIC-InterAct Study and the MedLey trial were used to derive the biomarker scores. For EPIC-Norfolk, 22 fatty acids were available which overlapped with both EPIC-InterAct and the MedLey trial.

5.3.1 The MedLey trial

The MedLey trial randomised 166 elderly, generally healthy, non-vegetarian participants in Australia to the Mediterranean diet intervention (n = 85) or continuation of habitual diet (n = 81), and 152 participants commenced the trial and had baseline serum carotenoids and erythrocyte fatty acids measured.[64,254] I excluded participants with missing dietary data and extreme values in nutritional biomarkers, defined as +/- 3 SD outside of the IQR of natural logarithm-transformed values (8 exclusions), leaving 143 participants available for derivation of the biomarker score in the baseline sample.

For validation of the biomarker scores, I applied the intention-to-treat analysis. No exclusions were made based on biomarker values. At the end of intervention 133 participants had complete biomarker data required for calculation of the biomarker scores, including 68 individuals in the Mediterranean diet and 65 in the control group.

5.3.2 The EPIC-InterAct study

The analytical sample was the same as in Chapter 4. It included 22,202 participants, with 9,453 participants who developed incident T2D and 13,313 subcohort participants, out of the original case-cohort of 27,779 individuals. The analysis in the UK was stratified by recruitment from the general population and health-conscious participants from the Oxford centre. The latter aimed to maximise recruitment of vegetarians who cannot adhere to some aspects of the Mediterranean diet and who were excluded from the MedLey trial. The UK health-conscious sample was excluded from derivation of the multi-country biomarker score. Following

exclusions of participants with missing biomarker and dietary data and implausible estimated energy intakes (Chapter 3), there were 12,495 participants available for derivation of the multi-country scores. Country-specific subcohort derivation samples included 499 participants in France, 1,763 in Italy, 3,288 in Spain, 960 in the UK general population, 213 in the UK health-conscious population, 1,367 in the Netherlands, 1,826 in Germany, 903 in Sweden, and 1,889 in Denmark.

5.3.3 The EPIC-Norfolk study

Dietary self-report was collected using a validated food frequency questionnaire (FFQ) and 7-day food diary at baseline and a follow-up visit.[305] Foods, energy, and nutrients were estimated using UK-specific portion sizes and food composition data.[306] I applied the InterAct MDS to estimated intakes from both the FFQ and 7-day food diary. The MDS-pyramid score included 15 components scored continuously between 0 and 1 based on pre-specified cut-offs (**Table 5.1**). Prior to calculation of the points, dietary intakes were estimated from the FFQ, residual-adjusted for energy and re-scaled to 2,000 kcal/day. The sum of the points represented the MDS-pyramid score. Information on demographic, medical and lifestyle covariates was collected at baseline and follow-up using questionnaires, and height, weight and waist circumference were measured by trained nurses.

Plasma carotenoids and phospholipid fatty acids measured at the Nutrition and Hormones Laboratory at the International Agency for Research on Cancer (Lyon, France) were used to derive the biomarker scores. They were available in a total of 7,241 participants out of the original cohort of 25,639 men and women. I excluded individuals who were EPIC-InterAct participants (718 exclusions), participants with repeat biomarker measurements during follow-up (used as test samples; further 127 exclusions), values in nutritional biomarkers +/- 3 SD outside of the IQR of natural logarithm-transformed values (further 211 exclusions), implausible estimated energy intakes or missing dietary or covariate data (further 604-749 exclusions depending on the diet score and dietary assessment method). Between 5,436-5,581 participants were included in derivation of the biomarker scores. A subsample of 432 non-InterAct participants had baseline and follow-up FFQ data and measurements of plasma carotenoids and phospholipid fatty acids based on the assays applied in EPIC-Interact.[307] I used this sample to test whether changes in self-reported MDS over an average of 3 years are associated with changes in biomarker score values.

Table 5.1 Scoring of dietary pattern indices

Component	Range of points	Minimum points	Maximum points
MDS (g/1,000 kcal)*			
Vegetable	0-2	<57.6	>100.3
Legumes	0-2	<0.49	>6.37
Fruits and nuts	0-2	<66.0	>133.8
Cereals	0-2	<81.3	>113.5
Fish and seafood	0-2	<9.48	>20.45
Meat and meat products	0-2	>59.8	<40.8
Dairy	0-2	>194	<102
Olive oil	0-2	Non-consumers	>6.85
Ethanol (g/day)	0 or 2	Intake outside of ranges for maximum points	Men: 10-50 Women: 5-25
MDS-pyramid (servings)†			
Vegetable	0-1	0/day	≥6/day
Legumes	0-1	0/week	≥2/week
Fruit	0-1	0/day	3-6/day
Nuts	0-1	0/day	1-2/day
Cereals	0-1	0/day	3-6/day
Dairy	0-1	0/day	1.5-2.5/day
Fish	0-1	0/week	≥2/week
Red meat	0-1	≥4/week	<2/week
Processed meat	0-1	≥2/week	≤1/week
White meat	0-1	0/week	1.5-2.5/week
Eggs	0-1	0/week	2-4/week
Potato	0-1	≥6/week	≤3/week
Sweets	0-1	≥4/week	≤2/week
Olive oil	0-1	Non-consumers	Consumers
Alcohol	0-1	Men: ≥4/day Women: ≥2/day	Men: 1.5-2.5/day Women: 0.5-1.5/day

Abbreviation: MDS – Mediterranean Diet Score

*Integer points were used. 1 point was assigned for estimated intakes between cut-offs for minimum and maximum points, except for ethanol for which either 0 or 2 points were assigned.

†Reported frequencies of intake in food frequency questionnaire were used to estimate numbers of servings. Estimated intakes were adjusted for energy using the residual-method and re-scaled to 2,000 kcal/day. Components with no upper limit on consumption were scored continuously. Continuous points proportional to intake were used for estimated intakes between cut-offs for minimum and maximum points. For foods for which moderate consumption was recommended (range given under Maximum points), overconsumption was defined as twice the mid-point of the recommended intake above the upper range of recommended intake. It was penalised with a score of 0.5. Intermediate values between the upper range of recommended intake and overconsumption were scored between 1 and 0.5 proportionally to intake. Binary scoring was applied to olive oil. For alcohol, 1 point was assigned for intake within recommended ranges, 0.5 for non-consumption, and 0 for overconsumption.

5.3.4 Literature-based biomarker score of the Mediterranean diet

The WHI-NPAAS feeding study of individualised habitual diets (Chapter 4, section 4.1.1) published biomarker score equations for adherence to dietary patterns, including the alternative Mediterranean diet (aMED) score.[50] The overlap of biomarkers included in the score with those available in EPIC-InterAct was incomplete. Thus, I considered the application of the WHI score to associations with incident T2D in EPIC-InterAct as an exploratory analysis.

The pre-specified internal validation criterion in the WHI-NPAAS feeding study was a cross-validated R^2 value ≥ 0.36 for regressing a given dietary exposure on biomarker concentrations, equivalent to a correlation of ≥ 0.60. The biomarker score of aMED had a $R^2 = 0.44$, and 0.36 after cross-validation. The score included seven fatty acids and two carotenoids available in EPIC-InterAct, as well as palmitoleic fatty acid, and urinary potassium and doubly labelled water which were not measured in EPIC-InterAct. The unavailable biomarkers contributed a total of 0.11 to the R^2 value in the empirical sample. Thus, their omission from calculation of the score would decrease the R^2 value in WHI-NPAAS to approximately 0.33, and the corresponding cross-validated R^2 would be expected to be below the validation criterion of 0.36. I calculated the partial WHI biomarker score of aMED for application in EPIC-InterAct as follows: log(β-carotene)*0.20 + log(α-carotene)*0.047 + log(C17:1n-9c)*-0.42 + log(C20:1n-9c)*0.45 + log(C24:1n-9c)*0.27 + log(C22:5n-3)*-0.31 + log(C22:6n-3)*0.25 + log (C22:4n-6)*-0.12.[50]

5.4 Statistical analysis

Derivation of biomarker scores and assessment of their associations with incidence of T2D in EPIC-InterAct followed the analytical workflow developed in Chapter 3. Briefly, the MDS indices were residual-adjusted for potential confounders between adherence to the Mediterranean diet and biomarker concentrations. Elastic net regression was run in 100 bootstrap samples, regressing the residual-adjusted MDS on nutritional biomarkers. Biomarkers selected in $\geq 90\%$ of samples were included in the scores, and model coefficients were estimated by means of ridge regression. For biomarker scores derived internally in EPIC-InterAct, a leave-one-out re-estimation of the ridge regression model was applied to each participant in the derivation sample to allow for estimation of their biomarker score value without the influence of their personal dietary self-report. Residual-adjustment of MDS in the

MedLey baseline sample was not applied due to homogeneity of the sample and unavailability of most of the covariates used for adjustment in EPIC-InterAct and EPIC-Norfolk.

The criterion for validation of biomarker scores in the MedLey trial was the detection of statistically significantly higher end-of-trial biomarker score values in the Mediterranean diet group than the control habitual diet arm. Differences in means of standardised biomarker values were estimated by mixed linear models with unstructured covariance. Correction for false discovery rate (FDR) was applied both to the validation stage in the MedLey trial and to testing of associations of biomarker scores with incidence of T2D in EPIC-InterAct.[308] I report the sharpened Q-values used to determine the statistical significance of results and conventional 95% CIs.

To inform analytical decisions in the subsequent Chapter, I repeated the validation procedure using biomarker scores established based on a computationally efficient single run of elastic net regression on the empirical derivation sample.

5.4.1 Additional information on derivation of the biomarker scores

My a priori approach to deriving the biomarker scores in this investigation was to include pairwise interactions between biomarkers in the set of predictor variables. However, this strategy turned out to be inefficient and ineffective in an interim analysis, as described in detail below. Thus, I used only linear terms in biomarker scores throughout the results presented.

The interactions-based approach turned out to be very computationally intensive. A single run of linear elastic net regression in the InterAct subcohort sample (~12,000 participants) took approximately one week, and 100 runs were needed in the bootstrapping procedure. I took a simplified approach to enable a more feasible interim analysis. I used only the empirical sample to select the optimal λ and α values by cross-validation. These fixed values were applied to bootstrap samples instead of selecting them internally in each sample. Furthermore, I used a coarse grid of α penalty values of 0.25, 0.5 and 0.75, as opposed to the pre-specified 0.1-0.9 in 0.1 increments. This simplified procedure was applied to the set of biomarker predictors both with and without pairwise interaction terms. The interactions-based biomarker score yielded materially similar correlations with MDS in the EPIC-InterAct subcohort as the biomarker scores with linear terms only (Δr 0.01-0.02; results not shown). The interactions-based scores did not differ in the MedLey trial between the Mediterranean and habitual diet groups. For

example, the difference in standardised means for the EPIC-InterAct multi-country score was -0.20 (95% CI: -0.52, 0.11) (other results not shown).

5.5 Results

5.5.1 Derivation and validation of biomarker scores

All the 29 nutritional biomarkers considered as predictors of adherence to MDSs were selected into any of the biomarker scores (**Table 5.2**). A single variable, rather than a combination of biomarkers, was identified in France, and the French biomarker score was not used in further analyses. The multi-country and Spanish scores, as well as those derived in EPIC-Norfolk, utilised most of the available biomarkers. Sparser solutions were identified in the remaining datasets, particularly for the biomarker score from the MedLey trial baseline sample. Overall, the biomarker scores were characterised by positive scoring of all carotenoids except for the negatively scored ß-carotene, negative coefficients for even-chain SFAs, and positive for docosahexaenoic acid (C22:6-n3). The remaining fatty acids were selected into the scores less consistently and with variable directionality of coefficients.

Correlation coefficients between the MDS and the biomarker scores in the InterAct subcohort ranged between 0.27-0.32, except for the score derived in British health-conscious participants (r = 0.19) and the MedLey trial baseline sample (r = 0.16) (**Table 5.3**). Country-specific correlations were generally the lowest in France and the health-conscious sample, and otherwise mostly in the range of ~0.2-0.4. Correlations tended to be the strongest for the InterAct biomarker scores within their derivation samples, suggesting lower out-of-sample performance in between-country comparisons. Correlation coefficients in non-subcohort incident type 2 diabetes cases, who did not participate in derivation of biomarker scores, were on average higher by 0.01 across all biomarker score-derivation sample comparisons, suggesting similar out-of-sample performance in within-country comparisons. There was substantial variation by country in median values of the biomarker scores standardised using the subcohort distributions (**Table 5.4**). Italy and Spain had consistently the highest values, and the Netherlands the lowest.

In the MedLey trial, eight of the 13 biomarker scores differed between the Mediterranean and habitual diet arms post-intervention with correction for multiple testing, including the InterAct multi-country score, several country-specific scores, the MedLey baseline score, and the EPIC-

Norfolk score of the MDS-pyramid (**Table 5.5**). The mean standardised difference for the EPIC multi-country score was 0.49 (95 % CI: 0.16, 0.82), and the effect size was similar for other scores which passed this validation criterion. FDR-corrected differences were not detected for the UK health-conscious biomarker score (-0.12; 95% CI: -0.44, 0.20), the Dutch (0.34; 95% CI: 0.04, 0.65), Danish 0.38 (0.05, 0.70), and the EPIC-Norfolk scores of FFQ-based MDS (0.39; 95% CI: 0.07, 0.70) and food diary-based MDS (0.41; 95% CI: 0.09, 0.72). Discriminatory performance of the biomarker scores between the trial arms was modest, with C-statistic values <0.70. Using biomarker scores from single runs of elastic net regression, as opposed to the above-described scores from bootstrapped elastic net regression, yielded materially similar estimates and identified FDR-corrected differences between the trial arms for the same biomarker scores (results not shown).

By contrast to the biomarker scores of MDS, biomarker scores of aHEI-2010 and DASH did not differ between the groups after FDR correction (Q-values > 0.11). Examples of end-of-trial differences in standardised means (95% CI) of the InterAct multi-country scores were 0.40 (0.09, 0.72) and 0.19 (-0.13, 0.50), respectively (other results not shown).

Dose-response analysis in the MedLey trial showed an overall pattern of increasing estimates of the validated biomarker scores of MDS across thirds of self-reported MDS in the Mediterranean diet arm, with evidence of a trend for the EPIC-InterAct scores from the UK (general population), Italy and Spain, and the MedLey trial baseline score (p trend < 0.05) (**Table 5.6**). These results were corroborated by positive longitudinal associations between self-reported MDS and biomarker scores in the Mediterranean diet arm (**Table 5.7**). Food components of the MDS had heterogeneous correlations with biomarker scores between different scores and across study samples, with substantial variation between EPIC-InterAct countries (**Table 5.8**).

In EPIC-Norfolk, the biomarker scores had high reliability ($r \geq 0.90$) between values calculated from the biomarker assays used to derive these scores and the biomarker assays used in EPIC-InterAct, as well as a fatty acids measured in erythrocyte membranes. Change in self-reported adherence to the Mediterranean diet over time was positively associated with values of the biomarker scores (**Table 5.9**). A 1 SD increase in the MDS over an average of 3.3 years was associated with ~0.3 SD increases in the InterAct and EPIC-Norfolk biomarker scores and a 0.08 SD (95% confidence interval: 0.02, 0.15) increase in the biomarker score from the MedLey baseline sample.

Table 5.2 Standardised coefficients of biomarker scores of self-reported Mediterranean Diet Score by sample of derivation*

Biomarker	EPIC-InterAct subcohort										ML	EPIC-Norfolk		
	Subc.	FR	IT	ES	UK-g	UK-h	NL	DE	SE	DK	V1	FFQ	7DD	FFQ_{pyr}
	n=12,495	n=499	n=1,763	n=3,288	n=960	n=213	n=1,367	n=1,826	n=903	n=1,889	n=143	n=5,581	n=5,525	n=5,436
α-carotene	0.09	-	-	0.07	-	-	0.11	0.08	0.20	0.17	-	0.11	0.12	0.06
ß-carotene	-	-	0.12	-	-	-0.11	-	-	-	-0.10	-	-0.04	-0.05	-
β-crypt.	0.09	-	0.06	0.10	0.12	-	0.10	-	0.05	0.07	-	0.10	0.09	0.13
Lycopene	0.04	-	-	0.00	0.08	-	0.06	0.10	0.08	0.08	-	0.08	0.12	0.10
Lut. & zeax	0.06	0.13	0.07	0.04	-	0.24	-	0.04	-	0.07	0.25	0.07	0.06	0.05
C14:0	0.01	-	-	0.04	-	-	-	-	-	-	-	0.08	0.06	0.06
C16:0	0.03	-	-	0.03	-	-	-	-	-	-	-	0.06	-	-
C18:0	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C15:0	-0.06	-	-0.07	-0.11	-0.12	-0.14	-	-	-0.15	-	-	-0.15	-0.13	-
C17:0	-0.04	-	-0.10	-0.08	-	-	-	-	-	0.06	-	-0.07	-0.06	-0.08
C20:0	-	-	-	-	-	0.12	-	-	-	-	-	-	-	-
C22:0	-0.07	-	-	-0.07	-0.23	-	-0.04	-0.11	-	-	-			
C24:0	0.06	-	-	0.02	0.20	-	-	0.09	0.05	-	-	-0.03	-0.02	-
C18:3n-3	0.02	-	-	0.04	-	-	-	-	0.05	-0.06	-	-	-0.02	-
C20:5n-3	0.01	-	-	-	-	-	-	-	-	-	-	0.05	0.1	0.05
C22:5n-3	-	-	-0.06	-0.05	-	-	-	-	-	-0.06	-	-0.07	-0.09	-0.12
C22:6n-3	0.07	-	0.14	0.14	0.11	-	0.08	0.12	0.15	0.10	0.45	0.16	0.15	0.18
C18:2n-6	-	-	-	-	-	-	-	-	-	-	0.29	-	-	-
C20:2n-6	-0.02	-	-	0.02	-	-	-0.06	-	-	-	-	-	-	-0.01
C20:3n-6	0.05	-	0.06	0.05	-	-	-	-	-	0.01	-	-0.03	-0.03	-
C20:4n-6	-	-	-	0.06	-	-	-	-	-	-0.09	-	-	-0.05	-0.06
C22:4n-6	-0.12	-	-0.11	-0.14	-0.16	-	-	-0.06	-	-	-	-	-	-
C22:5n-6	-0.02	-	-	-	-	-	-	-0.08	-0.07	-	-	-0.04	-0.04	-
C18:1n-9t	0.02	-	-0.04	0.01	-	-	-	-	-	-0.04	-			
C16:1	0.01	-	-	-0.09	-	-	-	0.06	-	0.08	-	0.03	0.06	-0.09
C17:1	-0.01	-	-	-0.02	-	-	-0.05	-	-	-	-	-	-	-
C18:1n-9c	0.02	-	-	0.16	-	-	-	-	-	-	0.25	-0.05	-0.07	-
C20:1	0.01	-	-	-0.04	-	-	-	-	-	-	-	-	-	-
C24:1	0.06	-	-	0.08	-	-	-	-	-	0.06	-	-	-	-

Abbreviations: ML – MedLey trial; V1 – first visit (baseline sample); Subc. – subcohort (multi-country score); FR – France; IT – Italy; ES – Spain; UK-g – United Kingdom, general population; UK-h – United Kingdom, health-conscious participants (recruitment targeting a large proportion of vegetarians); NL – the Netherlands; DE – Denmark; SE

– Sweden; DE – Germany, DK – Denmark; FFQ – food frequency questionnaire; 7DD – 7-day food diary; FFQ_{pyr} – Mediterranean Diet Score based on adherence to Mediterranean pyramid developed using FFQ data; n – number of participants in derivation samples; crypt – cryptoxanthin; Lut. & zeax – sum of lutein and zeaxanthin

*Red colour highlights positive coefficients, and blue colour highlights negative coefficients with intensity proportional to value. Blank cells represent variables not used for derivation of a given score. Dashed cells represent variables rejected by the bootstrap-enhanced elastic net regression. Standardised ordinary least squares coefficients are reported for comparability between the biomarker scores. Nutritional biomarkers were natural logarithm transformed.

Biomarker scores were derived using a combination of bootstrap selection stability with elastic net regression and post-selection ridge regression. Biomarker scores were developed by using as the outcome residuals from country-specific linear regressions of Mediterranean Diet Score (estimated from self-reported diet) on personal characteristics. In the MedLey trial, unadjusted values were used due to a homogenous sample. Adjustments included: age at blood draw, sex, recruitment centre, prevalent cancer, cardiovascular disease, hypertension and hyperlipidaemia ; familial history of type 2 diabetes, smoking status (never, former, current smoker), physical activity index (inactive, moderately inactive, moderately active, active), day of the year of the blood draw (restricted cubic spline with four knots), fasting status (<3, 3-6, >6 hours), current use of vitamin or mineral supplements, menopausal status (pre-, peri-, postmenopausal, bilateral ovariectomy), current hormone replacement therapy use, marital status (single, married or cohabiting, divorced or separated, widowed), educational attainment (none, primary school, technical or professional school, secondary school, post-secondary school education), current employment, estimated energy intake, body mass index and waist circumference. The multi-country EPIC-InterAct biomarker score was further adjusted for country prior to averaging of the residuals from country-specific regressions. Participants from the UK-h sample with a large proportion of vegetarians were excluded from derivation of the multi-country score.

Table 5.3 Pearson correlations between biomarker scores of Mediterranean diet and Mediterranean diet score estimated from self-report in EPIC-InterAct subcohort: pooled and country-specific estimates*

Biomarker score	Subcohort	France	Italy	Spain	UK-g	UK-h	Netherlands	Germany	Sweden	Denmark
	n=12,994	n=514	n=1,810	n=3,345	n=986	n=218	n=1,407	n=1,862	n=910	n=1,942
EPIC-InterAct, MDS										
Multi-country	0.32	**0.14**	**0.27**	**0.40**	**0.40**	0.17	**0.30**	**0.27**	**0.41**	**0.42**
Italy	0.30	0.14	**0.31**	0.38	0.37	0.17	0.27	0.21	0.40	0.39
Spain	0.28	0.13	0.29	**0.43**	0.36	0.15	0.23	0.18	0.38	0.34
UK-g	0.28	0.09	0.24	0.34	**0.45**	0.23	0.24	0.23	0.31	0.33
UK-h	0.19	0.10	0.16	0.17	0.23	**0.41**	0.13	0.12	0.24	0.22
Netherlands	0.29	0.08	0.22	0.29	0.36	0.14	**0.37**	0.26	0.37	0.45
Germany	0.29	0.09	0.24	0.28	0.40	0.17	0.27	**0.31**	0.38	0.40
Sweden	0.27	0.08	0.19	0.26	0.37	0.25	0.25	0.24	**0.43**	0.38
Denmark	0.28	0.10	0.24	0.28	0.30	0.15	0.31	0.26	0.37	**0.47**
MedLey trial										
MDS	0.16	0.12	0.17	0.30	0.08	-0.03	0.17	0.13	0.23	0.21
EPIC-Norfolk, MDS										
FFQ	0.29	0.09	0.25	0.31	0.40	0.27	0.25	0.23	0.41	0.38
7DD	0.29	0.10	0.24	0.30	0.40	0.24	0.26	0.25	0.39	0.40
FFQ_{pyr}	0.28	0.09	0.24	0.29	0.38	0.16	0.28	0.24	0.36	0.43

Abbreviations: UK-g – United Kingdom, general population; UK-h – United Kingdom, health-conscious participants (recruitment targeting a large proportion of vegetarians); n – number of participants; MDS – Mediterranean diet score; RCT – randomised controlled trial; FFQ – food frequency questionnaire; 7DD – 7-day food diary; FFQ_{pyr} – Mediterranean Diet Score based on adherence to Mediterranean pyramid developed using FFQ data

*Country-specific estimates were pooled using random-effects meta-analysis to estimate correlation coefficients for the subcohort. Fisher z-transformation was used to obtain standard errors of correlation coefficients prior to pooling. Values in bold indicate that a given sample participated in derivation of a given score. Correlation coefficients in non-subcohort incident type 2 diabetes cases, who did not participate in derivation of biomarker scores, were similar and on average higher by 0.01 (results not shown).

Table 5.4 Medians (interquartile ranges) of nutritional biomarker scores of Mediterranean diet in the MedLey trial post-intervention and EPIC-InterAct subcohort

Biomarker score*	MedLey trial		EPIC-InterAct							
	Hab-diet	Med-diet	France	Italy	Spain	UK	Netherlands	Germany	Sweden	Denmark
	n=65	n=68	n=529	n=1,910	n=3,423	n=1,230	n=1,426	n=1,890	n=924	n=1,981
Observational scores										
Standardised values										
EPIC-InterAct score, MDS	0.01	0.55	0.41	0.20	0.17	-0.03	-0.67	-0.20	0.26	0.15
	(-0.66, 0.56)	(-0.68, 1.17)	(-0.18, 0.95)	(-0.38, 0.77)	(-0.48, 0.81)	(-0.66, 0.67)	(-1.37, 0.05)	(-0.79, 0.35)	(-0.33, 0.84)	(-0.56, 0.84)
MedLey trial score, MDS	-0.06	0.73	0.02	0.37	0.45	-0.24	-0.89	-0.41	0.31	0.17
	(-0.67, 0.70)	(-0.02, 1.40)	(-0.59, 0.60)	(-0.23, 0.98)	(-0.43, 1.17)	(-0.83, 0.32)	(-1.48, -0.31)	(-0.93, 0.07)	(-0.09, 0.73)	(-0.26, 0.61)
EPIC-Norfolk, MDS-pyr.	-0.02	0.39	0.38	0.31	0.69	-0.22	-0.56	-0.25	-0.23	-0.41
	(-0.66, 0.61)	(-0.50, 1.41)	(-0.09, 0.92)	(-0.25, 0.84)	(0.11, 1.22)	(-0.78, 0.37)	(-1.18, 0.05)	(-0.85, 0.29)	(-0.80, 0.37)	(-1.12, 0.26)
Linear predictions										
EPIC-InterAct score, MDS	7.7	8.1	9.0	8.9	8.9	8.8	8.4	8.7	8.9	8.9
	(7.2, 8.1)	(7.2, 8.6)	(8.7, 9.3)	(8.6, 9.2)	(8.5, 9.3)	(8.4, 9.2)	(8.0, 8.8)	(8.3, 9.0)	(8.6, 9.3)	(8.4, 9.3)
MedLey trial score, MDS	9.5	10.5	4.3	5.0	5.2	3.7	2.3	3.3	4.9	4.6
	(8.7, 10.4)	(9.5, 11.3)	(2.9, 5.5)	(3.7, 6.3)	(3.3, 6.7)	(2.4, 4.9)	(1.1, 3.5)	(2.2, 4.4)	(4.0, 5.8)	(3.6, 5.5)
EPIC-Norfolk, MDS-pyr.	8.0	8.2	8.8	8.8	8.9	8.5	8.4	8.5	8.5	8.4
	(7.7, 8.3)	(7.8, 8.7)	(8.6, 9.0)	(8.5, 9.0)	(8.7, 9.2)	(8.3, 8.8)	(8.1, 8.6)	(8.2, 8.7)	(8.3, 8.8)	(8.1, 8.7)

n – number of participants; RCT – randomised controlled trial; Hab – habitual; Med – Mediterranean

*Standardised values are study specific. Med-/hab-diet probabilities are predicted probabilities of assignment to Med-diet or Hab-diet arms of the MedLey trial.

Table 5.5 Differences in standardised means of nutritional biomarker scores between the Mediterranean and habitual diet groups in the MedLey trial at 6 months

Biomarker score	Difference (95% CI)	Q-value	C-statistic
Derived in EPIC-InterAct, MDS			
Multi-country	0.49 (0.16, 0.82)	0.020	0.63
Italy	0.63 (0.31, 0.95)	0.005	0.67
Spain	0.60 (0.27, 0.93)	0.009	0.67
UK – general population	0.50 (0.17, 0.83)	0.020	0.63
UK – health-conscious	-0.12 (-0.44, 0.20)	0.999	0.53
Netherlands	0.34 (0.04, 0.65)	0.105	0.58
Germany	0.49 (0.18, 0.80)	0.020	0.64
Sweden	0.53 (0.21, 0.86)	0.018	0.64
Denmark	0.38 (0.05, 0.70)	0.096	0.59
Derived in EPIC-Norfolk, MDS			
FFQ	0.39 (0.07, 0.70)	0.075	0.61
7DD	0.41 (0.09, 0.72)	0.061	0.61
FFQ_{pyr}	0.49 (0.17, 0.81)	0.020	0.62
Derived in the MedLey trial			
Baseline sample, MDS	0.61 (0.28, 0.94)	<0.001	0.69

Abbreviations: CI – confidence interval; MDS – Mediterranean diet score; FFQ – food frequency questionnaire; 7DD – 7-day food diary; FFQ_{pyr} – Mediterranean Diet Score based on adherence to the Mediterranean pyramid developed using FFQ data

Table 5.6 Mean differences in standardised biomarker scores of the Mediterranean diet between the habitual diet group and tertiles of the Mediterranean Diet Score in the Mediterranean diet group in the MedLey trial after 6 months of intervention*

Biomarker score	Standardised difference (95 % CI)			p trend
	Tertile 1	Tertile 2	Tertile 3	
	n =21	n=23	n=24	
(Median MDS points, range 0-18)	13	14	15	
Derived in EPIC-InterAct, MDS				
Multi-country	0.26 (-0.21, 0.73)	0.53 (0.06, 1.00)	0.75 (0.29, 1.22)	0.133
Italy	0.32 (-0.13, 0.78)	0.62 (0.18, 1.07)	1.05 (0.60, 1.50)	0.012
Spain	0.33 (-0.15, 0.80)	0.58 (0.11, 1.06)	0.99 (0.52, 1.46)	0.040
UK – general population	0.16 (-0.31, 0.63)	0.50 (0.03, 0.96)	0.81 (0.34, 1.27)	0.032
UK – health-conscious	-0.22 (-0.67, 0.23)	-0.16 (-0.60, 0.28)	0.16 (-0.27, 0.60)	0.160
Netherlands	0.22 (-0.22, 0.66)	0.39 (-0.05, 0.82)	0.46 (0.02, 0.89)	0.459
Germany	0.27 (-0.17, 0.72)	0.55 (0.12, 0.99)	0.68 (0.25, 1.12)	0.175
Sweden	0.36 (-0.10, 0.83)	0.51 (0.04, 0.97)	0.78 (0.32, 1.24)	0.216
Denmark	0.20 (-0.27, 0.67)	0.39 (-0.08, 0.85)	0.62 (0.16, 1.09)	0.212
Derived in EPIC-Norfolk, MDS				
FFQ	0.20 (-0.26, 0.65)	0.40 (-0.05, 0.85)	0.66 (0.21, 1.11)	0.139
7DD	0.21 (-0.25, 0.66)	0.43 (-0.02, 0.88)	0.65 (0.20, 1.10)	0.159
FFQ_{pyr}	0.35 (-0.11, 0.82)	0.46 (0.00, 0.92)	0.77 (0.31, 1.23)	0.181
Derived in the MedLey trial				
Baseline sample, MDS	0.27 (-0.19, 0.73)	0.69 (0.24, 1.15)	1.06 (0.61, 1.51)	0.006

Abbreviations: CI – confidence interval; MDS – Mediterranean diet score; FFQ – food frequency questionnaire; 7DD – 7-day food diary; FFQ_{pyr} – Mediterranean Diet Score based on adherence to the Mediterranean pyramid developed using FFQ data

*Mixed linear models were used to estimate standardised changes. Standardised values were calculated using baseline means and standard deviations of Mediterranean Diet Score (MDS) and biomarker scores. Follow-up diet was measured at 2- and 4-months post-randomisation and nutritional biomarkers were measured 3- and 6-months post randomisation. Tertiles of MDS were derived using best unbiased linear predictions in participants who completed at least 2 out of 3 food diaries or completed the last food diary.

Example interpretation: Compared to the habitual diet group, participants randomised to the Mediterranean diet intervention in the lowest tertile of self-reported adherence to the Mediterranean diet had values of the multi-country InterAct score on average higher by 0.26 standard deviation (SD). Participants in the middle and upper tertile had values higher by 0.53 and 0.75 SD, respectively, than the habitual diet group. There was no evidence to suggest the presence of a dose-response across these thirds (p trend = 0.133).

Table 5.7 Longitudinal associations between nutritional biomarker scores of Mediterranean diet and self-reported Mediterranean Diet Score in the MedLey trial, stratified by trial arm*

Biomarker score	Standardised coefficient of change (95% CI)		*P* interaction
	Mediterranean diet	Habitual diet	
	n=68	n=63	
Derived in EPIC-InterAct, MDS			
Multi-country	0.12 (0.06, 0.18)	0.03 (-0.04, 0.11)	0.071
Italy	0.18 (0.11, 0.26)	0.11 (0.02, 0.21)	0.261
Spain	0.19 (0.13, 0.25)	0.07 (-0.00, 0.14)	0.008
UK – general population	0.15 (0.08, 0.22)	0.04 (-0.05, 0.12)	0.034
UK – health-conscious	0.19 (0.09, 0.28)	0.02 (-0.09, 0.13)	0.027
Netherlands	0.03 (-0.04, 0.09)	0.01 (-0.07, 0.09)	0.725
Germany	0.05 (-0.03, 0.13)	0.03 (-0.07, 0.12)	0.674
Sweden	0.09 (0.02, 0.16)	0.02 (-0.07, 0.11)	0.215
Denmark	0.11 (0.04, 0.17)	-0.00 (-0.08, 0.07)	0.030
Derived in EPIC-Norfolk, MDS			
FFQ	0.10 (0.04, 0.16)	0.01 (-0.07, 0.08)	0.057
7DD	0.09 (0.02, 0.15)	-0.00 (-0.08, 0.08)	0.096
FFQ_{pyr}	0.12 (0.05, 0.19)	0.03 (-0.05, 0.12)	0.115
Derived in the MedLey trial			
Baseline sample, MDS	0.21 (0.14, 0.28)	0.16 (0.07, 0.24)	0.339

Abbreviations: CI – confidence interval; MDS – Mediterranean diet score; FFQ – food frequency questionnaire; 7DD – 7-day food diary; FFQ_{pyr} – Mediterranean Diet Score based on adherence to Mediterranean pyramid developed using FFQ data

*Mixed linear models were used to estimate standardised changes. Standardised values were calculated using baseline means and standard deviations of Mediterranean Diet Score (MDS) and biomarker scores. Follow-up diet was measured at 2- and 4-months post-randomisation using 3-day weighed food diaries and nutritional biomarkers were measured 3- and 6-months post randomisation. Month 2 MDS and month 3 biomarker scores, and month 4 MDS and month 6 biomarker scores were fitted as collected at the same study visits.

Example interpretation: One standard deviation (SD) increase in self-reported MDS was associated with a 0.12 SD increase in the multi-country InterAct score in the Mediterranean diet group.

Table 5.8 Pearson correlation coefficients (minimum, maximum country-specific values in EPIC-InterAct) between biomarker scores of Mediterranean diet, Mediterranean Diet Score, and its energy-adjusted food components*

MDS or MDS component, g/1,000 kcal	Mean intake (SD)	MedLey trial baseline MDS biomarker score	EPIC-InterAct MDS biomarker score	EPIC-Norfolk MDS-pyr. biomarker score
MedLey trial, baseline sample (n = 133)				
MDS (points, max. 18)	9.6 (2.6)	0.51	0.22	0.25
Vegetable	81 (64)	0.30	0.04	0.16
Legumes	8.4 (17.5)	0.02	0.00	-0.01
Fruits and nuts	138 (75)	0.16	0.01	0.15
Cereals	71 (39)	0.13	0.13	0.13
Fish and seafood	22 (26)	0.09	0.03	0.03
Meat and meat products	34 (31)	0.09	0.13	0.07
Dairy	130 (94)	-0.15	0.02	0.01
Olive oil	4.1 (7.7)	0.22	0.06	0.10
Alcohol (g)	9.2 (10.6)	0.00	0.12	-0.06
MedLey trial, Mediterranean diet group post-intervention (n = 68)				
MDS (points, max. 18)	13.8 (1.9)	0.18	0.30	0.33
Vegetable	113 (56)	0.18	0.14	0.15
Legumes	23 (24)	-0.02	0.17	0.18
Fruits and nuts	211 (100)	0.21	0.13	0.22
Cereals	72 (26)	0.16	0.12	0.11
Fish and seafood	40 (26)	0.25	0.11	0.19
Meat and meat products	23 (24)	-0.01	-0.05	0.00
Dairy	161 (65)	0.05	0.15	0.12
Olive oil	34 (18)	0.01	-0.18	-0.17
Alcohol (g)	8.1 (8.1)	-0.10	0.16	0.09

MDS or MDS component, g/1,000 kcal	Mean intake (SD)	MedLey trial baseline MDS biomarker score		EPIC-InterAct MDS biomarker score		EPIC-Norfolk MDS-pyr. biomarker score	
EPIC-InterAct subcohort (n = 12,625)							
MDS (points, max. 18)	8.8 (3.1)	0.16	(-0.04, 0.30)	0.31	(0.13, 0.42)	0.28	(0.12, 0.43)
Vegetable	94 (65)	0.11	(0.05, 0.16)	0.24	(0.11, 0.36)	0.22	(0.09, 0.36)
Legumes	9.1 (13.2)	-0.03	(-0.15, 0.03)	-0.02	(-0.13, 0.07)	-0.06	(-0.22, 0.04)
Fruits and nuts	122 (99)	0.09	(0.05, 0.16)	0.22	(0.07, 0.35)	0.22	(0.07, 0.31)
Cereals	104 (41)	-0.04	(-0.17, 0.10)	0.03	(-0.08, 0.15)	0.05	(-0.08, 0.19)
Fish and seafood	18 (16)	0.17	(0.13, 0.36)	0.24	(0.20, 0.38)	0.19	(0.12, 0.32)
Meat and meat products	52 (24)	-0.02	(-0.12, 0.35)	-0.03	(-0.18, 0.33)	-0.04	(-0.20, 0.30)
Dairy	163 (114)	-0.02	(-0.08, 0.19)	-0.04	(-0.14, 0.05)	-0.02	(-0.15, 0.05)
Olive oil	9.1 (13.8)	0.09	(-0.13, 0.45)	0.06	(-0.16, 0.28)	0.03	(-0.15, 0.23)
Alcohol (g)	13.4 (18.7)	0.09	(-0.01, 0.18)	0.09	(-0.00, 0.22)	-0.05	(-0.19, 0.05)

Abbreviations: MDS – Mediterranean Diet Score; MDS-pyr. – MDS-pyramid

*Values in the EPIC-InterAct sample are pooled estimates of country-specific correlations. Fisher z-transformation was used to obtain standard errors of correlation coefficients for pooling via random-effects meta-analysis. Red colour highlights positive coefficients, and blue colour highlights negative coefficients with intensity proportional to value.

Table 5.9 Longitudinal associations between Mediterranean Diet Score and biomarker scores of Mediterranean diet in a subset of EPIC-Norfolk (n=432)*

Biomarker score and model†	Standardised coefficient of change (95% CI)
EPIC-InterAct score, MDS	
Age, follow-up time and sex adjusted	0.35 (0.29, 0.41)
Multivariable	0.30 (0.24, 0.36)
MedLey trial baseline, MDS	
Age, time, sex and energy-adjusted	0.10 (0.04, 0.16)
Multivariable	0.08 (0.02, 0.15)
EPIC-Norfolk, MDS-pyramid‡	
Age, time, sex and energy-adjusted	0.32 (0.26, 0.39)
Multivariable	0.27 (0.20, 0.33)

*432 participants had baseline and follow-up data on biomarker scores and MDS. In the multivariable model, 11% of participants had missing data. Results are presented for complete-case analysis. Mean follow-up time was 3.3 years (standard deviation 0.6, minimum-maximum: 2.2-5.7 years).

†Adjustment for age and time included linear and squared-terms. The multivariable model for was adjusted for the following covariates: age, follow-up time, sex, estimated energy intake, prevalent cancer, cardiovascular disease, hypertension and hyperlipidaemia; familial history of T2D, smoking status (never, former, current smoker), physical activity index (inactive, moderately inactive, moderately active, active), seasonality (sine and cosine function of the day of the year), fasting status (<3, 3-6, >6 hours), current use of vitamin or mineral supplements, menopausal status (pre-, peri-, postmenopausal, bilateral ovariectomy), current hormone replacement therapy use, marital status (single, married or cohabiting, divorced or separated, widowed), educational attainment (none, primary school, technical or professional school, secondary school, post-secondary school education), current employment, body mass index and waist circumference.

‡Coefficients estimated using the self-reported MDS-pyramid score were 0.32 (0.29, 0.42) and 0.29 (0.22-0.35), respectively, in the age, time, sex and energy-adjusted model, and the multivariable mode.

5.5.2 Associations of biomarker scores with incident T2D

Ten of the 13 biomarker scores of MDS were inversely associated with incident T2D in EPIC-InterAct (**Table 5.10**). The biomarker scores derived in Italian participants of EPIC-InterAct and British participants of EPIC-Norfolk were not associated with the outcome. In turn, the EPIC-Norfolk biomarker score of the MDS-pyramid had an inverse association. Biomarker scores inversely related to incidence of T2D had HR per 1 SD of approximately ~0.8-0.9 with high relative heterogeneity (I^2 ≥55%), and HR ~0.65-0.75 in comparisons of top versus bottom fifths of subcohort distributions of the scores. These associations were robust to multiple sensitivity analyses, independent from the influence of biomarkers constituting the scores, modified by baseline age and use of supplements without impacting statistical significance or directionality of HRs within strata, and the complete-case analysis yielded materially similar results as the primary multiply-imputed analysis (**Appendix 5.1-5.3**).

In terms of patterns of country-specific linear associations, the multi-country InterAct biomarker score had an inverse relationship with incident T2D in all datasets but Denmark. Among country-specific scores, the Swedish and Danish scores also had inverse associations in all but one country, and there was less consistency for the remaining scores (**Figures 5.1a and 5.1b**). Inverse associations were observed in all countries of derivation of a given biomarker score, except for Italy. The biomarker scores of MDS derived in non-InterAct participants of the EPIC-Norfolk study had inverse associations with incident T2D in EPIC-InterAct only in British participants recruited from the general population (**Figure 5.2**). By contrast, the biomarker score of the MDS-pyramid derived in EPIC-Norfolk had inverse associations in all but two countries. Only the biomarker score of MDS derived in the baseline sample of the MedLey trial had a 95% prediction interval which did not include the null for the prospective association with T2D in EPIC-InterAct (0.69-0.99).

The literature-based WHI biomarker score of the alternative Mediterranean diet calculated with omission of biomarkers unavailable in EPIC-InterAct was inversely associated with incident T2D in EPIC-InterAct (HR, 95% CI per 1 SD = 0.76, 0.69-0.85).

Within subsamples of EPIC-Norfolk with additional assays of circulating carotenoids and fatty acids, associations of the biomarker score of the MDS-pyramid were consistently inverse across the three sets of biomarkers (**Appendix 5.4**). In turn, the EPIC-Norfolk biomarker scores of MDS had inverse associations only in the EPIC-InterAct subsample of EPIC-Norfolk,

though the remaining subsamples had relatively limited numbers of cases, and the point estimates were consistently below the null.

Table 5.10 Associations of biomarker scores of Mediterranean diet scores with incidence of type 2 diabetes in EPIC-InterAct (n = 22,202): pooled Hazard Ratios (95% CI)*

Biomarker score	Quintiles					p_{trend}	Per 1 SD	Q-value†	I^2, % (95% CI)
	Q1	Q2	Q3	Q4	Q5				
InterAct scores									
Multi-country	1.0 (Ref.)	0.86 (0.73-1.00)	0.78 (0.69-0.87)	0.64 (0.52-0.79)	0.66 (0.56-0.78)	<0.001	0.82 (0.76-0.89)	<0.001	68 (34-85)
Italy	1.0 (Ref.)	0.79 (0.64-0.98)	0.94 (0.78-1.12)	0.81 (0.69-0.95)	0.86 (0.73-1.02)	0.042	0.94 (0.87-1.01)	0.261	68 (33-85)
Spain	1.0 (Ref.)	0.73 (0.55-0.97)	0.75 (0.62-0.91)	0.64 (0.51-0.82)	0.69 (0.58-0.81)	<0.001	0.85 (0.79-0.91)	<0.001	60 (13-82)
UK-GP	1.0 (Ref.)	0.79 (0.68-0.93)	0.80 (0.64-1.01)	0.79 (0.68-0.91)	0.75 (0.60-0.93)	<0.001	0.89 (0.83-0.96)	0.016	68 (33-85)
UK-HC	1.0 (Ref.)	0.91 (0.78-1.07)	0.76 (0.64-0.90)	0.74 (0.60-0.91)	0.72 (0.58-0.88)	<0.001	0.86 (0.79-0.93)	0.002	70 (41-85)
Netherlands	1.0 (Ref.)	0.81 (0.71-0.93)	0.77 (0.66-0.90)	0.75 (0.65-0.87)	0.65 (0.53-0.79)	<0.001	0.82 (0.75-0.89)	<0.001	77 (54-88)
Germany	1.0 (Ref.)	0.87 (0.73-1.03)	0.78 (0.65-0.93)	0.77 (0.65-0.91)	0.69 (0.56-0.83)	<0.001	0.86 (0.80-0.92)	<0.001	71 (41-86)
Sweden	1.0 (Ref.)	0.73 (0.55-0.97)	0.75 (0.62-0.91)	0.64 (0.51-0.82)	0.69 (0.58-0.81)	<0.001	0.81 (0.75-0.87)	<0.001	74 (48-87)
Denmark	1.0 (Ref.)	0.84 (0.72-0.99)	0.75 (0.66-0.86)	0.70 (0.61-0.80)	0.61 (0.53-0.70)	<0.001	0.81 (0.75-0.87)	<0.001	65 (26-84)
EPIC-Norfolk scores									
FFQ	1.0 (Ref.)	0.99 (0.82-1.20)	0.97 (0.82-1.16)	0.88 (0.74-1.05)	1.05 (0.91-1.21)	0.510	0.96 (0.89-1.03)	0.817	65 (28-64)
7-day diary	1.0 (Ref.)	1.01 (0.79-1.28)	0.93 (0.76-1.14)	0.94 (0.78-1.13)	0.99 (0.82-1.20)	0.925	0.96 (0.89-1.03)	0.817	77 (55-89)
MDS-pyramid (FFQ)	1.0 (Ref.)	0.78 (0.60-1.00)	0.72 (0.59-0.88)	0.62 (0.48-0.79)	0.60 (0.46-0.77)	0.002	0.79 (0.72-0.87)	<0.001	74 (50-87)
MedLey trial baseline	1.0 (Ref.)	0.87 (0.77-0.97)	0.72 (0.60-0.87)	0.71 (0.60-0.83)	0.65 (0.49-0.87)	0.098	0.83 (0.77-0.89)	<0.001	55 (1-80)

Abbreviations: CI – confidence interval; FFQ – food frequency questionnaire; MDS – Mediterranean Diet Score; SD – standard deviation; UK-GP – United Kingdom, general population; UK-HC – United Kingdom, health-conscious participants

*Associations were adjusted for: age (as timescale), sex, recruitment centre, prevalent cancer, cardiovascular disease, hypertension and hyperlipidaemia; familial history of type 2 diabetes, smoking status (never, former, current smoker), physical activity index (inactive, moderately inactive, moderately active, active), seasonality (sine and cosine function of the day of the year), fasting status (<3, 3-6, >6 hours), current use of vitamin or mineral supplements, marital status (single, married or cohabiting, divorced or separated, widowed), educational attainment (none, primary school, technical or professional school, secondary school, post-secondary school education), current employment, body mass index and waist circumference, and in women, menopausal status (pre-, peri-, postmenopausal, bilateral oophorectomy), current hormone replacement therapy use. There were 9,453 incident type 2 diabetes cases, including 564 cases overlapping with randomly selected subcohort participants as a feature of the case-cohort design.

†False discovery rate-corrected P values

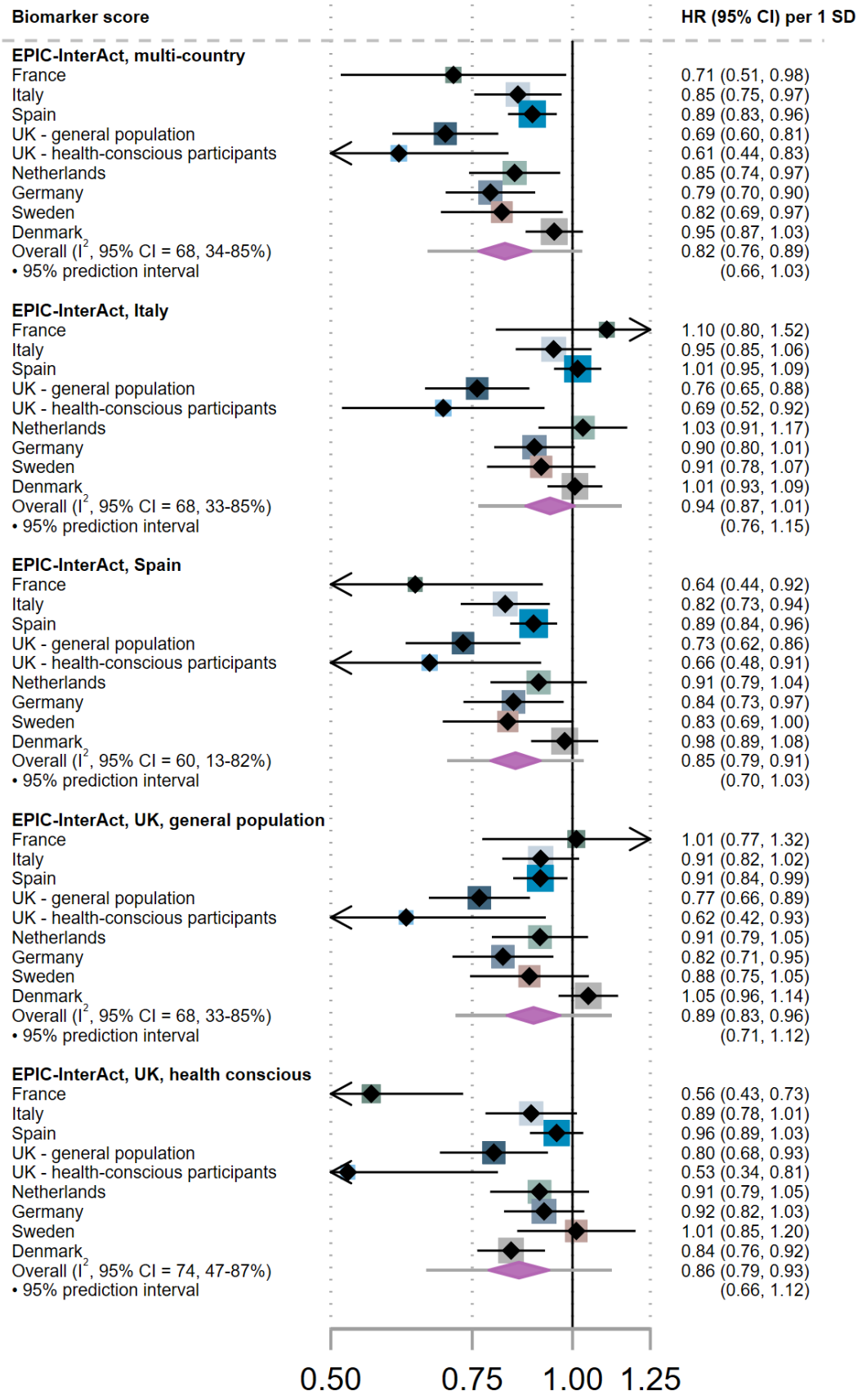


Figure 5.1a Associations between internally derived nutritional biomarker scores of the Mediterranean diet and incidence of type 2 diabetes in the EPIC-InterAct case-cohort study

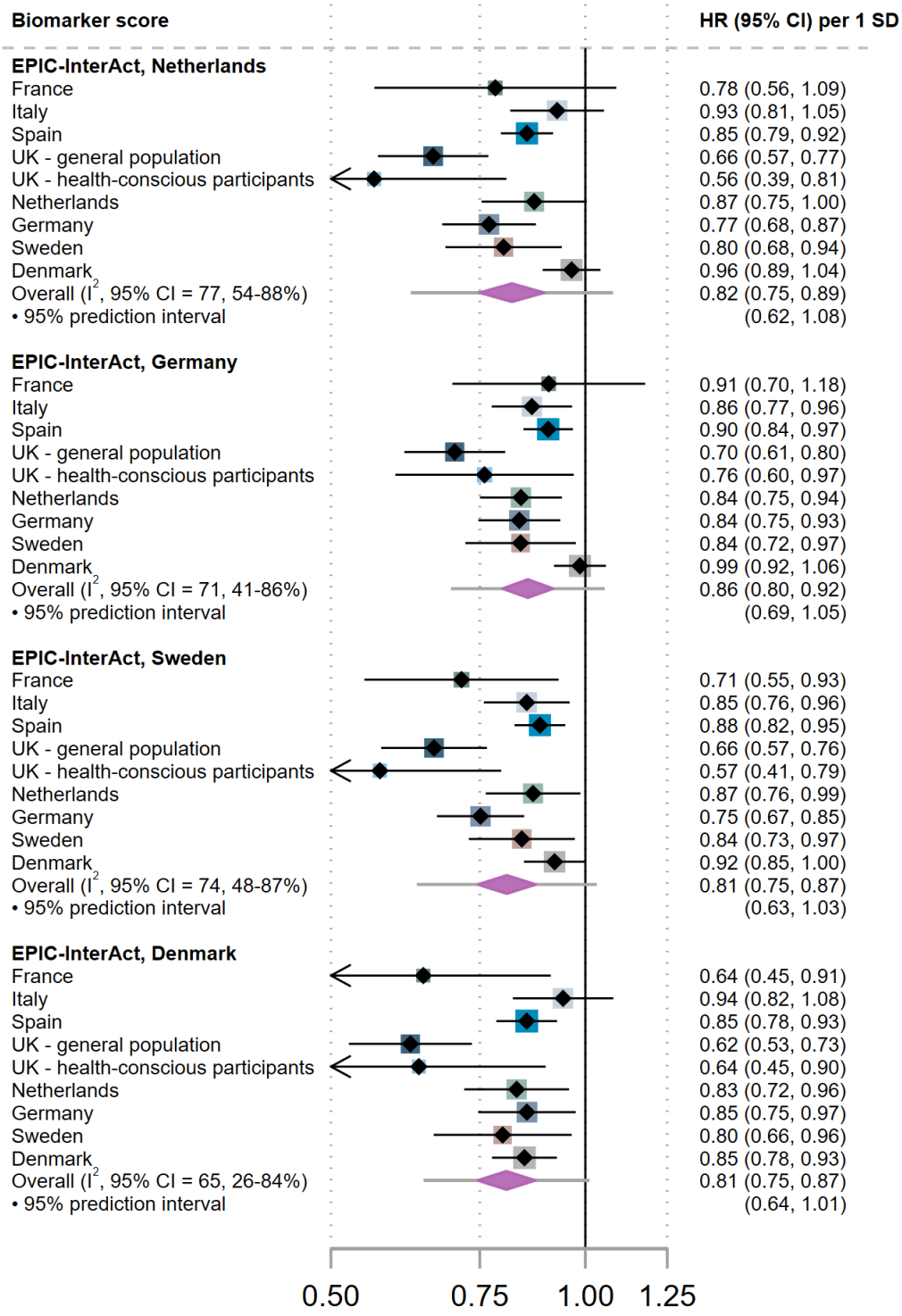


Figure 5.1b Associations between internally derived nutritional biomarker scores of the Mediterranean diet and incidence of type 2 diabetes in the EPIC-InterAct case-cohort study

Associations were adjusted for: age (as timescale), sex, recruitment centre, prevalent cancer, cardiovascular disease, hypertension and hyperlipidaemia; familial history of type 2 diabetes, smoking status (never, former, current smoker), physical activity index (inactive, moderately inactive, moderately active, active), seasonality (sine and cosine function of the day of the year), fasting status (<3, 3-6, >6 hours), current use of vitamin or mineral supplements, marital status (single, married or cohabiting, divorced or separated, widowed), educational attainment (none, primary school, technical or professional school, secondary school, post-secondary school education), current employment, body mass index and waist circumference, and in women, menopausal status (pre-, peri-, postmenopausal, bilateral oophorectomy), current hormone replacement therapy use.

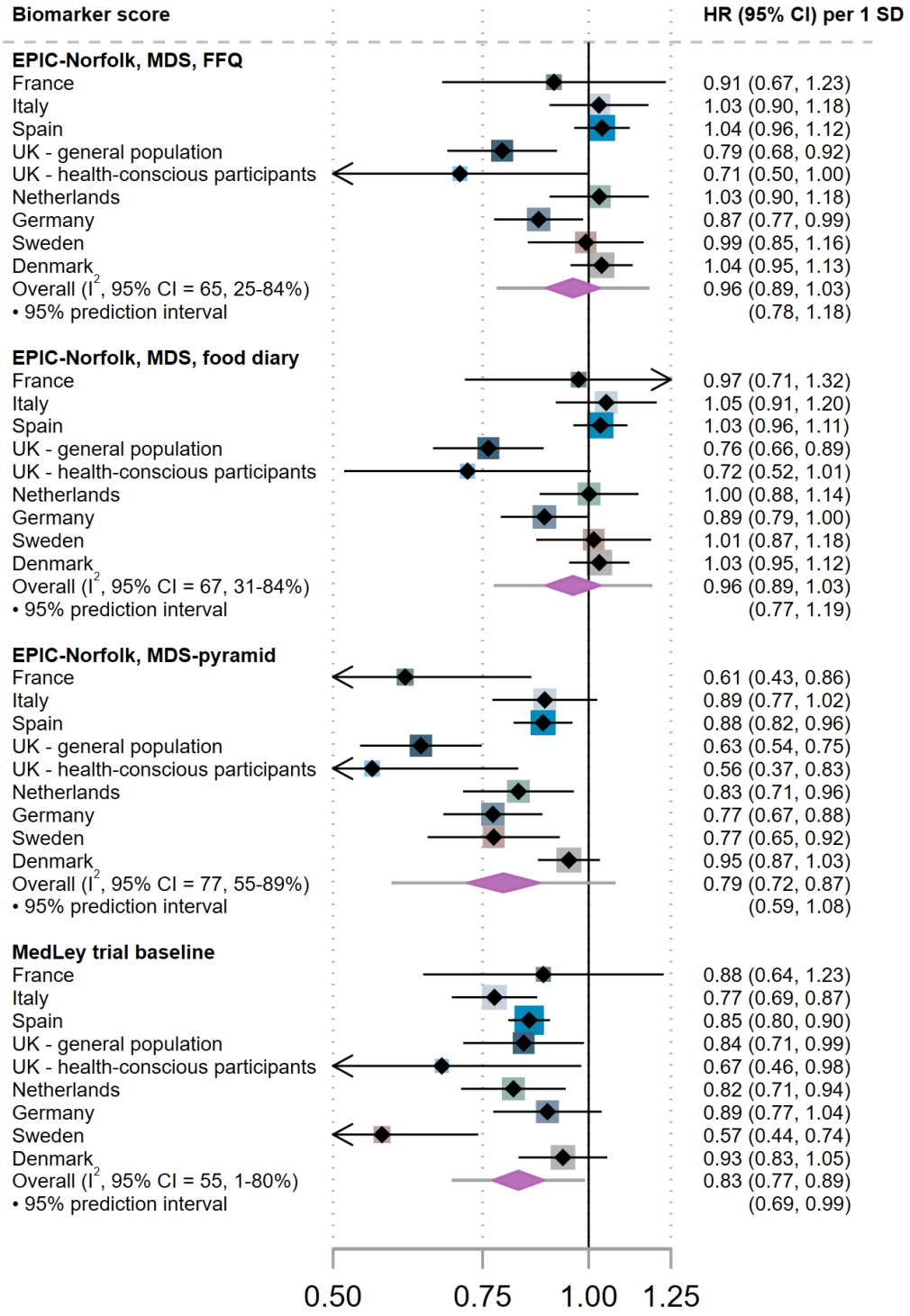


Figure 5.2 Associations between externally derived nutritional biomarker scores of the Mediterranean diet and incidence of type 2 diabetes in the EPIC-InterAct case-cohort study

Associations were adjusted for: age (as timescale), sex, recruitment centre, prevalent cancer, cardiovascular disease, hypertension and hyperlipidaemia; familial history of type 2 diabetes, smoking status (never, former, current smoker), physical activity index (inactive, moderately inactive, moderately active, active), seasonality (sine and cosine function of the day of the year), fasting status (<3, 3-6, >6 hours), current use of vitamin or mineral supplements, marital status (single, married or cohabiting, divorced or separated, widowed), educational attainment (none, primary school, technical or professional school, secondary school, post-secondary school education), current employment, body mass index and waist circumference, and in women, menopausal status (pre-, peri-, postmenopausal, bilateral oophorectomy), current hormone replacement therapy use.

5.6 Discussion

In the current research I tested validity of observationally-derived biomarker scores of the Mediterranean diet against an objective criterion of detecting differences between participants randomised to a dietary intervention with the Mediterranean diet or continuation of habitual diet. Eight of 13 biomarker scores considered met this criterion. By contrast, biomarker scores of aHEI-2010 and the DASH diet did not differ between the trial arms, thus supporting specificity of the biomarker scores of the Mediterranean diet to this dietary pattern. However, the magnitude of the differences in the scores of the Mediterranean diet and their discriminatory performance was similar to that of the best-performing individual biomarkers.

Ten of the 13 biomarker scores were inversely associated with incidence of T2D in EPIC-InterAct. Within the three biomarker which had null pooled results in EPIC-InterAct, two were derived externally in EPIC-Norfolk based on MDS estimated from FFQ and food diary, and they did not pass the trial validation benchmark. Interestingly, these two scores were inversely associated with incident T2D in EPIC-InterAct only in British participants recruited from the general population (Norfolk and Oxford recruitment centres), but not in any other meta-analytically pooled sample. This contrasts with the results for the biomarker score derived internally in the UK general population sample of InterAct, which both passed the trial validation criterion and was inversely related to incident T2D. Overall, generalisability of the biomarker scores as biomarkers of the Mediterranean diet and inverse correlates of new-onset T2D was high. However, correlation coefficients between the biomarker scores and components of the Mediterranean diet were heterogeneous across the biomarker scores and study samples.

5.6.1 Strengths and limitations

Strengths and limitations enumerated for biomarker score investigations in the MedLey and EPIC-InterAct studies in Chapters 3 and 4 equally apply to the current work. A strength specific to the research presented in this chapter was the derivation of biomarker scores in multiple datasets, some of which were external to the EPIC-InterAct study. The analyses were controlled for false-discovery rate, both at the stage of validation of biomarker scores and testing their associations with incident T2D. Evaluation of performance of the biomarker scores in the MedLey trial allowed to objectively evaluate their validity to a degree, however, it was

insufficient to determine whether the scores were valid and specific constructs of the Mediterranean diet. Moreover, I was not able to evaluate within the current analytical framework whether the biomarker score followed a classical measurement error, which is an important criterion of biomarker validity for applications to aetiological investigations.[226]

Among other limitations, the MedLey trial experienced moderate, non-differential drop-out rates between randomisation and end-of-trial assessment (20% in each arm, including missing biomarker data). I used mixed linear modelling for validation of biomarker scores which provides unbiased estimates under the assumption of missingness at random.[293] The analysis could be further strengthened by additional sensitivity analyses with multiple imputation approaches to test the robustness of this assumption.[263] The prospective analysis in EPIC-InterAct could be enhanced by formally testing whether the strength of association with incidence of T2D differed between nutritional biomarker scores.[36]

5.6.2 What this study adds and implications of this research

The measurement error of dietary self-report has motivated researchers to use complimentary, biomarker-based measures for evaluating diet-disease associations. There is a growing interest in the research community to apply objective exposure assessment to investigations of lifestyle risk factors at large for improved inference and enhancement of strategies of prevention of non-communicable diseases.[51,309,310] Until recently, biomarkers have been considered for exposure assessment in nutritional epidemiology on individual basis, and mostly in relation to single nutrients.[311] The field has since advanced towards application of multivariable statistical methods to characterise exposure to foods and nutrients by combining data from multiple analytes.[51] Feeding trials have demonstrated that such composite biomarkers can be developed with satisfactory performance for dietary patterns using nutritional biomarkers[178] or metabolomics.[50] However, to my knowledge no study to date has evaluated the validity of this approach when derivation of the biomarker model is based on subjective reporting. The current research suggests that observationally-derived biomarker scores of the Mediterranean diet can be valid measures of adherence to this dietary pattern. It also provides some evidence in support of their specificity to the Mediterranean diet by comparison with biomarker scores of aHEI-2010 and the DASH diet. The binary randomised assignment in the MedLey trial was the only source of objective information on dietary intakes which is insufficient to objectively evalute performance of a biomarker.[226] Feeding trials of the Mediterranean diet with several levels of

adherence are warranted to address this limitation.[50] Broader ranges of nutritional biomarkers, 'omics' approaches or combining these two sources of information should also be considered by future research. A methodological analytical contribution of this work to the field is the observation that computationally intensive bootstrap-enhanced elastic net regression did not outperform elastic net regression without bootstrapping for derivation of biomarker scores.

Most of the biomarker scores derived in this research passed the trial validation criterion and were inversely associated with incident T2D, suggesting that they were largely generalisable to external settings, both as biomarkers of the dietary pattern and risk factors for T2D. However, the associations were more consistently inverse within the countries of derivation than in the remaining datasets. Overall, 9 out of 12 country-specific biomarker scores considered had inverse pooled associations. This is a substantially higher proportion than in an investigation on generalisability of exploratory dietary patterns (based on dietary self-report) in EPIC-InterAct where only 3 out of 18 country-specific patterns had inverse associations.[312] Generalisability has been a longstanding concern in the field of data-driven methods to derivation of novel dietary patterns.[35,36] Some parallels can be drawn to limitations of development of biomarker scores. Namely, both processes seek to identify meaningful linear combinations of foods or biomarkers that aim to represent an underlying dietary pattern. In doing so, they use the interrelatedness of dietary intakes and biomarkers to train an algorithm. The undesirable consequence may be an overreliance on the data structure in the derivation sample and limited generalisability to external datasets.

5.6.3 Conclusions

Biomarker scores of the Mediterranean diet derived based on prediction of dietary self-report may be valid measures of adherence to the Mediterranean dietary pattern and largely generalisable inverse correlates of T2D risk. Further research is needed to confirm validity of biomarker scores derived in cross-sectional designs based on dietary self-report as biomarkers of the Mediterranean diet.

Chapter 6

Objective epidemiological assessment of the Mediterranean diet: nutritional biomarkers, metabolomics, or both?

Abstract

Background: Comparative performance of nutritional and metabolomic biomarkers for epidemiological assessment of the Mediterranean diet has not been evaluated. It is also unclear which groups of nutritional biomarkers should be used for this purpose.

Methods: I derived biomarker scores based on prediction of the Mediterranean Diet Score for adherence to pyramid-based guidelines (MDS-pyramid) and tested their associations with incident cardiovascular disease (CVD), cancer, type 2 diabetes (T2D), and mortality in subsets of participants of the Norfolk arm of the European Prospective Investigation into Cancer and Nutrition study (EPIC-Norfolk). Out of 25,639 participants, data were available in up to 7,017 individuals for nutritional biomarkers, 11,518 for untargeted metabolomics ($n_{metabolites}$ = 958) and 4,212 for both groups of biomarkers combined and following exclusions of 5% of participants with missing covariate data. Nutritional biomarkers included 5 plasma carotenoid variables and 22 phospholipid fatty acids (base set). Further sets of nutritional biomarkers were available for combining with the base set in samples between 583-6,882 participants, including urinary (spot sample) and serum cations and phytoestrogens, plasma vitamin C and tocopherols, iron status biomarkers, carbon and nitrogen stable isotopes, and urinary sugars. Performance of the scores was evaluated by calculating cross-validated Pearson correlation with MDS-pyramid. Additionally, a partial-feeding randomised controlled trial, the MedLey Trial, was used to test whether adding urinary sodium and potassium to the base set of biomarker predictors in EPIC-Norfolk results in greater differences in biomarker score values compared to the base set alone between participants after 6 months of a Mediterranean diet intervention or continuation of habitual diet (n = 131 out of 166 randomised).

Findings Biomarker score derived using the base set of nutritional biomarkers (blood carotenoids and fatty acids) had a moderate correlation with MDS-pyramid (cross-validated r = 0.40). Adding further sets of biomarkers did not impact meaningfully on the correlation or resulted in a decreased performance (r range: 0.30-0.41). The metabolomic and the joint

nutritional-metabolomic scores had cross-validated correlations with MDS-pyramid of 0.46 and 0.42, respectively. In the MedLey trial, mean differences between the Mediterranean diet and the control habitual diet group after 6 months were similar for the biomarker scores derived in EPIC-Norfolk using the base set only (0.52 standard deviation (SD), 95% CI: 0.21, 0.84) and the base set with urinary sodium and potassium (0.50 SD, 0.18, 0.82).

In EPIC-Norfolk, the base set nutritional biomarker score was inversely associated (hazard ratio (HR) per 1 SD, 95% CI) with incidence of CVD (0.96, 0.92-0.99) and T2D (0.78, 0.71-0.84), CVD mortality (0.93, 0.88-0.98), all-cause mortality (0.92, 0.89-0.96) and cancer mortality (0.93, 0.88-0.99) but not incidence (0.97, 0.92-1.02). Associations evaluated using the metabolomic and joint nutritional-metabolomic biomarker scores were similar in terms of the magnitude and statistical significance (p-values for the $HR/HR_{base\ set}$ ratios >0.05) except for incidence of cancer (0.99, 0.95-1.03 and 0.95,0.89-1.01, respectively). MDS-pyramid estimated directly from FFQ was associated only with incidence of CVD and T2D, and the strength of association for T2D was weaker compared to biomarker-based assessment (p-values for the $HR_{FFQ}/HR_{biomarker}$ ratios <0.05 in analyses with >500 cases). Adding urinary phytoestrogens to the base set of nutritional biomarkers resulted in a stronger inverse association with incidence of CVD (0.91, 0.84-0.98), and adding serum phytoestrogens yielded an inverse association with incidence of cancer (0.88, 0.78-0.98). Associations with CVD- and all-cause mortality were modestly stronger when adding several additional groups of biomarkers to the base set.

Conclusions: This study highlights the utility of plasma carotenoids and phospholipid fatty acids for the epidemiological assessment of the Mediterranean diet to study chronic disease risk and mortality, and it suggests a similar performance of these nutritional biomarkers compared with an untargeted metabolomics assay. Pre-selection of specific groups of biomarkers into the pool of candidate predictors to this dietary pattern may impact on performance of the ensuant biomarker scores for exposure assessment and associations with incident disease outcomes and mortality.

6.1 Background

Combining nutritional biomarkers into biomarker scores is a novel method of objective assessment of adherence to dietary patterns (Chapters 3-5).[50] It represents a complementary approach to dietary self-report for application to studying diet-disease associations. Metabolomic profiling is an alternative to objective measurement of dietary exposures via nutritional biomarkers. Biomarker-based dietary assessment in nutritional epidemiology removes the uncertainty due to subjective reporting of dietary intakes from the analytical framework. Concurrently, it may, however, introduce other potential sources of bias in diet-disease associations. These include measurement error of biomarkers, unclear measurement error structure of biomarker scores or metabolomic signatures,[313] and the potential for diet-disease associations to be driven by individual components of such composite biomarkers. The latter issue is particularly pertinent to investigations with nutritional biomarker scores in which the number of predictor variables (up to several dozens) is smaller than in studies using untargeted metabolomics (typically hundreds of metabolites), thereby increasing the likelihood of undue influence of individual analytes. Furthermore, investigations based on combining nutritional biomarkers entail selecting which of the targeted assays to perform. They may also necessitate use of samples from different blood fractions and have biomarker-specific storage requirements. In consequence, both financial and logistical considerations are likely to limit the number of biomarkers measured. This contrasts with untargeted metabolomic profiling – the mainstay of objective assessment of adherence to dietary patterns to date[60,69,178] – in which the number of metabolites is pre-determined by the choice of a given metabolomic platform or an analytical pipeline combining several platforms.

The impact of pre-selection of specific groups of analytes on performance of biomarker scores as biomarkers of dietary patterns and predictors of health outcomes has not been previously evaluated. A head-to-head comparison between nutritional biomarkers and metabolomics on these metrics, as well as the utility of combining the two sources of biomarker data, has likewise not been undertaken. To address this research gap, I analysed data from the EPIC-Norfolk population-based cohort study which measured a broad range of circulating and urinary nutritional biomarkers, as well as untargeted plasma metabolomics in large baseline subsamples of its participants. Using the example of the Mediterranean diet, previously shown to have inverse associations with CVD incidence and mortality in EPIC-Norfolk,[304] I have derived nutritional biomarker scores and metabolomic profiles of this well-established dietary

pattern. I then tested their associations with disease outcomes and mortality and evaluated whether use of different groups of biomarkers impacts on the magnitude of associations.

6.2 Aim

My first objective was to derive a nutritional biomarker score of the MDS-pyramid score (Chapter 5)[304] in the EPIC-Norfolk study using circulating carotenoids and fatty acids. Second, I aimed to assess the impact on the strength of its association with MDS-pyramid score when (i.) expanding this base set of biomarkers with all other nutritional biomarkers opportunistically available in EPIC-Norfolk, i.e., urinary sodium and potassium (biomarkers of their respective intakes[314]), urinary and circulating phytoestrogens (biomarkers of intake or exposure to dietary phytoestrogens and legumes[315,316]), and circulating biomarkers of vitamins, iron status, cations (biomarkers of nutrient status and candidate biomarkers of the Mediterranean diet; Chapters 2-5), and carbon and nitrogen stable isotopes (biomarkers of animal protein intake[317,318]); (ii.) replacing nutritional biomarkers with metabolomics and (iii.) combining nutritional biomarkers with metabolomics. Third, I aimed to evaluate the impact of the choice of analytes on the strength of association of the biomarker and metabolomic scores with incident CVD, cancer, T2D and mortality (CVD, cancer and all-cause), and whether they differ from FFQ-based estimates.

6.3 Methods

I used plasma carotenoids and fatty acids overlapping with the MedLey trial as the base set of nutritional biomarkers for derivation of the biomarker scores of the MDS-pyramid. This enabled performing the validation procedure in the MedLey trial previously applied in Chapter 5. Further groups of biomarkers were evaluated internally in the EPIC-Norfolk study based on adding them to this base set. Urinary sodium and potassium additionally overlapped between EPIC-Norfolk (spot urine) and the MedLey trial (24-hour excretion), thus allowing to test whether these biomarkers can improve performance of the biomarker score. Assays of urinary cations in the MedLey trial have been described in Chapter 4.

6.3.1 The EPIC-Norfolk study

The Norfolk arm of the European Prospective Investigation into Cancer and Nutrition (EPIC-Norfolk) is a population-based cohort study of middle-aged adults in the East of England.[253] Approximately 77,630 eligible men and women aged between 39-79 years were identified via registers of 35 general practice surgeries in Norfolk and surrounding areas, and 25,639 were recruited and attended the baseline assessment. Baseline data collection took place between the years 1993 and 1997, including in-person visits for health examination, blood draws and collection of urine samples. Venous blood samples were collected at varying times of the day from non-fasted participants and stored in liquid nitrogen at -196°C. Spot urine samples were requested from all participants and stored at -20°C. A subsample of 340 individuals took part in validation and calibration studies ancillary to EPIC-Norfolk and had up to six 24-h urine collections per participant over one year,[319,320] 335 of whom had spot urine samples collected at baseline. Questionnaires were used to collect information on health, behavioural exposures and diet. Participants were followed-up since the day of baseline data collection until the 31st of March 2018 for incidence of CVD, cancers, T2D diagnosed in hospital settings or stated in death certificates, and mortality. Additionally, a case-cohort study of incident T2D ascertained from multiple sources was available within the sample of participants with measurements of stable isotopes, with a follow-up until the end of 2016. I selected these outcomes based on power considerations, as well as facilitation of comparability of the results with findings from Chapters 3-5 for T2D. I considered CVD incidence as the primary endpoint based on interventional evidence of the effects of the Mediterranean diet on CVD[117] and prior observational work in EPIC-Norfolk.[304]

Sample sizes of participants varied markedly by availability of different sets of nutritional and metabolomic biomarkers. Following exclusions of participants with missing covariate data (~5%), investigations on mortality included between 524 T2D case-cohort participants, up to 7,017 quasi-randomly selected participants for nutritional biomarkers and 11,518 quasi-randomly selected individuals for metabolomics, 4,212 of whom had the base set of nutritional biomarkers available. The MDS-pyramid score estimated from the FFQ was available in 6,286 and 10,580 individuals with nutritional and metabolomic data, respectively. Participants with self-reported prevalent outcomes were excluded in analyses of disease incidence or cause-specific mortality for these outcomes, thereby decreasing the sample sizes by ~6% for CVD (history of myocardial infarction or stroke), and ~5% each for cancer and T2D. Application of this exclusion to mortality investigations was motivated by the potential of prior history of

associated non-fatal outcomes to introduce dietary changes for secondary prevention and effect modification by baseline disease status. Up to 13% of participants were excluded from derivation of the biomarker and metabolomic scores due to missing data on the MDS-pyramid score or covariates and extreme values in biomarker concentrations. Furthermore, derivation of biomarker scores including stable isotopes in the T2D case-cohort sample was restricted to the subcohort of randomly selected participants. Exact numbers of participants in each analysis are reported in the Results section.

The EPIC-Norfolk study had the ethical approval granted by the Norwich District Ethics Committee, and all participants gave written informed consent to take part in this research.

6.3.2 Nutritional biomarkers: counts and sampling schemes of participants

The base set of five plasma carotenoid variables (α-carotene, β-carotene, β-cryptoxanthin, lutein, and sum of lutein and zeaxanthin) and 22 phospholipid fatty acids was measured in 7,335 overlapping participants, or 7,489 and 7,384 non-overlapping participants, respectively. Further groups of biomarkers included (n overlapping with the base set sample/n total measured): additional five fatty acids (7,335/7,384), plasma α- and γ-tocopherol (7,326/7,486), canthaxantin and retinol (7,238/7,392), urinary sodium and potassium (7,195/25,116), plasma vitamin C (7,097/22,471), serum ferritin, transferrin, iron, calcium, magnesium (5,565/18,387); urinary sucrose and fructose (1,969/5,887), urinary daidzein, genistein, glycitein, *o*-desmethylangolensin (ODMA), equol, enterodiol and enterolactone (1,717/2,623); serum daidzein, genistein, ODMA, enterolactone and enterodiol (1,179/1,994); and serum stable isotopes $\delta^{15}N$ and $\delta^{13}C$ (607/1,518). All plasma carotenoids, tocopherols, vitamin C and urinary cations were concurrently measured in 7,007 participants, 6,862 of whom also had complete data on the base set of biomarkers, and thus I considered them as a single set of additional biomarkers available in the core sample of ~7,000 participants. The remaining sets of biomarkers had poor or no overlap of samples with the base set and could not be considered jointly for prediction of diet.

Plasma carotenoids, phospholipid fatty acids and tocopherols were measured in individuals who participated in multiple nested case-control studies of CVD and cancer in the late 2000s.[321] Up to four participants who did not develop the outcome of interest by the time of sampling were selected per each incident case. Further details and information on the sampling schemes of the individual studies were not available in the EPIC-Norfolk database at the time of

analysis. Compared to the remainder of the cohort, participants from this sample were somewhat less frequently of male sex (49% vs 57%), had moderately higher mean age (63.3 versus 57.7 years) and BMI (26.6 versus 26.3 kg/m^2), and had modestly lower plasma vitamin C (51.2 versus 54.3 μmol/L) The cumulative incidence of CVD within this sample as of 2018 was 68% compared to 60% in the overall EPIC-Norfolk cohort. The corresponding proportions for incident cancer were 29% and 23%, respectively. Overall, I judged the sample to approximate a random subcohort closely enough to apply standard methods for a prospective cohort study in the primary analysis.

Serum and urinary phytoestrogens have previously been reported to be measured in individuals who were participants of nested case-control studies of incident breast (n_{cases} = 237), prostate (n_{cases} = 193) and colorectal cancers (n_{cases} = 221) occurring by 2006 with 1:4 matching to non-cases.[322,323] However, the data deposited in the EPIC-Norfolk database did not appear to be enriched in incident cancer cases per these sampling schemes. Within the samples with data on biomarkers of phytoestrogens, the cumulative incidence of breast (women only), prostate (men only), and colon cancers was 2.2% (n_{cases} = 32), 4.7% (n_{cases} = 56) and 1.4%, (n_{cases} = 37) respectively, which was materially similar as cumulative incidences in the overall cohort of 2.1%, 4.8% and 1.2%, respectively. Thus, I considered the available subsample of participants with biomarker data on phytoestrogens to be a random subsample of the study in terms of implications for statistical analysis. Of note, the cumulative incidence of any first cancer in this subsample was higher by the end of 2018 (35%) than in the overall cohort (23%).

Stable isotopes were assayed in a nested case-cohort study of incident T2D,[318] and the remaining nutritional biomarkers were measured in random subsamples.

6.3.3 Nutritional biomarkers: laboratory assays

Reversed-phase high-performance liquid chromatography was used to measure plasma concentrations of carotenoids[193] and tocopherols[324] at IARC (Lyon, France) using the HPLC-1100 system (Hewlett Packard, Wilmington, IL, USA) coupled with a C18-Adsorbosphere column (Alltech, Deerfield, IL, USA). Serum total cholesterol, used for residual-adjustment of these groups of biomarkers, was assayed with the RA 1000 (Bayer Diagnostics, Basingstoke, UK). Plasma samples for vitamin C assays were stabilised with metaphosphoric acid and stored at -70°C. Measurements were performed with a fluorometric assay. Coefficients of variation ranged from 4.6% at the upper range of quantified values (mean, 102.3 μmol/L) to 5.6% at the

lower end (mean, 33.2 μmol/L). High-to-moderate reproducibility of carotenoids and vitamin C after long term storage was previously reported in EPIC-Norfolk.[193,194] Serum ferritin, iron, calcium and magnesium were assayed on an Olympus AU640 Chemistry Immuno Analyzer via a xylidyl blue-based colorimetric assay (Quotient Bioresearch, UK).

Plasma phospholipid fatty acids were assayed at IARC using a HP-5980 gas chromatograph (Agilent, Palo Alto, CA) with flame ionization detector. Absolute concentrations were quantified by comparisons of peak areas of individual fatty acids with those of internal standards. Coefficients of variation ranged between 3% for C16:0 to 13% for C18:3n-3.[257] Twenty-seven fatty acids were measured in total, of which 24 overlapped with the MedLey trial. Trans-fatty acids C16:1 and C18:1-n9t were not used in biomarker sets for validation in the MedLey trial due to concerns over accuracy stated in the EPIC-Norfolk study's data dictionary, but they were additionally evaluated internally in EPIC-Norfolk as predictors of MDS-pyramid. Absolute concentrations were converted to molar percentages of total fatty acids used in a given analysis.

Concentrations of sodium and potassium in spot urine samples collected from nearly all EPIC-Norfolk participants were measured between 1998-2002 by flame photometry (IL 943; Instrumentation Lab, Warrington, UK). Concentrations of creatinine were measured on a Roche Cobas Mira Plus analyser, and they were used to estimate 24-hour excretion of sodium and potassium. Urinary sucrose and fructose were measured on a Trace GC Ultra and a Trace DSQ quadrupole mass spectrometer (ThermoElectron, Hemel Hempstead, UK). Each analytical batch contained spot urine samples with added known amounts of these monosaccharides for quality control.[325] Tagatose was quantified in <25% of the samples with sucrose and fructose measured, and thus I excluded it from the analysis. Of note, baseline data collection preceded introduction of tagatose as a low-calorie sweetener in the European Union by approximately a decade,[326] so a heavily left-skewed distribution could not plausibly be reflective of a large proportion of non-consumers. Up to 4.4% of values for sucrose and fructose were quantified to be null or negative. As previously reported in detail, 45% of results for urinary sucrose were outside of the limits of quantification.[325]

Phytoestrogens were assayed in both spot urine and serum. Serum samples were assayed by liquid chromatography-mass spectrometry (LC-MS) with triply ^{13}C-labelled internal standards.[315] The mean intra-assay coefficient of variation ranged from 2.8% for enterolactone to 20.0% for glycitein. Isotope dilution LC-MS was used to measure urinary concentrations.

Coefficients of variation for urinary assays were <6% for glycetin and equol and otherwise <4%.[322,327] The proportion of missing values in the urinary measures was small (<6% in any biomarker). In serum, the proportion was considerable for glycetin (72%), equol (64%), genistein (36%), and otherwise <13%. I excluded these three serum biomarkers from the analysis. In lieu of serum genistein quantified with a ^{13}C-genistein standard, I used genistein quantified with a ^{13}C-daidzein standard which had 8% of missing data and was strongly correlated with the former measure (r = 0.90).

For carbon and nitrogen stable isotopes, thawed samples were dried in a speed vacuum concentrator (miVAC; Genevac Ltd). Isotopes were assayed on a Costech automated elemental analyzer interfaced with continuous-flow mode to an isotope ratio–monitoring mass spectrometer (Thermo Finnigan MAT253; Godwin Laboratory, Department of Earth Sciences, University of Cambridge). The isotopes were measured as ratios of carbon-13 to carbon-12 (δ^{13}C) and nitrogen-15 to nitrogen-14 (δ^{15}N) relative to external reference.[318] The units were parts per thousand of abundance of heavy isotopes (‰) calculated as $\delta = (R_{sample} - R_{reference})/(R_{reference}) \times 1000$ (‰) where R is the ratio of the heavy to light isotope. The reference values were internationally established standards of Vienna Pee Dee Belemnite for carbon ($^{13}C/^{12}C = 0.01124$) and mean atmospheric nitrogen for nitrogen ($^{15}N/^{14}N_{atm\text{-}N} = 0.003677$). Measurement errors were ≤0.2‰ for 92% of participants and otherwise ≤0.9‰.[318] Means of duplicate measurements were used in the statistical analysis.[318]

6.3.4 Metabolomics

A panel of metabolomic biomarkers was measured using LC-tandem MS on the Metabolon Discovery HD4 platform in two batches of plasma samples from subsets of 5,992 and 5,980 quasi-randomly selected participants, following exclusion from the sampling frame of incident T2D cases occurring prior to the end of 2006 (assayed for metabolomics earlier with a previous version Discovery HD4 which covered fewer metabolites or metabolites non-overlapping with the current investigation).[328] The assays were performed in January 2016 and March 2017, and included 1,168 and 1,219 metabolites per batch, respectively. There were 988 metabolites overlapping between these two batches, comprising of 891 endogenous or unannotated metabolites and 97 xenobiotics. The metabolites were measured semi-quantitatively whereby each analyte was corrected for day-to-day differences in instrument tuning and normalised by setting its median value to 1.00 and proportionately re-scaling all data points within run day

blocks. Identity of the metabolites was determined based on an extensive in-house Metabolon library of the retention time/index, mass to charge ratio and chromatographic data in authenticated standards.[328]

I excluded metabolites with missingness >90% (n = 30), yielding 958 metabolites available for analysis. Missingness over 50%, 30% and 10% was present in 99, 153 and 302 variables, respectively. None of the participants had non-missing data in all metabolites, thus precluding complete-case analysis without imputation of missing values.

6.3.5 Outcome ascertainment

Participants were followed-up for a median of 21 years and all outcomes were ascertained up to the 31st of March 2018. Incident disease outcomes were obtained via linkage to Hospital Episode Statistics and cancer registries.[329,330] Data on vital status and cause of death were acquired from the UK Office for National Statistics.[329,330] Incident CVD was defined as the first ever occurrence of fatal or non-fatal ischaemic heart disease, ischaemic stroke, haemorrhagic stroke, heart failure, peripheral vascular disease and other less common CVDs as defined by the ICD codes 401-448 (ICD9) or I10-I79 (ICD10). Incident cancer was defined as the first ever occurrence of fatal or non-fatal cancer (ICD9 140-208 or ICD10 C00-C97). T2D was ascertained based on the ICD9 250 or ICD10 E10-E14 codes in hospital admissions data or death certificates.

A case-cohort study of incident T2D was also available in which cases were adjudicated up to December 2006 based on additional sources of information: self-report of physician's diagnosis or use of diabetes-specific medications and record linkage with general practice registers. Within its randomly selected subcohort, there was a full overlap in classification of incident cases between incident T2D cases ascertained from hospital records or death certificates, and the more comprehensive case ascertainment. Among the oversampled incident cases, 90.8% were classified as cases based on hospital records or death certificate.

6.3.6 Covariates

Questionnaires and physical examination were conducted at baseline to collect information on covariates, including sociodemographic, medical and health behavioural factors and

anthropometry. Weight, height, waist circumference and blood pressure were measured by trained nurses. Blood pressure was measured twice after a 5-minute rest in a seated position, and averages of the two readings was used for analysis. Physical activity was evaluated using a validated questionnaire.[203] The Townsend Area Deprivation Index was used as one of the measures of socioeconomic status.[331] This index combines information on unemployment, car ownership, proportion of homes which are not owner-occupied and household crowding in the area of residence.

Prevalent cases of outcomes under investigation at baseline were defined based on self-reported physician's diagnosis of myocardial infarction or stroke (CVD), cancer, and use of diabetes-specific medications or following a diet modified due to diabetes (T2D).

6.4 Statistical analysis

Stata 16.1 was used for all analyses except for quantile regression imputation which was done in R, version 4.0.2 (package imputeLCMD). For analyses involving statistical significance testing, two-sided $\alpha = 0.05$ was used. For descriptive statistics, two-sided $\alpha = 0.10$ was used to test for trend across quintiles to capture both statistically significant and marginal associations.

6.4.1 Pre-treatment of biomarker variables

Biomarkers were natural logarithm-transformed to stabilise variance except for the negatively-skewed serum magnesium and calcium, and $\delta^{13}C$ and $\delta^{15}N$ which included negative values per their original scales of quantification. Missing values in fatty acids were imputed using quantile regression imputation.[207,208] Missing values or values quantified as ≤ 0 in biomarkers of phytoestrogens and urinary sugars were replaced with the smallest observed value and marked as missing in the original data with an indicator variable.[332] Both the continuous biomarker variable and the associated indicator of missingness were used jointly in statistical models incorporating these biomarkers.[332] Urinary biomarkers were normalised to specific gravity, and not to urinary creatinine, to avoid inducing confounding by cardiometabolic risk factors[325,333] and meat intake.[66] Carotenoids and tocopherols were residual-adjusted for total cholesterol concentrations to minimise the confounding effects of blood lipids on their associations with dietary intakes of these compounds.[44,209]

6.4.2 Estimation of 24-h urinary excretion of sodium and potassium

I estimated 24-h urinary excretion of sodium[334–342] and potassium[334–337] using several formulas based on spot urine concentrations combined with urinary creatinine and basic anthropometry. I used averages of up to six 24-hour urinary sodium and potassium outputs in a subsample as 'gold standard' measures of habitual excretion of these cations. I selected the best performing formula for each of the biomarkers based on intraclass correlation coefficients (ICC), Pearson correlation and Bland-Altman limits of agreement.

6.4.3 Imputation of metabolomic data

I imputed missing metabolite values using multiple imputation by chained equation with predictive mean matching (MICE-pmm) which was previously reported as the optimal approach in large datasets of metabolomic data measured by the Metabolon HD4 platform.[343,344] I performed the imputation in 10 datasets. I used the first dataset throughout the analyses upon confirming stable associations of metabolomic scores of the MDS-pyramid with disease outcomes across the datasets. Data were imputed separately in each batch, including non-overlapping metabolites as auxiliary variables.

First, I generated batch-specific treelet transform scores[345] to capture the underlying metabolomic profiles in a hypothesis-free manner for later application in MICE-pmm. This data reduction technique identifies linear combinations of variables (components) with the goal of maximising the explanatory power over variation in these variables, thus being conceptually and mathematically similar to principal component analysis. Unlike principal component analysis, treelet transform does not automatically select high-variance factors but performs variable selection to generate components based on sparser solutions. An advantage of this feature is that treelet transform scores may be possible to apply outside of the complete-case derivation samples in datasets with large proportions of missing data. The degree of sparsity is determined by the cut-level of a cluster-tree. I selected the optimal cut-levels by means of 10-fold cross-validation.[345] I visually examined scree plots of variances of scores plotted against treelet components, and I judged five components to be sufficient in each batch to be taken forwards to multiple imputation (based on eigenvalues ≥1 and/or presence of break points in the plots). I excluded metabolites with ≥1% proportion of missingness from the treelet

transform analysis which allowed for complete-case derivation of the treelet scores in approximately two thirds of participants in each batch.

Second, I performed the MICE-pmm procedure. I excluded metabolites with the proportion of missingness >90%. Predictive mean matching was used based on 10 nearest neighbours to draw the imputed values from. The chained equation for each metabolite included 10 metabolites with the highest absolute Pearson's correlation with that metabolite in addition to the basic imputation model which included: sex, age at blood draw, seasonality (sine cosine function of the day of the year), time since last meal prior to blood draw, five treelet transform components, non-metabolomic circulating biomarkers (HDL-cholesterol, triglycerides, C-reactive protein, vitamin C, ferritin), BMI, waist circumference, height, systolic and diastolic blood pressure, physical activity (inactive, moderately inactive, moderately active, active), smoking status (current, former, never-smoker) and number of cigarettes per day in current smokers, dietary exposures estimated from FFQ (MDS-pyramid, intakes of energy and alcohol), current use of dietary supplements (any supplement, vitamin C, vitamin D, fish oil), current use of medications (statins, antihypertensive medications), medical history (personal history of myocardial infarction, stroke, cancer, diabetes, family history of myocardial infarction and diabetes), highest educational attainment (none, O-levels, A-levels, degree), marital status (single, married, widowed separated, divorced), employment (currently employed, housewife, retired, unemployed, student), Townsend index, and, in women, hormone replacement therapy (current, former, never-user) and menopausal status (premenopausal, perimenopausal <1 year or 1-5 years, postmenopausal; women only).

Missingness in xenobiotics may have plausibly reflected true null or very small metabolite concentrations. MICE-pmm cannot impute values beyond the observed ranges in non-missing data. To address this issue, I performed in parallel quantile regression imputation of left-censored data (QRILC) which assumes missingness not at random due to concentrations below the limit of detection.[207,208] Values imputed by MICE-pmm below the first percentile of pre-imputation distributions were replaced with QRILC-imputed values. Additionally, I generated indicator variables of missingness of xenobiotics which were included in the set of candidate predictors of the MDS-pyramid for inclusion into metabolomic scores. This allowed the potential scenarios of true missingness in xenobiotics and concentrations below the limit of detection to be considered in further analysis in a data-driven manner.

6.4.4 Derivation of biomarker and metabolomic scores

I applied elastic net regression[212] to perform variable selection and penalisation of coefficients within the different sets of nutritional and metabolomic biomarkers for prediction of the MDS-pyramid score, following its residual-adjustment for personal characteristics (Chapter 5). I randomly split the data into five folds using participants with non-missing variables in the base set of nutritional biomarkers. The biomarker scores were derived by iteratively performing elastic net regression analysis in all possible combinations of four folds. Linear predictions from each model were applied to the fifth left-out fold and were standardised to mean = 0 and SD = 1, and values from all the left-out folds were combined into single variables which constituted the biomarker scores. Additionally, I performed elastic net regression modelling on the entire available samples for the purpose of validation in the MedLey trial and presentation of the biomarker scores. Participants with at least one derivation set-specific biomarker value outside of the 25th percentile minus 3 times the interquartile range (IQR) or the 75th percentile plus 3 times the IQR were excluded from derivation samples. For metabolomic scores or metabolomic scores combined with nutritional biomarkers, I used the two metabolomics batches to define the datasets for cross-validated derivation and application of the scores. I followed the same percentile and IQR criteria as for nutritional biomarkers to identify and Winsorise extreme values (exclusions of outliers based on these criteria would result in lack of participants with complete-case data).

The elastic net regression models were 10-fold cross-validated within the derivation samples.[212] The α values were tested in 0.1 increments between 0.1 and 0.9. A grid of 100 λ values was tested per each α value. Tuning parameters resulting in the lowest cross-validated prediction errors were applied in subsequent analyses. I applied the $\lambda_{+1\ SE}$ rule[222] to generate lower variance metabolomic scores to facilitate their presentation and qualitative interpretation. Standardised coefficients from ordinary least squares regression were used for comparative presentation of biomarker and metabolomic scores to avoid the impact of differential tuning parameters on coefficient shrinkage between scores. I additionally identified nutritional biomarker and metabolite variables used for derivation of the scores with an absolute correlation with the MDS-pyramid score (residual-adjusted for personal characteristics) of at least 0.10 and statistical significance of this association below the threshold of Bonferroni correction.

Performance of the biomarker scores was evaluated by computing 5-fold cross-validated Pearson correlation coefficients with the MDS-pyramid score. Between batch (2-fold) cross-validation was used for metabolomic scores and combined nutritional- metabolomic biomarker scores. Biomarker scores derived using the base set of nutritional biomarkers (carotenoids and fatty acids), as well as the base set combined with urinary sodium and potassium, were tested for differences in the MedLey RCT between the Mediterranean diet intervention and the control habitual diet arm using mixed linear modelling (Chapter 5). I tested whether inclusion of the urinary biomarkers in the set of candidate predictors would result in larger between-group standardised differences than when using the base set of biomarkers. I regressed biomarker score values on the randomised group assignment for each of the two biomarker sets. This was followed by combining regression estimates and variance-covariance matrices into a joint parameter vector and a `sandwich-type' variance-covariance matrix to directly compare the biomarker score coefficients from separate linear regression models.

6.4.5 Associations of biomarker and metabolomic scores with incident outcomes

Individual nutritional biomarkers and metabolites were Winsorised at 4 SDs below or above the study means and were then used to calculate the biomarker and metabolomic scores with the scoring algorithms developed as described above. I performed Cox regression analysis to estimate HRs for associations between the scores and incident outcomes. I applied a robust variance estimator to account for the random measurement error of the MDS-pyramid score and its influence on the process of derivation of the biomarker scores.[47] I utilised Prentice-weighting in analyses of incident T2D within the nested T2D case-cohort study.[216] For analyses of other outcomes in the case-cohort sample, I applied inverse probability weights[266,267] to account for oversampling of incident T2D cases and enable valid inference on secondary outcomes.[346]

Participants were censored at the end of follow-up or the date of loss to follow-up or death whichever occurred first. Thus, estimates from Cox regression models were cause-specific hazards accounting for competing risks from death from other causes than the outcome of a given analysis, except for all-cause mortality where there were no competing events.[347] Prevalent cases of outcomes of interest were excluded from their respective analyses on disease incidence and disease-specific mortality. Restricted cubic splines with five knots were used to assess potential non-linearity of the associations between biomarker scores and incident

outcomes. Heterogeneity in the magnitude of HRs between estimates for the base set of biomarkers and other sets of nutritional or metabolomic biomarkers, as well as heterogeneity in HRs between biomarker-based assessment and self-report, were evaluated by testing the null hypothesis of equality of their ratio to 1.[348] The tests for heterogeneity were performed in samples restricted to participants who had non-missing data on both exposures under comparison. Thus, reported ratios of HRs may deviate from simple divisions of individual HRs due to differences in samples used for their estimations. Analyses involving the MDS-pyramid score estimated from FFQ were adjusted for estimated energy intake via the residual model.[223]

The multivariable-adjusted model included the following covariates: age (as the underlying timescale), sex, prevalent comorbidity (CVD, cancer, T2D), current use of antihypertensive medications, personal history of CVD, cancer and diabetes, family history of CVD, cancer and diabetes; smoking status (never, former, current smoker), physical activity index (inactive, moderately inactive, moderately active, active), seasonality (sine and cosine function of the day of the year of blood draw), current use of dietary supplements (any supplement, vitamin C, vitamin D, fish oil), highest educational attainment (none, O-levels, A-levels, degree), marital status (single, married, widowed separated, divorced), employment (currently employed, housewife, retired, unemployed, student), Townsend index, adiposity (BMI, waist circumference), and, in women, hormone replacement therapy (current, former, never-user) and menopausal status (premenopausal, perimenopausal <1 year or 1-5 years, postmenopausal; women only). Adiposity measures and Townsend index were modelled using restricted cubic splines with five knots to account for potential non-linear associations with the outcomes under investigation and to prevent the information loss and amplification of residual confounding which may occur with categorisation.[349,350] Given the large number of exposure-outcome pairs and their formal comparisons, I selected a priori the above single multivariable model for clarity of presentation. It remains elusive and could not be tested in the available data whether adiposity was a confounder or mediator of diet-disease associations under investigation. I considered adjustment for adiposity as the primary approach because of its potential regulatory effects on concentrations of circulating lipid-soluble biomarkers.[271] I have additionally estimated HRs without adjustment for adiposity.

I tested the proportional hazard assumption for biomarker and FFQ-based measures using Schoenfeld residuals in multivariable-adjusted Cox models. In instances where this assumption was not met, I changed the timescale of analysis to time of follow-up, included age as a covariate in Cox models (restricted cubic splines with seven knots), and estimated HRs within

the time periods of <10, 10-<20, and 20-25 years of follow-up. Additionally, I estimated HRs within strata of baseline age of <60, 60-<70, ≥70 years.

6.4.6 Sensitivity analyses

Sensitivity analyses consisted in (i.) excluding the first two years of follow-up from analysis to test for reverse causation bias, (ii.) restricting analysis to non-users of dietary supplements to evaluate independence of diet-disease associations from potential confounding by use of dietary supplements, and (iii.) repeating the analysis of associations of metabolomic scores with incident outcomes in the nine multiply-imputed datasets which were not used for derivation of the scores to assess between-imputation variability. These analyses were limited to the base set nutritional biomarker score and the metabolomic score.

To test the impact of non-random sampling of participants in the subsample with the base set of nutritional biomarkers on validity of Cox regression modelling, I re-analysed the association of the base set biomarker score with incident CVD as (i.) a case-control sample of incident CVD (unconditional logistic regression)[321] and (ii.) by constructing a nested-case control study with 1:1 incidence density matching of cases to controls with replacement on the closest follow-up time available (weighted Cox regression).[351,352] I included age and duration of follow-up as additional covariates in the logistic regression, testing both linear effects and restricted cubic splines with seven knots.

I used the case-cohort study of incidence of T2D (ascertained up to December 2006) to assess associations of nutritional biomarker scores derived using the base set of nutritional biomarkers and stable isotopes with a more sensitive outcome ascertainment compared to the primary outcome definition (ascertained up to March 2018 and available for a broader range of nutritional biomarkers). For analyses involving urinary sugars, I repeated derivation of the biomarker score and associations with incident disease and mortality outcomes with exclusion of participants with values outside of the limits of quantification. Additionally, I tested the impact of using fatty acids in absolute concentrations on performance of the composite biomarker of the Mediterranean diet and its associations with incident outcomes. Finally, I used accidental deaths as a negative control outcome (ICD9 codes: 800–999, ICD10 codes: V01-V09, V20-X59) .

6.5 Results

6.5.1 Background characteristics

Most baseline characteristics of participants differed by quintiles of the Mediterranean diet as assessed either by the base set nutritional biomarker score or self-report (**Table 6.1**). The patterns of distribution were largely consistent between the two exposure assessment methods. Differences between the extreme quintiles (Q5 vs Q1) were somewhat more pronounced with the biomarker score, e.g., means of BMI (25.6 vs 27.0 kg/m^2 compared to 26.2 vs 26.5 kg/m^2 for FFQ-based assessment) or serum triglycerides (2.12 vs 1.68 mmol/L compared to 1.98 vs 1.74 mmol/L). By contrast, there was little difference between the metabolomic biomarker score and the FFQ-based patterns of distributions of these variables (**Table 6.2**). Both the nutritional and the metabolomic score were characterised by increasing levels of physical activity across their quintiles, which was not observed with the MDS-pyramid score estimated from the FFQ.

6.5.2 Estimation of absolute excretion of sodium and potassium

Prediction equations of urinary output of sodium and potassium from spot urine concentrations evaluated in EPIC-Norfolk are presented in **Table 6.3**. The two Intersalt study equations for estimation of 24-hour sodium excretion had the highest ICCs with measured 24-hour excretion. The ICC values (95% CI) were 0.33 (0.23-0.42) and 0.34 (0.24-0.43) for the formulae with and without concentrations of potassium in spot urine, respectively (**Table 6.4**). The corresponding Pearson correlation coefficients and mean differences (95% limits of agreement) were 0.37 and 0.38, and 8.5 (-95.3, 112.4) and 3.4 (-99.3, 106.2) mmol/day, respectively. These two equations introduced the smallest mean bias and yielded the narrowest limits of agreement among all methods evaluated (**Table 6.4**). Given their materially similar performance, I selected for use in further analyses the Intersalt equation for sodium which does not incorporate urinary potassium into the estimation[339,340] to maintain independence between the two biomarkers. I selected the Kawasaki[335] equation for estimation of urinary excretion of potassium, which had the best performance on the above metrics (**Table 6.4**).

Table 6.1 Baseline characteristics of EPIC-Norfolk participants included in the analyses of the Mediterranean diet and chronic disease risk and mortality by quintiles of adherence assessed by a nutritional biomarker score and food frequency questionnaire (n = 7,017)

Characteristic	Q1	Q2	Q3	Q4	Q5	p*
Age, y						
MDS-pyramid, biomarker	64.1 (8.3)	63.6 (8.5)	63.5 (8.2)	62.8 (8.4)	62.3 (8.4)	*
MDS-pyramid, FFQ	63.4 (8.3)	63.4 (8.4)	63.4 (8.1)	62.8 (8.4)	61.6 (8.6)	*
Men, %						
MDS-pyramid, biomarker	59	54	52	48	42	*
MDS-pyramid, FFQ	64	55	49	45	34	*
Tertiary education, %						
MDS-pyramid, biomarker	4.8	9.3	9.6	12	17	*
MDS-pyramid, FFQ	5.7	8.7	10	11	18	*
Townsend index, points						
MDS-pyramid, biomarker	-1.66 (2.40)	-2.02 (2.10)	-1.99 (2.12)	-2.18 (2.02)	-2.27 (1.99)	*
MDS-pyramid, FFQ	-1.92 (2.17)	-2.04 (2.14)	-1.97 (2.19)	-2.15 (2.14)	-2.17 (1.98)	*
Marital status, married (%)						
MDS-pyramid, biomarker	78	80	82	83	82	*
MDS-pyramid, FFQ	84	84	82	79	78	*
Current smokers, %						
MDS-pyramid, biomarker	22	13	8.4	7.1	4.0	*
MDS-pyramid, FFQ	18	12	10	7.8	7.0	*
≥ Moderately active, %						
MDS-pyramid, biomarker	33	35	36	39	41	*
MDS-pyramid, FFQ	38	39	35	35	43	
Family history of CVD, %						
MDS-pyramid, biomarker	52	51	50	53	56	*
MDS-pyramid, FFQ	51	50	52	54	55	*
Family history of cancer, %						
MDS-pyramid, biomarker	38	36	42	41	41	*
MDS-pyramid, FFQ	40	38	40	40	43	*
Family history of T2D, %						
MDS-pyramid, biomarker	14	13	12	14	12	
MDS-pyramid, FFQ	14	13	13	13	12	*
Anti-hypertensive drugs, %						
MDS-pyramid, biomarker	25	27	27	25	27	
MDS-pyramid, FFQ	22	23	23	24	24	
Hormone therapy use, %†						
MDS-pyramid, biomarker	13	13	18	21	26	*
MDS-pyramid, FFQ	13	17	19	19	23	*
Dietary supplements use, %						
MDS-pyramid, biomarker	30	37	46	50	59	*
MDS-pyramid, FFQ	35	42	44	49	58	*
Fish oil supplement use, %						
MDS-pyramid, biomarker	18	24	34	37	44	*
MDS-pyramid, FFQ	26	30	31	34	40	*

Characteristic	Q1	Q2	Q3	Q4	Q5	p*
Body mass index, kg/m^2						
MDS-pyramid, biomarker	27.0 (4.1)	26.9 (3.9)	26.9 (3.8)	26.4 (3.5)	25.6 (3.2)	*
MDS-pyramid, FFQ	26.5 (3.6)	26.6 (3.6)	26.7 (3.8)	26.5 (3.7)	26.2 (3.8)	*
Waist circumference, cm						
MDS-pyramid, biomarker	93 (13)	92 (12)	91 (12)	89 (12)	87 (12)	*
MDS-pyramid, FFQ	92 (12)	91 (12)	90 (12)	89 (12)	87 (12)	*
Systolic BP, mmHg						
MDS-pyramid, biomarker	141 (19)	139 (19)	139 (19)	138 (18)	137 (19)	*
MDS-pyramid, FFQ	140 (18)	139 (18)	140 (19)	139 (19)	136 (19)	*
HDL cholesterol, mmol/L						
MDS-pyramid, biomarker	1.30 (0.39)	1.34 (0.40)	1.39 (0.41)	1.39 (0.40)	1.46 (0.43)	*
MDS-pyramid, FFQ	1.32 (0.39)	1.36 (0.42)	1.39 (0.41)	1.41 (0.41)	1.47 (0.42)	*
Triglycerides, mmol/L						
MDS-pyramid, biomarker	2.12 (1.13)	2.02 (1.14)	1.93 (1.08)	1.88 (1.02)	1.68 (0.89)	*
MDS-pyramid, FFQ	1.98 (1.08)	1.97 (1.08)	1.91 (1.04)	1.89 (1.05)	1.74 (0.98)	*

Abbreviations: BP – blood pressure; CVD – cardiovascular disease; FFQ – food frequency questionnaire; HDL – high-density lipoprotein; MDS – Mediterranean diet score, T2D – type 2 diabetes

Ranges of participants per fifths were (FFQ/biomarker) 1,268/1,403 (Q1), 1,269/1,404 (Q2), 1,268/1,403 (Q3), 1,269/1,404 (Q4), 1,269/1,403 (Q5). Values are means (standard deviation) or percentages. Tests for trend were calculated by regressing the covariates on fifths of dietary patterns entered into linear or logistic regression models as continuous variables with values equal to quantile numbers.

*p trend < 0.10

†In women only

Table 6.2 Baseline characteristics of EPIC-Norfolk participants included in the analyses of the Mediterranean diet and chronic disease risk and mortality by quintiles of adherence assessed by a metabolomic biomarker score and food frequency questionnaire (n = 11,518)

Characteristic	Q1	Q2	Q3	Q4	Q5	p*
Age, y						
MDS-pyramid, MBS	60.0 (9.4)	59.6 (9.1)	59.7 (8.9)	59.8 (8.8)	59.5 (8.6)	
MDS-pyramid, FFQ	59.6 (9.0)	60.2 (9.0)	59.6 (8.9)	59.4 (8.8)	58.5 (8.6)	*
Men, %						
MDS-pyramid, MBS	49	47	45	46	45	*
MDS-pyramid, FFQ	60	52	45	38	31	*
Tertiary education, %						
MDS-pyramid, MBS	5.5	8.6	12	14	20	*
MDS-pyramid, FFQ	7.2	9.0	13	13	20	*
Townsend index, points						
MDS-pyramid, MBS	-1.77 (2.27)	-2.09 (2.11)	-2.12 (2.05)	-2.22 (2.02)	-2.21 (2.01)	*
MDS-pyramid, FFQ	-1.95 (2.18)	-2.16 (2.09)	-2.13 (2.12)	-2.10 (2.06)	-2.22 (1.94)	*
Marital status, married (%)						
MDS-pyramid, MBS	79	84	83	82	82	
MDS-pyramid, FFQ	85	84	83	82	79	*
Current smokers, %						
MDS-pyramid, MBS	17	13	11	8.4	7.3	*
MDS-pyramid, FFQ	18	13	9.7	8.5	6.2	*
≥ Moderately active, %						
MDS-pyramid, MBS	41	43	41	41	45	*
MDS-pyramid, FFQ	44	42	41	41	46	
Family history of CVD, %						
MDS-pyramid, MBS	46	51	51	52	54	*
MDS-pyramid, FFQ	48	49	51	53	52	*
Family history of cancer, %						
MDS-pyramid, MBS	37	38	40	39	42	*
MDS-pyramid, FFQ	40	39	37	40	41	
Family history of T2D, %						
MDS-pyramid, MBS	12	11	13	12	13	
MDS-pyramid, FFQ	12	12	12	12	12	
Anti-hypertensive drugs, %						
MDS-pyramid, MBS	19	18	19	21	16	
MDS-pyramid, FFQ	16	16	18	18	17	
Hormone therapy use, %†						
MDS-pyramid, MBS	17	20	20	23	25	*
MDS-pyramid, FFQ	17	19	21	22	25	*
Dietary supplements use, %						
MDS-pyramid, MBS	31	40	42	50	55	*
MDS-pyramid, FFQ	33	40	43	49	57	*
Fish oil supplement use, %						
MDS-pyramid, MBS	20	26	29	34	39	*
MDS-pyramid, FFQ	22	27	29	34	37	*

Characteristic	Q1	Q2	Q3	Q4	Q5	p*
Body mass index, kg/m^2						
MDS-pyramid, MBS	26.4 (3.9)	26.4 (3.8)	26.4 (3.7)	26.2 (3.6)	25.8 (3.6)	*
MDS-pyramid, FFQ	26.1 (3.5)	26.4 (3.7)	26.3 (3.8)	26.2 (3.8)	25.9 (3.7)	*
Waist circumference, cm						
MDS-pyramid, MBS	89.2 (12.3)	88.5 (12.2)	88.3 (12.1)	88.0 (11.7)	86.6 (11.7)	*
MDS-pyramid, FFQ	90 (12)	90 (12)	88 (12)	87 (12)	85 (12)	*
Systolic BP, mmHg						
MDS-pyramid, MBS	137 (18)	136 (18)	136 (19)	136 (18)	134 (18)	*
MDS-pyramid, FFQ	137 (18)	137 (18)	136 (18)	135 (18)	133 (18)	*
HDL cholesterol, mmol/L						
MDS-pyramid, MBS	1.39 (0.42)	1.39 (0.41)	1.41 (0.41)	1.41 (0.42)	1.43 (0.41)	*
MDS-pyramid, FFQ	1.34 (0.40)	1.38 (0.40)	1.41 (0.41)	1.44 (0.41)	1.49 (0.42)	*
Triglycerides, mmol/L						
MDS-pyramid, MBS	1.90 (1.04)	1.83 (1.04)	1.81 (1.01)	1.76 (1.00)	1.64 (0.91)	*
MDS-pyramid, FFQ	1.91 (1.04)	1.84 (1.01)	1.78 (1.00)	1.72 (0.97)	1.61 (0.92)	*

Abbreviations: BP – blood pressure; CVD – cardiovascular disease; FFQ – food frequency questionnaire; HDL – high-density lipoprotein; MBS – metabolomic biomarker score; MDS – Mediterranean diet score, T2D – type 2 diabetes

Ranges of participants per fifths were (FFQ/biomarker) 2,138/2,304 (Q1), 2,137/2,304 (Q2), 2,138/2,303 (Q3), 2,137/2,304 (Q4), 2,137/2,303 (Q5). Values are means (standard deviation) or percentages. Tests for trend were calculated by regressing the covariates on fifths of dietary patterns entered into linear or logistic regression models as continuous variables with values equal to quantile numbers.

*p trend < 0.10

†In women only

Table 6.3 Equations for estimations of 24-hour urinary output of sodium and potassium from spot urine samples

Reference	Equation
Sodium	
Tanaka[334]	$21.98\times[Na_{su}/Cr_{su}\times0.1\times(16.14\times height+14.89\times weight-2.04\times age-2244.45)]^{0.392}$
Kawasaki[335]	Male:$16.3\times[Na_{su}/Cr_{su}\times0.1\times(7.39\times height+15.12\times weight -12.63\times age-79.9)]^{0.5}$ Female:$16.3\times[Na_{su}/Cr_{su}\times0.1\times(5.09\times height+8.58\times weight -4.72\times age-74.95)]^{0.5}$
Mage[336,337]	Male: $Na_{su}/Cr_{su}\times0.1\times(0.00179\times(140 - age)\times(weight^{1.5}\times height^{0.5})\times(1.366-0.0159\times BMI)$ Female: $Na_{su}/Cr_{su}\times0.1\times(0.00163\times(140 - age)\times(weight^{1.5}\times height^{0.5})\times(1.429-0.0198\times BMI)$
Toft[338]	Male: $33.56\times[Na_{su}/Cr_{su}\times0.1\times(-7.54\times age+14.15\times weight+ 3.48\times height+ 423.15)]^{0.345}$ Female: $52.65\times[Na_{su}/Cr_{su}\times1/10\times(-6.13\times age+9.97\times weight+2.45\times height+ 342.73)]^{0.196}$
Intersalt[339]	Male: $(0.45\times Na_{su}+23.51)-3.09\times Cr_{su}+4.16\times BMI+0.22\times age+14.60$ Female: $(0.33\times Na_{su}+3.74)-2.44\times Cr_{su}+2.42\times BMI+2.34\times age-0.03\times age^2+11.38$
Intersalt with potassium[339,340]	Male: $(0.46\times Na_{su}+25.46)-2.75\times Cr_{su}-0.13\times K_{su}+4.10\times BMI +0.26\times age+23.17$ Female:$(0.34\times Na_{su}+5.07)-2.16\times Cr_{su}-0.09\times K_{su}+ 2.39\times BMI+2.35\times age-0.03\times age^2+15.73$
Uechi[341]	Male: $Na_{su}/Cr_{su}\times(0.139\times age-0.002\times age^2+0.127\times weight\ 0.0157\times height)$ Female: $Na_{su}/Cr_{su}\times(0.139\times age-0.002\times age^2+0.127\times weight\ 0.0157\times height-2.78)$
Uechi with potassium[341]	Male: $3.41\times age-0.028\times age^2+0.238\times Na_{su}-2.84\times Cr_{su}+0.128\times K_{su}+5.75\times BMI-33.52$ Female: $3.41\times age-0.028\times age^2+0.238\times Na_{su}-2.84\times Cr_{su}+0.128\times K_{su}+5.75\times BMI-65.58$
Whitton: AM[342]	$88.66+0.55\times Na_{su}-1.34\times Cr_{su}-1.05\times K_{su}-0.87\times age+2.10\times BMI+39.30$[male sex]
Whitton: PMa[342]	$53.30+0.49\times Na_{su}-4.51\times Cr_{su}-0.44\times K_{su}-0.26\times age+1.48\times BMI+41.58$[male sex]
Whitton: PMe[342]	$62.31+0.42\times Na_{su}-3.14\times Cr_{su}-0.06\times K_{su}-0.49\times age+2.19\times BMI+40.13$[male sex]
Potassium	
Tanaka[334]	$7.59\times[Na_{su}/Cr_{su}\times1/10\times(16.14\times height+14.89\times weight-2.04\times age-2244.45)]^{0.431}$
Kawasaki[335]	Male:$7.2\times[K_{su}/Cr_{su}\times0.1\times(7.39\times height+15.12\times weight -12.63\times age-79.9)]^{0.5}$ Female:$7.2\times[K_{su}/Cr_{su}\times0.1\times(5.09\times height+8.58\times weight -4.72\times age-74.95)]^{0.5}$
Mage[336,337]	Male: $K_{su}/Cr_{su}\times0.1\times(0.00179\times(140 - age)*(weight^{1.5}\times height^{0.5})\times(1.366-0.0159\times BMI)$ Female: $K_{su}/Cr_{su}\times0.1\times(0.00163\times(140 - age)*(weight^{1.5}\times height^{0.5})\times(1.429-0.0198\times BMI)$

Abbreviations: AM – morning samples; BMI – body mass index; Cr_{su} – spot urinary creatinine, mg/dL; K_{su} – spot urinary potassium; Na_{su} – spot urinary sodium; PMa – afternoon samples; PMe – evening samples

Units: BMI – kg/m^2; Cr_{su} – mg/dL; height – cm; K_{su} and Na_{su} – mmol/L; weight – kg

Table 6.4 Performance of prediction equations of 24-hour urinary output of sodium and potassium from spot urine samples in the EPIC-Norfolk study: comparison against repeated 24-h collections (n = 335)*

Reference	Pearson correlation coefficient	Intraclass correlation coefficient (95% CI)	Mean difference, mmol/day (95% limits of agreement)
Sodium			
Tanaka[334]	0.15	0.13 (0.03-0.23)	18.4 (-100, 136.7)
Kawasaki[335]	0.18	0.18 (0.07-0.28)	58.8 (-79.2, 196.9)
Mage[336,337]	0.21	0.20 (0.10-0.30)	-11.4 (-167.3, 144.5)
Toft[338]	0.32	0.31 (0.21-0.40)	23.9 (-87.6, 135.3)
Intersalt without potassium[339]	0.37	0.34 (0.24-0.43)	8.5 (-95.3, 112.4)
Intersalt with potassium[339,340]	0.38	0.33 (0.23-0.42)	3.4 (-99.3, 106.2)
Uechi[341]	0.31	0.25 (0.15-0.35)	59.7 (-45.6, 165.0)
Uechi with potassium[341]	0.15	0.14 (0.03-0.24)	17.5 (-168.5, 203.6)
Whitton – morning[342]	0.23	0.23 (0.12-0.33)	-53.9 (-174.9, 67.2)
Whitton – afternon[342]	0.21	0.19 (0.09-0.29)	-65.5 (-181.2, 50.2)
Whitton – evening[342]	0.34	0.28 (0.17-0.37)	-24.2 (-127.0, 78.6)
Potassium			
Tanaka[334]	0.31	0.25 (0.14-0.35)	-19.6 (-63.0, 23.7)
Kawasaki[335]	0.31	0.29 (0.19-0.39)	-4.5 (-50.9, 42.0)
Mage[336,337]	0.32	0.27 (0.17-0.37)	7.7 (-71.9, 87.4)

*Means of from up to six 24-h urine collections per participant over one year were used as the reference measurement. Mean measured 24-h excretions of sodium and potassium were 141.1 and 75.6 mmol, respectively.

6.5.3 Nutritional and metabolomic biomarker scores

Composition of the nutritional biomarker scores is presented in **Table 6.5.** There was a consistent pattern of inverse scoring for docosapentaenoic (C22:5n-3) and positive scoring for docosahexaenoic fatty acid (C22:6n-3) with substantially larger effect sizes than for most remaining biomarkers. Relatively large coefficients were also observed for plasma $\delta^{15}N$ (negative), presence of quantifiable amounts of genistein in serum (positive) and sucrose in urine (negative). Expanding the pool of candidate predictor biomarkers for derivation of biomarker scores of the MDS-pyramid score beyond the base set of plasma carotenoids and phospholipid fatty acids resulted in inclusion of most of the additional analytes into the scores.

In the MedLey trial, the mean of the base set nutritional biomarker score was higher in the Mediterranean diet group than in the control habitual diet participants by 0.52 SD (95% CI: 0.21, 0.84) after six months of the partial-feeding intervention. The biomarker score derived by additionally using estimated urinary excretion of sodium and potassium yielded a materially similar result (0.50 SD; 95% CI: 0.18, 0.82; p difference from the previous score = 0.42). In EPIC-Norfolk, both biomarker scores had the same cross-validated Pearson coefficient of 0.40 for their correlation with the MDS-pyramid score. Adding further sets of nutritional biomarkers into the derivation procedure did not materially improve this correlation or it yielded moderately lower values in the smaller subsets of participants with stable isotopes, phytoestrogens, and urinary sugars (r range: 0.30-0.37; **Table 6.5**).

The metabolomic score derived using both batches combined comprised of 456 predictors (334 and 373 stratified by batch), including 45 indicator variables for detection of xenobiotics, selected based on the a priori approach of minimising the cross-validated prediction error in elastic net regression. I confirmed materially similar performance of this score in terms of cross-validated correlation with the MDS-pyramid score as an analogous metabolomic score generated under the increased $\lambda_{+1\ SE}$ penalisation (r = 0.46 for both levels of penalisation). The $\lambda_{+1\ SE}$ score was based on 102 metabolites, including 53 annotated analytes. The same was applicable to the biomarker score which combined metabolomics with targeted measurements of nutritional biomarkers (r = 0.42 for both levels of penalisation). I present the more parsimonious $\lambda_{+1\ SE}$ scores in **Table 6.6** to facilitate interpretability. The metabolomic score encompassed a broad range of metabolites, including lipids (plasmalogens, lysophospholipids, phosphatydylcholines, fatty acids), xenobiotics (plant foods-derived and medications), and metabolites of carbohydrate and amino acid metabolism. Directionality of coefficients was

heterogeneous both between and within these groups of metabolites and effect sizes were overall small. The plant foods-derived xenobiotics had the most consistent, positive relationship with the MDS-pyramid score. There was substantial overlap in terms of selection of metabolites for inclusion into the score between batches. Adding nutritional biomarkers into the set of candidate predictors led to selection of a more parsimonious list of metabolomic biomarkers, and α- and ß-carotene, lycopene, lutein and zeaxanthin, and C22:6n-3m C22:4 and C18:1n-9c fatty acids among nutritional biomarkers.

In the analysis which considered nutritional biomarkers and metabolites on one-by-one basis, 40 analytes had a statistically significant absolute $r \geq 0.10$ with the MDS-pyramid score (residual-adjusted for personal characteristics) after Bonferroni correction (**Appendix 6.1**). This included 9 nutritional biomarkers from targeted measurements (carotenoids, C20:5n-3, C22:6n-3 and vitamin C), and 18 identified and 13 unidentified metabolites. The metabolomic biomarkers included C22:6n3 and an acyl choline involved in its metabolism (docosahexaenoylcholine), and threonate and oxalate (vitamin C metabolism), thereby having some overlap with nutritional biomarkers. Four plant-derived xenobiotics were identified: 4-allylphenol sulfate, ergothioneine, methyl glucopyranoside (α + β) and stachydrine. The remaining metabolites included products of amino acid metabolism (indolepropionate, tryptophan betaine, dopamine 4-sulfate, N-methylproline), glycolysis (glycerate), and lipids (1-docosahexaenoyl-GPC (22:6), 1-(1-enyl-palmitoyl)-GPC (P-16:0), 1-(1-enyl-palmitoyl)-GPE (P-16:0), 1-palmitoyl-2-docosahexaenoyl-GPC (16:0/22:6) and 1-stearoyl-2-docosahexaenoyl-GPC (18:0/22:6)). Correlation coefficients for the above analytes were all positive and did not exceed 0.19.

Table 6.5 Standardised coefficients of biomarker scores of the Mediterranean diet in the EPIC-Norfolk study

Biomarker*	Base set	+urinary Na & K	+urinary Na & K, vitamins	+Fe status & Mg, Ca	+stable isotopes	+serum phytoest rogens	+urinary phytoest rogens	+urinary sugars
N	*6,116*	*6,028*	*5,589*	*4,464*	*483*	*983*	*1,227*	*1,638*
Cross-validated r	*0.40*	*0.40*	*0.41*	*0.39*	*0.35*	*0.30*	*0.32*	*0.37*
α-carotene	0.07	0.06	0.07	0.06	0.08	0.01	0.08	0.13
ß-carotene	-0.03	-	-0.02	-	0.04	0.03	-	-0.08
ß-cryptoxanthin	0.12	0.12	0.13	0.09	0.05	0.06	0.11	0.13
Lycopene	0.10	0.09	0.08	0.10	0.09	0.06	0.03	0.10
Lutein & zeaxanthin	0.04	0.04		0.04	0.01	0.06	0.02	0.02
Lutein			0.07					
Zeaxanthin			-0.05					
Canthaxanthin			0.02					
Retinol			0.00					
C14:0	0.07	0.07	0.02	0.05	0.22	0.12	0.05	0.07
C15:0	-0.02	-0.02	0.04	-	-0.03	-	0.04	-0.01
C17:0	-0.06	-0.06	-0.03	-0.08	-0.06	-0.07	-0.08	-0.04
C20:0	-0.01	-	0.01	-0.01	-0.05	-0.02	-0.06	0.02
C24:0	-0.02	-0.03	0.02	-0.04	-0.09	-0.04	-	-0.04
C18:3n-3	0.00	-	-0.01	-	-0.10	-	0.04	-0.01
C20:5n-3	0.06	0.05	0.04	0.05	0.06	0.01	0.03	0.06
C22:5n-3	-0.14	-0.14	-0.14	-0.12	-0.14	-0.07	-0.15	-0.15
C22:6n-3	0.20	0.20	0.20	0.19	0.36	0.17	0.21	0.20
C18:3n-6			0.04					
C20:2	-0.01	-0.01	-0.01	0.00	0.00	-0.02	-0.03	-0.04
C20:4n-6	-0.06	-0.04	-0.06	-0.06	-0.03	-0.04	-0.05	-0.06
C22:4n-6	0.03	-	0.01	-	-0.10	0.01	0.03	0.02
C14:1			0.02					
C16:1	-0.09	-0.07	-0.08	-0.08	-0.11	-0.13	-0.15	-0.10
C18:1n-7c	0.00	0.00	0.00	-0.01	-0.01	-	0.00	0.03
C18:1n-9c	0.00	0.00	-	-	0.02	-0.02	-	0.00
C20:1	-0.01	-0.01	-	0.00	-0.02	-0.01	-0.03	-0.03
C24:1	0.00	-0.01	0.01					
C16:1n-9t			-0.09					
C18:1n-9t			-0.03					
Potassium		0.02	0.01					
Sodium		0.03	0.04					
Vitamin C			0.03					
α-tocopherol			0.03					
γ-tocopherol			0.03					
Ferritin				0.01				
Transferrin				-0.02				
Magnesium				0.00				
$\delta^{13}C$					-0.08			
$\delta^{15}N$					-0.21			
Genistein, continuous						-0.11		
Genistein, detectable						0.26		
Daidzein, continuous						0.07		
Daidzein, detectable						-0.04		
Enterolactone						0.04	0.02	
Enterodiol, continuous						0.07	0.05	
Enterodiol, detectable						0.14		
ODMA, continuous						0.07		
ODMA, detectable							0.01	
Equol, detectable							-0.07	
Glycitin							-0.06	
Fructose								-0.04
Sucrose, continuous								-0.04
Sucrose, detectable								-0.20

Abbreviations: ODMA – *o*-desmethylangolensin; r – Pearson correlation coefficient (5-fold cross-validated)

*Variables are continuous and standardized unless otherwise stated. "Detectable" refers to a binary, non-standardised indicator of non-missing and non-zero values in assayed samples. Red colour highlights positive coefficients, and blue colour highlights negative coefficients with intensity proportional to value.

Table 6.6 Standardised coefficients of metabolomic scores of the Mediterranean diet in the EPIC-Norfolk study

Analyte†	Metabolomics only			Metabolomics and nutritional biomarkers		
	All‡	Batch 1	Batch 2	All‡	Batch 1	Batch 2
Number of participants	10,544	5,278	5,266	3,788	1,922	1,870
Nutritional biomarkers (targeted measurement)						
α-carotene				0.01		0.01
ß-carotene				0.07	0.09	0.06
lycopene				0.05	0.08	0.04
lutein and zeaxanthin				0.02		
C22:6n-3				0.03	0.06	0.00
C22:4				-0.02		
C18:1n-9c				-0.01		0.01
Carbohydrate metabolism						
1,5-anhydroglucitol (1,5-AG) (glycolysis & gluconeogenesis)	-0.02	-0.03	-0.01			
Gucuronate (aminosugar metabolism)	-0.02		-0.04			
Mannitol/sorbitol (fructose, mannose & galactose metabolism)				0.04		
Lipids, plasmalogens						
1-(1-enyl-palmitoyl)-2-docosahexaenoyl-GPC (P-16:0/22:6)*	0.07	0.03	0.08			
1-(1-enyl-palmitoyl)-2-docosahexaenoyl-GPE (P-16:0/22:6)*	0.00	0.01		0.05		
1-(1-enyl-palmitoyl)-2-oleoyl-GPE (P-16:0/18:1)*	0.04	0.00				
1-(1-enyl-stearoyl)-2-arachidonoyl-GPC (P-18:0/20:4)	-0.04	-0.03	-0.06			
1-(1-enyl-stearoyl)-2-arachidonoyl-GPE (P-18:0/20:4)*	-0.02	-0.02	-0.04	-0.06		
1-(1-enyl-stearoyl)-2-linoleoyl-GPC (P-18:0/18:2)*	-0.05					
1-(1-enyl-stearoyl)-2-linoleoyl-GPE (P-18:0/18:2)*	-0.02	-0.03	-0.02	0.00		
1-(1-enyl-stearoyl)-2-oleoyl-GPC (P-18:0/18:1)	0.00		-0.06	-0.05		
Lipids, lysophospholipids						
1-adrenoyl-GPC (22:4)*	0.04	0.04	0.03	0.01		
1-arachidonoyl-GPE (20:4n6)*	0.03		0.03			
1-docosahexaenoyl-GPC (22:6)*	0.02	-0.01	0.03			
1-docosapentaenoyl-GPC (22:5n3)*	-0.07					
1-docosapentaenoyl-GPC (22:5n6)*	0.07	0.01				
1-palmitoleoyl-GPC (16:1)*	-0.05	-0.06	-0.05	-0.06		-0.05
1-palmityl-2-arachidonoyl-GPC (O-16:0/20:4)*	-0.04	-0.05				
2-oleoyl-GPC (18:1)*		-0.05				
Lipids, phosphatidylcholines						
1-palmitoyl-2-docosahexaenoyl-GPC (16:0/22:6)	0.03	0.07	0.03	0.07	-0.04	0.09
1-stearoyl-2-docosahexaenoyl-GPC (18:0/22:6)	0.06	0.06	0.00	-0.06	0.07	
1-stearoyl-2-docosapentaenoyl-GPC (18:0/22:5n6)*	-0.04	-0.04				
1-stearoyl-2-meadoyl-GPC (18:0/20:3n9)*	-0.03	-0.03	-0.03	0.02		-0.09
1-stearoyl-2-oleoyl-GPC (18:0/18:1)	-0.01					
phosphatidylcholine (16:0/22:5n3, 18:1/20:4)*	-0.03	-0.04	-0.04			
Lipids, fatty acids and fatty acid metabolism						
3-carboxy-4-methyl-5-propyl-2-furanpropanoate	-0.02		-0.02	0.02		
5-hydroxyhexanoate (monohydroxy)		-0.04				
docosahexaenoate (DHA; 22:6n3)	0.04	0.00	0.08	0.03	0.06	
sebacate (decanedioate) (dicarboxylate)				-0.04		
docosahexaenoylcholine (acyl choline)	0.02	0.04				
stearoylcarnitine (acyl carnitine)		0.02		0.01		
linoleoylcarnitine* (acyl carnitine)			0.07			
Lipids, other						
5α-pregnan-3(α or ß),20ß-diol disulfate (progestin steroid)		-0.04				
linoleoyl ethanolamide (endocannabinoid)				0.03		
4-androsten-3ß,17ß-diol monosulfate (2) (steroid)	-0.04		-0.06			
Xenobiotics, food component/plant source						
4-allylphenol sulfate	0.03	0.04	0.03	0.03	0.04	0.05
4-vinylphenol sulfate	0.04	0.04	0.04	0.06		
acesulfame, detectable (yes/no)	0.11		0.14	0.15		
dimethyl sulfone	0.04		0.02			

Analyte†	Metabolomics only			Metabolomics and nutritional biomarkers		
	All‡	Batch 1	Batch 2	All‡	Batch 1	Batch 2
ergothioneine, detectable (yes/no)	0.13	0.14				
ergothioneine, continuous	0.04	0.03	0.03	0.01	0.04	0.03
eugenol sulfate				0.03		
homostachydrine*	0.02		0.04			
indolin-2-one				-0.02		
N-acetylalliin, detectable (yes/no)	0.02	0.01	0.05	0.06		0.11
S-allylcysteine, detectable (yes/no)	0.04	0.03	0.02	-0.02		0.00
stachydrine, detectable (yes/no)	0.04	0.04	-0.01	0.03		0.02
thymol sulfate, detectable (yes/no)	0.07					
Xenobiotics, medications						
2-methoxyacetaminophen glucuronide*	-0.01					
2-methoxyacetaminophen sulfate			-0.05			
4-acetamidophenol	-0.03	-0.04				
4-hydroxycoumarin	0.03	0.04	0.02	0.03		
hydroquinone sulfate	0.02	0.03				
quinine	-0.04	-0.08			-0.09	
Amino acid metabolism						
1-methylhistidine (histidine)	-0.02	0.01				
3-hydroxy-2-ethylpropionate (BCAA)	-0.03					
3-indoxyl sulfate, (tryptophan)				-0.04		
dopamine sulfate (2) (phenylalanine and tyrosine)	0.01	0.02	0.01	-0.02		0.07
thyroxine (phenylalanine and tyrosine)	-0.03					
trans-4-hydroxyproline (urea cycle/arginine and proline)	-0.05	-0.05	-0.08	-0.07	-0.08	
tryptophan betaine (tryptophan)	0.08	0.07	0.09	0.07	0.06	0.08
homocitrulline (arginine and proline)		-0.04				
indolepropionate (tryptophan)	-0.02	-0.05	0.00	0.02		
N2,N5-diacetylornithine (urea cycle/arginine and proline)			0.03			
N-acetyl-3-methylhistidine* (histidine)	0.02		0.04	0.04		
N-acetylglutamate (glutamate)				-0.02		-0.07
p-cresol-glucuronide* (phenylalanine and tyrosine)	-0.04	-0.04				
Miscellaneous						
citrate (Krebs cycle)		0.05		0.06		
N1-methylinosine (purine metabolism)	0.01	-0.02				
oxalate (ethanedioate) (ascorbate and aldarate metabolism)			0.02			
threonate (ascorbate and aldarate metabolism)	0.01	0.02	-0.01	0.03		

†Variables are continuous and standardized unless otherwise stated. “Detectable” refers to a binary, non-standardised indicator of non-missing values of xenobiotics. There were up to 49 unidentified metabolites included in the scores with coefficients in the range +/- 0.10. Metabolites identified via in-silico predictions (not by direct confirmation against a matching purified standard), are followed by a “*” suffix. Red colour highlights positive coefficients, and blue colour highlights negative coefficients with intensity proportional to value.

‡Metabolomic scores derived jointly in two subsamples of participants with metabolomic assays undertaken in two batches.

6.5.4 Diet-disease associations: main results

The nutritional biomarker score derived using the base set of analytes (carotenoids and fatty acids) and validated in the MedLey trial was inversely associated with incidence of CVD in the EPIC-Norfolk study. The hazard ratio (HR) per 1 standard deviation (95% confidence interval) was: 0.96 (0.92-0.99) (**Table 6.7**). Adding additional groups of biomarkers to sets of predictors for derivation of the biomarker score, displacing nutritional biomarkers with metabolomics, or combining metabolomics with nutritional biomarkers did not overall affect this association beyond the impact of varying samples on precision and statistical significance. There was some indication that the biomarker score combining the base set with urinary phytoestrogens resulted in a marginally lower risk estimate (HR = 0.91, 95% CI: 0.84-0.98), whereby its HR ratio to the HR for the base set (HR/HR_{base}) was 0.96 (95% CI: 0.92-1.00, p = 0.03). The biomarker-based measures yielded materially similar associations with incident CVD as adherence to the Mediterranean diet estimated from dietary self-report (p difference ≥ 0.35 across all comparisons).

Among the secondary outcomes of disease incidence, neither the base set biomarker score nor the FFQ-based score were statistically significantly associated with incidence of cancer. (**Table 6.7**) However, the biomarker score had an inverse association following its derivation with serum phytoestrogens (HR per 1 SD = 0.88; 95% CI: 0.78-0.98) and marginally inverse with urinary sodium and potassium and circulating vitamins (HR per 1 SD = 0.95; 95% CI: 0.90-1.00). For incidence of T2D, both the base set biomarker score and the FFQ-based score had inverse associations (95% CI) of 0.78 (0.71-0.84) and 0.90 (0.82-0.97) HR per 1 SD, respectively. Biomarker-based assessment with varying sets of biomarker predictors yielded consistently lower HR estimates compared to FFQ in samples with >500 incident T2D cases (p difference ≤ 0.04) There was no evidence to suggest that any of the combinations of nutritional biomarkers and/or metabolomics resulted in a materially different magnitude of association than that estimated from the base set of nutritional biomarkers (p values for $HR/HR_{base} \geq 0.12$).

For mortality, the nutritional biomarker-based assessment generally yielded inverse associations between the Mediterranean diet and deaths from CVD, cancer, and all-cause mortality in samples over ~5,000 participants (**Table 6.8**). For example, the base set nutritional biomarker score had HRs (95% CI) per SD of 0.93 (0.88-0.98), 0.93 (0.88-0.99) and 0.92 (0.89-0.96), respectively. Adding additional nutritional biomarkers to the base set tended to result in

stronger inverse associations with CVD and all-cause mortality. For CVD mortality, the HR/HR_{base} (95% CI) ratios were 0.99 (0.98-0.99), 0.97 (0.96-0.99) and 0.96 (0.93-0.99) when adding to the base set urinary sodium and potassium alone, urinary sodium and potassium with circulating vitamins, and urinary sugars, respectively. For all-cause mortality, the first two additional sets of biomarkers also yielded stronger associations (HR/HR_{base} = 0.99, 95% CI: 0.99-1.00 and 0.98, 95% CI: 0.97-0.99, respectively), as well as serum phytoestrogens (HR/HR_{base} = 0.94, 95% CI: 0.89-0.99). Metabolomic scores were inversely associated with CVD and all-cause mortality, but not cancer mortality, and the inverse associations were materially similar as those estimated using nutritional biomarkers (**Table 6.8**).

Corresponding associations with mortality of the MDS-pyramid score assessed by FFQ were largely null, except for all-cause mortality in the largest sample in the analysis of participants with metabolomics data (HR = 0.96; 95% CI = 0.93-0.99). However, there was formal evidence to suggest that the magnitude of the FFQ-based associations was smaller than that of the biomarker-based assessment only for a small number of specific pairs of biomarker sets and mortality outcomes. These included the base set of biomarkers combined with circulating vitamins, and urinary sodium and potassium in relation to cancer-specific and all-cause mortality (p-values <0.02), as well as the combined metabolomics-nutritional biomarkers score for all-cause mortality only (p = 0.04; **Table 6.8**).

Removing BMI and waist circumference from the multivariable model resulted in a small degree of deattenuation of the associations assessed using nutritional biomarker scores and had no appreciable impact on the results for metabolomic scores. For example, the HR (95% CI) per SD of the base set nutritional biomarker score for incidence of CVD was 0.96 (0.92-0.99) and 0.94 (0.91-0.97) before and after removing the adiposity measures from the model. The corresponding results for the metabolomics-only score were 0.95 (0.91-0.99) and 0.96 (0.93-0.98), respectively. For T2D, the estimates changed from 0.77 (0.71-0.84) to 0.72 (0.67-0.78) for the nutritional biomarker score, and 0.85 (0.78-0.92) to 0.82 (0.76-0.89) for the metabolomic score.

Table 6.7 Associations between the Mediterranean diet pyramid score and disease incidence per standard deviation of adherence: biomarker-based assessment and self-report in the EPIC-Norfolk study

Outcome and set of biomarkers*	n/N	Biomarker score		FFQ	
		HR (95% CI)	HR ratio to base set (95% CI)	HR (95% CI)	p†
CVD					
Base set	4,466/6,554	0.96 (0.92-0.99)‡	-	0.95 (0.92-0.99)	0.70
+urin. Na & K	4,381/6,433	0.95 (0.92-0.99)‡¶	1.00 (0.99-1.00)	0.95 (0.92-0.99)	0.77
+vitamins	4,177/6,142	0.96 (0.92-0.99)‡	1.00 (0.99-1.01)	0.95 (0.92-0.98)	0.65
+Fe status & Mg	3,283/4,889	0.96 (0.93-1.00)‡	1.00 (0.99-1.01)	0.96 (0.92-1.00)	0.86
+$\delta^{13}C$ & $\delta^{15}N$	359/472	1.10 (0.93-1.30)‡	1.08 (0.97-1.21)	1.09 (0.91-1.32)	0.73
+serum phytoestr.	698/1,034	0.96 (0.88-1.05)	0.96 (0.91-1.01)	0.97 (0.89-1.05)	0.84
+urin. phytoestr.	963/1,455	0.91 (0.84-0.98)‡	0.96 (0.92-1.00)#	0.95 (0.88-1.02)	0.41
+urin. sugars	1,152/1,754	0.92 (0.86-0.98)	0.99 (0.97-1.01)	0.96 (0.90-1.02)	0.40
Metabolomics	6,689/11,032	0.97 (0.94-0.99)‡	0.97 (0.94-1.02)	0.96 (0.94-0.99)	0.35
+base set	2,696/3,961	0.95 (0.91-0.99)¶	0.98 (0.95-1.02)	0.95 (0.91-0.99)	0.90
Cancer					
Base set	1,941/6,672	0.97 (0.92-1.02)‡	-	1.00 (0.95-1.06)	0.18
+urin. Na & K	1,912/6,544	0.96 (0.92-1.01)‡	1.00 (0.99-1.00)	1.00 (0.95-1.06)	0.19
+vitamins	1,820/6,243	0.95 (0.90-1.00)‡	1.00 (0.98-1.01)	1.01 (0.96-1.06)	0.08
+Fe status & Mg	1,451/5,050	0.97 (0.92-1.03)	1.01 (1.00-1.02)	0.99 (0.93-1.05)	0.69
+$\delta^{13}C$ & $\delta^{15}N$	166/564	1.08 (0.75-1.54)‡	1.04 (0.83-1.30)	1.01 (0.84-1.23)	0.39
+serum phytoestr.	383/1,085	0.88 (0.78-0.98)	0.93 (0.87-0.99)#	0.94 (0.83-1.05)	0.51
+urin. phytoestr.	560/1,522	0.92 (0.84-1.01)¶	0.97 (0.91-1.02)	0.91 (0.83-1.00)	0.94
+urin. sugars	611/1,756	0.97 (0.89-1.06)	0.99 (0.96-1.02)	1.04 (0.95-1.14)	0.14
Metabolomics	2,548/10,937	0.99 (0.95-1.03)	1.01 (0.95-1.07)	1.00 (0.96-1.04)	0.94
+base set	1,170/4,020	0.95 (0.89-1.01)‡	0.97 (0.92-1.03)	0.97 (0.91-1.04)	0.30
Type 2 diabetes					
Base set	752/6,675	0.78 (0.71-0.84)‡	-	0.90 (0.82-0.97)‡	<0.01
+urin. Na & K	736/6,551	0.78 (0.72-0.85)‡	1.00 (0.99-1.01)	0.90 (0.82-0.98)‡	<0.01
+vitamins	711/6,260	0.76 (0.70-0.83)‡¶	1.01 (0.98-1.03)	0.90 (0.82-0.98)‡	<0.01
+Fe status & Mg	564/5,014	0.83 (0.75-0.91)‡	1.02 (1.00-1.03)	0.93 (0.84-1.03)	0.03
+$\delta^{13}C$ & $\delta^{15}N$	299/524	0.63 (0.47-0.84)‡	1.00 (0.85-1.18)	0.71 (0.51-0.99)‡¶	0.60
+serum phytoestr.	120/1,070	0.84 (0.66-1.06)‡	0.95 (0.84-1.07)	0.91 (0.74-1.14)‡	0.73
+urin. phytoestr.	146/1,503	0.84 (0.70-1.02)‡	0.91 (0.80-1.04)	1.04 (0.84-1.28)‡	0.04
+urin. sugars	155/1,783	0.86 (0.71-1.05)	1.01 (0.95-1.08)	0.92 (0.76-1.10)	0.73
Metabolomics	749/11,134	0.84 (0.78-0.91)	0.98 (0.86-1.12)	0.95 (0.88-1.03)	0.04
+base set	266/3,989	0.87 (0.75-1.00)‡	1.01 (0.89-1.15)	0.93 (0.81-1.07)	0.12

Abbreviations: CVD – cardiovascular disease; FFQ – food frequency questionnaire; HR – hazard ratio; n – number of cases; N – number of participants; phytoestr – phytoestrogens; urin – urinary

*The base set comprised of plasma carotenoids and phospholipid fatty acids. Associations were adjusted for age, sex, smoking status, education, marital status, Townsend index, current employment, physical activity, personal and family history of chronic disease, use of antihypertensive medication, use of dietary supplements, seasonality, time since last meal prior to blood draw, body mass index and waist circumference, and in women: menopausal status and use of hormone replacement therapy. There were 4-10% fewer participants in the FFQ-based analysis.

†P-value for difference between HRs for biomarker score versus FFQ.

‡Proportional hazards assumption not met (Schoenfeld residuals test p-value for the main exposure <0.05)

¶Evidence of departure from a linear association (p-value for non-linearity <0.05)

#HR ratio statistically significantly different from 1; see Methods 6.4.5 for details

Table 6.8 Associations between the Mediterranean diet pyramid score and mortality per standard deviation of adherence: biomarker-based assessment and self-report in the EPIC-Norfolk study

Outcome and set of biomarkers*	n/N	Biomarker score		FFQ	
		HR (95% CI)	HR ratio to base set (95% CI)	HR (95% CI)	p†
CVD mortality					
Base set	1,641/6,555	0.93 (0.88-0.98)	-	0.98 (0.93-1.03)	0.27
+urin. Na & K	1,601/6,434	0.92 (0.87-0.97)	0.99 (0.98-0.99)#	0.98 (0.92-1.03)	0.18
+vitamins	1,518/6,143	0.90 (0.85-0.96)	0.97 (0.96-0.99)#	0.97 (0.92-1.03)	0.07
+Fe status & Mg	1,188/4,890	0.92 (0.87-0.99)	1.00 (0.99-1.01)	0.97 (0.91-1.03)	0.21
+$\delta^{13}C$ & $\delta^{15}N$	149/527	0.96 (0.73-1.26)	1.08 (0.88-1.33)	1.27 (0.87-1.86)	0.32
+serum phytoestr.	240/1,035	0.97 (0.83-1.14)	0.98 (0.90-1.08)	1.07 (0.92-1.26)	0.31
+urin. phytoestr.	308/1,456	0.88 (0.77-1.01)	0.99 (0.92-1.07)	1.01 (0.88-1.17)	0.11
+urin. sugars	464/1,754	0.90 (0.81-0.99)	0.96 (0.93-0.99)#	0.99 (0.90-1.10)	0.41
Metabolomics	1,878/11,032	0.94 (0.89-0.99)	0.99 (0.92-1.06)	0.97 (0.92-1.02)	0.46
+base set	993/3,961	0.90 (0.84-0.96)¶	0.97 (0.92-1.04)	0.96 (0.90-1.03)	0.11
Cancer mortality					
Base set	1,197/6,672	0.93 (0.88-0.99)	-	1.01 (0.95-1.08)	0.06
+urin. Na & K	1,175/6,544	0.93 (0.87-0.99)	1.00 (0.99-1.01)	1.01 (0.95-1.08)	0.07
+vitamins	1,121/6,243	0.90 (0.85-0.97)	0.99 (0.97-1.01)	1.02 (0.95-1.08)	0.02
+Fe status & Mg	899/5,050	0.93 (0.86-1.00)	1.00 (0.99-1.02)	1.00 (0.93-1.08)	0.19
+$\delta^{13}C$ & $\delta^{15}N$	101/564	0.97 (0.63-1.50)	1.05 (0.77-1.41)	1.05 (0.68-1.64)‡	0.33
+serum phytoestr.	215/1,085	0.85 (0.73-0.99)	0.92 (0.84-1.01)	0.95 (0.81-1.10)	0.66
+urin. phytoestr.	320/1,522	0.93 (0.82-1.05)¶	0.98 (0.91-1.06)	0.87 (0.76-1.00)	0.45
+urin. sugars	393/1,756	0.93 (0.83-1.04)	0.98 (0.94-1.01)	1.01 (0.91-1.13)	0.23
Metabolomics	1,424/10,937	0.97 (0.92-1.03)	1.02 (0.95-1.11)	1.01 (0.95-1.07)	0.54
+base set	740/4,020	0.94 (0.87-1.02)	0.99 (0.93-1.07)	1.02 (0.93-1.10)	0.07
All-cause					
Base set	3,724/7,017	0.92 (0.89-0.96)‡	-	0.96 (0.93-1.00)	0.13
+urin. Na & K	3,634/6,882	0.92 (0.88-0.95)‡	0.99 (0.99-1.00)#	0.96 (0.93-1.00)	0.09
+vitamins	3,458/6,572	0.90 (0.86-0.93)‡	0.98 (0.97-0.99)#	0.97 (0.93-1.00)	<0.01
+Fe status & Mg	2,760/5,277	0.94 (0.90-0.98)‡	1.00 (1.00-1.01)	0.97 (0.93-1.01)	0.14
+$\delta^{13}C$ & $\delta^{15}N$	315/583	0.98 (0.80-1.20)	1.07 (0.92-1.24)	1.09 (0.87-1.36)	0.17
+serum phytoestr.	587/1,110	0.95 (0.86-1.04)	0.94 (0.89-0.99)#	1.01 (0.92-1.11)	0.34
+urin. phytoestr.	780/1,558	0.92 (0.85-1.00)	0.97 (0.92-1.01)	0.97 (0.89-1.06)	0.52
+urin. sugars	1,105/1,845	0.89 (0.83-0.95)¶	0.98 (0.96-1.00)	0.95 (0.89-1.02)	0.36
Metabolomics	4,421/11,518	0.94 (0.91-0.97)	1.00 (0.95-1.05)	0.96 (0.93-0.99)	0.50
+base set	2,258/4,212	0.90 (0.86-0.95)	0.98 (0.94-1.03)	0.96 (0.91-1.00)	0.04

Abbreviations: CVD – cardiovascular disease; FFQ – food frequency questionnaire; HR – hazard ratio; n – number of cases; N – number of participants; phytoestr – phytoestrogens; urin – urinary

*The base set comprised of plasma carotenoids and phospholipid fatty acids. Associations were adjusted for age, sex, smoking status, education, marital status, Townsend index, current employment, physical activity, personal and family history of chronic disease, use of antihypertensive medication, use of dietary supplements, seasonality, time since last meal prior to blood draw, body mass index and waist circumference, and in women: menopausal status and use of hormone replacement therapy. There were 4-10% fewer participants in the FFQ-based analysis.

†P-value for difference between HRs for biomarker score versus FFQ.

‡Proportional hazards assumption not met (Schoenfeld residuals test p-value for the main exposure <0.05)

¶Evidence of departure from a linear association (p-value for non-linearity <0.05)

#HR ratio statistically significantly different from 1; see Methods 6.4.5 for details

6.5.5 Diet-disease associations: model assumptions

The assumption of proportionality of hazards throughout the follow-up was violated for multiple associations between the biomarker scores and incident disease outcomes, as well as for all-cause mortality. By contrast, it was not met only for incident T2D in case of the FFQ-based MDS-pyramid score (**Tables 6.7 and 6.8**).

Table 6.9 shows the results for the base set nutritional biomarker score and metabolomic scores by different durations of follow-up. The association of the nutritional biomarker score with incident CVD was stably inverse across the time bands of <10 years, 10-<20 years, and 20-25 years. The HRs (95% CI) per SD were 0.94 (0.89-<1.00), 0.94 (0.88-0.99), 0.93 (0.93-0.97), thus suggesting that the violation of proportional hazards assumption was inconsequential for interpretation of the average HR and may have arisen due to interaction with age when modelling the association on the timescale of participants' attained age. The results stratified by categories of baseline age of <60, 60-<70 and ≥70 years were consistent with this possibility, with stratum-specific HRs per SD (95% CI) of 0.93 (0.88-0.98), 0.91 (0.86-0.96) and 0.96 (0.90-1.03), respectively. For the remaining combinations of types of biomarker scores and incident outcomes, the inverse associations were stronger during the first 10 years of follow-up, and weaker or null thereafter (**Table 6.9**), and there was no indication of effect modification by age (results not shown). The combinations of nutritional biomarker scores and outcomes identified in Tables 6.7-6.8 as having violated the proportional hazards assumption which are not shown in Table 6.9 followed similar patterns of HR point estimates across the time bands of the follow-up as the base set biomarker score for a given outcome.

The assumption of linear associations was not met for several results (**Tables 6.7 and 6.8**). These results followed a similar pattern of a linear inverse association up to approximately the 50th percentile of biomarker score distribution, followed by levelling off of the inverse relationship, and the upper bound of the 95% CI eventually exceeding the null towards the end of the biomarker score distribution. **Figure 6.1** shows an example of such a non-linear association for the biomarker score comprised of the base set of circulating nutritional biomarkers and urinary sodium and potassium as the result with the highest number of participants among those with evidence of departure from linearity (other results not shown).

Table 6.9 Associations between biomarker scores of the Mediterranean diet and incident outcomes with violated proportional hazards assumption in the EPIC-Norfolk study

Outcome and biomarker score*	Main result	Duration of follow-up		
		<10 years	10-<20 years	20-25 years
CVD incidence				
Nutritional				
n/N	4,466/6,554	1,566/6,554	923/4,507	1,977/3,334
HR per 1 SD (95% CI)	0.93 (0.90-0.97)	0.94 (0.89-<1.00)	0.94 (0.88-0.99)	0.93 (0.93-0.97)
Metabolomic				
n/N	6,689/11,032	1,589/11,032	1,363/ 9,026	3,737/ 7,244
HR per 1 SD (95% CI)	0.98 (0.95-1.01)	0.93 (0.88-0.98)	0.98 (0.94-1.03)	1.00 (0.97- 1.04)
Cancer incidence				
Nutritional				
n/N	1,941/6,672	774/6,672	550/5,157	617/3,637
HR per 1 SD (95% CI)	0.96 (0.91-1.01)	0.90 (0.83-0.97)	1.00 (0.92-1.09)	1.02 (0.93-1.12)
Metabolomic-nutritional				
n/N	1,170/4,020	459/4,020	334/3,114	377/2,197
HR per 1 SD (95% CI)	0.94 (0.88-<1.00)	0.89 (0.81-0.98)	0.99 (0.89-1.10)	0.96 (0.86-1.07)
T2D incidence				
Nutritional				
n/N	752/6,675	178/6,675	211/5,263	363/3,686
HR per 1 SD (95% CI)	0.74 (0.68-0.80)	0.68 (0.58-0.80)	0.81 (0.69-0.94)	0.77 (0.67-0.87)
Metabolomic-nutritional				
n/N	266/3,989	26/3,989	37/3,203	143/2,240
HR per 1 SD (95% CI)	0.82 (0.71-0.95)	0.64 (0.43-0.94)	0.87 (0.70-1.07)	0.86 (0.69-1.07)
All-cause mortality				
Nutritional				
n/N	3,724/7,017	1,422/7,017	1,711/5,595	591/3,881
HR per 1 SD (95% CI)	0.92 (0.90-0.95)	0.90 (0.85-0.95)	0.95 (0.92-0.99)	0.92 (0.84-1.02)

Abbreviations: CVD – cardiovascular disease; HR – hazard ratio; n – number of cases; N – number of participants; SD – standard deviation; T2D – type 2 diabetes

*The nutritional biomarker score was derived using the base set of plasma carotenoids and phospholipid fatty acids to predict adherence to the Mediterranean diet. The metabolomic score was derived using untargeted plasma metabolomics. The metabolomic-nutritional biomarker score combined the two sources of biomarkers. Results are presented only for exposure-outcome pairs for which the proportional hazards assumption of the Cox regression model was violated (Schoenfeld residuals test p-value <0.05). The violation was not detected for CVD-specific or cancer-specific mortality. Associations were adjusted for age, sex, smoking status, education, marital status, Townsend index, current employment, physical activity, personal and family history of disease, use of antihypertensive medication, use of dietary supplements, seasonality, time since last meal prior to blood draw, body mass index and waist circumference, and in women: menopausal status and use of hormone replacement therapy. Continuous covariates were entered into the model as restricted cubic splines with five knots, except for time since last meal (log-transformed) and age (splines with seven knots). Duration of the follow-up was used as the underlying time variable.

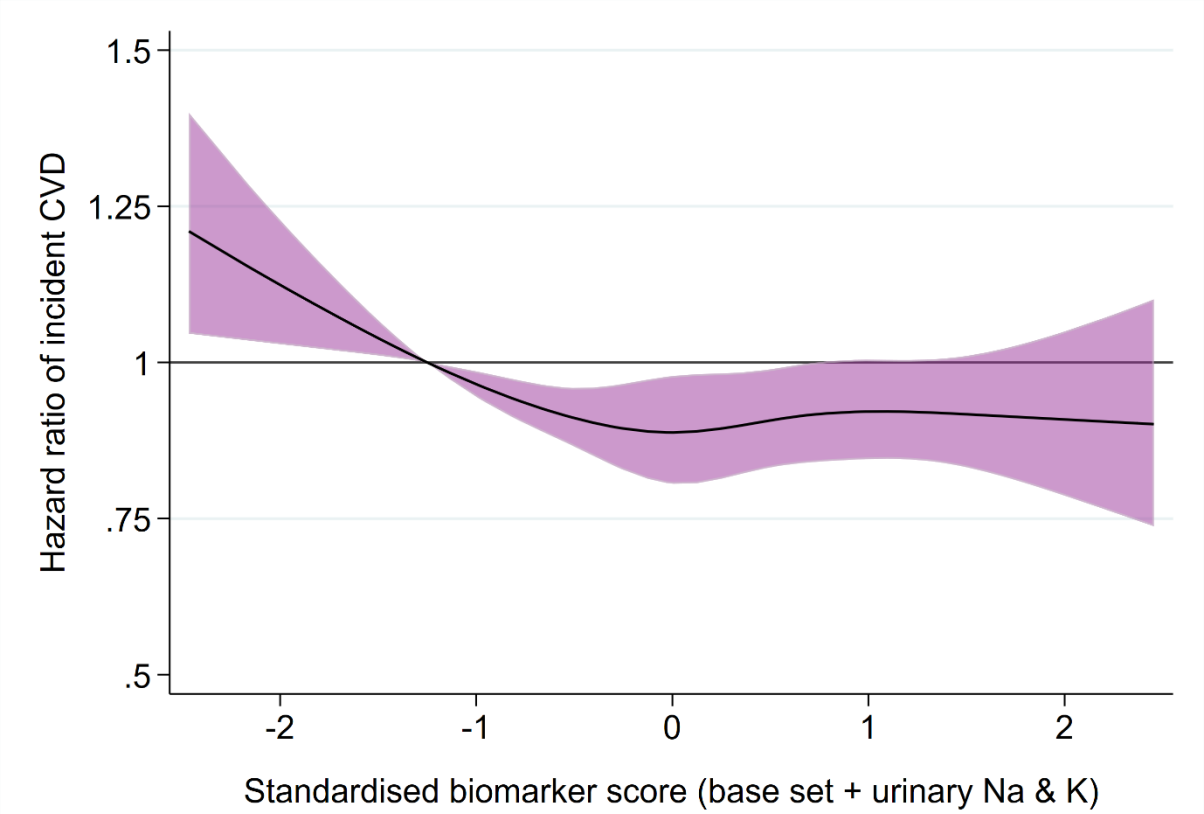


Figure 6.1 Non-linear association between a nutritional biomarker score of the Mediterranean diet and incidence of cardiovascular disease in the EPIC-Norfolk study

The biomarker score comprised of a base set of plasma carotenoids and phospholipid fatty acids and urinary sodium and potassium. Associations were adjusted for age, sex, smoking status, education, marital status, Townsend index, current employment, physical activity, personal and family history of chronic disease, use of antihypertensive medication, use of dietary supplements, seasonality, time since last meal prior to blood draw, body mass index and waist circumference, and in women: menopausal status and use of hormone replacement therapy. The analysis included 4,381 incident cases which occurred in a sample of 6,433 participants. The 10^{th} percentile of the biomarker score distribution was used as reference. The p-value for non-linearity was 0.030.

6.5.6 Sensitivity analyses

Analyses of the association between the base set biomarker score and incidence of CVD as case-control studies replicated the inverse association from the main analysis. The odds ratio per 1 SD of the biomarker score was 0.93 (95% CI: 0.87-0.99) for an unmatched case-control sample and was stable across all combinations of modelling baseline age and follow-up time as linear terms or restricted cubic splines. The HR was 0.92 (0.89-0.96) when modelling the data as a nested case-control study.

The inverse association of the base set biomarker score with incidence of T2D (HR per SD = 0.78, 95% CI: 0.71-0.84) was robustly replicated in the largely non-overlapping sample of a T2D case-cohort study with outcome ascertainment from additional sources (HR per SD = 0.81, 95% CI: 0.73-0.91; n cases/n total: 421/699). Within the T2D case-cohort sample, the inverse association for the biomarker score additionally incorporating stable isotopes from the primary analysis (HR per SD = 0.63, 95% CI: 0.47-0.84; n cases/n total: 299/524) became marginally significant when using the more comprehensive outcome ascertainment (HR per SD = 0.75, 95% CI: 0.56-1.00, p-value = 0.051; seven cases recoded to non-cases).

Neither the base set nutritional biomarker score nor the metabolomic score were associated with the negative control outcome of accidental deaths. The HRs (95% CI) per SD were 1.23 (0.96-1.57) and 0.96 (0.75-1.22), respectively, based on 52 and 70 deaths.

Associations between the base set nutritional biomarker score and the metabolomic score with incidence of CVD, cancer, T2D, and all cause-mortality were robust to sensitivity analyses which evaluated the impact of confounding by use of dietary supplements, reverse causation, use of higher penalisation in derivation of the scores ($\lambda_{+1\ SE}$), and impact of imputation of missing metabolomic data on stability of associations with incident outcomes. The inverse association between the nutritional biomarker score with serum phytoestrogens and incidence of cancer remained statistically significant after exclusion of the first two years of follow-up (HR per SD = 0.85, 95% CI: 0.76-0.95). Some of these analyses resulted in attenuation to the null of the results for CVD-specific and cancer-specific mortality (**Table 6.10**). Associations for the metabolomic score were reliably replicated across multiply-imputed datasets which were not used for derivation of the score in all but two out of 54 dataset-outcome combinations (CVD mortality).

Adding fatty acids in absolute concentrations to the set of predictors of plasma carotenoids and fatty acids in molar percentages did not have a meaningful impact on biomarker

performance (cross-validated r with the MDS-pyramid score = 0.41) or its association with incident CVD (HR per SD = 0.96, 95% CI: 0.92-0.99). Re-deriving the biomarker score containing urinary sugars after exclusion of participants with concentrations outside of limits of quantification modestly improved biomarker performance (cross-validated r = 0.40 versus r = 0.37 without exclusion). It did not materially affect the results for any of the outcomes, except for incident T2D for which the inverse association became statistically significant (HR per SD = 0.77, 95% CI: 0.61-0.96; n cases/n total: 99/1,050).

Table 6.10 Associations of biomarker scores of the Mediterranean diet with incidence of cardiovascular disease, cancer, type 2 diabetes, and mortality in the EPIC-Norfolk study: sensitivity analyses

Outcome and analysis	Nutritional biomarker score†		Metabolomic biomarker score	
	n/N	HR per 1 SD (95% CI)	n/N	HR per 1 SD (95% CI)
Cardiovascular disease incidence				
Main result	4,466/6,554	0.96 (0.92-0.99)	6,689/11,032	0.97 (0.94-0.99)
Imputation datasets (range)‡		n/a		0.96 (0.93-0.98), 0.97 (0.94-0.99)
Increased penalisation ($\lambda_{+1\ SE}$)¶		0.93 (0.90-0.96)		0.96 (0.93-0.98)
Excluding users of supplements	2,459/3,592	0.94 (0.90-0.99)	3,752/ 6,191	0.95 (0.92-0.98)
Excluding first 2 years of follow-up	4,298/6,296	0.96 (0.92-0.99)	6,583/ 10,865	0.97 (0.94-0.99)
Cancer incidence				
Main result	1,941/6,672	0.97 (0.92-1.02)	2,548/10,937	0.99 (0.95-1.03)
Imputation datasets (range)‡		n/a		0.98 (0.94-1.02), 0.99 (0.95-1.03)
Increased penalisation ($\lambda_{+1\ SE}$)¶		0.93 (0.89-0.98)		0.99 (0.95-1.03)
Excluding users of supplements	1,103/3,717	0.97 (0.91-1.03)	1,494/6,208	0.96 (0.91-1.01)
Excluding first 2 years of follow-up	1,824/6,436	0.97 (0.92-1.02)	2,482/10,802	0.98 (0.94-1.02)
Type 2 diabetes incidence				
Main result	752/6,675	0.78 (0.71-0.84)	749/11,134	0.84 (0.78-0.91)
Imputation datasets (range)‡		n/a		0.82 (0.76-0.88), 0.85 (0.78-0.92)
Increased penalisation ($\lambda_{+1\ SE}$)¶		0.69 (0.63-0.75)		0.83 (0.77-0.90)
Excluding users of supplements	450/3,683	0.77 (0.69-0.85)	427/6,260	0.84 (0.76-0.94)
Excluding first 2 years of follow-up	742/6,476	0.78 (0.72-0.85)	745/11,013	0.85 (0.78-0.91)
Cardiovascular disease mortality				
Main result	1,641/6,555	0.93 (0.88-0.98)	1,878/11,032	0.94 (0.89-0.99)
Imputation datasets (range)‡				0.94 (0.90-0.99), 0.96 (0.91-1.00)#
Increased penalisation ($\lambda_{+1\ SE}$)¶		0.92 (0.87-0.98)		0.93 (0.88-0.98)
Excluding users of supplements	930/3,592	0.94 (0.87-1.01)	1,061/6,191	0.92 (0.86-0.99)
Excluding first 2 years of follow-up	1,563/6,383	0.93 (0.87-0.98)	1,831/10,923	0.94 (0.90-0.99)
Cancer mortality				
Main result	1,197/6,672	0.93 (0.88-0.99)	1,424/10,937	0.97 (0.92-1.03)
Imputation datasets (range)‡		n/a		0.97 (0.91-1.03), 0.96 (0.91-1.00)
Increased penalisation ($\lambda_{+1\ SE}$)¶		0.90 (0.84-0.96)		0.96 (0.90-1.01)
Excluding users of supplements	680/3,717	0.92 (0.85-1.00)	825/6,208	0.95 (0.88-1.02)
Excluding first 2 years of follow-up	1,145/6,501	0.93 (0.87-1.00)	1,391/10,835	0.97 (0.92-1.03)
All-cause mortality				
Main result	3,724/7,017	0.92 (0.89-0.96)	4,421/11,518	0.94 (0.91-0.97)
Imputation datasets (range)‡				0.94 (0.91-0.97), 0.95 (0.92-0.98)
Increased penalisation ($\lambda_{+1\ SE}$)¶		0.91 (0.87-0.94)		0.93 (0.90-0.96)
Excluding users of supplements	2,122/3,902	0.93 (0.88-0.97)	2,518/6,517	0.93 (0.89-0.97)
Excluding first 2 years of follow-up	3,515/6,808	0.92 (0.89-0.96)	4,290/11,387	0.94 (0.91-0.97)

Abbreviations: HR – hazard ratio; n – number of cases; N – number of participants n/a – not applicable; SD – standard deviation

*Associations were adjusted for age, sex, smoking status, education, marital status, Townsend index, current employment, physical activity, personal and family history of disease, use of antihypertensive medication, use of dietary supplements, seasonality, time since last meal prior to blood draw, body mass index and waist circumference, and in women: menopausal status and use of hormone replacement therapy. Continuous covariates were entered into the model as restricted cubic splines with five knots, except for time since last meal (log-transformed) and age (splines with seven knots). Duration of the follow-up was used as the underlying time variable.

†The nutritional biomarker score was derived using the base set of plasma carotenoids and phospholipid fatty acids to predict adherence to the Mediterranean diet.

‡Minimum and maximum estimates in imputation datasets which were not used for derivation of the metabolomic score.

¶The $\lambda_{+1\ SE}$ penalty selects the λ penalty as the largest λ value that is within 1 standard of the minimum of the cross-validation function, as opposed to the optimal λ selected via cross-validation, leading to selection of more parsimonious biomarker score equations.

#Associations were not statistically significant in two imputation datasets.

6.6 Discussion

This analysis consisted in evaluating the utility of different combinations of a broad range of nutritional and metabolomic biomarkers as composite biomarkers of the Mediterranean diet. Specifically, I tested how different groups of biomarkers might impact on associations of objectively measured adherence to this dietary pattern with disease incidence and mortality outcomes in subsamples of up to ~11,000 participants of the population-based EPIC-Norfolk study (n ~26,000).

The key findings were that a nutritional biomarker score of plasma carotenoids and phospholipid fatty acids validated in the external MedLey partial-feeding RCT was inversely associated with CVD incidence and mortality, T2D, all-cause mortality, and cancer mortality but not cancer incidence. Except for cancer mortality, these associations were replicated in partially overlapping subsamples using a metabolomic and a joint metabolomic-nutritional biomarker score. FFQ-based assessment of the Mediterranean diet revealed these associations only for incident CVD, T2D and all cause-mortality within the subsamples overlapping with the biomarker-based analysis. The FFQ-based associations were consistently weaker for incidence of T2D compared to a range of biomarker scores, for all-cause mortality in some comparisons, and were otherwise of similar magnitude as the biomarker-based associations. The effect sizes of association for a given outcome were broadly similar between different sets of biomarkers, though some outcome-specific differences emerged. Adding biomarkers of phytoestrogens to the base set of carotenoids and fatty acids modestly strengthened the inverse associations for incidence of CVD, all-cause and CVD mortality, and introduced an inverse association with incidence of cancer. Similarly, stronger inverse associations were observed for CVD- and all-cause mortality when adding urinary sodium and potassium and circulating vitamins, and urinary sugars for CVD mortality only. The magnitude of these differences was, however, mostly small. The most pronounced differences in effect sizes were driven by the outcome under investigation, where the inverse associations were substantially stronger for T2D than for the remaining endpoints. Performance of the biomarker scores as biomarkers of exposure was moderate, as indicated by cross-validated correlations with self-reported adherence to the Mediterranean diet up to 0.46 for the metabolomic score. Adding urinary sodium and potassium to plasma carotenoids and phospholipid fatty acids for derivation of the biomarker score in the EPIC-Norfolk study did not result in larger differences between the Mediterranean and habitual diet arms in the external validation in the MedLey trial than carotenoids and fatty acids alone.

6.6.1 Strengths

The primary strength of this study was the broad range of candidate nutritional biomarkers of the Mediterranean diet available, as well as the inclusion of metabolomic biomarkers. To my knowledge, this was the first investigation of comparative utility of these groups of biomarkers or their concurrent use as composite biomarkers of a dietary pattern and predictors of major disease outcomes. A further strength was the availability of external interventional data which allowed for objective evaluation of a base set of nutritional biomarkers for its validity as a biomarker score of the Mediterranean diet. This increased the credibility of diet-disease and diet-biomarker associations evaluated using this score and provided for an internal benchmark against which the impact of incorporation of other groups of analytes into biomarker scores could be investigated. The biomarker scores were derived within a cross-validation framework which allowed for analytical independence of the biomarker scores from personal dietary self-report. The measure of self-reported adherence to the Mediterranean diet used throughout the current study has been previously extensively evaluated in a cohort-wide analysis of EPIC-Norfolk which demonstrated its content validity through inverse associations with incident CVD and all-cause mortality.[304] Additionally, previous research suggested its suitability as a target for prediction from concentrations of nutritional (Chapter 5) and metabolomic biomarkers.[353] The EPIC-Norfolk study collected detailed information on lifestyle exposures and disease risk factors which enabled comprehensive adjustment for potential confounding factors in the above analyses.

Most of the analytes were measured at scale in large subsamples, thus allowing to robustly test associations of biomarker scores with major disease and mortality outcomes. The long follow-up of up to 25 years facilitated accrual of cases and increased the statistical power to evaluate the hypotheses. This included up to nearly 7,000 incident CVD cases in metabolomic analysis which was an order of magnitude larger than previously used in a limited number of published investigations on metabolomic profiling of dietary patterns and cardiometabolic outcomes.[60,354]

6.6.2 Limitations

This study was limited by non-random sampling of participants included in the core subsamples of participants with biomarker measurements. Namely, the metabolomics subsample was

assembled quasi-randomly with omission of most of the incident T2D cases occurring in approximately the first 15 years of follow-up. Conversely, the sample with carotenoids and fatty acids was enriched in incident CVD and cancer cases through 1:4 case-control matched sampling, though differences in baseline characteristics with the remaining EPIC-Norfolk participants were modest. Sensitivity analysis which entailed analysing the data as an unmatched or a nested case-control study confirmed the primary result of an inverse association between the biomarker score based on carotenoids and fatty acids and incidence of CVD. Generalisability of these findings to populations down-weighted within the sampling schemes remains unclear. Another limitation related to sampling of participants was the variable sample size for different groups of biomarkers and the incomplete overlap of individuals across different biomarker assays. This limited the statistical power in some analyses and may have resulted in a comparatively worse performance of elastic net regression in the derivation samples with smaller numbers of participants.[225] The widely varying sample sizes motivated my decision not to correct for multiple testing which would have disproportionately inflated the rate of false-negative results in the smaller samples. In turn, lack of correction of type I error rate increased the probability of false-positive findings.

Only data from baseline assessment were available for biomarkers and for covariates measured in all participants. Thus, I was unable to account for their changes during follow-up. Analyses stratified by follow-up time suggested attenuation of inverse associations between biomarker scores and some incident outcomes with increasing follow-up. This may have been in part driven by random within-person variation in biomarker concentrations and insufficiency of a single assessment of biomarkers, particularly given the long study duration. Single measurements of urinary biomarkers and circulating phytoestrogens used in the current study were likely inadequate to reflect habitual dietary exposures, and within-person variation may have exceeded between person-variation.[355–359] Three 24-h urine collections have been suggested as sufficient for epidemiological applications to capture long-term exposures for most nutritional biomarkers, and a larger number may be required with spot urine samples which are subject to diurnal variation in addition to day-to-day variation.[360] Consistent with this concern, biomarker scores incorporating phytoestrogen and urinary biomarkers mostly had poorer predictive performance for the Mediterranean diet outside of their derivation samples than biomarker scores derived using the more stable, lipid-soluble circulating biomarkers alone.[361]

Other potential sources of error include non-dietary regulation of concentrations of nutritional[43,80,235] and metabolomic biomarkers,[362] residual confounding and batch effects, degradation of biomarkers and metabolites with long-term storage, and reverse causality early in the follow-up in the prospective analyses. The latter was overall unlikely given the results of sensitivity analyses. However, it could not be ruled out for cancer mortality, for which exclusion of the first two years of follow-up attenuated inverse associations to the null. Standardisation of effect sizes may have limited the quantitative comparison of associations measured with different biomarker scores and the FFQ-based MDS-pyramid score,[236] though this was the only common scale that could be applied to all measures of exposure given the differential penalisation in the elastic net regression models.

6.6.3 Comparison with previous research: biomarkers of the Mediterranean diet

Expanding the pool of candidate biomarker predictors in the current study beyond plasma carotenoids and phospholipid fatty acids typically resulted in their selection into the biomarker scores. This contrasts with the results of the NPAAS-FS feeding study of individual habitual diet in participants of the WHI trial, which favoured selection of more parsimonious models focused on carotenoids and fatty acids, and rejected circulating tocopherols, retinol, folate, vitamin B_{12} and urinary nitrogen (novel study design described in Chapter 4.1.1).[50] The discrepancy may be attributable to differences in the analytical approaches (Chapter 3.6.3), measurement error of dietary self-report in EPIC-Norfolk, use of different indices of the Mediterranean diet, and divergent, population-specific associations between this dietary pattern and nutritional biomarkers. Of note, the WHI biomarker score included 24-hour urinary excretion of potassium which accounted for a sizeable 6% out of the total 44% of variance in adherence to the Mediterranean diet explained by biomarker concentrations. Adding urinary potassium and sodium to the base set of plasma carotenoids and fatty acids did not improve the cross-validated correlation of the biomarker score with the MDS-pyramid score in EPIC-Norfolk or external performance of the score in the MedLey trial. Application of urinary excretion of these cations in EPIC-Norfolk was limited by their estimation from single spot urine samples which had poor agreement with measured habitual excretion. The equations used for estimation have been designed for prediction of 24-h urinary output on the day of collection of the spot sample, so using habitual excretion as the reference instrument was likely to underestimate their performance for the intended purpose.[356] Correlation coefficients between

estimated urinary excretion and the reference in EPIC-Norfolk were moderately lower than those between estimated urinary excretion and average intakes during two-weeks of the feeding study in WHI for sodium (0.37 versus 0.45) and potassium (0.31 versus 0.44).[314] Another potential explanation for the lack of added value of these biomarkers in EPIC-Norfolk was the introduction or amplification of confounding by energy intake when using estimated absolute daily outputs. Notably, the WHI biomarker score incorporated energy expenditure measured by the doubly labelled water technique, thus effectively adjusting 24-h urinary potassium for energy intake. Remaining groups of nutritional biomarkers evaluated in EPIC-Norfolk for enhancing performance of the biomarker scores have not been previously evaluated as markers of adherence to the Mediterranean diet in adults. Urinary sucrose has been reported to be negatively associated with this dietary pattern in a pan-European study in children,[363] which is consistent with inclusion of this biomarker into the score in EPIC-Norfolk when expanding the pool of candidate predictors to urinary sugars.

The metabolomic score shared some common features with previous reports of plasma metabolomic profiling of the Mediterranean diet.[60,364] This included an abundance of phosphatidylcholines, lysophospholipids plasmalogens and metabolites of amino acid metabolism with variable directionality of scoring, and positively scored metabolites of nutritional biomarkers (docosahexaenoate, oxalate, threonate), glycerate, and xenobiotics stachydrine (proline betaine, a biomarker of citrus fruits[365,366]) and ergothioneine. The latter is synthesised by fungi and available in high amounts in edible mushrooms.[367] Ergothioneine is also ubiquitously present in plants, albeit in an order-of-magnitude lower concentrations.[367] It has been identified as the most robust correlate of a 'healthy' data-driven dietary pattern out of 111 metabolites in the Malmö Diet and Cancer Study and a negative predictor of incidence of CVD and CVD-specific and all-cause mortality in this cohort.[368] Among reports on metabolomic profiling of the Mediterranean diet, only two studies reported on the performance of composite metabolomic biomarkers in terms of correlation with dietary self-report. The Fenland Study used the MDS-pyramid score and found a similar moderate correlation (r = 0.43)[353] as identified in the current study (r = 0.46) for the same dietary index. A metabolomic score derived in the PREDIMED trial baseline sample using the Mediterranean Diet Adherence Screener (MEDAS) had a correlation with self-reported adherence of 0.37 in the derivation sample and 0.24-0.37 in test samples of individuals after 12 months of interventions with the Mediterranean diet and participants of American cohort studies.[297] In contrast to the investigation in EPIC-Norfolk, the metabolites available in previous research were only[353] or

almost exclusively[297] endogenous and the overall number of metabolites was substantially smaller (175[353] and 67[297] versus 958). Implications of these differences in metabolomic platforms are unclear, though a larger number of exogenous analytes may plausibly improve content validity of dietary metabolomic scores.

Among metabolites not previously reported on in humans in relation to the Mediterranean diet, there were several plant-derived xenobiotics which were either incorporated into the EPIC-Norfolk metabolomic score with positive weights or had positive multivariable-adjusted correlations with the MDS-pyramid score following Bonferroni correction, including: 4-vinylphenol sulfate (candidate biomarker of nuts[369]), dimethyl sulfone (potential dietary sources: rye, onions, asparagus and cabbage[370]), eugenol sulfate (ubiquitous in plant oils, non-specific;[371,372] increased in urine after consumption of curry in UK participants, potentially driven by the use of clove as a spice[373]), homostachydrine (pipecolic acid betaine; increased after a whole-grains intervention[374]), indolin-2-one (higher in mammary glands after a Mediterranean versus Western diet in a primate model[375]), N-acetylalliin (potential dietary sources: beans, mushrooms and onions[376]), S-allylcysteine (candidate urinary biomarker of garlic;[376] 4-fold increase in plasma after an intervention with navy beans[377]), thymol sulfate (plausible biomarker of thyme intake[378]) and 4-allylphenol sulfate and methyl glucopyranoside ($\alpha + \beta$). The latter two have been previously associated with fruit intake and were increased in serum following DASH diet interventions relative to control diets.[379–381] Some metabolites of amino acid metabolism which were related to the Mediterranean diet in EPIC-Norfolk have also been linked to specific dietary exposures. N-methylproline and tryptophan betaine increased in response to the DASH diet intervention, and they are present in citrus fruit and lentils, respectively.[380]

The finding of detectable acesulfame in plasma of EPIC-Norfolk participants being related to higher adherence to the Mediterranean diet is surprising, as this non-nutritive sweetener is typically present in processed foods which are not compatible with the Mediterranean dietary pattern. This may have been a spurious association driven by a complex mechanism of missingness or poor accuracy of the assay for this metabolite as only 14% of participants in the metabolomics sample had non-missing data on acesulfame. Assuming identification of a true relationship, more frequent use of acesulfame-containing low-calorie sweeteners in tea and coffee may have been associated with higher adherence to the Mediterranean diet. Of note, the metabolomic scores incorporated a small number of biomarkers of exposure to medications, including several metabolites of acetaminophen, 4-hydroxycoumarin, and quinine and one of

its metabolites. Therapeutic uses of quinine had been limited to treatment of malaria and nocturnal leg cramps, thus presence and variation in quinine levels in plasma of EPIC-Norfolk participants was likely attributable to its consumption as a food additive, primarily in soft drinks.[382] The remaining metabolites were likely related to therapeutic use of acetaminophen and anticoagulants, thereby facilitating prediction of diet based on non-causal associations. Alternatively, veterinary residues in foods of animal origin may have contributed to exposure to acetaminophen.[383,384] Incorporation of drug metabolites into biomarker scores of dietary exposures is not unique to the current research[385] and it is unlikely to negatively impact the internal performance of a composite biomarker, however, it may decrease its external generalisability.

6.6.4 Diet-disease associations: implications and comparison with previous research

To my knowledge, metabolomic profiling of dietary patterns with applications to associations with disease risk have so far been published for non-Mediterranean dietary patterns and incidence of T2D (Chapter 3.6.4) and for the Mediterranean diet and risk of CVD in the joint analysis of case-cohort samples of the PREDIMED study and case-control studies nested within prospective cohorts in the USA.[297] The study found substantially stronger inverse associations of the metabolomic score with incidence of CVD (HR range per 1 SD: 0.71-0.86; n_{cases} range: 143-351)[297] than the current result in EPIC-Norfolk. However, derivation of the score in case-cohort samples of CVD and T2D without exclusion of the oversampled incident cases may have amplified the magnitude of the inverse association (Chapter 4.6.3). Use of a targeted metabolomic assay of endogenous metabolites and cotinine may have further contributed to this issue by focusing on biomarkers with potentially higher aetiological relevance compared to use of a metabolomic platform with a sizeable proportion of exogenous metabolites.

Nutritional biomarkers have not been used at scale for epidemiological assessment of dietary patterns beyond the work presented in this PhD thesis. The WHI investigators have recently published on associations of the Healthy Eating Index-2010 (HEI-2010) with chronic disease risk[386] following calibration of HEI-2010 with an equation developed in the WHI ancillary feeding study (NPAAS-FS) based on a nutritional biomarker score of HEI-2010 and personal characteristics (BMI, race, education, dietary supplement use, recreational physical activity).[50] Multivariable-adjusted HRs (95% CI) per a 20% increment in HEI-2010 were substantially

stronger with than without calibration. The estimates for CVD incidence were 0.82 (0.69, 0.98) versus 0.95 (0.93, 0.97) for calibrated versus uncalibrated analysis, 0.74 (0.63, 0.87) versus 0.94 (0.92, 0.96) for CVD mortality, 0.81 (0.76, 0.87) versus 0.96 (0.94, 0.98) for cancer incidence, 0.77 (0.71, 0.85) versus 0.94 (0.91, 0.96) for cancer mortality, and 0.45 (0.36, 0.55) versus 0.91 (0.90, 0.93) for incidence of T2D.[386] These models did not include measures of adiposity, and further adjustment for BMI substantially attenuated the associations of calibrated HEI-2010, and only the results for cancer incidence, cancer mortality and T2D incidence remained statistically significant at 0.92 (0.85, 0.99), 0.88 (0.78, 0.98) and 0.85 (0.74 0.97), respectively. This contrasts with the results of the current work, both in terms of the striking differences between calibrated and uncalibrated estimates (as compared to the differences between FFQ-based and biomarker-based results in EPIC-Norfolk), and the impact of further adjustment for adiposity of a multivariable model, which was small or negligible in EPIC-Norfolk. Several lines of reasoning can be pursued to explain these discrepancies. First, MDS-pyramid and HEI-2010 scores may have had study-specific relationships with outcomes of interest and the associated confounding structures in relation to adiposity. Of note, however, neither the Mediterranean diet in EPIC-Norfolk,[387] nor HEI-2010 in the WHI[388] were associated with prevention of weight gain. Second, the relationship between the biomarker score and BMI in EPIC-Norfolk may have been mis-specified because uncorrelated errors between dietary self-report and biomarkers could not have been ensured (Chapter 3.6.2). Third, derivation of the WHI calibration equation was a two-stage process in which the biomarker score was first derived by regressing known HEI-2010 values (calculated from individualised diets provided to participants of the feeding study) on nutritional biomarkers and personal characteristics (cross-validated $R^2 = 0.40$), followed by establishing the calibration equation by regressing the biomarker score on HEI-2010 estimated from FFQ and personal characteristics ($R^2 = 0.40$; R^2 adjusted for correlation between repeated FFQs = 0.64).[50] Given that the overall aim of the calibration process is to estimate true values of HEI-2010, not the biomarker score per se, the proportion of variance explained in the true values of HEI-2010 by the calibrated estimates can be as low as the product of the R^2 parameters from the two equations (0.4^2 or 0.4×0.64). Thus, the overall R^2 for the two-stage process may have been below the internal biomarker score validation criterion of 0.36 (established based on the performance of urinary nitrogen as a biomarker of protein intake),[50] potentially decreasing the credibility of inference on risk estimates for calibrated HEI-2010. Fourth, application of participants' personal characteristics for calibration of HEI-2010 raises concerns about introducing unintended dependencies between the exposure and these characteristics when they are considered as

confounders in subsequent modelling of disease risk.[251] Notably, both the biomarker score and the calibration equation for HEI-2010 included BMI with negative coefficients and a R^2 of 0.12 was attributable to BMI in the calibration equation (30% of total R^2), which may have amplified the degree of attenuation of disease risk estimates upon additional adjustment for BMI in the WHI. Disentangling the role of adiposity in the relationship between dietary quality and risk of chronic disease has major consequences for translation of aetiological research into actionable evidence to inform public health interventions and policy. Given the early stage of research into objective measures of dietary quality, it may be prudent to focus on development of methods which do not incorporate established risk factors for cardiometabolic disease and cancer, and to utilise study designs with biomarker measurements undertaken in all participants.

Of note, I identified stronger inverse associations between the Mediterranean diet and incidence of T2D than for incidence or mortality from CVD and cancer or all-cause mortality. This was consistent across all types of exposure assessment, albeit the difference was less pronounced with FFQ-based assessment compared to biomarkers. Similar observations were previously reported in an analysis of case-control studies nested within Swedish cohorts which applied metabolomic profiling to a data-driven 'healthy' dietary pattern.[354] The HRs of incident coronary artery disease per 1 SD (95% CI) of the metabolomic score in the Malmö Diet and Cancer and the Malmö Preventive Project studies were 0.87 (0.77–0.99) and 0.86 (0.74–1.00), respectively. The corresponding estimates for T2D were 0.58 (0.52–0.66) and 0.54 (0.44–0.65).[354] It is unclear whether these differences and those reported in the current study reflect stronger underlying inverse associations between dietary quality and incidence of T2D than for other outcomes, or whether metabolomic and nutritional biomarker scores have stronger aetiological links with T2D via specific analytes incorporated into the scores. A cohort-wide analysis from the EPIC study in eight European countries found similar strength of inverse associations between the self-reported relative MDS (Chapter 3) and incidence of T2D and CVD, and a somewhat weaker inverse association for incident cancer.[389] By contrast, the EPIC-Potsdam centre reported inverse associations of similar magnitude as the current study between the relative MDS (HR per 1 SD = 0.92; 95% CI: 0.87–0.97) and the MDS-pyramid score (0.93; 0.88–0.98) with incidence of T2D, but null associations with incidence of cancer, myocardial infarction and stroke.[390] Meta-analytical evidence from prospective cohort studies on the Mediterranean diet suggests that high versus low adherence is inversely associated with the

above-discussed outcomes to a similar degree,[31,391–393] though this comparison is limited by lack of standardisation in the analyses between the outcomes.

Among other contributions, the current study demonstrated suitability of the MICE-pmm approach for imputation of metabolomic data under the scenario of multivariable predictive modelling.[343] I applied several enhancements to the previously published procedure,[343] such as derivation of treelet transform factors as auxiliary variables, which may have increased robustness of the imputation process. Correlations between metabolomic scores derived using the imputed data with dietary self-report were relatively high when applied to separately imputed test datasets with metabolomic measurements conducted as a separate batch. Diet-disease or diet-mortality associations evaluated using the metabolomic scores were remarkably stable across imputation datasets which were not used for derivation of the scores. These results raise the possibility that a single dataset imputed by MICE-pmm may be sufficient for complex analyses of metabolomic data, which would substantially decrease the computational and coding burden. However, this should be evaluated in formal investigations specifically addressing this issue and suitable sensitivity analyses should be exercised in future research.

6.6.5 Conclusions

These findings suggest that plasma carotenoids and phospholipid fatty acids, metabolomics, or a combination of these nutritional biomarkers and metabolomics have similar utility for derivation of objective measures of adherence to the Mediterranean diet and their application to associations with chronic disease risk or mortality. Adding additional nutritional biomarkers to plasma carotenoids and phospholipid fatty acids may yield stronger inverse associations for some outcome-biomarker group combinations. However, it may not improve the performance of the biomarker score for exposure assessment, and the performance may decrease when using biomarkers with known high within-person variation such as urinary biomarkers measured in spot samples.

Chapter 7

Outcome-wide associations of the Mediterranean diet with chronic diseases

Background: The Mediterranean diet has been investigated extensively in relation to incidence of major cardiovascular diseases (CVD) based on dietary self-report. Prospective evidence is limited on most CVD subtypes and non-CVD endpoints. Objective biomarkers have rarely been used to assess adherence to this dietary pattern in epidemiological studies.

Methods: I assessed associations of the Mediterranean diet with 27 incident disease outcomes, including cardiometabolic, respiratory, neurodegenerative, and ophthalmic conditions, site-specific cancers and total fractures in the Norfolk arm of the European Prospective Investigation into Cancer and Nutrition (n = 25,639). I used four measures of adherence to the Mediterranean diet: (i.) a nutritional biomarker score comprised of plasma carotenoids and phospholipid fatty acids derived in the MedLey randomised controlled trial (RCT) based on discrimination between participants under an intervention with the Mediterranean diet or continuation of habitual diet, (ii.) a score comprised of the same nutritional biomarkers ($n \leq$ 6,994 after exclusion of prevalent cases and participants with missing covariate data) and a (iii.) metabolomic biomarker score derived based on prediction of the Mediterranean Diet Score for adherence to pyramid-based guidelines (MDS-pyramid) in EPIC-Norfolk ($n \leq$ 11,555), and (iv.) the MDS-pyramid score derived from a food frequency questionnaire ($n \leq$ 22,850).

Findings Adjusting for potential confounders, inverse Bonferroni-corrected associations were detected by at least two measures of adherence to the Mediterranean diet for incidence of (HR range per 1 SD; measures): type 2 diabetes (0.78-0.92; all measures), chronic obstructive pulmonary disease (0.81-0.87; all measures except the RCT score), and heart failure (0.95-0.95; metabolomic score and MDS-pyramid). At the nominal two-sided $\alpha = 0.05$, inverse associations were detected with ≥2 measures for incidence of: ischaemic heart disease, renal disease, oesophageal and stomach cancers, and cataracts. There was no indication of a relationship between the Mediterranean diet and risk of disease across all methods of exposure assessment at the nominal p value for: atrial fibrillation, venous thrombosis, stroke, abdominal

aneurysm, cancers of the prostate, colon, ovaries and endometrium, Parkinson's disease, bone fractures and glaucoma.

Conclusions: These findings suggest that there is a degree of specificity of the Mediterranean diet to associations with incident disease outcomes. Exposure assessment with multiple methods with uncorrelated or partially uncorrelated errors consistently identified robust prospective associations of sizeable magnitude of adherence to the Mediterranean diet with type 2 diabetes and chronic obstructive pulmonary disease.

7.1 Background

Results from the previous chapter have shown that measures of adherence to the Mediterranean diet were inversely associated with risk of CVD and cancer mortality in the EPIC-Norfolk cohort. Such composite outcomes were useful for maximising statistical power in comparative analyses of different biomarker scores and FFQ-based assessment of this dietary pattern. However, they are of limited use for generating evidence of aetiological, clinical and public health relevance because associations of risk factors with specific outcomes of distinct pathophysiology may be heterogeneous.[394] Thus, investigations of biomarker scores of the Mediterranean diet with CVD-subtypes and site-specific cancers are warranted. Major new-onset CVDs and cancers have been studied extensively in prospective cohort studies in relation to the Mediterranean diet.[31,391,392] However, there have been comparatively fewer investigations into specific CVD and cancer outcomes with lower incidence rates,[31,391,392] as well as other less common diseases such as neurodegenerative[395] and respiratory conditions.[396] Concurrent analyses of risk of a range of heterogeneous diseases have rarely been considered in nutritional epidemiology.[397] This approach, known as outcome-wide epidemiology, offers additional benefits over investigations focusing on a single outcome or a group of closely related coditions.[398] It allows for a direct comparison of the role of a given exposure as a risk factor or an inverse correlate across multiple unrelated outcomes, which is helpful in identifying public health priorities, as well as testing hypotheses about specificity of exposure-outcome associations.

7.2 Aim

The aim of this chapter was to perform an outcome-wide analysis of prospective associations of all measures of adherence to the Mediterranean diet developed or used within this PhD thesis, with risk of chronic diseases.

7.3 Methods

Methods described in Chapter 6 were used throughout the current chapter.

I used four primary measures of adherence to the Mediterranean diet: the MedLey RCT nutritional biomarker score of discrimination between participants under intervention with the

Mediterranean diet and individuals who continued their habitual diet (Chapter 4), the nutritional and metabolomic biomarker scores of prediction of the FFQ-based MDS-pyramid (Chapter 6) and the MDS-pyramid score as previously developed in EPIC-Norfolk.[304] The nutritional biomarker scores were derived using circulating carotenoids and fatty acids. Urinary sodium and potassium additionally overlapped between the MedLey trial and EPIC-Norfolk; however, I made the a priori decision to exclude them from the primary analysis due to concerns over performance of single spot measures in EPIC-Norfolk (Chapter 6). The RCT biomarker score was based on pairwise interactions between biomarkers (Chapter 4), and the remaining biomarker scores were derived using linear terms (Chapter 5). All participants with non-missing MDS-pyramid score (n = 23,692) were eligible for inclusion in the analysis, and up to 22,850 individuals were available for prospective associations with disease risk following exclusions of baseline cases and missingness in covariate data. Sampling of participants with nutritional biomarkers and metabolomics has been described in Chapter 6.

I assessed associations of the above exposures with incidence of 27 common diseases, encompassing cardiovascular, metabolic, respiratory, ophthalmic and neurologic diseases, and cancers and fractures. The list of specific outcomes, definitions, and exclusion criteria of baseline cases are shown in **Table 7.1**. These diseases were previously selected for an outcome-wide analysis in EPIC-Norfolk based on maximising case numbers and availability in data linkage between Hospital Episode Statistics, cancer and death registries.[328] Analyses of incidence of breast and prostate cancers were restricted to women and men, respectively. For reference, I have additionally used composite outcomes of any first incident CVD and any first incident cancer.

Secondary measures of adherence to the Mediterranean diet included (i.) the RCT biomarker score re-derived using urinary sodium and potassium in the pool of candidate biomarkers for variable selection, (ii.) the RCT biomarker score based on linear biomarker terms only, and (iii.) the WHI biomarker score for prediction of the alternative Mediterranean diet score[50] calculated using the partially overlapping biomarkers between the WHI feeding study and EPIC-Norfolk (Chapter 5).

Table 7.1 Outcome definitions and criteria for exclusion of prevalent cases

Outcome	Definition, ICD codes	Criteria for exclusion of prevalent cases*
Cardiovascular		
Coronary heart disease	ICD-9: 410-414; ICD-10: I20-I25	Self-reported history of myocardial infarction or angina
Atrial fibrillation	ICD-9: 427.3; ICD-10: I48	Self-reported use of drugs consistent with treatment of atrial fibrillation: digitalis or vitamin K antagonists
Heart failure	ICD-9: 428; ICD-10: I50	Self-reported use of drugs consistent with treatment of heart failure: (i.) loop diuretics and digitalis or (ii.) angiotensin-converting enzyme inhibitors
Venous thrombosis	ICD-9: 451-453; ICD-10: I80-I82	Self-reported history of pulmonary embolism or deep vein thrombosis
Peripheral arterial disease	ICD-9: 440-448; ICD-10: I70-I79	Self-reported history of vascular disease
Cerebral stroke	ICD-9: 433-435; ICD-10: I63, I65, I66	Self-reported history of stroke
Haemorrhagic stroke	ICD-9: 430-432; ICD-10: I60-I62	Self-reported history of stroke
Abdominal aortic aneurysm	ICD-9: 441; ICD-10: I71	Self-reported history of myocardial infarction or stroke
Metabolic		
Type 2 diabetes	ICD-9: 250; ICD-10: E10-E14	Self-reported physician's diagnosis of type 2 diabetes, use of diabetic medications or following a diet modified due to diabetes
Renal disease	ICD-9: 580-589, 593; ICD-10: N00-N19,N25-N29	Estimated glomerular filtration rate < 50 ml/min/1.73m^2
Liver disease		Self-reported physician's diagnosis of a liver disease
Liver cancer	ICD-9: 155; ICD-10: C22	
Alcoholic disease	ICD-9: 571.0-571.3 ; ICD-10: K70	
Fibrosis/cirrhosis	ICD-9: 571.40, 571.5, 571.6; ICD-10: K74	
Other non-viral and non-toxic liver disease	ICD-9: 570, 571.41, 571.42, 571.49, 571.8, 572.0, 572.1, 572.2, 572.3, 572.4, 572.8; ICD-10: K72-73, K75-76	
Viral hepatitis	ICD-9: 070; ICD-10: B15-19	
Metabolism disorders	ICD-9: 275.0, 275.1, 273.4; ICD-10: E83.11, E83.01, E88.01	
Indicators of liver disease	ICD-9: 456.0-456.2, 789.5, V42.7; ICD-10: I85, R18, Z94.4	
Cancers		
Breast	ICD-9: 174; ICD-10: C50	Self-reported history of any cancer
Prostate	ICD-9: 185; ICD.10: C61	Self-reported history of any cancer
Non-melanoma skin	ICD-9: 173; ICD-10: C44	Self-reported history of any cancer
Colon	ICD-9: 153.0-153.9; ICD-10: C18	Self-reported history of any cancer
Rectal	ICD-9: 154.0-154.1, 159.0; ICD-10: C19-C20	Self-reported history of any cancer
Lung	ICD-9: 162.0-162.9; ICD-10: C33-C34	Self-reported history of any cancer
Oesophageal	ICD-9: 150.0-150.9; ICD-10: C15	Self-reported history of any cancer
Stomach	ICD-9: 151; ICD-10: C16	Self-reported history of any cancer
Ovarian	ICD-9: 183; ICD-10: C56-C57	Self-reported history of any cancer
Endometrial	ICD-9: 182.0; ICD-10: C54.1	Self-reported history of any cancer

Outcome	Definition, ICD codes	Criteria for exclusion of prevalent cases*
Other diseases		
Asthma	ICD-9: 493; ICD-10: J45-J46	Self-reported history of asthma
COPD	ICD-9: 490-496; ICD-10: J40-J44,J47	Self-reported history of bronchitis or emphysema
Parkinson's disease	ICD-9: 332; ICD-10: G20-G22	Self-reported use of drugs consistent with treatment of Parkinson's disease
Fracture, any site	ICD-10 codes: S02, S12, S22, S32, S42, S52, S62, S72, S82, S92, S120-S122, S127-S129, S220-S225, S228, S229, S320-SS325, S327, S328, S520-S529, S620-S627, S720-S729, S820-S829, S920-S929, T02, T08, T10	Self-reported history of fracture
Glaucoma	ICD-9 codes: 365; ICD-10 codes: H40	Self-reported history of glaucoma
Cataracts	ICD-10 codes: H25, H26	Self-reported history of cataracts

Abbreviations: COPD – chronic obstructive pulmonary disease; ICD – International Classification of Diseases

*Participants whose date of diagnosis preceded the date of entry into the EPIC-Norfolk cohort were also excluded from analyses of a given outcome (<1% of all excluded prevalent cases).

7.4 Statistical analysis

Statistical significance of the results is reported and described in relation to both the nominal two-sided $\alpha = 0.05$ and a Bonferroni-corrected threshold for analyses of 27 outcomes (two-sided $\alpha = 0.00185$) to facilitate identification of results which are unlikely to be chance findings arising because of multiple testing.

Cox regression modelling and multivariable adjustment described in Chapter 6 were used with modifications for family history of disease. Family histories of myocardial infarction, stroke, cancer and diabetes were removed from the model and adjustments were re-introduced in an outcome-specific manner as follows: family history of myocardial infarction for incident ischaemic heart disease, family history of stroke for incident cerebral and haemorrhagic strokes, family history of myocardial infarction and stroke for the remaining incident cardiovascular outcomes (**Table 7.1**), family history of diabetes for incident T2D, family histories of breast, prostate, colorectal cancer for corresponding incident cancers, family history of any cancer for the remaining incident cancers, and family history of osteoporosis for incident fractures. In addition to continuous modelling per 1 SD of measures of adherence to the Mediterranean diet, I categorised exposures into quartiles and estimated HRs for the top versus bottom quartiles of adherence. I computed p values for the test of trend of HRs across quartiles by entering ordinal quartile variables into the model as continuous variables.

As sensitivity analyses, I excluded the first two years of follow-up to assess the robustness of the main findings to reverse causation. I repeated the analyses for pulmonary diseases and diseases of the aero- and upper-digestive tract (cancers of the lung, oesophagus and stomach, asthma and COPD) in never-smokers to minimise the impact of residual confounding by smoking status.

7.5 Results

7.5.1 Primary results

Adherence to the Mediterranean diet was associated with incidence of several diseases per each of the methods of exposure assessment in a partially overlapping manner (**Table 7.2**). Consistently inverse relationships with at least one Bonferroni-corrected significant result were detected for T2D (HR, 95% CI range per 1 SD: 0.78, 0.71-0.84 to 0.92, 0.88-0.97; all significant after Bonferroni correction), COPD (0.81, 0.74-0.90 to 0.91, 0.83-0.99; RCT biomarker score not significant after Bonferroni correction), and renal disease (0.92, 0.86-0.98 to 0.94, 0.90-0.97; only MDS-pyramid significant after Bonferroni correction). For some of the outcomes, there were patterns of Bonferroni-corrected inverse associations detected by two measures of adherence to the Mediterranean diet and null results for the remaining two, including cataracts (inverse predictors: RCT biomarker score and MDS-pyramid), ischaemic heart disease, oesophageal cancer, stomach cancer, and heart failure following Bonferroni correction (metabolomic score, MDS-pyramid). There was no indication of a relationship between the Mediterranean diet and risk of disease across all methods of exposure assessment for: atrial fibrillation, venous thrombosis, stroke, abdominal aneurysm, cancers of the prostate, colon, ovaries and endometrium, Parkinson's disease, bone fracture and glaucoma.

Among associations detected by single methods of exposure assessment, the RCT biomarker score was inversely associated (HR per 1 SD; 95% CI) with incidence of liver disease (0.84; 0.75-0.94) and non-melanoma skin cancer (0.90; 0.84-0.97), and positively related to new-onset cancers of the breast (1.20; 1.05-1.37) and rectum following Bonferroni correction (1.37; 1.13-1.67). The EPIC-Norfolk nutritional biomarker score was inversely associated with incidence of lung cancer (0.83; 0.70-0.97) and peripheral arterial disease following Bonferroni correction (0.88; 0.82-0.95). The metabolomic score was inversely associated with rectal cancer (0.82; 0.68-0.98) at the nominal significance level.

The continuously-modelled associations were generally reflected in comparisons of the hazards between the fourth and the first quartile of distributions of measures of the Mediterranean diet or p values for a monotone trend in HRs across quartiles (**Table 7.3**), except for the RCT biomarker score and COPD (HR_{Q4vsQ1} = 0.82, 95% CI: 0.64-1.04; p-trend = 0.18) and the metabolomic biomarker score and ischaemic heart disease (HR_{Q4vsQ1} = 0.89, 95% CI: 0.78-1.01; p-trend = 0.062).

Removing BMI and waist circumference from the multivariable model materially affected some of the results (HR per 1 SD; 95% CI). For the RCT biomarker score, the positive association with rectal cancer was no longer statistically significant following Bonferroni correction but remained significant at the nominal p value (1.37; 1.12-1.67). For the EPIC-Norfolk nutritional biomarker score, associations became significant at the nominal p value for ischaemic heart disease (0.94; 0.89-0.99), heart failure (0.91; 0.85-0.98), cerebral stroke (0.91; 0.83-1.00; p = 0.045) and liver disease (0.87; 0.76-1.00; p = 0.049), and the association with renal disease was modestly deattenuated (0.89; 0.83-0.95) and became significant following Bonferroni correction. The latter was also observed for the metabolomic score (0.91; 0.86-0.96). There was no material change caused by removing adjustment for adiposity from the model for any other associations of the metabolomic score or the FFQ-based MDS-pyramid score in terms of the impact on statistical significance or effect sizes (results not shown).

Risk of the composite outcome of any first incident cardiovascular disease was inversely related to all measures of adherence to the Mediterranean diet at the nominal p value (HR per 1 SD; 95% CI): the RCT biomarker score (0.97; 0.94-1.00; p = 0.044), the EPIC-Norfolk nutritional biomarker score (0.94; 0.91-0.98), the metabolomic score (0.97; 0.94-0.99) and the MDS-pyramid score (0.95; 0.93-0.97). Neither measure was associated with any first incident cancer (results not shown).

Table 7.2 Outcome-wide analysis of the Mediterranean diet using nutritional and metabolomic biomarker scores and the Mediterranean pyramid score in EPIC-Norfolk

Incident outcome	Nutritional biomarker scores			Metabolomic biomarker score		Mediterranean pyramid score	
	n/N	HR (95% CI) per 1 SD		n/N	HR (95% CI) per 1 SD	n/N	HR (95% CI) per 1 SD
		MedLey RCT	EPIC-Norfolk				
Ischaemic heart disease	1,772/6,401	0.97 (0.93-1.02)	0.96 (0.91-1.02)	2,230/10,923	0.95 (0.90-0.99)*	4,313/22,067	0.95 (0.92-0.99)*
Atrial fibrillation	1,334/6,962	0.99 (0.93-1.05)	1.02 (0.96-1.09)	1,852/11,559	0.99 (0.95-1.05)	3,369/22,608	0.98 (0.94-1.01)
Heart failure	1,129/6,804	1.03 (0.97-1.10)	0.94 (0.88-1.01)	1,304/11,360	0.90 (0.85-0.96)†	2,313/22,226	0.92 (0.88-0.97)†
Venous thrombosis	163/6,976	0.97 (0.82-1.15)	0.91 (0.77-1.08)	227/11,479	0.92 (0.81-1.05)	434/22,378	1.03 (0.93-1.14)
Peripheral arterial disease	834/6,904	0.95 (0.89-1.02)	0.88 (0.82-0.95)†	1,102/11,405	0.95 (0.89-1.02)	1,957/22,299	0.98 (0.94-1.03)
Cerebral stroke	514/6,994	1.07 (0.98-1.18)	0.91 (0.83-1.00)	681/11,555	0.95 (0.88-1.04)	1,220/22,850	0.95 (0.89-1.01)
Haemorrhagic stroke	209/6,994	0.94 (0.82-1.08)	0.89 (0.76-1.04)	246/11,555	1.05 (0.92-1.20)	480/22,850	1.01 (0.92-1.12)
Abdominal aneurysm	77/6,663	1.10 (0.86-1.42)	0.91 (0.70-1.20)	102/11,219	0.84 (0.68-1.04)	182/22,850	0.97 (0.83-1.13)
Type 2 diabetes	752/6,675	0.87 (0.81-0.93)†	0.78 (0.71-0.84)†	749/11,134	0.84 (0.78-0.91)†	1,941/22,225	0.92 (0.88-0.97)†
Renal disease	1,065/6,451	0.92 (0.86-0.98)*	0.90 (0.84-0.96)*	1,441/10,988	0.92 (0.87-0.98)*	2,711/21,916	0.93 (0.89-0.97)†
Liver disease	265/6,958	0.84 (0.75-0.94)*	0.90 (0.79-1.03)	371/11,440	0.98 (0.88-1.10)	767/22,322	0.98 (0.91-1.06)
Cancer, breast	2,46/3,243	1.20 (1.05-1.37)*	0.98 (0.86-1.12)	299/5,801	0.93 (0.83-1.04)	611/11,748	0.96 (0.88-1.05)
Cancer, prostate	343/3,413	0.95 (0.85-1.06)	1.08 (0.96-1.21)	430/5,156	0.98 (0.88-1.09)	782/9,573	1.01 (0.93-1.08)
Cancer, non-melanoma skin	631/6,656	0.90 (0.84-0.97)*	1.03 (0.94-1.12)	976/10,957	1.04 (0.97-1.10)	1,768/21,321	1.04 (1.00-1.09)
Cancer, colon	212/6,656	1.15 (0.99-1.33)	1.11 (0.96-1.28)	229/10,957	1.00 (0.88-1.12)	444/21,321	1.04 (0.94-1.15)
Cancer, rectal	98/6,656	1.37 (1.13-1.67)†	0.86 (0.67-1.10)	127/10,957	0.82 (0.68-0.98)*	216/21,321	0.93 (0.81-1.06)
Cancer, lung	187/6,656	0.95 (0.84-1.09)	0.83 (0.70-0.97)*	214/10,957	0.91 (0.78-1.06)	391/21,321	0.95 (0.86-1.05)
Cancer, oesophageal	51/6,656	1.08 (0.81-1.43)	0.90 (0.65-1.22)	64/10,957	0.72 (0.57-0.92)*	104/21,321	0.80 (0.67-0.95)*
Cancer, stomach	39/6,656	0.94 (0.67-1.32)	0.89 (0.65-1.23)	43/10,957	0.65 (0.51-0.82)†	84/21,321	0.79 (0.64-0.98)*
Cancer, ovarian	50/3,243	0.98 (0.71-1.34)	1.23 (0.90-1.69)	65/5,801	0.83 (0.63-1.10)	125/11,748	1.04 (0.85-1.26)
Cancer, endometrial	47/3,243	0.85 (0.65-1.12)	0.89 (0.66-1.20)	61/5,801	0.90 (0.71-1.13)	140/11,748	1.01 (0.84-1.22)
Asthma	345/6,528	0.96 (0.86-1.07)	0.95 (0.84-1.06)	594/10,799	0.94 (0.86-1.03)	1,155/20,941	0.94 (0.88-1.00)
COPD	524/6,467	0.91 (0.83-0.99)*	0.81 (0.74-0.90)†	761/10,635	0.85 (0.78-0.92)†	13,84/20,763	0.87 (0.82-0.92)†
Parkinson's disease	161/7,107	0.97 (0.83-1.13)	0.95 (0.80-1.14)	230/11,679	0.99 (0.86-1.13)	416/22,780	1.07 (0.97-1.19)
Fracture, any site	805/6,581	1.00 (0.93-1.07)	0.94 (0.87-1.02)	1,203/10,875	0.94 (0.88-1.00)	2,332/21,274	1.01 (0.96-1.05)
Glaucoma	418/6,946	0.91 (0.82-1.00)	1.07 (0.97-1.19)	682/11,476	0.99 (0.91-1.07)	1,368/22,415	1.05 (1.00-1.11)
Cataracts	1,908/6,656	0.95 (0.90-0.99)*	0.99 (0.94-1.04)	2,864/11,150	0.98 (0.94-1.02)	5,302/21,786	0.96 (0.93-0.99)*

Abbreviations: COPD – chronic obstructive pulmonary disease; HR – hazard ratio; SD – standard deviation, n/N – cases/participants; RCT – randomised controlled trial

*$p < 0.05$; †$p < 0.00185$ (Bonferroni-corrected for 27 comparisons)

Associations were adjusted for age, sex, lifestyle and socioeconomic indicators, personal and family medical history, body mass index and waist circumference.

Table 7.3 Outcome-wide analysis of the Mediterranean diet using nutritional and metabolomic biomarker scores and the Mediterranean pyramid score in EPIC-Norfolk: hazard ratios for Q4 versus Q1 comparisons

	RCT biomarker score		EPIC-N biomarker score		Metabolomic biomarker score		Mediterranean pyramid score	
Incident outcome	HR (95% CI)	p trend	HR (95% CI)	p trend	HR (95% CI)	p trend	HR (95% CI)	p trend
Ischaemic heart disease	0.97 (0.85-1.11)	0.914	0.89 (0.77-1.03)	0.148	0.89 (0.78-1.01)	0.062	0.90 (0.83-0.99)	0.019
Atrial fibrillation	0.93 (0.80-1.09)	0.748	1.01 (0.86-1.18)	0.765	0.97 (0.85-1.11)	0.670	0.98 (0.88-1.08)	0.639
Heart failure	1.10 (0.92-1.30)	0.226	0.84 (0.70-1.01)	0.038	0.80 (0.68-0.95)	0.007	0.81 (0.72-0.92)	0.001*
Venous thrombosis	1.09 (0.70-1.70)	0.942	0.74 (0.46-1.19)	0.244	0.75 (0.50-1.12)	0.131	1.20 (0.91-1.58)	0.887
Peripheral arterial disease	0.90 (0.74-1.09)	0.432	0.64 (0.51-0.79)	<0.001*	0.85 (0.71-1.02)	0.137	0.94 (0.83-1.08)	0.578
Cerebral stroke	1.21 (0.94-1.56)	0.101	0.84 (0.64-1.09)	0.060	0.86 (0.69-1.08)	0.240	0.86 (0.73-1.01)	0.030
Haemorrhagic stroke	0.86 (0.59-1.26)	0.451	0.71 (0.47-1.07)	0.110	1.15 (0.78-1.68)	0.564	1.04 (0.80-1.36)	0.818
Abdominal aneurysm	1.55 (0.76-3.17)	0.246	0.80 (0.39-1.65)	0.465	0.55 (0.28-1.07)	0.020	0.87 (0.55-1.37)	0.567
Type 2 diabetes	0.68 (0.56-0.83)	<0.001*	0.60 (0.48-0.75)	<0.001*	0.72 (0.58-0.89)	0.002	0.81 (0.71-0.92)	0.002
Renal disease	0.81 (0.68-0.95)	0.010	0.77 (0.65-0.92)	0.003	0.81 (0.70-0.94)	0.005	0.85 (0.76-0.95)	0.002
Liver disease	0.56 (0.39-0.80)	0.004	0.88 (0.61-1.27)	0.307	0.89 (0.66-1.20)	0.461	0.86 (0.69-1.07)	0.337
Cancer, breast	1.61 (1.09-2.39)	0.009	0.94 (0.64-1.38)	0.950	0.83 (0.58-1.19)	0.394	0.87 (0.69-1.10)	0.331
Cancer, prostate	0.90 (0.66-1.23)	0.590	1.27 (0.91-1.78)	0.162	0.91 (0.69-1.21)	0.772	1.07 (0.88-1.31)	0.475
Cancer, non-melanoma skin	0.78 (0.63-0.97)	0.006	1.11 (0.88-1.41)	0.626	1.05 (0.88-1.26)	0.461	1.11 (0.97-1.26)	0.121
Cancer, colon	1.47 (0.99-2.17)	0.038	1.26 (0.83-1.91)	0.281	1.08 (0.72-1.61)	0.838	1.15 (0.88-1.50)	0.575
Cancer, rectal	1.69 (0.93-3.08)	0.034	0.88 (0.48-1.61)	0.533	0.60 (0.35-1.02)	0.041	0.79 (0.53-1.17)	0.125
Cancer, lung	0.95 (0.64-1.42)	0.424	0.61 (0.39-0.97)	0.015	0.81 (0.55-1.21)	0.251	0.93 (0.69-1.24)	0.396
Cancer, oesophageal	1.23 (0.58-2.61)	0.492	0.97 (0.46-2.06)	0.642	0.32 (0.13-0.74)	0.010	0.51 (0.29-0.89)	0.007
Cancer, stomach	1.02 (0.46-2.26)	0.676	0.62 (0.22-1.76)	0.451	0.11 (0.03-0.47)	<0.001*	0.53 (0.28-1.02)	0.024
Cancer, ovarian	1.15 (0.53-2.50)	0.907	1.91 (0.76-4.78)	0.269	0.68 (0.34-1.36)	0.373	1.26 (0.73-2.18)	0.264
Cancer, endometrial	0.54 (0.23-1.26)	0.229	1.01 (0.41-2.50)	0.471	0.87 (0.41-1.83)	0.557	0.77 (0.44-1.35)	0.289
Asthma	0.91 (0.68-1.22)	0.610	0.98 (0.71-1.37)	0.456	0.87 (0.68-1.12)	0.211	0.87 (0.74-1.03)	0.065
COPD	0.82 (0.64-1.04)	0.184	0.63 (0.48-0.83)	<0.001*	0.67 (0.54-0.84)	0.001*	0.73 (0.62-0.87)	<0.001*
Parkinson's disease	0.70 (0.44-1.11)	0.404	0.87 (0.52-1.43)	0.646	0.99 (0.68-1.45)	0.896	1.11 (0.83-1.49)	0.300
Fracture, any site	1.06 (0.87-1.30)	0.687	0.90 (0.73-1.12)	0.187	0.82 (0.70-0.98)	0.082	1.05 (0.93-1.18)	0.529
Glaucoma	0.78 (0.59-1.03)	0.078	1.15 (0.87-1.54)	0.123	0.97 (0.77-1.21)	0.668	1.20 (1.03-1.41)	0.047
Cataracts	0.84 (0.74-0.96)	0.056	0.95 (0.83-1.10)	0.511	0.94 (0.84-1.05)	0.244	0.89 (0.82-0.96)	0.003

Abbreviations: COPD – chronic obstructive pulmonary disease; HR – hazard ratio; N – Norfolk; Q – quartile; RCT – randomised controlled trial

*$p < 0.00185$ in test of trend of hazard ratios across quartiles (Bonferroni-corrected for 27 comparisons)

Associations were adjusted for age, sex, lifestyle and socioeconomic indicators, personal and family medical history, body mass index and waist circumference.

7.5.2 Sensitivity analyses

Excluding the first two years of follow-up from the analysis did not materially influence most of the results (HR per 1 SD; 95% CI). Exceptions were the associations between the RCT biomarker score and COPD (0.91; 0.84-1.00; p = 0.053), the metabolomic score and lung cancer (0.88; 0.76-1.03), and the MDS-pyramid score and stomach cancer (0.81; 0.64-1.01).

The main analyses identified several associations between measures of the Mediterranean diet and incident diseases of aero and upper-digestive sites. In analyses restricted to never-smokers, the overall inverse association (HR per 1 SD; 95% CI, n_{cases}/n_{total}) between the EPIC-Norfolk biomarker score and lung cancer was reversed based on a very limited number of cases (2.32; 1.36-3.94, 17/3,033). For oesophageal cancer, the inverse associations in never-smokers of the metabolomic score (0.79; 0.50-1.25, 22/5,369) and the MDS-pyramid score (0.87; 0.65-1.17, 47/10,746) were compatible with decreased risk though no effect or increased risk could not be ruled out. This was also applicable to the association between the MDS-pyramid score and stomach cancer (0.64; 0.40-1.05, 24/10,746). In turn, the association with this outcome evaluated using the metabolomic score in never-smokers was highly statistically significant (0.35; 0.19-0.63, 14/5,369). A similar pattern emerged for incidence of COPD and the MDS-pyramid score (0.90; 0.80-1.02, 278/9,930) and the metabolomic score (0.77; 0.65-0.92, 146/4,968). By contrast, associations with this outcome evaluated using the RCT biomarker score (1.11; 0.89-1.39, 97/2,814) and the EPIC-Norfolk nutritional biomarker score (0.92; 0.72-1.17, 97/2,814) suggested null associations between these biomarker scores and incident COPD in never-smokers. No novel inverse associations were detected when restricting the analysis to participants without a history of smoking.

7.5.3 Secondary nutritional biomarker scores

Use of the secondary RCT biomarker score derived with urinary cations as predictors (selecting urinary potassium as positively scored interaction terms with β-carotene and lycopene) did not materially affect any of the estimates compared to the results for the corresponding biomarker score without urinary measures (results not shown). The secondary RCT biomarker score derived using only linear terms had an inverse association with incidence of haemorrhagic stroke (0.86; 0.75-0.98) and a positive association with atrial fibrillation (1.07; 1.01-1.13), and it was otherwise not associated with any of the outcomes. The partial-WHI nutritional biomarker score was inversely associated with incidence of ischaemic heart disease (0.94; 0.89-

1.00, p = 0.044), haemorrhagic stroke (0.83; 0.70-0.99), renal disease (0.90; 0.83-0.99), COPD (0.88; 0.78-1.00, p = 0.046), and following Bonferroni correction for multiple testing, T2D (0.71; 0.60-0.84) and liver disease (0.71; 0.60-0.84).

7.6 Discussion

This investigation in a population-based prospective cohort combined the strengths of several methods of objective exposure assessment in sizeable subsamples of a population-based cohort and a study-wide use of FFQ to facilitate identification of robust associations between the Mediterranean diet and 27 diseases. The most consistent findings were the inverse associations between measures of adherence to this dietary pattern and incidence of T2D and COPD. A decreased risk of heart failure was the only remaining finding with at least two primary methods of exposure assessment (metabolomic score, self-report) detecting the association following Bonferroni correction. Several inverse associations specific to single measures of adherence to the Mediterranean diet were also detected, though predominantly at the nominal significance level. Associations were consistently null at the nominal α level across methods of exposure assessment for 11 out of the 27 outcomes, including most of the CVDs, Parkinson's disease, bone fracture, glaucoma, and prostate, colon, ovarian and endometrial cancers.

7.6.1 Strengths and limitations

In addition to considerations outlined in Chapter 6, strengths of this research included several methods of exposure assessment with unrelated or partly unrelated sources of bias, and use of multiple endpoints ascertained via record linkage with routine data on hospitalisations, cancer, and death registries.

Limitations encompassed the availability of biomarker-based measures of adherence to the Mediterranean diet in partially-overlapping subsamples, multiple testing, and concerns over accuracy of case ascertainment for outcomes other than cancers, major cardiovascular events or fractures. Participants with well-controlled conditions which do not manifest overt symptoms, are not life threatening or are managed solely in primary care may have not been ascertained as cases using hospital records alone. Alternatively, they may have had a delayed date of diagnosis compared to their true transition to the case status or the timing of diagnosis of otherwise similar participants with more severe manifestations. Conversely, the pool of

cases in this research may have comprised of a relative over-representation of severe cases. Additionally, multimorbidity may have been associated with accelerated diagnoses based on hospital records alone as comorbidities are likely to be recorded during hospital admissions which are not directly related to the primary cause of a given hospitalisation. Practical implications of such ascertainment bias are likely to be minor for most of the outcomes and relate primarily to the timing of recorded diagnosis, rather than incident case status per se, given the long follow-up in EPIC-Norfolk. A comparison of risk estimates for incident T2D in Chapter 6 confirmed similar results between outcome ascertainment based on multiple sources and hospital records and death certificates alone.

I addressed the issue of multiple testing by using Bonferroni correction which may have been an overly conservative threshold leading to inflation of the rate of false negative findings. However, it appears to be well suited for outcome-wide analyses given the sample size of EPIC-Norfolk.[399]

7.6.2 Implications and comparison with previous research

To my knowledge, this is the first outcome-wide analysis of the Mediterranean diet, as well as the first report on associations of biomarker scores of adherence to this dietary pattern and most of the endpoints included in the analysis. Patterns of associations had the largest overlap between the analyses based on metabolomics and self-report. The metabolomic score detected a similar number of associations as the FFQ-based assessment, albeit at approximately half the sample size. In instances when the results were consistent between biomarker-based assessment and the FFQ-derived MDS-pyramid, the magnitude of associations was similar for most of the outcomes. The RCT biomarker score yielded the most divergent pattern of risk estimates, including increased risks of breast and colorectal cancers which were unlikely to be plausible.[400,401] Given the novelty of the derivation method of this score (Chapter 4), this somewhat erratic pattern should be interpreted with caution and evaluated in future research with larger numbers of cases. All observational literature-based results discussed below in this section relate to exposure assessment based on self-reported dietary intakes.

The most consistent finding of the current analysis was an inverse association between the Mediterranean diet and incidence of T2D. It reinforces prior interventional[17] and observational evidence (Chapters 3-6).[181] Investigations on dietary patterns and incidence of COPD have been scant but supportive of a potential preventative role of dietary quality, and the evidence

on the Mediterranean diet has been limited to one nested case-control study.[402,403] Consistency of the inverse association between this dietary pattern and incidence of COPD in EPIC-Norfolk between several methods of exposure assessment is an important contribution to the field. Confirmatory analyses are warranted, particularly in studies with substantially larger sample sizes to allow for more robust inference in analyses stratified by smoking status. Prospective studies on the Mediterranean diet and incidence of asthma have been limited and mostly focused on children with overall null findings.[404–407] Investigations into other dietary patterns yielded inconsistent results, suggesting that dietary quality may not be an important risk factor for asthma.[404,405]

I report mostly null findings at the nominal p value level on specific CVDs, though an inverse association was detected for the composite outcome of any first incident CVD across all exposure assessment methods. These results are in line with interventional evidence[408,409] and meta-analytical estimates from observational studies on the Mediterranean diet and composite CVD outcomes.[391] For CVD subtypes, only ischaemic heart disease has been investigated extensively in relation to this dietary pattern (including myocardial infarction as its common manifestation).[391,410] The current results confirm the previously reported inverse associations,[391,410] though only at the nominal significance level in analyses with larger sample sizes. The relationship between the Mediterranean diet and risk of other non-composite CVD outcomes has been evaluated less frequently. Atrial fibrillation has been investigated only in an intervention as a post-hoc analysis of the PREDIMED trial.[411,412] A 38% risk reduction was found in the Mediterranean diet group supplemented with olive oil relative to a lower-fat control diet and no effect was found in the Mediterranean diet group supplemented with nuts.[411,412] The latter intervention was likely more closely aligned with high adherence to the Mediterranean diet in EPIC-Norfolk participants, and thus the null result in EPIC-Norfolk can be interpreted as compatible with interventional evidence from the PREDIMED trial.

An inverse association between adherence to the Mediterranean diet and incidence of heart failure has been a largely consistent finding of prospective cohort studies.[413,414] The PREDIMED trial reported a statistically non-significant result on this relationship, though directionally consistent with a potential decreased risk in the Mediterranean diet arms.[415,416] There was a favourable effect of the PREDIMED intervention on plasma risk factors for heart failure (decreased plasma N-terminal pro-brain natriuretic peptide, oxidised LDL and lipoprotein(a))[417] which identifies potential biological pathways underpinning the observational results. Of note, the inverse association with heart failure was detected in the

current report using the FFQ-based MDS-pyramid score (n ~22,000) and the metabolomic score (n~11,000) but not the nutritional biomarker scores (n ~7,000). The non-significant inverse result for the EPIC-Norfolk biomarker score may have been due to a limited sample size, and the convincingly null result for the MedLey RCT biomarker score may have been driven by the positive impact of the MedLey intervention on circulating long-chain MUFAs (Chapter 4), which in turn have been positively associated with incidence of heart failure in American cohorts.[418]

Previous research on healthy dietary patterns and incidence of venous thrombosis has identified mostly null results and positive associations tended to be detected for data-driven 'Western' patterns.[419] EPIC centres from the Netherlands (EPIC-NL) have reported an inverse association between the Mediterranean diet and incidence of pulmonary embolism[420] which contrasts with the null findings in EPIC-Norfolk. This discrepancy may have been in part due to the overall stronger associations of the Mediterranean diet with CVD outcomes in the former study.[420] EPIC-NL has also reported a non-significantly inverse result for incidence of peripheral arterial disease. The only other observational study which reported on this outcome, the WHI, revealed an inverse association of the Mediterranean diet at four times the number of incident cases of EPIC-NL.[421] Peripheral arterial disease was the only pre-specified non-composite CVD endpoint in the PREDIMED trial for which a risk reduction was reported in the Mediterranean diet arms.[422–424] The current study identified this relationship only by using the EPIC-Norfolk nutritional biomarker score, and reasons for specificity of this exposure assessment method to peripheral arterial disease are unclear. Association of the Mediterranean diet with incidence of abdominal aneurysm has only been evaluated in Swedish population-based cohorts, finding an inverse relationship.[425] This analysis was based on a 10-fold larger number of incidence cases than the current investigation (1,781 versus 182), and null findings in EPIC-Norfolk may have been related to insufficient statistical power. Meta-analytical estimates on the Mediterranean diet and incidence of stroke suggest an inverse association for total and cerebral stroke but not haemorrhagic stroke.[391] Neither of the measures of adherence was related in the current study to incidence of cerebral and haemorrhagic stroke, though a marginally inverse association at the nominal p value level was noted for the MDS-pyramid score, and a previous analysis from EPIC-Norfolk on total stroke reported an inverse association.[391]

The analysis of incidence of cancer in the current investigation has been limited by small numbers of cases available for most sites. Results have been largely null but directionally consistent with published meta-analytical estimates which yielded inverse association between

adherence to the Mediterranean diet and incidence of colorectal, respiratory and stomach cancers.[392] Pooled effect sizes were mostly modest and substantially stronger for stomach cancer than for other sites.[392] The association for stomach cancer was also notably strongly inverse in EPIC-Norfolk based on the FFQ and metabolomic scores but not nutritional biomarker scores.

Three cohorts from the USA and Sweden found an inverse association between the Mediterranean diet and incidence of Parkinson's disease[426] and the Rotterdam Study reported null results with point estimates compatible with decreased risk.[427] Outcome ascertainment in these studies was based on linkage with both hospital data and primary care registries or cases were ascertained based on other multiple sources. Lack of an association between the Mediterranean diet and Parkinson's disease in EPIC-Norfolk may have been, in part, driven by reliance on cases identified only based on hospital visits and death certificates. Hospital records may miss on average 27% of cases in the UK with higher under-reporting in admissions for which Parkinson's disease is not the primary cause.[428] Thus, null findings from the current research should be interpreted with caution.

Two studies were identified by a 2018 meta-analysis on the Mediterranean diet and risk of fracture which investigated fractures at any site as an outcome, yielding a relative risk of 1.00 (95% CI 0.99–1.02) for high versus low adherence based on 28,873 incident fractures.[429] A later investigation in EPIC-Norfolk found an inverse association between the original Trichopoulou MDS and aMED scores estimated from 7-day food diaries and incidence of any first fracture of the hip, spine or wrist.[430] These sites accounted for 62% of fractures in the current analysis which reported overall null findings and a marginally inverse association when using the metabolomic score. Beyond the outcome definition, use of different indices of the Mediterranean diet and the food diary as the source of dietary self-report may have contributed to divergent results. Published findings suggest that adherence to this dietary pattern may be associated with lower incidence of hip fracture[429] which was not a pre-specified endpoint of the current outcome-wide study.

For ophthalmic diseases, two cohorts found no association between the Mediterranean diet and incidence of glaucoma[431,432] which is consistent with the null results in EPIC-Norfolk. The PREDIMED trial found no impact of the intervention on the risk of cataract surgery.[433] No prospective cohort to date has evaluated this relationship. The current results of an inverse association of the MedLey RCT biomarker score and the MDS-pyramid score with incidence

of cataracts at the nominal p value level merits evaluation in other cohorts, particularly given the inverse relationship between the Mediterranean diet and dry age-related macular degeneration.[434]

The current study used unique composite outcomes of renal and liver diseases which limits comparability of the findings with the published literature. Inverse associations of the Mediterranean diet with incidence of renal disease in EPIC-Norfolk is consistent with the overall body of evidence suggesting a lower risk of chronic kidney disease and slower decline of kidney function over time with higher adherence.[435] For liver disease, previous research has suggested potential benefits of the Mediterranean diet for prevention of liver cancer[392] and accumulation of liver fat.[436,437] Moreover, the Mediterranean diet is a therapeutic option for treatment of non-alcoholic fatty liver disease.[272] Only the RCT biomarker score was inversely related in EPIC-Norfolk to incidence of the composite outcome of liver disease which included cancer, alcoholic and non-alcoholic disease, as well as viral hepatitis. Such outcome definition may have been suboptimal for investigations of nutritional factors but it has been pre-specified for consistency with a prior outcome-wide analysis in EPIC-Norfolk[328]

7.6.3 Conclusions

The Mediterranean diet may be associated with lower risk of multiple chronic diseases in middle-aged and older adults in the UK. Dietary self-report and different methods of biomarker-based exposure assessment identified partially overlapping sets of diseases which may be targets for dietary prevention by increasing adherence to the Mediterranean dietary pattern. A high degree of agreement between the methods in detecting highly significant inverse associations with incidence of T2D and COPD, and to a lesser extent heart failure, increase the credibility of the findings and highlight potential specificity of the Mediterranean diet to these outcomes in the British population.

Chapter 8

Discussion

8.1 Summary of findings

This PhD focused on the link between the Mediterranean diet and incidence of T2D. It aimed to appraise the impact of dietary patterns on nutritional biomarkers, evaluate whether they can be usefully combined into composite measures reflecting dietary quality, and test their associations with incident disease outcomes by comparison with dietary self-report.

8.1.1 Effects of the Mediterranean diet on nutritional biomarkers

I conducted a systematic review with meta-analyses on the effects of Mediterranean diet interventions on nutritional biomarkers (**Chapter 2**). A range of circulating biomarkers responded to such interventions, including β-carotene, lycopene, retinol, vitamin C, and several monounsaturated and n-3 and n-6 polyunsaturated fatty acids. In the process of appraising the literature, I observed a common limitation of using nutritional biomarkers as measures of adherence to dietary patterns in a univariate manner. I additionally identified a suitable trial for development of a composite biomarker of the Mediterranean diet, the MedLey trial.

I undertook primary data analysis of the MedLey trial, and I reported several novel findings on the effects of the Mediterranean diet on circulating fatty acids. These included decreases in long-chain saturated fatty acids and the plasmalogen C18:0 dimethyl acetal and increases in long-chain monounsaturated fatty acids. The results likely reflected across changes in intakes of foods and metabolic effects of the intervention.

8.1.2 Derivation of biomarker scores of dietary patterns

I derived nutritional (**Chapters 3-6**) and metabolomic (**Chapter 6**) biomarker scores of adherence to dietary patterns using interventional and observational study designs. Observationally-derived biomarker scores based on prediction of self-reported adherence to dietary patterns had modest correlations with their respective dietary patterns. The correlation coefficients were in the range of approximately 0.3-0.5 which is in line with the strength of

correlation between established biomarkers of intake and their corresponding measures from self-report.[229,230] However, this is not an indication of satisfactory performance of these biomarker scores as validity cannot be judged by using subjectively measured diet as a reference measure.

I took two approaches to derive or validate nutritional biomarker scores of the Mediterranean diet within analytical frameworks which were free from reporting errors of dietary self-report. First, I derived a biomarker score within the partial-feeding MedLey RCT, which compared the effects of six months of an intensive Mediterranean diet intervention with continuation of habitual diet in elderly Australians on concentrations of circulating carotenoids and fatty acids (**Chapter 4**). The score was based on prediction of the randomised assignment from biomarker concentrations, and it had a high discriminatory performance to distinguish participants in the Mediterranean and control diet arms (cross-validated C-statistic = 0.88). The second approach consisted in observationally deriving the biomarker scores in thirteen cross-sectional samples of the EPIC study, applying the prediction equations to the MedLey trial, and testing whether values of the biomarker scores were increased in the Mediterranean diet group post-intervention. Controlling for multiple testing, eight of the thirteen biomarker scores were higher by approximately 0.5-0.6 standard deviation in the intervention arm relative to the control habitual diet group (**Chapter 5**). The objective derivation and validation of the biomarker scores was limited to the set of circulating biomarkers which overlapped between the MedLey trial and EPIC studies (carotenoids and fatty acids). Separate investigations in the EPIC study (**Chapter 3**) and the Norfolk arm of EPIC (**Chapter 6**) considered further groups of biomarkers, including: untargeted metabolomics, vitamins C and 25(OH)D, tocopherols, iron status biomarkers, carbon and nitrogen stable isotopes, serum and urinary cations and phytoestrogens, and urinary sugars. Expanding the pool of biomarker predictors beyond carotenoids and fatty acids did not overall meaningfully improve the performance of biomarker scores in prediction of dietary self-report. However, no objective criterion was available to evaluate their potential additive value on top of plasma carotenoids and phospholipid fatty acids.

8.1.3 Associations of biomarker scores of dietary patterns with incident disease outcomes

Prospective diet-disease associations in this thesis primarily focused on T2D (**Chapters 3-7**). Using observationally-derived biomarker scores of the MDS, aHEI-2010 and the DASH diet score, I demonstrated inverse associations of these dietary patterns with incidence of T2D in the EPIC-InterAct case-cohort study. By contrast, adherence to aHEI-2010 and the DASH diet estimated from self-report was not associated with risk of T2D (**Chapter 3**), raising the possibility that the measurement error of self-report may have biased the results towards the null in prior published work from this study.[183] The RCT-derived biomarker score was strongly inversely associated with incidence of T2D in EPIC-InterAct. My modelling indicated that a modest 10 percentile shift of adherence to the Mediterranean diet, as measured by the RCT biomarker score, could avert 11% of cases of T2D in the underlying pan-European population under the assumption of a causal relationship (**Chapter 4**). Observationally derived-biomarker scores of the MDS and MDS-pyramid scores had high generalisability as inverse correlates of risk of T2D. Ten out of 13 scores developed in heterogeneous cross-sectional samples were inversely associated with incidence of T2D in EPIC-InterAct (**Chapter 5**). In the above investigations, the magnitude of associations of the Mediterranean diet with new-onset T2D was approximately 2-3-fold stronger when using biomarker scores compared to self-report.

Chapter 6 aimed to formally test whether combinations of different groups of nutritional or metabolomic biomarkers as composite measures of adherence to the MDS-pyramid yield materially stronger associations with incident outcomes (CVD, cancer, T2D, all-cause and cause-specific mortality) than FFQ-based assessment of the MDS-pyramid score in EPIC-Norfolk. Such pattern of risk estimates was confirmed for incident T2D and additionally all-cause mortality but not incident CVD, cancer, or CVD- and cancer-specific mortality. The Mediterranean diet was not associated with cancer incidence with either the biomarker-based or the FFQ-based assessment, and for cause-specific mortality associations tended to be null based on the FFQ and inverse when applying biomarker scores. Supplementing the base set of plasma carotenoids and fatty acids with additional groups of nutritional biomarkers had overall little impact on risk estimates, however, inverse associations were modestly stronger when the biomarker scores incorporated some sets of biomarkers in an outcome-specific manner (e.g., urinary and serum phytoestrogens for incidence of CVD and cancer, respectively). Metabolomics-based estimates did not differ meaningfully from associations assessed by nutritional biomarkers.

In **Chapter 7**, I leveraged the availability of multiple objective measures of adherence to the Mediterranean diet developed throughout the PhD (RCT and observational nutritional biomarker scores and a metabolomic score) measured in up to half of EPIC-Norfolk participants and the cohort-wide availability of self-report to perform an outcome-wide analysis of 27 incident disease outcomes. Incident T2D and COPD emerged as the outcomes most consistently identified by all four measures of adherence as inversely related with adherence to this dietary pattern. Several additional inverse associations were detected with incident cardiometabolic diseases, oesophageal and stomach cancers and cataracts when not controlling for multiple testing or considering diet-disease associations not detected by some of the methods of exposure assessment.

8.2 Strengths and limitations

Detailed discussions of strengths and limitations were included in the Discussion sections of each chapter. The major strength of this thesis overall was the derivation of novel objective measures of adherence to the Mediterranean diet, and their application to epidemiological investigations by integrating data from a partial-feeding intervention and large-scale prospective cohorts.

The major general limitation was the possibility of residual confounding in observational analyses, despite comprehensive adjustments for broad ranges of covariates. It may have introduced bias both towards and away from the null, thus precluding causal inference based on these analyses alone. Limitations specific to this work pertained to unclear validity and specificity of the biomarker scores as biomarkers of the Mediterranean diet. The biomarker scores derived based on prediction of self-reported adherence may have been partly affected by systematic errors introduced by subjective reporting with potential implications of directionally unpredictable bias in diet-disease associations. The experimentally derived biomarker score may have been driven not only by changes in dietary intakes but also by a favourable metabolic response to improved dietary quality. This metabolic response may have not been specific to the Mediterranean diet or even dietary exposures at large, thus raising concerns about specificity as a dietary biomarker.

This work was undertaken in European and Australian study participants of predominantly white descent. Generalisability to other geographical settings and ethnicities remains unknown.

This pertains to both the potential for effect modification of diet on nutritional and metabolomic biomarkers by ethnicity, as well as the applicability and cultural acceptability of the Mediterranean diet and other 'West-centric' healthy dietary patterns in non-Western countries. Derivation and validation of nutritional biomarker scores were undertaken in a RCT in elderly Australians using different biomarker assays than in their subsequent application to diet-disease associations. Moderating effects of population and study characteristics on results of the trial could not be ruled out, and their potential impact on external generalisability to the younger European populations remains unknown.

8.3 Interpretation of findings

The Mediterranean diet has arguably been the most extensively researched dietary pattern in nutritional epidemiology, including the rarely available evidence of efficacy of dietary factors in primary prevention of chronic disease.[26,438] Despite the considerable interventional body of evidence on its effects on CVD incidence and mortality, the quality of evidence is insufficient to reach reasonable certainty about presence of an effect or its precise quantification for most CVD endpoints.[424] A recent systematic assessment of stability of findings on Mediterranean diet interventions has highlighted the limited robustness of findings from RCTs, as indicated by a fragility index of 5.0 (95% CI: 3.9-8.3) for CVD outcomes.[439] This suggests that adding on average five incident CVD cases to treatment arms would attenuate to the null the evidence on risk-lowering effects of the Mediterranean diet.[439] Trials investigating the impact of this dietary pattern on incidence of T2D have been limited to two secondary analyses of RCTs with exclusions of approximately half of participants who were prevalent cases at baseline (Chapter 4.6.3).[17,298] Logistical and financial difficulties of large-scale primary prevention RCTs render unlikely in the foreseeable future further accumulation of interventional evidence on the Mediterranean diet beyond the PREDIMED trial. Feasibility of implementation of long-term changes in adherence to dietary patterns in thousands of participants has been called into question based on the experience of the PREDIMED trial[19] and to a larger degree the Women's Health Initiative[440] of not achieving the desired dietary contrasts between the intervention and control groups. Regardless of this or other methodological limitations of food-based RCTs (e.g. inability to fully blind participants), agreement should be sought between interventional, observational and mechanistic evidence to robustly infer causality.[20] Key recommendations for enhanced inference from prospective cohort studies include improving measurement of dietary

intakes and reducing confounding.[441] This thesis has been focused on the former by employing objective exposure assessment, and it has generated a line of evidence on diet-disease associations complimentary to and compatible with prior evidence based on dietary self-report.

The aetiological nature of most of my investigations (Chapters 3-7) motivates evaluation for causality of the key finding of an inverse association between biomarker scores of the Mediterranean diet and incidence of T2D. Revisiting the Bradford-Hill criteria through the lens of modern epidemiology can be helpful in establishing causal links.[442] **Strength of association** has been demonstrated by associations which were of similar magnitude as those of other strongly related risk or protective factors, not including measures of glycaemia and insulin resistance.[187,328,443] **Consistency** of the inverse association of the RCT biomarker score was apparent across eight European countries, though with considerable heterogeneity, and indicative of expected inverse associations in other similar populations. The inverse association was reliably replicated by using different types of biomarker scores throughout the thesis, as well as dietary self-report. **Specificity** was partially demonstrated in the outcome-wide analysis (Chapter 7) by comparison with other disease outcomes. However, given the multi-factorial nature of T2D and other investigated conditions, and the likely impact of diet on multiple disease-related pathways, specificity was not a strongly plausible, a priori consideration. Falsification endpoints may be better suited to establish non-spurious associations in presence of multiple risk factors.[442] I employed accidental deaths as a pre-specified negative control outcome which was not associated with measures of adherence to the Mediterranean diet. The **temporality** assumption was fulfilled by using prospective designs with long durations of follow-up. The **dose-response** assumption relates to a linear relationship which was fulfilled when using the observationally-derived nutritional and metabolomic biomarker scores but not the RCT biomarker score for which a plateau effect was apparent at the upper extreme of the distribution. However, a non-linear association may arise causally due to presence of threshold effects.[442] There was considerable evidence to support **plausibility**, **coherence** and **experimental evidence** based on the results of the PREDIMED trial[19] and multiple short- to mid-term RCTs robustly demonstrating favourable effects of the Mediterranean diet on glycaemic control and multiple other cardiometabolic risk factors.[415] The criterion of **analogy** is of limited use for assessment of causality[442] and almost a contraindication of the specificity assumption. It was fulfilled by detecting inverse associations of biomarker scores of other dietary patterns with incidence of T2D (Chapter 3), consistent inverse associations of the

Mediterranean diet detected by four different methods of exposure assessment (Chapter 7), and inverse associations of this dietary pattern with other disease outcomes (Chapters 6-7).

Overall, most of the criteria for assessment of causality were met. The primary finding of this thesis of an inverse association between the Mediterranean diet and incidence of T2D is consistent with the PREDIMED trial which compared its effects against a lower-fat diet,[19] and observational[393] and mechanistic[58,444] studies. Whether or not this constitutes sufficient evidence to infer causality is a matter of ongoing, epistemological debate.[445,446] On a practical note, the evidence generated within this thesis strengthens the actionable evidence[20] for emphasising the Mediterranean diet in dietary guidelines. Comparative analyses with other dietary patterns using analytical approaches developed within the current thesis are warranted.

8.4 Implications and directions for future research

The systematic review I conducted on the effects of the Mediterranean diet and nutritional biomarkers (Chapter 2) and the development of a robust biomarker of a Mediterranean diet intervention (that I led in Chapter 4) are informative for trialists to improve monitoring of compliance. The latter is an important advancement over the commonplace practice of assessing adherence by using single biomarkers – which can plausibly capture only individual components of the intervention – or dietary self-report which in this context is particularly susceptible to desirability bias as non-compliant participants may subconsciously overestimate their compliance to decrease cognitive dissonance.[440]

Biomarkers of dietary pattern interventions could be adopted for application in dietetic practice to ensure and motivate sufficient adherence to achieve the desired clinical goals. It is unclear whether such biomarker scores could be used to effectively monitor habitual dietary quality in free-living individuals, and it should be explored in future research. The RCT biomarker score of the Mediterranean diet was weakly correlated with adherence to this dietary pattern estimated from dietary self-report. This may have been caused by measurement error due to subjective reporting, poor external generalisability of the biomarker score or the biomarker score and the self-reported score constituting two different constructs. It is plausible that the former may have captured a 'true' Mediterranean diet, and the latter may have been a more generic dietary quality index. It would be prudent for future research to evaluate external generalisability of RCT-derived biomarker scores to cross-sectional settings by using more

tangibly quantifiable dietary exposures. Major foods or foods groups may be a desirable target to avoid the issues of construct validity and lack of a natural scale inherent in dietary pattern indices, though similar research on dietary patterns is also warranted, particularly those that are more culturally universal than the Mediterranean diet.

The observational work on biomarker scores of dietary patterns is of relevance to researchers interested in development of objective measures of exposure (Chapters 3, 5-6). The finding of a high rate of external validity of biomarker scores of the Mediterranean diet when applied to a partial-feeding RCT (Chapter 5) provides a sense of reassurance that the measurement error of dietary self-report may typically not be overwhelming when the self-report is used for derivation of composite dietary biomarkers. Data from feeding interventions are rarely available, and hence observational derivation of biomarker scores or metabolomic profiling is common in the literature.[51] The comparative assessment of different groups of nutritional biomarkers and untargeted metabolomics for derivation of composite biomarkers of the Mediterranean diet is also informative for shaping future research (Chapter 6). It found a similar performance of nutritional and metabolomic biomarkers and identified plasma carotenoids and phospholipid fatty acids as a potentially sufficient set of biomarkers to characterise adherence to the Mediterranean diet. These findings may provide an impetus for a more widespread use of nutritional biomarkers in applications typically reserved to date for metabolomics[51] and open avenues for objective assessment of dietary patterns in novel settings, for example in nutrition surveys.[447]

In the longer run, measurement of biomarker scores of dietary patterns could be incorporated into wearable sensors both for the purpose of personalised dietary assessment and data collection in research settings.[448,449] Perspectives for measurement of fat-soluble nutrients in small, portable devices remain elusive,[448,449] however, other minimally invasive methods can be employed instead. For example, metabolomic profiling of postprandially collected dried blood spots has recently been shown to discriminate well between a diet high in meat and a high-carbohydrate vegan diet with similar performance as that of 24-hour urine collections.[450]

Diet-disease associations reported throughout this thesis (Chapters 3-7) serve as a proof-of-concept of applying nutritional biomarker scores of dietary patterns for exposure assessment in epidemiological investigations. Practical implications for aetiological epidemiology include strengthening inference on inverse associations between the Mediterranean diet and incidence of T2D and CVD, and identification of a novel link with

new-onset COPD. The most important contribution of this thesis to informing public health practice was the modelling of preventability of T2D by increasing exposure to the Mediterranean diet which suggested that a 10-percentile upward population shift in adherence, as assessed by the RCT biomarker score, could potentially avert 11% of incident cases. This finding was limited by the lack of absolute quantification of the required dietary changes to achieve the 10% increase. This was not possible given the derivation of the biomarker score in a different setting than the prospective investigation of T2D and use of different biomarker assays, which resulted in miscalibration of the RCT biomarker score when applied to data from the EPIC study. Future research should explore the integration of experimental data for exposure assessment and longitudinal health data with application of the same biomarker assays in both types of studies to enable reliable quantification of diet-disease associations in absolute terms.

8.5 Conclusions

This PhD identified combinations of nutritional biomarkers as plausible biomarkers of the Mediterranean diet and other dietary patterns suitable for application in epidemiological investigations. Utilising the randomised controlled trial design for derivation and validation of such biomarkers has been the major contribution of this thesis to the field of study. Application of experimentally derived or validated biomarker scores of the Mediterranean diet to investigations of incidence of T2D yielded substantially stronger inverse associations than when using self-reported adherence to this dietary pattern. These findings provide additional support for incorporating advice on increasing adherence to the Mediterranean diet into dietary guidelines. Development of robust methods of objective assessment of dietary patterns and other dietary exposures is warranted to address limitations of self-report and to improve the credibility of findings in nutritional epidemiology.

References

1. Sun H, Saeedi P, Karuranga S, et al. IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes Res Clin Pract*. 2022;183. doi:10.1016/j.diabres.2021.109119 pmid:34879977

2. Forouzanfar MH, Alexander L, Bachman VF, et al. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. *Lancet*. 2015;386(10010):2287–323. doi:10.1016/S0140-6736(15)00128-2

3. Stanaway JD, Afshin A, Gakidou E, et al. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: A systematic analysis for the Global Burden of Disease Stu. *Lancet*. 2018;1923–94. doi:10.1016/S0140-6736(18)32225-6

4. Afshin A, Sur PJ, Fay KA, et al. Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2019;393(10184):1958–72. doi:10.1016/S0140-6736(19)30041-8

5. Forouhi NG, Misra A, Mohan V, Taylor R, Yancy W. Dietary and nutritional approaches for prevention and management of type 2 diabetes. *Bmj*. 2018;2234(June):k2234. doi:10.1136/bmj.k2234 pmid:29898883

6. Neuenschwander M, Ballon A, Weber KS, et al. Role of diet in type 2 diabetes incidence: umbrella review of meta-analyses of prospective observational studies. *BMJ*. 2019;366:l2368. doi:10.1136/bmj.l2368 pmid:31270064

7. Subar AF, Freedman LS, Tooze JA, et al. Addressing Current Criticism Regarding the Value of Self-Report Dietary Data. *J Nutr*. 2015;145(12):2639–45. doi:10.3945/jn.115.219634 pmid:26468491

8. Satija A, Yu E, Willett WC, Hu FB. Understanding Nutritional Epidemiology and Its Role in Policy. *Adv Nutr An Int Rev J*. 2015;6(1):5–18. doi:10.3945/an.114.007492 pmid:25593140

9. Ioannidis JPA. Implausible results in human nutrition research. *BMJ*. 2013;347(November):1–2. doi:10.1136/bmj.f6698 pmid:24231028

10. Ioannidis JPA. The Challenge of Reforming Nutritional Epidemiologic Research. *JAMA*. 2018;320(10):969–70. doi:10.1001/jama.2018.11025 pmid:30422271

11. Nathan DM, Barrett-Connor E, Crandall JP, et al. Long-term effects of lifestyle intervention or metformin on diabetes development and microvascular complications over 15-year follow-up: the Diabetes Prevention Program Outcomes Study. *lancet Diabetes Endocrinol*. 2015;3(11):866–75. doi:10.1016/S2213-8587(15)00291-0 pmid:26377054

12. Li G, Zhang P, Wang J, et al. Cardiovascular mortality, all-cause mortality, and diabetes incidence after

lifestyle intervention for people with impaired glucose tolerance in the Da Qing Diabetes Prevention Study: A 23-year follow-up study. *Lancet Diabetes Endocrinol*. 2014;2(6):474–80. doi:10.1016/S2213-8587(14)70057-9 pmid:24731674

13. Lindström J, Peltonen M, Eriksson JG, et al. Improved lifestyle and decreased diabetes risk over 13 years: long-term follow-up of the randomised Finnish Diabetes Prevention Study (DPS). *Diabetologia*. 2013;56(2):284–93. doi:10.1007/S00125-012-2752-5 pmid:23093136

14. WC K, E B-C, SE F, et al. Reduction of the incidence of type 2 diabetes with lifestyle intervention or metformin [Internet]. Vol. 34, International Urology and Nephrology. N Engl J Med; 2002. p. 162–3. doi:10.1097/01.ogx.0000055759.75837.e7 pmid:11832527

15. Tinker LF, Bonds DE, Margolis KL, et al. Low-fat dietary pattern and risk of treated diabetes mellitus in postmenopausal women: the Women's Health Initiative randomized controlled dietary modification trial. *Arch Intern Med*. 2008;168(14):1500–11. doi:10.1001/ARCHINTE.168.14.1500 pmid:18663162

16. Howard B V., Aragaki AK, Tinker LF, et al. A low-fat dietary pattern and diabetes: A secondary analysis from the women's health initiative dietary modification trial. *Diabetes Care*. 2018;41(4):680–7. doi:10.2337/dc17-0534 pmid:29282203

17. Salas-Salvadó J, Bulló M, Estruch R, et al. Prevention of diabetes with Mediterranean diets: A subgroup analysis of a randomized trial. *Ann Intern Med*. 2014;160(1):1–10. doi:10.7326/M13-1725 pmid:24573661

18. Patterson RE, Prentice RL, Beresford S, et al. Dietary adherence in the women's health initiative dietary modification trial. *J Am Diet Assoc*. 2004;104(4):654–8. doi:10.1016/j.jada.2004.01.014 pmid:15054353

19. Appel LJ, Van Horn L. Did the PREDIMED Trial Test a Mediterranean Diet? *N Engl J Med*. 2013;368(14):1353–4. doi:10.1056/nejme1301582

20. Mozaffarian D, Forouhi NG. Dietary guidelines and health - Is nutrition science up to the task? *BMJ*. 2018;360(March):1–9. doi:10.1136/bmj.k822 pmid:29549076

21. Dao MC, Subar AF, Warthon-Medina M, et al. Dietary assessment toolkits: an overview. *Public Health Nutr*. 2018;(2):1–15. doi:10.1017/S1368980018002951

22. Bennett DA, Landry D, Little J, Minelli C. Systematic review of statistical approaches to quantify, or correct for, measurement error in a continuous exposure in nutritional epidemiology. *BMC Med Res Methodol*. 2017;17(1):1–22. doi:10.1186/s12874-017-0421-6 pmid:28927376

23. Naska A, Lagiou A, Lagiou P. Dietary assessment methods in epidemiological research: current state of the art and future prospects. *F1000Research*. 2017;6(0):926. doi:10.12688/f1000research.10703.1 pmid:28690835

24. Tapsell LC, Neale EP, Satija A, Hu FB. Foods, Nutrients, and Dietary Patterns: Interconnections and Implications for Dietary Guidelines. *Adv Nutr*. 2016;7(3):445–54. doi:10.3945/an.115.011718 pmid:27184272

25. Mozaffarian D. Dietary and Policy Priorities for Cardiovascular Disease, Diabetes, and Obesity. *Circulation*. 2016;133(2):187–225. doi:10.1161/CIRCULATIONAHA.115.018585 pmid:26746178

26. Schulze MB, Martínez-González MA, Fung TT, Lichtenstein AH, Forouhi NG. Food based dietary patterns and chronic disease prevention. *BMJ*. 2018;361:k2396. doi:10.1136/bmj.k2396 pmid:29898951

27. Gebreslassie M, Sampaio F, Nystrand C, Ssegonja R, Feldman I. Economic evaluations of public health interventions for physical activity and healthy diet: A systematic review [Internet]. Vol. 136, Preventive Medicine. Prev Med; 2020. doi:10.1016/j.ypmed.2020.106100 pmid:32353572

28. Jacobs-Van Der Bruggen MAM, Bos G, Bemelmans WJ, et al. Lifestyle interventions are cost-effective in people with different levels of diabetes risk: Results from a modeling study. *Diabetes Care*. 2007;30(1):128–34. doi:10.2337/dc06-0690 pmid:17192345

29. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol*. 2002;13(1):3–9. doi:10.1097/00041433-200202000-00002 pmid:11790957

30. Menotti A, Puddu PE. How the Seven Countries Study contributed to the definition and development of the Mediterranean diet concept: a 50-year journey. *Nutr Metab Cardiovasc Dis*. 2015;25(3):245–52. doi:10.1016/J.NUMECD.2014.12.001 pmid:25650160

31. Dinu M, Pagliai G, Casini A, Sofi F. Mediterranean diet and multiple health outcomes: An umbrella review of meta-analyses of observational studies and randomised trials [Internet]. Vol. 72, European Journal of Clinical Nutrition. Eur J Clin Nutr; 2018. p. 30–43. doi:10.1038/ejcn.2017.58 pmid:28488692

32. Galbete C, Schwingshackl L, Schwedhelm C, Boeing H, Schulze MB. Evaluating Mediterranean diet and risk of chronic disease in cohort studies: an umbrella review of meta-analyses. *Eur J Epidemiol*. 2018;33(10):909–31. doi:10.1007/s10654-018-0427-3

33. Imamura F, Jacques PF. Invited commentary: Dietary pattern analysis. *Am J Epidemiol*. 2011;173(10):1105–8. doi:10.1093/aje/kwr063 pmid:21474588

34. Zhao J, Li Z, Gao Q, et al. A review of statistical methods for dietary pattern analysis [Internet]. Vol. 20, Nutrition Journal. Nutr J; 2021. doi:10.1186/s12937-021-00692-7 pmid:33874970

35. Edefonti V, De Vito R, Dalmartello M, et al. Reproducibility and Validity of A Posteriori Dietary Patterns: A Systematic Review. *Adv Nutr*. 2019; doi:10.1093/advances/nmz097

36. Imamura F, Lichtenstein AH, Dallal GE, Meigs JB, Jacques PF. Generalizability of dietary patterns associated with incidence of type 2 diabetes mellitus. *Am J Clin Nutr*. 2009;90(4):1075–83. doi:10.3945/AJCN.2009.28009 pmid:19710193

37. Liese AD, Krebs-Smith SM, Subar AF, et al. The Dietary Patterns Methods Project: Synthesis of Findings across Cohorts and Relevance to Dietary Guidance. *J Nutr*. 2015;145(3):393–402. doi:10.3945/jn.114.205336 pmid:25733454

38. Potischman N, Freudenheim JL. Biomarkers of Nutritional Exposure and Nutritional Status: An

Overview. *J Nutr*. 2003;133(3):873S-874S. doi:10.1093/jn/133.3.873S pmid:12612172

39. Corella D, Ordovás JM. Biomarkers: background, classification and guidelines for applications in nutritional epidemiology. *Nutr Hosp*. 2015;31:177–88. doi:10.3305/nh.2015.31.sup3.8765

40. Tasevska N, Runswick SA, McTaggart A, Bingham SA. Urinary Sucrose and Fructose as Biomarkers for Sugar Consumption. *Cancer Epidemiol Biomarkers Prev*. 2005;14(5):1287–94. doi:10.1158/1055-9965.EPI-04-0827 pmid:15894688

41. Tasevska N, Midthune D, Potischman N, et al. Use of the Predictive Sugars Biomarker to Evaluate Self-Reported Total Sugars Intake in the Observing Protein and Energy Nutrition (OPEN) Study. *Cancer Epidemiol Biomarkers Prev*. 2011;20(3):490–500. doi:10.1158/1055-9965.EPI-10-0820 pmid:21278329

42. Tasevska N, Pettinger M, Kipnis V, et al. Associations of Biomarker-Calibrated Intake of Total Sugars with the Risk of Type 2 Diabetes and Cardiovascular Disease in the Women's Health Initiative Observational Study. *Am J Epidemiol*. 2018;187(10):2126–35. doi:10.1093/aje/kwy115 pmid:29868784

43. Jenab M, Slimani N, Bictash M, Ferrari P, Bingham SA. Biomarkers in nutritional epidemiology: Applications, needs and new horizons. *Hum Genet*. 2009;125(5–6):507–25. doi:10.1007/s00439-009-0662-5 pmid:19357868

44. Lampe JW, Huang Y, Neuhouser ML, et al. Dietary biomarker evaluation in a controlled feeding study in women from the Women's Health Initiative cohort. *Am J Clin Nutr*. 2017;105(April):466–75. doi:10.3945/ajcn.116.144840.1 pmid:28031191

45. Couillard C, Lemieux S, Vohl MC, Couture P, Lamarche B. Carotenoids as biomarkers of fruit and vegetable intake in men and women. *Br J Nutr*. 2016;116(7):1206–15. doi:10.1017/S0007114516003056 pmid:27572625

46. Tinker LF, Sarto GE, Howard B V, et al. Biomarker-calibrated dietary energy and protein intake associations with diabetes risk among postmenopausal women from the Women's Health Initiative. *Am J Clin Nutr*. 2011;94(6):1600. doi:10.3945/AJCN.111.018648 pmid:22071707

47. Prentice RL, Pettinger M, Neuhouser ML, et al. Application of blood concentration biomarkers in nutritional epidemiology: example of carotenoid and tocopherol intake in relation to chronic disease risk. *Am J Clin Nutr*. 2019;109(4):1–8. doi:10.1093/ajcn/nqy360 pmid:30915444

48. Imamura F, Fretts A, Marklund M, et al. Fatty acid biomarkers of dairy fat consumption and incidence of type 2 diabetes: a pooled analysis of prospective cohort studies. *PLoS Med*. 2018;15(10). doi:10.1371/JOURNAL.PMED.1002670 pmid:30303968

49. Zheng JS, Sharp SJ, Imamura F, et al. Association of plasma biomarkers of fruit and vegetable intake with incident type 2 diabetes: EPIC-InterAct case-cohort study in eight European countries. *BMJ*. 2020;370. doi:10.1136/bmj.m2194 pmid:32641421

50. Neuhouser ML, Pettinger M, Lampe JW, et al. Novel Application of Nutritional Biomarkers From a Controlled Feeding Study and an Observational Study to Characterization of Dietary Patterns in

Postmenopausal Women. *Am J Epidemiol*. 2021;190(11):2461–73. doi:10.1093/aje/kwab171 pmid:34142699

51. Garcia-Aloy M, Rabassa M, Casas-Agustench P, et al. Novel strategies for improving dietary exposure assessment: Multiple-data fusion is a more accurate measure than the traditional single-biomarker approach. *Trends Food Sci Technol*. 2017;69:220–9. doi:10.1016/j.tifs.2017.04.013

52. Song X, Huang Y, Neuhouser ML, et al. Dietary long-chain fatty acids and carbohydrate biomarker evaluation in a controlled feeding study in participants from the Women's Health Initiative cohort. *Am J Clin Nutr*. 2017;105(6):1272–82. doi:10.3945/ajcn.117.153072 pmid:28446501

53. Souverein OW, De Vries JHM, Freese R, et al. Prediction of fruit and vegetable intake from biomarkers using individual participant data of diet-controlled intervention studies. *Br J Nutr*. 2015;113(9):1396–409. doi:10.1017/S0007114515000355 pmid:25850683

54. Garcia-Aloy M, Andres-Lacueva C. Food Intake Biomarkers for Increasing the Efficiency of Dietary Pattern Assessment through the Use of Metabolomics: Unforeseen Research Requirements for Addressing Current Gaps. *J Agric Food Chem*. 2018;66(1):5–7. doi:10.1021/acs.jafc.7b05586 pmid:29272125

55. Brennan L, Hu FB. Metabolomics-Based Dietary Biomarkers in Nutritional Epidemiology-Current Status and Future Opportunities. *Mol Nutr Food Res*. 2018;1701064:1–5. doi:10.1002/mnfr.201701064 pmid:29688616

56. Satija A, Stampfer MJ, Rimm EB, Willett W, Hu FB. Perspective: Are Large, Simple Trials the Solution for Nutrition Research? *Adv Nutr*. 2018;9(4):378–87. doi:10.1093/advances/nmy030

57. Schwingshackl L, Balduzzi S, Beyerbach J, et al. Evaluating agreement between bodies of evidence from randomised controlled trials and cohort studies in nutrition research: Meta-epidemiological study [Internet]. Vol. 374, The BMJ. BMJ; 2021. doi:10.1136/bmj.n1864 pmid:34526355

58. Papadaki A, Nolen-Doerr E, Mantzoros CS. The effect of the Mediterranean diet on metabolic health: A systematic review and meta-analysis of controlled trials in adults [Internet]. Vol. 12, Nutrients. MDPI AG; 2020. p. 1–21. doi:10.3390/nu12113342 pmid:33143083

59. Arnett DK, Blumenthal RS, Albert MA, et al. 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation*. 2019;140(11):e563–95. doi:10.1161/CIR.0000000000000677 pmid:30879339

60. Kim H, Rebholz CM. Metabolomic Biomarkers of Healthy Dietary Patterns and Cardiovascular Outcomes. *Curr Atheroscler Rep*. 2021;23(6):26. doi:10.1007/s11883-021-00921-8 pmid:33782776

61. de Lorgeril M, Renaud S, Salen P, et al. Mediterranean alpha-linolenic acid-rich diet in secondary prevention of coronary heart disease. *Lancet*. 1994;343(8911):1454–9. doi:10.1016/S0140-6736(94)92580-1 pmid:7911176

62. De Lorgeril M, Salen P, Martin JL, et al. Mediterranean dietary pattern in a randomized trial: Prolonged survival and possible reduced cancer rate. *Arch Intern Med*. 1998;158(11):1181–7. doi:10.1001/archinte.158.11.1181 pmid:9625397

63. Renaud S, de Lorgeril M, Delaye J, et al. Cretan Mediterranean diet for prevention of coronary heart disease. *Am J Clin Nutr*. 1995;61(6 Suppl):1360S-1367S. doi:10.1093/ajcn/61.6.1360S pmid:7754988

64. Davis C, Hodgson J, Bryan J, et al. Older Australians Can Achieve High Adherence to the Mediterranean Diet during a 6 Month Randomised Intervention; Results from the Medley Study. *Nutrients*. 2017;9(6):534. doi:10.3390/nu9060534

65. Sotos-Prieto M, Christophi C, Black A, et al. Assessing Validity of Self-Reported Dietary Intake within a Mediterranean Diet Cluster Randomized Controlled Trial among US Firefighters. *Nutrients*. 2019;11(9):2250. doi:10.3390/nu11092250

66. Cuparencu C, Praticó G, Hemeryck LY, et al. Biomarkers of meat and seafood intake: An extensive literature review [Internet]. Vol. 14, Genes and Nutrition. Genes Nutr; 2019. doi:10.1186/s12263-019-0656-4 pmid:31908682

67. Jackson KH, Harris WS. Blood Fatty Acid Profiles: New Biomarkers for Cardiometabolic Disease Risk [Internet]. Vol. 20, Current Atherosclerosis Reports. Springer US; 2018. p. 22. doi:10.1007/s11883-018-0722-1

68. Davis C, Bryan J, Hodgson J, Murphy K. Definition of the mediterranean diet: A literature review [Internet]. Vol. 7, Nutrients. Nutrients; 2015. p. 9139–53. doi:10.3390/nu7115459 pmid:26556369

69. Guasch-Ferré M, Bhupathiraju SN, Hu FB. Use of Metabolomics in Improving Assessment of Dietary Intake. *Clin Chem*. 2018;64(1):82–98. doi:10.1373/clinchem.2017.272344 pmid:29038146

70. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372. doi:10.1136/BMJ.N71 pmid:33782057

71. Campbell M, McKenzie JE, Sowden A, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. *BMJ*. 2020;368. doi:10.1136/BMJ.L6890 pmid:31948937

72. National Heart, Lung and BI (NHLBI). Study quality assessment tools. Bethesda: U.S. Department of Health & Human Services [Internet]. [cited 2022 Apr 7]. Available from: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools

73. Hedges L V. Distribution Theory for Glass's Estimator of Effect size and Related Estimators. *J Educ Stat*. 1981;6(2):107–28. doi:10.3102/10769986006002107

74. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Methodol*. 2014;14(1). doi:10.1186/1471-2288-14-135 pmid:25524443

75. Higgins JPT, White IR, Anzures-Cabrera J. Meta-analysis of skewed data: Combining results reported

on log-transformed or raw scales. *Stat Med*. 2008;27(29):6072–92. doi:10.1002/sim.3427 pmid:18800342

76. Higgins JPT, Li T, Deeks JJ. Choosing effect measures and computing estimates of effect. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ WV, editor. Cochrane Handbook for Systematic Reviews of Interventions version 62 (updated February 2021). Cochrane; 2021. p. 143–76.

77. Altman DG, Bland JM. Statistics notes: How to obtain the P value from a confidence interval. *BMJ*. 2011;343(7825). doi:10.1136/bmj.d2304 pmid:22803193

78. Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ*. 2011;342(7804):964–7. doi:10.1136/bmj.d549 pmid:21310794

79. Da Costa BR, Nüesch E, Rutjes AW, et al. Combining follow-up and change data is valid in meta-analyses of continuous outcomes: A meta-epidemiological study. *J Clin Epidemiol*. 2013;66(8):847–55. doi:10.1016/j.jclinepi.2013.03.009 pmid:23747228

80. 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. *Prog Lipid Res*. 2008;47(5):348–80. doi:10.1016/j.plipres.2008.03.003 pmid:18435934

81. Mayneris-Perxachs J, Sala-Vila A, Chisaguano M, et al. Effects of 1-year intervention with a Mediterranean diet on plasma fatty acid composition and metabolic syndrome in a population at high cardiovascular risk. *PLoS One*. 2014;9(3):1–11. doi:10.1371/journal.pone.0085202

82. Perona JS, Covas MI, Fitó M, et al. Reduction in systemic and VLDL triacylglycerol concentration after a 3-month Mediterranean-style diet in high-cardiovascular-risk subjects. *J Nutr Biochem*. 2010;21(9):892–8. doi:10.1016/j.jnutbio.2009.07.005

83. Djuric Z, Ren J, Blythe J, VanLoon G, Sen A. A Mediterranean dietary intervention in healthy American women changes plasma carotenoids and fatty acids in distinct clusters. *Nutr Res*. 2009;29(3):156–63. doi:10.1016/j.nutres.2009.03.001

84. Furuya-Kanamori L, Barendregt JJ, Doi SAR. A new improved graphical and quantitative method for detecting bias in meta-analysis. *Int J Evid Based Healthc*. 2018;16(4):195–203. doi:10.1097/XEB.0000000000000141 pmid:29621038

85. Elbourne DR, Altman DG, Higgins JPT, et al. Meta-analyses involving cross-over trials: Methodological issues. *Int J Epidemiol*. 2002;31(1):140–9. doi:10.1093/ije/31.1.140 pmid:11914310

86. Li T, Yu T, Hawkins BS, Dickersin K. Design, analysis, and reporting of crossover trials for inclusion in a meta-analysis. *PLoS One*. 2015;10(8). doi:10.1371/journal.pone.0133023 pmid:26284684

87. Fuentes F, López-Miranda J, Pérez-Martínez P, et al. Chronic effects of a high-fat diet enriched with virgin olive oil and a low-fat diet enriched with α-linolenic acid on postprandial endothelial function in healthy men. *Br J Nutr*. 2008;100(1):159–65. doi:10.1017/S0007114508888708

88. Itsiopoulos C, Brazionis L, Kaimakamis M, et al. Can the Mediterranean diet lower HbA1c in type 2 diabetes? Results from a randomized cross-over study. *Nutr Metab Cardiovasc Dis*. 2011;21(9):740–7. doi:10.1016/j.numecd.2010.03.005 pmid:20674309

89. Pérez-Jiménez F, López-Miranda J, Pinillos MD, et al. A mediterranean and a high-carbohydrate diet improve glucose metabolism in healthy young persons. *Diabetologia*. 2001;44(11):2038–43. doi:10.1007/s001250100009

90. Marin C, Ramirez R, Delgado-Lista J, et al. Mediterranean diet reduces endothelial damage and improves the regenerative capacity of endothelium. *Am J Clin Nutr*. 2011;93(2):267–74. pmid:21123460

91. Barona J, Jones JJ, Kopec RE, et al. A Mediterranean-style low-glycemic-load diet increases plasma carotenoids and decreases LDL oxidation in women with metabolic syndrome. *J Nutr Biochem*. 2012;23(6):609–15. doi:10.1016/j.jnutbio.2011.02.016

92. Bihuniak JD, Ramos A, Huedo-Medina T, et al. Adherence to a Mediterranean-Style Diet and Its Influence on Cardiovascular Risk Factors in Postmenopausal Women. *J Acad Nutr Diet*. 2016;116(11):1767–75. doi:10.1016/j.jand.2016.06.377

93. Michalsen A, Lehmann N, Pithan C, et al. Mediterranean diet has no effect on markers of inflammation and metabolic risk factors in patients with coronary artery disease. *Eur J Clin Nutr*. 2006;60(4):478–85. doi:10.1038/SJ.EJCN.1602340 pmid:16306923

94. Seethaler B, Basrai M, Vetter W, et al. Fatty acid profiles in erythrocyte membranes following the Mediterranean diet – data from a multicenter lifestyle intervention study in women with hereditary breast cancer (LIBRE). *Clin Nutr*. 2019; doi:10.1016/j.clnu.2019.10.033 pmid:31735538

95. Ambring A, Johansson M, Axelsen M, et al. Mediterranean-inspired diet lowers the ratio of serum phospholipid n–6 to n–3 fatty acids, the number of leukocytes and platelets, and vascular endothelial growth factor in healthy subjects. *Am J Clin Nutr*. 2006;83(3):575–81. doi:10.1093/AJCN.83.3.575

96. Micallef MA, Garg ML. Anti-inflammatory and cardioprotective effects of n-3 polyunsaturated fatty acids and plant sterols in hyperlipidemic individuals. *Atherosclerosis*. 2009;204(2):476–82. doi:10.1016/j.atherosclerosis.2008.09.020 pmid:18977480

97. Perticone M, Maio R, Sciacqua A, et al. Ketogenic diet-induced weight loss is associated with an increase in Vitamin D levels in obese adults. *Molecules*. 2019;24(13). doi:10.3390/molecules24132499

98. Bemelmans WJE, Broer J, Feskens EJM, et al. Effect of an increased intake of alpha-linolenic acid and group nutritional education on cardiovascular risk factors: the Mediterranean Alpha-linolenic Enriched Groningen Dietary Intervention (MARGARIN) study. *Am J Clin Nutr*. 2002;75(2):221–7. doi:10.1093/AJCN/75.2.221 pmid:11815311

99. Hagfors L, Leanderson P, Sköldstam L, Andersson J, Johansson G. Antioxidant intake, plasma antioxidants and oxidative stress in a randomized, controlled, parallel, Mediterranean dietary

intervention study on patients with rheumatoid arthritis. *Nutr J*. 2003;2:1–11. doi:10.1186/1475-2891-2-5

100. Hagfors L, Nilsson I, Sköldstam L, Johansson G. Fat intake and composition of fatty acids in serum phospholipids in a randomized, controlled, Mediterranean dietary intervention study on patients with rheumatoid arthritis. *Nutr Metab*. 2005;2:26. doi:10.1186/1743-7075-2-26

101. Griffin LE, Djuric Z, Angiletta CJ, et al. A Mediterranean diet does not alter plasma trimethylamine: N -oxide concentrations in healthy adults at risk for colon cancer. *Food Funct*. 2019;10(4):2138–47. doi:10.1039/c9fo00333a pmid:30938383

102. Li Y, Sen A, Ren J, et al. Effects of vitamin E from supplements and diet on colonic α-and γ-tocopherol concentrations in persons at increased colon cancer risk. *Nutr Cancer*. 2015;67(1):73–81. doi:10.1080/01635581.2015.965333 pmid:25372556

103. Porenta SR, Ko YA, Gruber SB, et al. Interaction of fatty acid genotype and diet on changes in colonic fatty acids in a Mediterranean diet intervention study. *Cancer Prev Res*. 2013;6(11):1212–21. doi:10.1158/1940-6207.CAPR-13-0131

104. Sen A, Ren J, Ruffin MT, et al. Relationships between serum and colon concentrations of carotenoids and fatty acids in randomized dietary intervention trial. *Cancer Prev Res*. 2013;6(6):558–65. doi:10.1158/1940-6207.CAPR-13-0019

105. Hjerkinn EM, Abdelnoor M, Breivik L, et al. Effect of diet or very long chain ω-3 fatty acids on progression of atherosclerosis, evaluated by carotid plaques, intima-media thickness and by pulse wave propagation in elderly men with hypercholesterolaemia. *Eur J Cardiovasc Prev Rehabil*. 2006;13(3):325–33. doi:10.1097/01.HJR.0000209817.28444.FB

106. Jaacks LM, Sher S, De Staercke C, et al. Pilot randomized controlled trial of a Mediterranean diet or diet supplemented with fish oil, walnuts, and grape juice in overweight or obese US adults. *BMC Nutr*. 2018;4(1):26. doi:10.1186/s40795-018-0234-y

107. Jula A, Marniemi J, Huupponen R, et al. Effects of diet and simvastatin on serum lipids, insulin, and antioxidants in hypercholesterolemic men: a randomized controlled trial. *JAMA*. 2002;287(5):598–605. doi:10.1001/JAMA.287.5.598 pmid:11829698

108. Muzsik A, Bajerska J, Jeleń HH, et al. FADS1 and FADS2 polymorphism are associated with changes in fatty acid concentrations after calorie-restricted Central European and Mediterranean diets. *Menopause*. 2019;26(12):1415–24. doi:10.1097/GME.0000000000001409 pmid:31567872

109. Jennings A, Berendsen AM, De Groot LCPGM, et al. Mediterranean-style diet improves systolic blood pressure and arterial stiffness in older adults: Results of a 1-year european multi-center trial. *Hypertension*. 2019;73(3):578–86. doi:10.1161/HYPERTENSIONAHA.118.12259

110. O'Neill C. The effects of whole-diet interactions on vascular health and inflammatory and fatty acid status (Doctoral thesis). University of East Anglia; 2015.

111. Parcina M, Brune M, Kaese V, et al. No short-term effects of calorie-controlled Mediterranean or fast food dietary interventions on established biomarkers of vascular or metabolic risk in healthy individuals. *Nutr Res Pract*. 2015;9(2):165–73. doi:10.4162/nrp.2015.9.2.165

112. Leighton F, Cuevas A, Guasch V, et al. Plasma polyphenols and antioxidants, oxidative DNA damage and endothelial function in a diet and wine intervention study in humans. In: Drugs under Experimental and Clinical Research. 1999. p. 133–41. pmid:10370876

113. Mezzano D, Leighton F, Strobel P, et al. Mediterranean diet, but not red wine, is associated with beneficial changes in primary haemostasis. *Eur J Clin Nutr*. 2003;57(3):439–46. doi:10.1038/sj.ejcn.1601558

114. Urquiaga I, Guasch V, Marshall G, et al. Effect of Mediterranean and Occidental diets, and red wine, on plasma fatty acids in humans. An intervention study. *Biol Res*. 2004;37(2):253–61. doi:10.4067/s0716-97602004000200012 pmid:15455655

115. Urquiaga I, Strobel P, Perez D, et al. Mediterranean diet and red wine protect against oxidative damage in young volunteers. *Atherosclerosis*. 2010;211(2):694–9. doi:10.1016/j.atherosclerosis.2010.04.020

116. Bulló M, Amigó-Correig P, Márquez-Sandoval F, et al. Mediterranean diet and high dietary acid load associated with mixed nuts: Effect on bone metabolism in elderly subjects. *J Am Geriatr Soc*. 2009;57(10):1789–98. doi:10.1111/j.1532-5415.2009.02481.x

117. Estruch R, Ros E, Salas-Salvadó J, et al. Primary Prevention of Cardiovascular Disease with a Mediterranean Diet Supplemented with Extra-Virgin Olive Oil or Nuts. *N Engl J Med*. 2018;378(25):e34. doi:10.1056/NEJMoa1800389 pmid:29897866

118. Guasch-Ferré M, Hu FB, Ruiz-Canela M, et al. Plasma metabolites from choline pathway and risk of cardiovascular disease in the PREDIMED (Prevention with Mediterranean Diet) study. *J Am Heart Assoc*. 2017;6(11). doi:10.1161/JAHA.117.006524

119. Medina-Remón A, Tresserra-Rimbau A, Pons A, et al. Effects of total dietary polyphenols on plasma nitric oxide and blood pressure in a high cardiovascular risk cohort. The PREDIMED randomized trial. *Nutr Metab Cardiovasc Dis*. 2015;25(1):60–7. doi:10.1016/j.numecd.2014.09.001 pmid:25315667

120. Ruiz-Canela M, Toledo E, Clish CB, et al. Plasma branched-chain amino acids and incident cardiovascular disease in the PREDIMED Trial. *Clin Chem*. 2016;62(4):582–92. doi:10.1373/clinchem.2015.251710

121. Richard C, Couture P, Desroches S, et al. Effect of the Mediterranean diet with and without weight loss on surrogate markers of cholesterol homeostasis in men with the metabolic syndrome. *Br J Nutr*. 2012;107(5):705–11. doi:10.1017/S0007114511003436

122. Skouroliakou M, Grosomanidis D, Massara P, et al. Serum antioxidant capacity, biochemical profile and body composition of breast cancer survivors in a randomized Mediterranean dietary intervention study. *Eur J Nutr*. 2018;57(6):2133–45. doi:10.1007/s00394-017-1489-9

123. Sofi F, Dinu M, Pagliai G, et al. Low-calorie vegetarian versus mediterranean diets for reducing body weight and improving cardiovascular risk profile. *Circulation*. 2018;137(11):1103–13. doi:10.1161/CIRCULATIONAHA.117.030088 pmid:29483085

124. Stachowska E, Wesołowska T, Olszewska M, et al. Elements of Mediterranean diet improve oxidative status in blood of kidney graft recipients. *Br J Nutr*. 2005;93(3):345–52. doi:10.1079/bjn20051374

125. Thomazella MCD, Góes MFS, Andrade CR, et al. Effects of high adherence to mediterranean or low-fat diets in medicated secondary prevention patients. *Am J Cardiol*. 2011;108(11):1523–9. doi:10.1016/j.amjcard.2011.07.008

126. Tutino V, De Nunzio V, Caruso MG, et al. Aerobic physical activity and a low glycemic diet reduce the AA/EPA ratio in red blood cell membranes of patients with NAFLD. *Nutrients*. 2018;10(9):1299. doi:10.3390/nu10091299

127. Tuttle KR, Shuler LA, Packard DP, et al. Comparison of low-fat versus Mediterranean-style dietary intervention after first myocardial infarction (from The Heart Institute of Spokane Diet Intervention and Evaluation Trial). *Am J Cardiol*. 2008;101(11):1523–30. doi:10.1016/J.AMJCARD.2008.01.038 pmid:18489927

128. Vincent-Baudry S, Defoort C, Gerber M, et al. The Medi-RIVAGE study: Reduction of cardiovascular disease risk factors after a 3-mo intervention with a Mediterranean-type diet or a low-fat diet. *Am J Clin Nutr*. 2005;82(5):964–71. doi:10.1093/ajcn/82.5.964 pmid:16280426

129. Lindman AS, Pedersen JI, Hjerkinn EM, et al. The effects of long-term diet and omega-3 fatty acid supplementation on coagulation factor VII and serum phospholipids with special emphasis on the R353Q polymorphism of the FVII gene. *Thromb Haemost*. 2004;91(6):1097–104. doi:10.1160/TH03-12-0773 pmid:15175795

130. Trøseid M, Arnesen H, Hjerkinn EM, Seljeflot I. Serum levels of interleukin-18 are reduced by diet and n-3 fatty acid intervention in elderly high-risk men. *Metabolism*. 2009;58(11):1543–9. doi:10.1016/j.metabol.2009.04.031

131. Estruch R, Martínez-González MA, Corella D, et al. Effect of a high-fat Mediterranean diet on bodyweight and waist circumference: a prespecified secondary outcomes analysis of the PREDIMED randomised controlled trial. *Lancet Diabetes Endocrinol*. 2019;7(5):e6–17. doi:10.1016/S2213-8587(19)30074-9 pmid:31003626

132. Burrows TL, Williams R, Rollo M, et al. Plasma carotenoid levels as biomarkers of dietary carotenoid consumption: A systematic review of the validation studies. *J Nutr Intermed Metab*. 2015;2(1–2):15–64. doi:10.1016/j.jnim.2015.05.001

133. Mantzioris E, B BSM, Villani A. Impact of the Mediterranean Dietary pattern on n-3 fatty acid tissue levels–A systematic review. *Prostaglandins, Leukot Essent Fat Acids*. 2021;176:102387. doi:10.1016/j.plefa.2021.102387 pmid:34929617

134. Cleophas TJM. A simple method for the estimation of interaction bias in crossover studies. *J Clin Pharmacol*. 1990;30(11):1036–40. doi:10.1002/J.1552-4604.1990.TB03591.X pmid:2243151

135. Pennant M, Steur M, Moore C, Butterworth A, Johnson L. Comparative validity of vitamin C and carotenoids as indicators of fruit and vegetable intake: a systematic review and meta-analysis of randomised controlled trials. *Br J Nutr*. 2015;114(9):1331–40. doi:10.1017/S0007114515003165 pmid:26349405

136. Lin PY, Huang SY, Su KP. A meta-analytic review of polyunsaturated fatty acid compositions in patients with depression. *Biol Psychiatry*. 2010;68(2):140–7. doi:10.1016/j.biopsych.2010.03.018 pmid:20452573

137. Cummings P. Arguments for and against standardized mean differences (effect sizes) [Internet]. Vol. 165, Archives of Pediatrics and Adolescent Medicine. Arch Pediatr Adolesc Med; 2011. p. 592–6. doi:10.1001/archpediatrics.2011.97 pmid:21727271

138. Brenna JT, Plourde M, Stark KD, Jones PJ, Lin YH. Best practices for the design, laboratory analysis, and reporting of trials involving fatty acids [Internet]. Vol. 108, American Journal of Clinical Nutrition. Oxford University Press; 2018. p. 211–27. doi:10.1093/ajcn/nqy089 pmid:29931035

139. Wanders AJ, Alssema M, Van Greevenbroek MJ, et al. Comment on Sergeant et al.: Impact of methods used to express levels of circulating fatty acids on the degree and direction of associations with blood lipids in humans [Internet]. Vol. 115, British Journal of Nutrition. Br J Nutr; 2016. p. 2077–8. doi:10.1017/S0007114516000933 pmid:27079885

140. Zwetsloot PP, Van Der Naald M, Sena ES, et al. Standardized mean differences cause funnel plot distortion in publication bias assessments. *Elife*. 2017;6. doi:10.7554/ELIFE.24260 pmid:28884685

141. Deeks JJ, Higgins JPT, Altman DG, Group CSM. Analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ WV, editor. Cochrane Handbook for Systematic Reviews of Interventions version 62 (updated February 2021). Cochrane; 2021. p. 241–84.

142. Arballo J, Amengual J, Erdman JW. Lycopene: A critical review of digestion, absorption, metabolism, and excretion [Internet]. Vol. 10, Antioxidants. Antioxidants (Basel); 2021. p. 1–16. doi:10.3390/antiox10030342 pmid:33668703

143. Ahuja KDK, Pittaway JK, Ball MJ. Effects of olive oil and tomato lycopene combination on serum lycopene, lipid profile, and lipid oxidation. *Nutrition*. 2006;22(3):259–65. doi:10.1016/j.nut.2005.07.015 pmid:16413753

144. Rinaldi de Alvarenga JF, Tran C, Hurtado-Barroso S, et al. Home cooking and ingredient synergism improve lycopene isomer production in Sofrito. *Food Res Int*. 2017;99(Pt 2):851–61. doi:10.1016/J.FOODRES.2017.01.009 pmid:28847422

145. Tanumihardjo SA, Russell RM, Stephensen CB, et al. Biomarkers of Nutrition for Development (BOND)-Vitamin A Review. *J Nutr*. 2016;146(9):1816S-1848S. doi:10.3945/JN.115.229708

pmid:27511929

146. Lietz G, Lange J, Rimbach G. Molecular and dietary regulation of β,β-carotene 15,15′-monooxygenase 1 (BCMO1) [Internet]. Vol. 502, Archives of Biochemistry and Biophysics. Arch Biochem Biophys; 2010. p. 8–16. doi:10.1016/j.abb.2010.06.032 pmid:20599666

147. Waniek S, Di Giuseppe R, Esatbeyoglu T, et al. Vitamin E (α-and γ-tocopherol) levels in the community: Distribution, clinical and biochemical correlates, and association with dietary patterns. *Nutrients*. 2018;10(1). doi:10.3390/nu10010003 pmid:29267223

148. Henríquez-Sánchez P, Sánchez-Villegas A, Doreste-Alonso J, et al. Dietary assessment methods for micronutrient intake: a systematic review on vitamins. *Br J Nutr*. 2009;102 Suppl 1. doi:10.1017/S0007114509993126 pmid:20100364

149. Traber MG. Vitamin E inadequacy in humans: causes and consequences. *Adv Nutr*. 2014;5(5):503–14. doi:10.3945/AN.114.006254 pmid:25469382

150. El-Sohemy A, Baylin A, Ascherio A, et al. Population-based study of alpha- and gamma-tocopherol in plasma and adipose tissue as biomarkers of intake in Costa Rican adults. *Am J Clin Nutr*. 2001;74(3):356–63. doi:10.1093/AJCN/74.3.356 pmid:11522560

151. Abdulla KA, Um CY, Gross MD, Bostick RM. Circulating γ-Tocopherol Concentrations Are Inversely Associated with Antioxidant Exposures and Directly Associated with Systemic Oxidative Stress and Inflammation in Adults. *J Nutr*. 2018;148(9):1453–61. doi:10.1093/JN/NXY132 pmid:30184224

152. Tulleken JE, Limburg PC, Muskiet FAJ, Rijswijk MHV. Vitamin E status during dietary fish oil supplementation in rheumatoid arthritis. *Arthritis Rheum*. 1990;33(9):1416–9. doi:10.1002/ART.1780330914 pmid:2206140

153. Imamura F, Fretts AM, Marklund M, et al. Fatty acids in the de novo lipogenesis pathway and incidence of type 2 diabetes: A pooled analysis of prospective cohort studies. *PLoS Med*. 2020;17(6). doi:10.1371/JOURNAL.PMED.1003102 pmid:32530938

154. Pranger IG, Joustra ML, Corpeleijn E, et al. Fatty acids as biomarkers of total dairy and dairy fat intakes: A systematic review and meta-analysis. *Nutr Rev*. 2019;77(1):46–63. doi:10.1093/nutrit/nuy048 pmid:30307550

155. Fretts A, Imamura F, Marklund M, et al. Associations of circulating very-long chain saturated fatty acids and incident type 2 diabetes: a pooled analysis of prospective cohort studies. *Am J Clin Nutr*. 2019;(in press).

156. Garcia-Aloy M, Hulshof PJM, Estruel-Amades S, et al. Biomarkers of food intake for nuts and vegetable oils: an extensive literature search [Internet]. Vol. 14, Genes and Nutrition. BioMed Central Ltd.; 2019. doi:10.1186/s12263-019-0628-8

157. Freisling H, Fahey MT, Moskal A, et al. Region-specific nutrient intake patterns exhibit a geographical gradient within and between European countries. *J Nutr*. 2010;140(7):1280–6.

doi:10.3945/JN.110.121152 pmid:20484545

158. Hodson L, Eyles HC, McLachlan KJ, et al. Plasma and erythrocyte fatty acids reflect intakes of saturated and n-6 PUFA within a similar time frame. *J Nutr*. 2014;144(1):33–41. doi:10.3945/JN.113.183749 pmid:24225449

159. Skeaff CM, Hodson L, McKenzie JE. Dietary-induced changes in fatty acid composition of human plasma, platelet, and erythrocyte lipids follow a similar time course. *J Nutr*. 2006;136(3):565–9. doi:10.1093/JN/136.3.565 pmid:16484525

160. Pertiwi K, Kok DE, Wanders AJ, et al. Circulating n-3 fatty acids and linoleic acid as indicators of dietary fatty acid intake in post-myocardial infarction patients. *Nutr Metab Cardiovasc Dis*. 2019;29(4):343–50. doi:10.1016/j.numecd.2018.12.010 pmid:30718141

161. Forsythe CE, Phinney SD, Fernandez ML, et al. Comparison of low fat and low carbohydrate diets on circulating fatty acid composition and markers of inflammation. *Lipids*. 2008;43(1):65–77. doi:10.1007/S11745-007-3132-7 pmid:18046594

162. Simon JA, Fong J, Bernert JT, Browner WS. Relation of smoking and alcohol consumption to serum fatty acids. *Am J Epidemiol*. 1996;144(4):325–34. doi:10.1093/oxfordjournals.aje.a008933 pmid:8712189

163. Kawashima H. Intake of arachidonic acid-containing lipids in adult humans: dietary surveys and clinical trials. *Lipids Health Dis*. 2019;18(1). doi:10.1186/S12944-019-1039-Y pmid:30992005

164. Rett BS, Whelan J. Increasing dietary linoleic acid does not increase tissue arachidonic acid content in adults consuming Western-type diets: a systematic review. *Nutr Metab (Lond)*. 2011;8. doi:10.1186/1743-7075-8-36 pmid:21663641

165. Hussein N, Ah-Sing E, Wilkinson P, et al. Long-chain conversion of [13C]linoleic acid and alpha-linolenic acid in response to marked changes in their dietary intake in men. *J Lipid Res*. 2005;46(2):269–80. doi:10.1194/JLR.M400225-JLR200 pmid:15576848

166. Calder PC, Campoy C, Eilander A, et al. A systematic review of the effects of increasing arachidonic acid intake on PUFA status, metabolism and health-related outcomes in humans. *Br J Nutr*. 2019;121(11):1201–14. doi:10.1017/S0007114519000692 pmid:31130146

167. Cicero AFG, Nascetti S, Nyyssönen K, et al. Changes in LDL fatty acid composition as a response to olive oil treatment are inversely related to lipid oxidative damage: The EUROLIVE study. *J Am Coll Nutr*. 2008;27(2):314–20. doi:10.1080/07315724.2008.10719705 pmid:18689564

168. Simopoulos AP. Omega-3 fatty acids in health and disease and in growth and development. *Am J Clin Nutr*. 1991;54(3):438–63. doi:10.1093/AJCN/54.3.438 pmid:1908631

169. Simopoulos AP. The importance of the ratio of omega-6/omega-3 essential fatty acids. *Biomed Pharmacother*. 2002;56(8):365–79. doi:10.1016/S0753-3322(02)00253-6 pmid:12442909

170. Harris WS, Mozaffarian D, Rimm E, et al. Omega-6 fatty acids and risk for cardiovascular disease: A science advisory from the American Heart Association nutrition subcommittee of the council on nutrition, physical activity, and metabolism; council on cardiovascular nursing; and council on epidem. *Circulation*. 2009;119(6):902–7. doi:10.1161/CIRCULATIONAHA.108.191627 pmid:19171857

171. Stanley JC, Elsom RL, Calder PC, et al. UK Food Standards Agency Workshop Report: the effects of the dietary n-6:n-3 fatty acid ratio on cardiovascular health. *Br J Nutr*. 2007;98(6):1305–10. doi:10.1017/S000711450784284X pmid:18039412

172. Donahue SMA, Rifas-Shiman SL, Olsen SF, et al. Associations of maternal prenatal dietary intake of n-3 and n-6 fatty acids with maternal and umbilical cord blood levels. *Prostaglandins Leukot Essent Fatty Acids*. 2009;80(5–6):289–96. doi:10.1016/J.PLEFA.2009.02.007 pmid:19380219

173. Laviolle B, Froger-Bompas C, Guillo P, et al. Relative validity and reproducibility of a 14-item semi-quantitative food frequency questionnaire for cardiovascular prevention. *Eur J Cardiovasc Prev Rehabil*. 2005;12(6):587–95. doi:10.1097/01.HJR.0000186625.98604.5D pmid:16319550

174. Brantsæter AL, Haugen M, Hagve TA, et al. Self-reported dietary supplement use is confirmed by biological markers in the Norwegian Mother and Child Cohort Study (MoBa). *Ann Nutr Metab*. 2007;51(2):146–54. doi:10.1159/000103275 pmid:17536192

175. Hutchins-Wiese HL, Bales CW, Starr KNP. Mediterranean diet scoring systems: understanding the evolution and applications for Mediterranean and non-Mediterranean countries. *Br J Nutr*. 2021;1–22. doi:10.1017/S0007114521002476

176. Marklund M, Magnusdottir OK, Rosqvist F, et al. A Dietary Biomarker Approach Captures Compliance and Cardiometabolic Effects of a Healthy Nordic Diet in Individuals with Metabolic Syndrome. *J Nutr*. 2014;144(10):1642–9. doi:10.3945/jn.114.193771 pmid:25080537

177. Vázquez-Fresno R, Llorach R, Urpi-Sarda M, et al. Metabolomic pattern analysis after Mediterranean diet intervention in a nondiabetic population: A 1- and 3-year follow-up in the PREDIMED study. *J Proteome Res*. 2015;14(1):531–40. doi:10.1021/pr5007894 pmid:25353684

178. Garcia-Perez I, Posma JM, Gibson R, et al. Objective assessment of dietary patterns by use of metabolic phenotyping: a randomised, controlled, crossover trial. *Lancet Diabetes Endocrinol*. 2017;5(3):184–95. doi:10.1016/S2213-8587(16)30419-3 pmid:28089709

179. Langenberg C, Sharp S, Forouhi NG, et al. Design and cohort description of the InterAct Project: An examination of the interaction of genetic and lifestyle factors on the incidence of type 2 diabetes in the EPIC Study. *Diabetologia*. 2011;54(9):2272–82. doi:10.1007/s00125-011-2182-9 pmid:21717116

180. Schwingshackl L, Bogensberger B, Hoffmann G. Diet Quality as Assessed by the Healthy Eating Index, Alternate Healthy Eating Index, Dietary Approaches to Stop Hypertension Score, and Health Outcomes: An Updated Systematic Review and Meta-Analysis of Cohort Studies. *J Acad Nutr Diet*. 2018;118(1):74-100.e11. doi:10.1016/j.jand.2017.08.024 pmid:25680825

181. Jannasch F, Kröger J, Schulze MB. Dietary Patterns and Type 2 Diabetes: A Systematic Literature Review and Meta-Analysis of Prospective Studies. *J Nutr*. 2017;147(6):1174–82. doi:10.3945/jn.116.242552 pmid:28424256

182. Romaguera D, Guevara M, Norat T, et al. Mediterranean diet and type 2 diabetes risk in the European Prospective Investigation into Cancer and Nutrition (EPIC) study: the InterAct project. *Diabetes Care*. 2011;34(9):1913–8. doi:10.2337/dc11-0891 pmid:21788627

183. The InterAct Consortium. Adherence to predefined dietary patterns and incident type 2 diabetes in European populations: EPIC-InterAct Study. *Diabetologia*. 2014;57(2):321–33. doi:10.1007/s00125-013-3092-9 pmid:24196190

184. Shirani F, Salehi-Abargouei A, Azadbakht L. Effects of Dietary Approaches to Stop Hypertension (DASH) diet on some risk for developing type 2 diabetes: A systematic review and meta-analysis on controlled clinical trials. Vol. 29, Nutrition. Elsevier; 2013. p. 939–47. doi:10.1016/j.nut.2012.12.021 pmid:23473733

185. Ard JD, Grambow SC, Liu D, et al. The effect of the PREMIER interventions on insulin sensitivity. *Diabetes Care*. 2004;27(2):340–7. doi:10.2337/DIACARE.27.2.340 pmid:14747211

186. Appel LJ, Moore TJ, Obarzanek E, et al. A clinical trial of the effects of dietary patterns on blood pressure. DASH Collaborative Research Group. *N Engl J Med*. 1997;336(16):1117–24. doi:10.1056/NEJM199704173361601 pmid:9099655

187. Bellou V, Belbasis L, Tzoulaki I, Evangelou E. Risk factors for type 2 diabetes mellitus: An exposure-wide umbrella review of meta-analyses. *PLoS One*. 2018;13(3):1–27. doi:10.1371/journal.pone.0194127 pmid:29558518

188. Chiuve SE, Fung TT, Rimm EB, et al. Alternative Dietary Indices Both Strongly Predict Risk of Chronic Disease. *J Nutr*. 2012;142(6):1009–18. doi:10.3945/jn.111.157222

189. Chen GC, Koh WP, Neelakantan N, et al. Diet Quality Indices and Risk of Type 2 Diabetes Mellitus: The Singapore Chinese Health Study. *Am J Epidemiol*. 2018;187(12):2651–61. doi:10.1093/AJE/KWY183 pmid:30165478

190. Hodge AM, Karim MN, Hébert JR, Shivappa N, de Courten B. Association between Diet Quality Indices and Incidence of Type 2 Diabetes in the Melbourne Collaborative Cohort Study. *Nutrients*. 2021;13(11). doi:10.3390/NU13114162 pmid:34836416

191. Brevik A, Andersen LF, Karlsen A, et al. Six carotenoids in plasma used to assess recommended intake of fruits and vegetables in a controlled feeding study. *Eur J Clin Nutr*. 2004;58(8):1166–73. doi:10.1038/sj.ejcn.1601945 pmid:15054430

192. Karlsen A, Blomhoff R, Gundersen TE. High-throughput analysis of Vitamin C in human plasma with the use of HPLC with monolithic column and UV-detection. *J Chromatogr B Anal Technol Biomed Life Sci*. 2005;824(1–2):132–8. doi:10.1016/j.jchromb.2005.07.008

193. Cooper AJM, Sharp SJ, Luben RN, et al. The association between a biomarker score for fruit and vegetable intake and incident type 2 diabetes: The EPIC-Norfolk study. *Eur J Clin Nutr*. 2015;69(4):449–54. doi:10.1038/ejcn.2014.246 pmid:25387899

194. Harding AH, Wareham NJ, Bingham SA, et al. Plasma vitamin C level, fruit and vegetable consumption, and the risk of new-onset type 2 diabetes mellitus The European prospective investigation of Cancer-Norfolk prospective study. *Arch Intern Med*. 2008;168(14):1493–9. doi:10.1001/archinte.168.14.1493 pmid:18663161

195. Wang L, Summerhill K, Rodriguez-Canas C, et al. Development and validation of a robust automated analysis of plasma phospholipid fatty acids for metabolic phenotyping of large epidemiological studies. *Genome Med*. 2013;5(4):39. doi:10.1186/gm443

196. Forouhi NG, Koulman A, Sharp SJ, et al. Differences in the prospective association between individual plasma phospholipid saturated fatty acids and incident type 2 diabetes: The EPIC-InterAct case-cohort study. *Lancet Diabetes Endocrinol*. 2014;2(10):810–8. doi:10.1016/S2213-8587(14)70146-9 pmid:25107467

197. Zheng J-S, Imamura F, Sharp SJ, et al. Association of plasma vitamin D metabolites with incident type 2 diabetes: EPIC-InterAct case-cohort study. *J Clin Endocrinol Metab*. 2018; doi:10.1210/jc.2018-01522 pmid:30418614

198. McCullough ML, Feskanich D, Hu FB, et al. Diet quality and major chronic disease risk in men and women: moving toward improved dietary guidance. *Am J Clin Nutr*. 2002;76(6):1261–71. doi:10.1093/ajcn/76.6.1261 pmid:12450892

199. Trichopoulou A, Kouris-Blazos A, Wahlqvist ML, et al. Diet and overall survival in elderly people. *BMJ*. 1995;311(7018):1457. doi:10.1136/bmj.311.7018.1457 pmid:8520331

200. Trichopoulou A, Costacou T, Bamia C, Trichopoulos D. Adherence to a Mediterranean Diet and Survival in a Greek Population. *N Engl J Med*. 2003;348(26):2599–608. doi:10.1056/nejmoa025039 pmid:12826634

201. Buckland G, Agudo A, Luján L, et al. Adherence to a Mediterranean diet and risk of gastric adenocarcinoma within the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort study. *Am J Clin Nutr*. 2010;91(2):381–90. doi:10.3945/ajcn.2009.28209

202. Günther ALB, Liese AD, Bell RA, et al. Association between the dietary approaches to hypertension diet and hypertension in youth with diabetes mellitus. *Hypertension*. 2009;53(1):6–12. doi:10.1161/HYPERTENSIONAHA.108.116665 pmid:19029488

203. Wareham NJ, Jakes RW, Rennie KL, et al. Validity and repeatability of a simple index derived from the short physical activity questionnaire used in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *Public Health Nutr*. 2003;6(4):407–13. doi:10.1079/phn2002439 pmid:12795830

204. Riboli E, Hunt K, Slimani N, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr*. 2002;5(6b):1113. doi:10.1079/PHN2002394 pmid:12639222

205. Slimani N, Deharveng G, Unwin I, et al. The EPIC nutrient database project (ENDB): A first attempt to standardize nutrient databases across the 10 European countries participating in the EPIC study. *Eur J Clin Nutr*. 2007;61(9):1037–56. doi:10.1038/sj.ejcn.1602679 pmid:17375121

206. Davey GK, Spencer EA, Appleby PN, et al. EPIC–Oxford: lifestyle characteristics and nutrient intakes in a cohort of 33 883 meat-eaters and 31 546 non meat-eaters in the UK. *Public Health Nutr*. 2003;6(3):259–68. doi:10.1079/phn2002430 pmid:12740075

207. Antonelli J, Claggett BL, Henglin M, et al. Statistical workflow for feature selection in human metabolomics data. *Metabolites*. 2019;9(7). doi:10.3390/metabo9070143

208. Wei R, Wang J, Su M, et al. Missing Value Imputation Approach for Mass Spectrometry-based Metabolomics Data. *Sci Rep*. 2018;8(1). doi:10.1038/s41598-017-19120-0 pmid:29330539

209. Stryker WS, Kaplan LA, Stein EA, et al. The relation of diet, cigarette smoking, and alcohol consumption to plasma beta-carotene and alpha-tocopherol levels. *Am J Epidemiol*. 1988;127(2):283–96. doi:10.1093/oxfordjournals.aje.a114804

210. Degerud E, Hoff R, Nygård O, et al. Cosinor modelling of seasonal variation in 25-hydroxyvitamin D concentrations in cardiovascular patients in Norway. *Eur J Clin Nutr*. 2016;70(4):517–22. doi:10.1038/EJCN.2015.200 pmid:26603883

211. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med*. 2011;30(4):377–99. doi:10.1002/sim.4067 pmid:21225900

212. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Ser B*. 2005;67(2):301–20. doi:10.1111/j.1467-9868.2005.00503.x

213. Cho S, Kim K, Kim YJ, et al. Joint Identification of Multiple Genetic Variants via Elastic-Net Variable Selection in a Genome-Wide Association Analysis. *Ann Hum Genet*. 2010;74(5):416–28. doi:10.1111/j.1469-1809.2010.00597.x

214. Ahrens A, Hansen CB, Schaffer ME, Ahrens A, Hansen CB. lassopack : Model Selection and Prediction with Regularized Regression in Stata Iassopack : Model Selection and Prediction with Regularized Regression in Stata. 2019;(12081).

215. Hoerl AE, Kennard RW. Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics*. 1970;12(1):69–82. doi:10.1080/00401706.1970.10488635

216. Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*. 1986;73(1):1–11. doi:10.1093/biomet/73.1.1

217. Orsini N, Bellocco R, Greenland S. Generalized least squares for trend estimation of summarized dose-

response data [Internet]. Vol. 6, Stata Journal. DPC Nederland; 2006. p. 40–57. doi:10.1177/1536867x0600600103

218. White IR. Multivariate random-effects meta-analysis. *Stata J*. 2009;9(1):40–56. doi:10.1177/1536867x0900900103

219. Tilling K, Williamson EJ, Spratt M, Sterne JAC, Carpenter JR. Appropriate inclusion of interactions was needed to avoid bias in multiple imputation. *J Clin Epidemiol*. 2016;80:107–15. doi:10.1016/j.jclinepi.2016.07.004 pmid:27445178

220. Seaman SR, Bartlett JW, White IR. Multiple imputation of missing covariates with non-linear effects and interactions: An evaluation of statistical methods. *BMC Med Res Methodol*. 2012;12. doi:10.1186/1471-2288-12-46 pmid:22489953

221. Li K-H, Meng H-L, Raghunathan TE, Rubin DB. Significance levels from repeated p-values with multiply-imputed data. *Stat Sin*. 1991;1:65–92.

222. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1–22. doi:10.18637/jss.v033.i01 pmid:20808728

223. Hu FB, Stampfer MJ, Rimm E, et al. Dietary fat and coronary heart disease: a comparison of approaches for adjusting for total energy intake and modeling repeated dietary measurements. *Am J Epidemiol*. 1999;149(6):531–40. doi:10.1093/oxfordjournals.aje.a009849 pmid:10084242

224. Sellem L, Jackson KG, Paper L, Givens ID, Lovegrove JA. Can individual fatty acids be used as functional biomarkers of dairy fat consumption in relation to cardiometabolic health? A narrative review. *Br J Nutr*. 2022; doi:10.1017/S0007114522000289 pmid:35086579

225. Cui Z, Gong G. The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features. *Neuroimage*. 2018;178:622–37. doi:10.1016/j.neuroimage.2018.06.001 pmid:29870817

226. Prentice RL, Huang Y. Nutritional epidemiology methods and related statistical challenges and opportunities. *Stat Theory Relat Fields*. 2018;2(1):2–10. doi:10.1080/24754269.2018.1466098

227. Prentice RL. Criteria for Acceptable Dietary Intake Biomarkers. *Cancer Epidemiol Biomarkers Prev*. 2022;31(6):1151–3. doi:10.1158/1055-9965.EPI-22-0180 pmid:35642392

228. Usher-Smith JA, Sharp SJ, Griffin SJ. The spectrum effect in tests for risk prediction, screening, and diagnosis. *BMJ*. 2016;353. doi:10.1136/bmj.i3139 pmid:27334281

229. Freedman LS, Commins JM, Moler JE, et al. Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for potassium and sodium intake. *Am J Epidemiol*. 2015;181(7):473–87. doi:10.1093/aje/kwu325 pmid:24918187

230. Freedman LS, Commins JM, Moler JE, et al. Pooled Results From 5 Validation Studies of Dietary Self-Report Instruments Using Recovery Biomarkers for Energy and Protein Intake. *Am J Epidemiol*.

2014;180(2):172–88. doi:10.1093/aje/kwu116 pmid:24918187

231. Freedman LS, Schatzkin A, Midthune D, Kipnis V. Dealing with dietary measurement error in nutritional cohort studies. *J Natl Cancer Inst*. 2011;103(14):1086–92. doi:10.1093/jnci/djr189 pmid:21653922

232. Jurek AM, Greenland S, Maldonado G, Church TR. Proper interpretation of non-differential misclassification effects: Expectations vs observations. *Int J Epidemiol*. 2005;34(3):680–7. doi:10.1093/ije/dyi060 pmid:15802377

233. Abay KA, Abate GT, Barrett CB, Bernard T. Correlated non-classical measurement errors, 'Second best' policy inference, and the inverse size-productivity relationship in agriculture. *J Dev Econ*. 2019;139:171–84. doi:10.1016/j.jdeveco.2019.03.008

234. Trijsburg L, Geelen A, Hollman PCH, et al. BMI was found to be a consistent determinant related to misreporting of energy, protein and potassium intake using self-report and duplicate portion methods. *Public Health Nutr*. 2017;20(4):598–607. doi:10.1017/S1368980016002743 pmid:27724995

235. Ferreri C, Masi A, Sansone A, et al. Fatty Acids in Membranes as Homeostatic, Metabolic and Nutritional Biomarkers: Recent Advancements in Analytics and Diagnostics. *Diagnostics*. 2016;7(1):1. doi:10.3390/diagnostics7010001

236. Greenland S, Schlesselman JJ, Criqui MH. The fallacy of employing standardized regression coefficients and correlations as measures of effect. *Am J Epidemiol*. 1986;123(2):203–8. doi:10.1093/OXFORDJOURNALS.AJE.A114229 pmid:3946370

237. Fung TT, Chiuve SE, McCullough ML, et al. Adherence to a DASH-style diet and risk of coronary heart disease and stroke in women. *Arch Intern Med*. 2008;168(7):713–20. doi:10.1001/archinte.168.7.713 pmid:18413553

238. Gerber M. Qualitative methods to evaluate Mediterranean diet in adults. *Public Health Nutr*. 2006;9(1a):147–51. doi:10.1079/phn2005937

239. Neuhouser ML, Patterson RE, King IB, Horner NK, Lampe JW. Selected nutritional biomarkers predict diet quality. *Public Health Nutr*. 2003;6(7):703–9. doi:Doi 10.1079/Phn2003486 pmid:14552672

240. Hidalgo-Liberona N, Meroño T, Zamora-Ros R, et al. Adherence to the Mediterranean diet assessed by a novel dietary biomarker score and mortality in older adults: the InCHIANTI cohort study. *BMC Med*. 2021;19(1). doi:10.1186/s12916-021-02154-7 pmid:34814922

241. Ardisson Korat A V., Malik VS, Furtado JD, et al. Circulating Very-Long-Chain SFA Concentrations Are Inversely Associated with Incident Type 2 Diabetes in US Men and Women. *J Nutr*. 2020;150(2):340–9. doi:10.1093/JN/NXZ240 pmid:31618417

242. Rayman MP. Food-chain selenium and human health: emphasis on intake. *Br J Nutr*. 2008;100(2):254–68. doi:10.1017/S0007114508939830 pmid:18346308

243. Watanabe F. Vitamin B12 sources and bioavailability. *Exp Biol Med (Maywood)*. 2007;232(10):1266–74. doi:10.3181/0703-MR-67 pmid:17959839

244. Brouwer-Brolsma EM, Dhonukshe-Rutten RAM, van Wijngaarden JP, et al. Dietary Sources of Vitamin B-12 and Their Association with Vitamin B-12 Status Markers in Healthy Older Adults in the B-PROOF Study. *Nutrients*. 2015;7(9):7781–97. doi:10.3390/NU7095364 pmid:26389945

245. Shi L, Brunius C, Johansson I, et al. Plasma metabolites associated with healthy Nordic dietary indexes and risk of type 2 diabetes—a nested case-control study in a Swedish population. *Am J Clin Nutr*. 2018;108(3):564–75. doi:10.1093/ajcn/nqy145

246. Wang F, Baden MY, Guasch-Ferré M, et al. Plasma metabolite profiles related to plant-based diets and the risk of type 2 diabetes. *Diabetologia*. 2022;65(7):1119–32. doi:10.1007/S00125-022-05692-8 pmid:35391539

247. Ocké MC. Evaluation of methodologies for assessing the overall diet: Dietary quality scores and dietary pattern analysis. *Proc Nutr Soc*. 2013;72(2):191–9. doi:10.1017/S0029665113000013 pmid:23360896

248. Yun HY, Lampe JW, Tinker LF, et al. Serum Nitrogen and Carbon Stable Isotope Ratios Meet Biomarker Criteria for Fish and Animal Protein Intake in a Controlled Feeding Study of a Women's Health Initiative Cohort. *J Nutr*. 2018;1931–7. doi:10.1093/jn/nxy168 pmid:30239866

249. Yun HY, Tinker LF, Neuhouser ML, et al. The Carbon Isotope Ratios of Serum Amino Acids in Combination with Participant Characteristics can be Used to Estimate Added Sugar Intake in a Controlled Feeding Study of US Postmenopausal Women. *J Nutr*. 2020; doi:10.1093/jn/nxaa195

250. Zheng C, Gowda GAN, Raftery D, et al. Evaluation of potential metabolomic-based biomarkers of protein, carbohydrate and fat intakes using a controlled feeding study. *Eur J Nutr*. 2021;60(8):4207–18. doi:10.1007/s00394-021-02577-1 pmid:33991228

251. Appel LJ, Jacobs DR. Invited Commentary: Can Estimation of Sodium Intake Be Improved by Borrowing Information from Other Variables? *Am J Epidemiol*. 2017;186(9):1044–6. doi:10.1093/aje/kwx236

252. King IB, Lemaitre RN, Kestin M. Effect of a low-fat diet on fatty acid composition in red cells, plasma phospholipids, and cholesterol esters: Investigation of a biomarker of total fat intake. *Am J Clin Nutr*. 2006;83(2):227–36. doi:10.1093/ajcn/83.2.227 pmid:16469979

253. Day N, Oakes S, Luben R, et al. EPIC-Norfolk: Study design and characteristics of the cohort. *Br J Cancer*. 1999;80 Suppl 1:95–103. doi:10466767 pmid:10466767

254. Davis CR, Bryan J, Hodgson JM, et al. A randomised controlled intervention trial evaluating the efficacy of an Australianised Mediterranean diet compared to the habitual Australian diet on cognitive function, psychological wellbeing and cardiovascular health in healthy older adults (MedLey study): Protocol paper. *BMC Nutr*. 2015;1(1):35. doi:10.1186/s40795-015-0033-7

255. Barua AB, Kostic D, Olson JA. New simplified procedures for the extraction and simultaneous high-

performance liquid chromatographic analysis of retinol, tocopherols and carotenoids in human serum. *J Chromatogr B Biomed Sci Appl*. 1993;617(2):257–64. doi:10.1016/0378-4347(93)80496-Q pmid:8408391

256. Tu WC, Mühlhäusler BS, Yelland LN, Gibson RA. Correlations between blood and tissue omega-3 LCPUFA status following dietary ALA intervention in rats. *Prostaglandins Leukot Essent Fat Acids*. 2013;88(1):53–60. doi:10.1016/j.plefa.2012.04.005 pmid:22521090

257. Patel PS, Sharp SJ, Jansen E, et al. Fatty acids measured in plasma and erythrocyte-membrane phospholipids and derived by food-frequency questionnaire and the risk of new-onset type 2 diabetes: A pilot study in the European Prospective Investigation into Cancer and Nutrition (EPIC)-Norfolk cohort. *Am J Clin Nutr*. 2010;92(5):1214–22. doi:10.3945/ajcn.2010.29182 pmid:20861175

258. Ounpraseuth S, Lensing SY, Spencer HJ, Kodell RL. Estimating misclassification error: A closer look at cross-validation based methods. *BMC Res Notes*. 2012;5. doi:10.1186/1756-0500-5-656 pmid:23190936

259. Sauerbrei W, Meier-Hirmer C, Benner A, Royston P. Multivariable regression model building by using fractional polynomials: Description of SAS, STATA and R programs. *Comput Stat Data Anal*. 2006;50(12):3464–85. doi:10.1016/j.csda.2005.07.015

260. Vinknes KJ, Elshorbagy AK, Nurk E, et al. Plasma stearoyl-CoA desaturase indices: Association with lifestyle, diet, and body composition. *Obesity*. 2013;21(3). doi:10.1002/oby.20011 pmid:23404690

261. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143(1):29–36. doi:10.1148/RADIOLOGY.143.1.7063747 pmid:7063747

262. Wit E, van den Heuvel E, Romeijn JW. 'All models are wrong...': an introduction to model uncertainty. *Stat Neerl*. 2012;66(3):217–36. doi:10.1111/j.1467-9574.2012.00530.x

263. Li P, Stuart EA. Best (but oft-forgotten) practices: missing data methods in randomized controlled nutrition trials. *Am J Clin Nutr*. 2019;109(3):504–8. doi:10.1093/ajcn/nqy271

264. Newson RB. Attributable and unattributable risks and fractions and other scenario comparisons. *Stata J*. 2013;13(4):672–98. doi:10.1177/1536867x1301300402

265. Bellavia A, Discacciati A, Bottai M, Wolk A, Orsini N. Using Laplace Regression to Model and Predict Percentiles of Age at Death When Age Is the Primary Time Scale. *Am J Epidemiol*. 2015;182(3):271–7. doi:10.1093/aje/kwv033 pmid:26093508

266. Kalbfleisch JD, Lawless JF. Likelihood analysis of multi-state models for disease incidence and mortality. *Stat Med*. 1988;7(1–2):149–60. doi:10.1002/sim.4780070116 pmid:3353602

267. Connolly SE. Beyond Parameter Estimation: Analysis of the Case-Cohort Design in Cox Models (Doctoral thesis). University of Cambridge; 2020. doi:10.17863/CAM.51383

268. IntHout J, Ioannidis JPA, Rovers MM, Goeman JJ. Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open*. 2016;6(7):e010247. doi:10.1136/bmjopen-2015-010247 pmid:27406637

269. Davis CR, Bryan J, Hodgson JM, Wilson C, Murphy KJ. Older Australians can adhere to a traditional Mediterranean style diet over two weeks: A pilot dietary intervention study. *BMC Nutr*. 2015;1(1):1–10. doi:10.1186/s40795-015-0021-y

270. Jacobs S, Jäger S, Jansen E, et al. Associations of Erythrocyte Fatty Acids in the De Novo Lipogenesis Pathway with Proxies of Liver Fat Accumulation in the EPIC-Potsdam Study. *PLoS One*. 2015;10(5). doi:10.1371/JOURNAL.PONE.0127368 pmid:25984792

271. Bethancourt HJ, Schmidt KA, Cromer G, et al. Assessing the validity of plasma phospholipid fatty acids as biomarkers of dairy fat intake using data from a randomized controlled intervention trial. *Am J Clin Nutr*. 2022;115(6):1577–88. doi:10.1093/ajcn/nqac029 pmid:35134818

272. Romero-Gómez M, Zelber-Sagi S, Trenell M. Treatment of NAFLD with diet, physical activity and exercise. *J Hepatol*. 2017;67(4):829–46. doi:10.1016/J.JHEP.2017.05.016 pmid:28545937

273. Jenkins B, Aoun M, Feillet-Coudray C, et al. The dietary total-fat content affects the in vivo circulating C15:0 and C17:0 fatty acid levels independently. *Nutrients*. 2018;10(11):1646. doi:10.3390/nu10111646 pmid:30400275

274. Münger LH, Garcia-aloy M, Vázquez-fresno R, et al. Biomarker of food intake for assessing the consumption of dairy and egg products. *Genes Nutr*. 2018;1–18. doi:10.1186/s12263-018-0615-5

275. Alves SP, Marcelino C, Portugal P V., Bessa RJB. Short communication: The nature of heptadecenoic acid in ruminant fats. *J Dairy Sci*. 2006;89(1):170–3. doi:10.3168/JDS.S0022-0302(06)72081-1 pmid:16357280

276. Vicario IM, Malkova D, Lund EK, Johnson IT. Olive oil supplementation in healthy adults: effects in cell membrane fatty acid composition and platelet function. *Ann Nutr Metab*. 1998;42(3):160–9. doi:10.1159/000012729 pmid:9657460

277. Perona JS, Vögler O, Sánchez-Domínguez JM, et al. Consumption of virgin olive oil influences membrane lipid composition and regulates intracellular signaling in elderly adults with type 2 diabetes mellitus. *J Gerontol A Biol Sci Med Sci*. 2007;62(3):256–63. doi:10.1093/GERONA/62.3.256 pmid:17389722

278. Lemaitre RN, King IB. Very long-chain saturated fatty acids and diabetes and cardiovascular disease. *Curr Opin Lipidol*. 2022;33(1):76–82. doi:10.1097/MOL.0000000000000806 pmid:34907969

279. Lam C, Wong D, Cederbaum S, Lim B, Qu Y. Peanut consumption increases levels of plasma very long chain fatty acids in humans. *Mol Genet Metab*. 2012;107(3):620–2. doi:10.1016/J.YMGME.2012.07.015 pmid:22864056

280. Garg ML, Blake RJ, Wills RBH. Macadamia nut consumption lowers plasma total and LDL cholesterol levels in hypercholesterolemic men. *J Nutr*. 2003;133(4):1060–3. doi:10.1093/JN/133.4.1060 pmid:12672919

281. Kihara A. Very long-chain fatty acids: elongation, physiology and related disorders. *J Biochem*.

2012;152(5):387–95. doi:10.1093/JB/MVS105 pmid:22984005

282. Tsutsumi R, Yamasaki Y, Takeo J, et al. Long-chain monounsaturated fatty acids improve endothelial function with altering microbial flora. *Transl Res*. 2021;237:16–30. doi:10.1016/J.TRSL.2021.03.016 pmid:33775867

283. Yang ZH, Emma-Okon B, Remaley AT. Dietary marine-derived long-chain monounsaturated fatty acids and cardiovascular disease risk: a mini review. *Lipids Health Dis*. 2016;15(1):1–9. doi:10.1186/S12944-016-0366-5 pmid:27876051

284. Hyde PN, Sapper TN, Crabtree CD, et al. Dietary carbohydrate restriction improves metabolic syndrome independent of weight loss. *JCI insight*. 2019;4(12). doi:10.1172/JCI.INSIGHT.128308 pmid:31217353

285. Zheng Y, Yu B, Alexander D, Steffen LM, Boerwinkle E. Human metabolome associates with dietary intake habits among African Americans in the atherosclerosis risk in communities study. *Am J Epidemiol*. 2014;179(12):1424–33. doi:10.1093/aje/kwu073 pmid:24801555

286. Nielsen FH, Johnson LAK. Data from Controlled Metabolic Ward Studies Provide Guidance for the Determination of Status Indicators and Dietary Requirements for Magnesium. *Biol Trace Elem Res*. 2017;177(1):43–52. doi:10.1007/S12011-016-0873-2 pmid:27778151

287. Nielsen FH. Guidance for the determination of status indicators and dietary requirements for magnesium. *Magnes Res*. 2016;29(4):154–60. doi:10.1684/mrh.2016.0416 pmid:28132953

288. Moser AB, Steinberg SJ, Watkins PA, et al. Human and great ape red blood cells differ in plasmalogen levels and composition. *Lipids Health Dis*. 2011;10. doi:10.1186/1476-511X-10-101 pmid:21679470

289. Sanders TAB, Ellis FR, Dickerson JWT. Studies of vegans: the fatty acid composition of plasma choline phosphoglycerides, erythrocytes, adipose tissue, and breast milk, and some indicators of susceptibility to ischemic heart disease in vegans and omnivore controls. *Am J Clin Nutr*. 1978;31(5):805–13. doi:10.1093/AJCN/31.5.805 pmid:645628

290. Stender S, Astrup A, Dyerberg J. Ruminant and industrially produced trans fatty acids: Health aspects [Internet]. Vol. 52, Food and Nutrition Research. Food Nutr Res; 2008. doi:10.3402/fnr.v52i0.1651 pmid:19109659

291. Rotondo A, La Torre GL, Dugo G, et al. Oleic acid is not the only relevant mono-unsaturated fatty ester in olive oil. *Foods*. 2020;9(4). doi:10.3390/foods9040384 pmid:32224949

292. Potischman N. Biologic and methodologic issues for nutritional biomarkers. *J Nutr*. 2003;133(Suppl 3):875S-880S. doi:10.1093/jn/133.3.875s pmid:12612173

293. Bell ML, Kenward MG, Fairclough DL, Horton NJ. Differential dropout and bias in randomised controlled trials: when it matters and when it may not. *BMJ*. 2013;346. doi:10.1136/bmj.e8668 pmid:23338004

294. Verbeek JH, Whaley P, Morgan RL, et al. An approach to quantifying the potential importance of residual confounding in systematic reviews of observational studies: A GRADE concept paper. *Environ Int*. 2021;157. doi:10.1016/j.envint.2021.106868 pmid:34530289

295. Obeid CA, Gubbels JS, Jaalouk D, Kremers SPJ, Oenema A. Adherence to the Mediterranean diet among adults in Mediterranean countries: a systematic literature review. *Eur J Nutr*. 2022;61(7):3327–44. doi:10.1007/S00394-022-02885-0 pmid:35451614

296. Veronese N, Notarnicola M, Cisternino AM, et al. Trends in adherence to the Mediterranean diet in South Italy: A cross sectional study. *Nutr Metab Cardiovasc Dis*. 2020;30(3):410–7. doi:10.1016/J.NUMECD.2019.11.003 pmid:31822430

297. Li J, Guasch-Ferré M, Chung W, et al. The Mediterranean diet, plasma metabolome, and cardiovascular disease risk. *Eur Heart J*. 2020;1–14. doi:10.1093/eurheartj/ehaa209

298. Roncero-Ramos I, Alcala-Diaz JF, Rangel-Zuñiga OA, et al. Prediabetes diagnosis criteria, type 2 diabetes risk and dietary modulation: The CORDIOPREV study. *Clin Nutr*. 2020;39(2):492–500. doi:10.1016/J.CLNU.2019.02.027 pmid:30852029

299. de Koning L, Chiuve SE, Fung TT, et al. Diet-quality scores and the risk of type 2 diabetes in men. *Diabetes Care*. 2011;34(5):1150–6. doi:10.2337/dc10-2352 pmid:21464460

300. Trepanowski JF, Ioannidis JPA. Perspective: Limiting dependence on nonrandomized studies and improving randomized trials in human nutrition research: Why and how. *Adv Nutr*. 2018;9(4):367–77. doi:10.1093/ADVANCES/NMY014 pmid:30032218

301. Dhurandhar N V., Schoeller D, Brown AW, et al. Energy balance measurement: When something is not better than nothing. *Int J Obes*. 2015;39(7):1109–13. doi:10.1038/ijo.2014.199 pmid:25394308

302. Kipnis V, Subar AF, Midthune D, et al. Structure of dietary measurement error: Results of the OPEN biomarker study. *Am J Epidemiol*. 2003;158(1):14–21. doi:10.1093/aje/kwg091 pmid:12835281

303. Shi L, Brunius C, Johansson I, et al. Plasma metabolite biomarkers of boiled and filtered coffee intake and their association with type 2 diabetes risk. *J Intern Med*. 2020;287(4):405–21. doi:10.1111/JOIM.13009 pmid:31814205

304. Tong TYN, Wareham NJ, Khaw KT, Imamura F, Forouhi NG. Prospective association of the Mediterranean diet with cardiovascular disease incidence and mortality and its population impact in a non-Mediterranean population: The EPIC-Norfolk study. *BMC Med*. 2016;14(1):1–11. doi:10.1186/s12916-016-0677-4 pmid:27679997

305. Bingham SA, Welch AA, McTaggart A, et al. Nutritional methods in the European Prospective Investigation of Cancer in Norfolk. *Public Health Nutr*. 2001;4(03):847. doi:10.1079/PHN2000102 pmid:11415493

306. Mulligan AA, Luben RN, Bhaniani A, et al. A new tool for converting food frequency questionnaire data into nutrient and food group values: FETA research methods and availability. *BMJ Open*.

2014;4(3). doi:10.1136/bmjopen-2013-004503 pmid:24674997

307. Zheng JS, Imamura F, Sharp SJ, et al. Changes in plasma phospholipid fatty acid profiles over 13 years and correlates of change: European Prospective Investigation into Cancer and Nutrition-Norfolk Study. *Am J Clin Nutr*. 2019;109(6):1527–34. doi:10.1093/ajcn/nqz030 pmid:30997506

308. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann Stat*. 2001;29(4):1165–88. doi:10.1214/aos/1013699998

309. Maruvada P, Lampe JW, Wishart DS, et al. Perspective: Dietary Biomarkers of Intake and Exposure - Exploration with Omics Approaches. *Adv Nutr*. 2020;11(2):200–15. doi:10.1093/advances/nmz075 pmid:31386148

310. Prentice RL, Willet WC, Greenwald P, et al. Nutrition and physical activity and chronic disease prevention: research strategies and recommendations. *J Natl Cancer Inst*. 2004;96(17):1276–87. doi:10.1093/JNCI/DJH240 pmid:15339966

311. Sébédio JL. Metabolomics, Nutrition, and Potential Biomarkers of Food Quality, Intake, and Health Status. In: Advances in Food and Nutrition Research. Academic Press Inc.; 2017. p. 83–116. doi:10.1016/bs.afnr.2017.01.001 pmid:28427537

312. Jannasch F, Kröger J, Agnoli C, et al. Generalizability of a Diabetes-Associated Country-Specific Exploratory Dietary Pattern Is Feasible Across European Populations. 2019;1–9.

313. Freedman LS, Shaw PA. On the formation and use of calibration equations in nutritional epidemiology–Discussion of the Paper by R. L. Prentice and Y. Huang. *Stat Theory Relat Fields*. 2018;2(1):11–3. doi:10.1080/24754269.2018.1493631

314. Tinker LF, Huang Y, Johnson KC, et al. Estimating 24-Hour Urinary Excretion of Sodium and Potassium Is More Reliable from 24-Hour Urine Than Spot Urine Sample in a Feeding Study of US Older Postmenopausal Women. *Curr Dev Nutr*. 2021;5(11). doi:10.1093/cdn/nzab125 pmid:34761160

315. Grace PB, Taylor JI, Low YL, et al. Phytoestrogen concentrations in serum and spot urine as biomarkers for dietary phytoestrogen intake and their relation to breast cancer risk in European Prospective Investigation of Cancer and Nutrition-Norfolk. *Cancer Epidemiol Biomarkers Prev*. 2004;13(5):698–708. doi:10.1158/1055-9965.698.13.5 pmid:15159299

316. Sri Harsha PSC, Wahab RA, Garcia-Aloy M, et al. Biomarkers of legume intake in human intervention and observational studies: a systematic review. *Genes Nutr*. 2018;13(1). doi:10.1186/s12263-018-0614-6 pmid:30214640

317. O'Brien DM, Sagi-Kiss V, Palma-Duran SA, et al. An evaluation of the serum carbon isotope ratio as a candidate predictive biomarker of the dietary animal protein ratio (animal protein/total protein) in a 15-day controlled feeding study of US adults. *Am J Clin Nutr*. 2022;115(4):1134–43. doi:10.1093/ajcn/nqac004 pmid:35030258

318. Patel PS, Cooper AJM, O'Connell TC, et al. Serum carbon and nitrogen stable isotopes as potential

biomarkers of dietary intake and their relation with incident type 2 diabetes: The EPIC-Norfolk study. *Am J Clin Nutr*. 2014;100(2):708–18. doi:10.3945/ajcn.113.068577 pmid:24990425

319. Runswick SA, McTaggart A, Mulligan AA, et al. Use of biological markers to validate self-reported dietary intake in a random sample of the European Prospective Investigation into Cancer United Kingdom Norfolk cohort. *Am J Clin Nutr*. 2018;74(2):188–96. doi:10.1093/ajcn/74.2.188

320. Slimani N, Bingham S, Runswick S, et al. Group level validation of protein intakes estimated by 24-hour diet recall and dietary questionnaires against 24-hour urinary nitrogen in the European Prospective Investigation into Cancer and Nutrition (EPIC) calibration study. *Cancer Epidemiol Biomarkers Prev*. 2003;12(8):784–95. pmid:12917211

321. Khaw K-T, Friesen MD, Riboli E, Luben R, Wareham N. Plasma Phospholipid Fatty Acid Concentration and Incident Coronary Heart Disease in Men and Women: The EPIC-Norfolk Prospective Study. Katan MB, editor. *PLoS Med*. 2012;9(7):e1001255. doi:10.1371/journal.pmed.1001255

322. Ward H, Chapelais G, Kuhnle GGC, et al. Breast cancer risk in relation to urinary and serum biomarkers of phytoestrogen exposure in the European Prospective into Cancer-Norfolk cohort study. *Breast Cancer Res*. 2008;10(2). doi:10.1186/BCR1995 pmid:18419813

323. Ward H, Chapelais G, Kuhnle GGC, et al. Lack of prospective associations between plasma and urinary phytoestrogens and risk of prostate or colorectal cancer in the European Prospective into Cancer-Norfolk study. *Cancer Epidemiol Biomarkers Prev*. 2008;17(10):2891–4. doi:10.1158/1055-9965.EPI-08-0335 pmid:18843034

324. Mulligan AA, Hayhoe RPG, Luben RN, Welch AA. Positive Associations of Dietary Intake and Plasma Concentrations of Vitamin E with Skeletal Muscle Mass, Heel Bone Ultrasound Attenuation and Fracture Risk in the EPIC-Norfolk Cohort. *Antioxidants (Basel, Switzerland)*. 2021;10(2):1–30. doi:10.3390/ANTIOX10020159 pmid:33499166

325. Kuhnle GGC, Tasevska N, Lentjes MAH, et al. Association between sucrose intake and risk of overweight and obesity in a prospective sub-cohort of the European Prospective Investigation into Cancer in Norfolk (EPIC-Norfolk). *Public Health Nutr*. 2015;18(15). doi:10.1017/S1368980015000300 pmid:25702697

326. Tandel KR. Sugar substitutes: Health controversy over perceived benefits. *J Pharmacol Pharmacother*. 2011;2(4):236–43. doi:10.4103/0976-500X.85936 pmid:22025850

327. Grace PB, Taylor JI, Botting NP, et al. Quantification of isoflavones and lignans in serum using isotope dilution liquid chromatography/tandem mass spectrometry. *Rapid Commun Mass Spectrom*. 2003;17(12):1350–7. doi:10.1002/rcm.1059 pmid:12811759

328. Pietzner M, Stewart ID, Raffler J, et al. Plasma metabolites to profile pathways in noncommunicable disease multimorbidity. *Nat Med*. 2021;27(3):471–9. doi:10.1038/s41591-021-01266-0 pmid:33707775

329. Sinha S, Myint PK, Luben RN, Khaw KT. Accuracy of death certification and hospital record linkage

for identification of incident stroke. *BMC Med Res Methodol*. 2008;8. doi:10.1186/1471-2288-8-74 pmid:19000303

330. Chamnan P, Simmons RK, Jackson R, et al. Non-diabetic hyperglycaemia and cardiovascular risk: Moving beyond categorisation to individual interpretation of absolute risk. *Diabetologia*. 2011;54(2):291–9. doi:10.1007/s00125-010-1914-6 pmid:20859613

331. Luben R, Hayat S, Khawaja A, et al. Residential area deprivation and risk of subsequent hospital admission in a British population: The EPIC-Norfolk cohort. *BMJ Open*. 2019;9(12). doi:10.1136/bmjopen-2019-031251 pmid:31848162

332. Buis ML. Stata tip 135: Leaps and bounds [Internet]. Vol. 20, Stata Journal. SAGE Publications Inc.; 2020. p. 244–9. doi:10.1177/1536867X20909707

333. Ottaviani JI, Britten A, Lucarelli D, et al. Biomarker-estimated flavan-3-ol intake is associated with lower blood pressure in cross-sectional analysis in EPIC Norfolk. *Sci Rep*. 2020;10(1). doi:10.1038/S41598-020-74863-7 pmid:33087825

334. Tanaka T, Okamura T, Miura K, et al. A simple method to estimate populational 24-h urinary sodium and potassium excretion using a casual urine specimen. *J Hum Hypertens*. 2002;16(2):97–103. doi:10.1038/sj.jhh.1001307 pmid:11850766

335. Kawasaki T, Itoh K, Uezono K, Sasaki H. A SIMPLE METHOD FOR ESTIMATING 24 H URINARY SODIUM AND POTASSIUM EXCRETION FROM SECOND MORNING VOIDING URINE SPECIMEN IN ADULTS. *Clin Exp Pharmacol Physiol*. 1993;20(1):7–14. doi:10.1111/j.1440-1681.1993.tb01496.x pmid:8432042

336. Mage DT, Allen RH, Kodali A. Creatinine corrections for estimating children's and adult's pesticide intake doses in equilibrium with urinary pesticide and creatinine concentrations. *J Expo Sci Environ Epidemiol*. 2008;18(4):360–8. doi:10.1038/sj.jes.7500614 pmid:17878925

337. Huber DR, Blount BC, Mage DT, et al. Estimating perchlorate exposure from food and tap water based on US biomonitoring and occurrence data. *J Expo Sci Environ Epidemiol*. 2011;21(4):395–407. doi:10.1038/JES.2010.31 pmid:20571527

338. Toft U, Cerqueira C, Andreasen AH, et al. Estimating salt intake in a Caucasian population: Can spot urine substitute 24-hour urine samples? *Eur J Prev Cardiol*. 2014;21(10):1300–7. doi:10.1177/2047487313485517 pmid:23559538

339. Brown IJ, Dyer AR, Chan Q, et al. Estimating 24-hour urinary sodium excretion from casual urinary sodium concentrations in Western populations: the INTERSALT study. *Am J Epidemiol*. 2013;177(11):1180–92. doi:10.1093/AJE/KWT066 pmid:23673246

340. Elliott P, Brown IJ, Dyer AR, et al. Elliott et al. respond to "quantifying urine sodium excretion." *Am J Epidemiol*. 2013;177(11):1196–8. doi:10.1093/aje/kwt065

341. Uechi K, Asakura K, Ri Y, Masayasu S, Sasaki S. Advantage of multiple spot urine collections for

estimating daily sodium excretion: Comparison with two 24-h urine collections as reference. *J Hypertens*. 2016;34(2):204–14. doi:10.1097/HJH.0000000000000778 pmid:26599218

342. Whitton C, Gay GMW, Lim RBT, et al. Evaluation of equations for predicting 24-hour urinary sodium excretion from casual urine samples in asian adults. *J Nutr*. 2016;146(8):1609–15. doi:10.3945/jn.116.232108 pmid:27358415

343. Faquih T, van Smeden M, Luo J, et al. A workflow for missing values imputation of untargeted metabolomics data. *Metabolites*. 2020;10(12):1–23. doi:10.3390/metabo10120486 pmid:33256233

344. Do KT, Wahl S, Raffler J, et al. Characterization of missing values in untargeted MS-based metabolomics data and evaluation of missing data handling strategies. *Metabolomics*. 2018;14(10). doi:10.1007/s11306-018-1420-2 pmid:30830398

345. Lee AB, Nadler B, Wasserman L. Treelets-an adaptive multi-scale basis for sparse unordered data. *Ann Appl Stat*. 2008;2(2):435–71. doi:10.1214/07-AOAS137

346. Kim RS, Kaplan RC. Analysis of secondary outcomes in nested case-control study designs. *Stat Med*. 2014;33(24):4215–26. doi:10.1002/sim.6231 pmid:24919979

347. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. *Am J Epidemiol*. 2009;170(2):244–56. doi:10.1093/AJE/KWP107 pmid:19494242

348. Hoffmann K, Pischon T, Schulz M, et al. A statistical test for the equality of differently adjusted incidence rate ratios. *Am J Epidemiol*. 2008;167(5):517–22. doi:10.1093/AJE/KWM357 pmid:18230678

349. Gauthier J, Wu Q V., Gooley TA. Cubic splines to model relationships between continuous variables and outcomes: a guide for clinicians [Internet]. Vol. 55, Bone Marrow Transplantation. Bone Marrow Transplant; 2020. p. 675–80. doi:10.1038/s41409-019-0679-x pmid:31576022

350. Groenwold RHH, Klungel OH, Altman DG, et al. Adjustment for continuous confounders: An example of how to prevent residual confounding. *C Can Med Assoc J*. 2013;185(5):401–6. doi:10.1503/cmaj.120592 pmid:23401401

351. Langholz B, Thomas DC. Nested case-control and case-cohort methods of sampling from a cohort: A critical comparison. *Am J Epidemiol*. 1990;131(1):169–76. doi:10.1093/oxfordjournals.aje.a115471 pmid:2403467

352. Delcoigne B, Støer NC, Reilly M. Valid and efficient subgroup analyses using nested case-control data. *Int J Epidemiol*. 2018;47(3):841–9. doi:10.1093/ije/dyx282 pmid:29390147

353. Tong TYN, Koulman A, Griffin JL, et al. A Combination of Metabolites Predicts Adherence to the Mediterranean Diet Pattern and Its Associations with Insulin Sensitivity and Lipid Homeostasis in the General Population: The Fenland Study, United Kingdom. *J Nutr*. 2019; doi:10.1093/jn/nxz263

354. Smith E, Ericson U, Hellstrand S, et al. A healthy dietary metabolic signature is associated with a lower risk for type 2 diabetes and coronary artery disease. *BMC Med*. 2022;20(1). doi:10.1186/S12916-022-

02326-Z pmid:35443726

355. Abreu TC, Hulshof PJM, Boshuizen HC, et al. Validity coefficient of repeated measurements of urinary marker of sugar intake is comparable to urinary nitrogen as marker of protein intake in free-living subjects. *Cancer Epidemiol Biomarkers Prev*. 2021;30(1):193–202. doi:10.1158/1055-9965.EPI-20-0271 pmid:32998945

356. He J, Obst K. Estimating dietary sodium intake using spot urine samples: correlation and bias. *J Hypertens*. 2017;35(3):466–7. doi:10.1097/HJH.0000000000001230 pmid:28121836

357. Van Der Velpen V, Hollman PC, Van Nielen M, et al. Large inter-individual variation in isoflavone plasma concentration limits use of isoflavone intake data for risk assessment. *Eur J Clin Nutr*. 2014;68(10):1141–7. doi:10.1038/ejcn.2014.108 pmid:24939433

358. Hausner H, Johnsen NF, Hallund J, Tetens I. A Single Measurement Is Inadequate to Estimate Enterolactone Levels in Danish Postmenopausal Women Due to Large Intraindividual Variation. *J Nutr*. 2004;134(5):1197–200. doi:10.1093/jn/134.5.1197 pmid:15113970

359. Lampe JW. Isoflavonoid and lignan phytoestrogens as dietary biomarkers. *J Nutr*. 2003;133(3 SUPPL.). doi:10.1093/jn/133.3.956s pmid:12612182

360. Sun Q, Bertrand KA, Franke AA, et al. Reproducibility of urinary biomarkers in multiple 24-h urine samples. *Am J Clin Nutr*. 2017;105(1):159–68. doi:10.3945/ajcn.116.139758 pmid:28049663

361. Yuan C, Spiegelman D, Rimm EB, et al. Relative Validity of Nutrient Intakes Assessed by Questionnaire, 24-Hour Recalls, and Diet Records as Compared with Urinary Recovery and Plasma Concentration Biomarkers: Findings for Women. *Am J Epidemiol*. 2018;187(5):1051–63. doi:10.1093/aje/kwx328 pmid:28490509

362. Bar N, Korem T, Weissbrod O, et al. A reference map of potential determinants for the human serum metabolome. *Nature*. 2020;588(7836):135–40. doi:10.1038/s41586-020-2896-2 pmid:33177712

363. Stratakis N, Siskos AP, Papadopoulou E, et al. Urinary metabolic biomarkers of diet quality in European children are associated with metabolic health. *Elife*. 2022;11. doi:10.7554/ELIFE.71332 pmid:35076016

364. Jin Q, Black A, Kales SN, et al. Metabolomics and Microbiomes as Potential Tools to Evaluate the Effects of the Mediterranean Diet. *Nutrients*. 2019;11(1):207. doi:10.3390/nu11010207

365. Gibbons H, Michielsen CJR, Rundle M, et al. Demonstration of the utility of biomarkers for dietary intake assessment; proline betaine as an example. *Mol Nutr Food Res*. 2017;61(10). doi:10.1002/mnfr.201700037 pmid:28556565

366. Atkinson W, Downer P, Lever M, Chambers ST, George PM. Effects of orange juice and proline betaine on glycine betaine and homocysteine in healthy male subjects. *Eur J Nutr*. 2007;46(8):446–52. doi:10.1007/S00394-007-0684-5 pmid:18060588

367. Halliwell B, Cheah IK, Tang RMY. Ergothioneine - a diet-derived antioxidant with therapeutic

potential. *FEBS Lett*. 2018;592(20):3357–66. doi:10.1002/1873-3468.13123 pmid:29851075

368. Smith E, Ottosson F, Hellstrand S, et al. Ergothioneine is associated with reduced mortality and decreased risk of cardiovascular disease. *Heart*. 2019; doi:10.1136/heartjnl-2019-315485 pmid:31672783

369. Langenau J, Oluwagbemigun K, Brachem C, et al. Blood Metabolomic Profiling Confirms and Identifies Biomarkers of Food Intake. *Metabolites*. 2020;10(11):1–17. doi:10.3390/METABO10110468 pmid:33212857

370. He X, Slupsky CM. Metabolic fingerprint of dimethyl sulfone (DMSO2) in microbial-mammalian co-metabolism [Internet]. Vol. 13, Journal of Proteome Research. J Proteome Res; 2014. p. 5281–92. doi:10.1021/pr500629t pmid:25245235

371. Lahlou S, Interaminense LFL, Magalhães PJC, Leal-Cardoso JH, Duarte GP. Cardiovascular Effects of Eugenol, A Phenolic Compound Present in Many Plant Essential Oils, in Normotensive Rats. *J Cardiovasc Pharmacol*. 2004;43(2):250–7. doi:10.1097/00005344-200402000-00013 pmid:14716213

372. Ulaszewska M, Vázquez-Manjarrez N, Garcia-Aloy M, et al. Food intake biomarkers for apple, pear, and stone fruit. *Genes Nutr*. 2018;13(1). doi:10.1186/S12263-018-0620-8 pmid:30519365

373. Willis ND, Lloyd AJ, Xie L, et al. Design and Characterisation of a Randomized Food Intervention That Mimics Exposure to a Typical UK Diet to Provide Urine Samples for Identification and Validation of Metabolite Biomarkers of Food Intake. *Front Nutr*. 2020;7. doi:10.3389/fnut.2020.561010 pmid:33195362

374. Kärkkäinen O, Lankinen MA, Vitale M, et al. Diets rich in whole grains increase betainized compounds associated with glucose metabolism. *Am J Clin Nutr*. 2018;108(5):971–9. doi:10.1093/ajcn/nqy169 pmid:30256894

375. Shively CA, Register TC, Appt SE, et al. Consumption of Mediterranean versus Western Diet Leads to Distinct Mammary Gland Microbiome Populations. *Cell Rep*. 2018;25(1):47-56.e3. doi:10.1016/J.CELREP.2018.08.078 pmid:30282037

376. Wang Y, Hodge RA, Stevens VL, Hartman TJ, McCullough ML. Identification and Reproducibility of Urinary Metabolomic Biomarkers of Habitual Food Intake in a Cross-Sectional Analysis of the Cancer Prevention Study-3 Diet Assessment Sub-Study. *Metabolites*. 2021;11(4). doi:10.3390/METABO11040248 pmid:33920694

377. Zarei I, Baxter BA, Oppel RC, et al. Plasma and Urine Metabolite Profiles Impacted by Increased Dietary Navy Bean Intake in Colorectal Cancer Survivors: A Randomized-Controlled Trial. *Cancer Prev Res (Phila)*. 2021;14(4):497–508. doi:10.1158/1940-6207.CAPR-20-0270 pmid:33361317

378. Vázquez-Fresno R, Rosana ARR, Sajed T, et al. Herbs and Spices- Biomarkers of Intake Based on Human Intervention Studies - A Systematic Review [Internet]. Vol. 14, Genes and Nutrition. Genes Nutr; 2019. doi:10.1186/s12263-019-0636-8 pmid:31143299

379. Kim H, Hu EA, E Wong K, et al. Serum Metabolites Associated with Healthy Diets in African Americans and European Americans. *J Nutr*. 2021;151(1):40–9. doi:10.1093/jn/nxaa338 pmid:33244610

380. Rebholz CM, Lichtenstein AH, Zheng Z, Appe LJ, Coresh J. Serum untargeted metabolomic profile of the Dietary Approaches to Stop Hypertension (DASH) dietary pattern. *Am J Clin Nutr*. 2018;108(2):243–55. doi:10.1093/ajcn/nqy099

381. Derkach A, Sampson J, Joseph J, Playdon MC, Stolzenberg-Solomon RZ. Effects of dietary sodium on metabolites: The Dietary Approaches to Stop Hypertension (DASH)–Sodium Feeding Study. *Am J Clin Nutr*. 2017;106(4):1131–41. doi:10.3945/ajcn.116.150136 pmid:28855223

382. Koleva II, van Beek TA, Soffers AEMF, Dusemund B, Rietjens IMCM. Alkaloids in the human food chain - natural occurrence and possible adverse effects [Internet]. Vol. 56, Molecular Nutrition and Food Research. Mol Nutr Food Res; 2012. p. 30–52. doi:10.1002/mnfr.201100165 pmid:21823220

383. Baron PA, Love DC, Nachman KE. Pharmaceuticals and personal care products in chicken meat and other food animal products: A market-basket pilot study. *Sci Total Environ*. 2014;490:296–300. doi:10.1016/j.scitotenv.2014.04.112 pmid:24858227

384. Rana MS, Lee SY, Kang HJ, Hur SJ. Reducing Veterinary Drug Residues in Animal Products: A Review. *Food Sci Anim Resour*. 2019;39(5):687–703. doi:10.5851/KOSFA.2019.E65 pmid:31728441

385. Li C, Imamura F, Wedekind R, et al. Development and validation of a metabolite score for red meat intake: an observational cohort study and randomized controlled dietary intervention. *Am J Clin Nutr*. 2022;116(2):511–22. doi:10.1093/ajcn/nqac094 pmid:35754192

386. Neuhouser ML, Pettinger M, Tinker LF, et al. Associations of Biomarker-Calibrated HEI-2010 Scores With Chronic Disease Risk and Their Dependency On Energy Intake and Body Mass Index in Postmenopausal Women. *J Nutr*. 2022; doi:10.1093/JN/NXAC199 pmid:36040344

387. Romaguera D, Norat T, Vergnaud AC, et al. Mediterranean dietary patterns and prospective weight change in participants of the EPIC-PANACEA project. *Am J Clin Nutr*. 2010;92(4):912–21. doi:10.3945/ajcn.2010.29482 pmid:20810975

388. Ford C, Chang S, Vitolins MZ, et al. Evaluation of diet pattern and weight gain in postmenopausal women enrolled in the Women's Health Initiative Observational Study. *Br J Nutr*. 2017;117(8):1189–97. doi:10.1017/S0007114517000952 pmid:28509665

389. Freisling H, Viallon V, Lennon H, et al. Lifestyle factors and risk of multimorbidity of cancer and cardiometabolic diseases: a multinational cohort study. *BMC Med*. 2020;18(1). doi:10.1186/S12916-019-1474-7 pmid:31918762

390. Galbete C, Kröger J, Jannasch F, et al. Nordic diet, Mediterranean diet, and the risk of chronic diseases: the EPIC-Potsdam study. *BMC Med*. 2018;16(1). doi:10.1186/S12916-018-1082-Y pmid:29945632

391. Rosato V, Temple NJ, La Vecchia C, et al. Mediterranean diet and cardiovascular disease: a systematic

review and meta-analysis of observational studies. *Eur J Nutr*. 2019;58(1):173–91. doi:10.1007/s00394-017-1582-0 pmid:29177567

392. Morze J, Danielewicz A, Przybyłowicz K, et al. An updated systematic review and meta-analysis on adherence to Mediterranean diet and risk of cancer. *Eur J Nutr*. 2021;60(3):1561–86. doi:10.1007/s00394-020-02346-6 pmid:32770356

393. Sarsangi P, Salehi-Abargouei A, Ebrahimpour-Koujan S, Esmaillzadeh A. Association between Adherence to the Mediterranean Diet and Risk of Type 2 Diabetes: An Updated Systematic Review and Dose-Response Meta-Analysis of Prospective Cohort Studies. *Adv Nutr*. 2022;13(5). doi:10.1093/ADVANCES/NMAC046 pmid:35472102

394. Ferreira-González I, Permanyer-Miralda G, Busse JW, et al. Methodologic discussions for using and interpreting composite endpoints are limited, but still identify major concerns [Internet]. Vol. 60, Journal of Clinical Epidemiology. J Clin Epidemiol; 2007. p. 651–7. doi:10.1016/j.jclinepi.2006.10.020 pmid:17573977

395. Solch RJ, Aigbogun JO, Voyiadjis AG, et al. Mediterranean diet adherence, gut microbiota, and Alzheimer's or Parkinson's disease risk: A systematic review [Internet]. Vol. 434, Journal of the Neurological Sciences. J Neurol Sci; 2022. doi:10.1016/j.jns.2022.120166 pmid:35144237

396. Dominguez LJ, Di Bella G, Veronese N, Barbagallo M. Impact of Mediterranean Diet on Chronic Non-Communicable Diseases and Longevity. *Nutrients*. 2021;13(6). doi:10.3390/NU13062028 pmid:34204683

397. Papier K, Fensom GK, Knuppel A, et al. Meat consumption and risk of 25 common conditions: outcome-wide analyses in 475,000 men and women in the UK Biobank study. *BMC Med*. 2021;19(1). doi:10.1186/s12916-021-01922-9 pmid:33648505

398. Vanderweele TJ. Outcome-wide epidemiology. *Epidemiology*. 2017;28(3):399–402. doi:10.1097/EDE.0000000000000641 pmid:28166102

399. VanderWeele TJ, Mathur MB. SOME DESIRABLE PROPERTIES OF THE BONFERRONI CORRECTION: IS THE BONFERRONI CORRECTION REALLY SO BAD? [Internet]. Vol. 188, American Journal of Epidemiology. Am J Epidemiol; 2019. p. 617–8. doi:10.1093/aje/kwy250 pmid:30452538

400. Toledo E, Salas-Salvado J, Donat-Vargas C, et al. Mediterranean Diet and Invasive Breast Cancer Risk Among Women at High Cardiovascular Risk in the PREDIMED Trial: A Randomized Clinical Trial. *JAMA Intern Med*. 2015;175(11):1752–60. doi:10.1001/JAMAINTERNMED.2015.4838 pmid:26365989

401. Murphy N, Moreno V, Hughes DJ, et al. Lifestyle and dietary environmental factors in colorectal cancer susceptibility [Internet]. Vol. 69, Molecular Aspects of Medicine. Mol Aspects Med; 2019. p. 2–9. doi:10.1016/j.mam.2019.06.005 pmid:31233770

402. Fischer A, Johansson I, Blomberg A, Sundström B. Adherence to a Mediterranean-like Diet as a Protective Factor Against COPD: A Nested Case-Control Study. *COPD J Chronic Obstr Pulm Dis*. 2019;16(3–4):272–7. doi:10.1080/15412555.2019.1634039 pmid:31405301

403. Van Iersel LEJ, Beijers RJHCG, Gosker HR, Schols AMWJ. Nutrition as a modifiable factor in the onset and progression of pulmonary function impairment in COPD: a systematic review. *Nutr Rev*. 2022;80(6):1434–44. doi:10.1093/NUTRIT/NUAB077 pmid:34537848

404. Hlaing-Hlaing H, Dolja-Gore X, Tavener M, et al. Diet Quality and Incident Non-Communicable Disease in the 1946-1951 Cohort of the Australian Longitudinal Study on Women's Health. *Int J Environ Res Public Health*. 2021;18(21). doi:10.3390/IJERPH182111375 pmid:34769892

405. Alwarith J, Kahleova H, Crosby L, et al. The role of nutrition in asthma prevention and treatment. *Nutr Rev*. 2020;78(11):928–38. doi:10.1093/NUTRIT/NUAA005 pmid:32167552

406. Parvizian MK, Li J, Dhaliwal M, Satia I, Kurmi OP. Relationship between dietary patterns and asthma: A systematic review and meta-analysis. *Can J Respir Crit Care, Sleep Med*. 2021;5(6):362–73. doi:10.1080/24745332.2020.1716647

407. Zhang Y, Lin J, Fu W, et al. Mediterranean diet during pregnancy and childhood for asthma in children: A systematic review and meta-analysis of observational studies [Internet]. Vol. 54, Pediatric Pulmonology. Pediatr Pulmonol; 2019. p. 949–61. doi:10.1002/ppul.24338 pmid:30997754

408. Liyanage T, Ninomiya T, Wang A, et al. Effects of the Mediterranean diet on cardiovascular outcomes-a systematic review and meta-analysis [Internet]. Vol. 11, PLoS ONE. PLoS One; 2016. doi:10.1371/journal.pone.0159252 pmid:27509006

409. Delgado-Lista J, Alcala-Diaz JF, Torres-Peña JD, et al. Long-term secondary prevention of cardiovascular disease with a Mediterranean diet and a low-fat diet (CORDIOPREV): a randomised controlled trial. *Lancet*. 2022;399(10338):1876–85. doi:10.1016/S0140-6736(22)00122-2 pmid:35525255

410. Shan Z, Li Y, Baden MY, et al. Association Between Healthy Eating Patterns and Risk of Cardiovascular Disease. *JAMA Intern Med*. 2020;180(8):1090–100. doi:10.1001/JAMAINTERNMED.2020.2176 pmid:32539102

411. Martínez-González MÁ, Toledo E, Arós F, et al. Extravirgin olive oil consumption reduces risk of atrial fibrillation: the PREDIMED (Prevención con Dieta Mediterránea) trial. *Circulation*. 2014;130(1):18–26. doi:10.1161/CIRCULATIONAHA.113.006921 pmid:24787471

412. Martínez-González MÁ, Toledo E, Arós F, et al. Correction: Extravirgin olive oil consumption reduces risk of atrial fibrillation: The PREDIMED (Prevención con Dieta Mediterránea) trial (Circulation (2014) 130 (18-26) DOI: 10.1161/CIRCULATIONAHA.113.006921). *Circulation*. 2019;140(24):E939. doi:10.1161/CIR.0000000000000716 pmid:31815542

413. Sanches Machado D'Almeida K, Ronchi Spillere S, Zuchinali P, Corrêa Souza G. Mediterranean diet

and other dietary patterns in primary prevention of heart failure and changes in cardiac function markers: A systematic review [Internet]. Vol. 10, Nutrients. Nutrients; 2018. doi:10.3390/nu10010058 pmid:29320401

414. Strengers JG, den Ruijter HM, Boer JMA, et al. The association of the Mediterranean diet with heart failure risk in a Dutch population. *Nutr Metab Cardiovasc Dis*. 2021;31(1):60–6. doi:10.1016/j.numecd.2020.08.003 pmid:33127254

415. Papadaki A, Martínez-González MÁ, Alonso-Gómez A, et al. Mediterranean diet and risk of heart failure: results from the PREDIMED randomized controlled trial. *Eur J Heart Fail*. 2017;19(9):1179–85. doi:10.1002/ejhf.750 pmid:28133855

416. Papadaki A, Martínez-González MÁ, Alonso-Gómez A, et al. Erratum to: Mediterranean diet and risk of heart failure: results from the PREDIMED randomized controlled trial: Mediterranean diet and heart failure risk (European Journal of Heart Failure, (2017), 19, 9, (1179-1185), 10.1002/ejhf.750). *Eur J Heart Fail*. 2019;21(3):389–91. doi:10.1002/ejhf.1346 pmid:30883998

417. Fitõ M, Estruch R, Salas-Salvadõ J, et al. Effect of the Mediterranean diet on heart failure biomarkers: A randomized sample from the PREDIMED trial. *Eur J Heart Fail*. 2014;16(5):543–50. doi:10.1002/ejhf.61 pmid:24574190

418. Imamura F, Lemaitre RN, King IB, et al. Long-chain monounsaturated fatty acids and incidence of congestive heart failure in 2 prospective cohorts. *Circulation*. 2013;127(14):1512–21. doi:10.1161/CIRCULATIONAHA.112.001197 pmid:23487436

419. Yi SY, Steffen LM, Lutsey PL, Cushman M, Folsom AR. Contrasting Associations of Prudent and Western Dietary Patterns with Risk of Developing Venous Thromboembolism. *Am J Med*. 2021;134(6):763-768.e3. doi:10.1016/j.amjmed.2021.01.016 pmid:33609530

420. Hoevenaar-Blom MP, Nooyens ACJ, Kromhout D, et al. Mediterranean Style Diet and 12-Year Incidence of Cardiovascular Diseases: The EPIC-NL Cohort Study. *PLoS One*. 2012;7(9). doi:10.1371/journal.pone.0045458 pmid:23029021

421. Chen GC, Arthur R, Mossavar-Rahmani Y, et al. Adherence to Recommended Eating Patterns Is Associated with Lower Risk of Peripheral Arterial Disease: Results from the Women's Health Initiative. *Hypertension*. 2021;78(2):447–55. doi:10.1161/HYPERTENSIONAHA.121.17432 pmid:34176290

422. Ruiz-Canela M, Estruch R, Corella D, Salas-Salvadó J, Martínez-González MA. Association of Mediterranean diet with peripheral artery disease: the PREDIMED randomized trial. *JAMA*. 2014;311(4):415–7. doi:10.1001/JAMA.2013.280618 pmid:24449321

423. Martínez-González MA. Protocol Deviations, Reanalyses, and Corrections to PREDIMED Trial Derivative Study on Peripheral Artery Disease. *JAMA - J Am Med Assoc*. 2018;320(21):2272. doi:10.1001/jama.2018.16553 pmid:30398616

424. Rees K, Takeda A, Martin N, et al. Mediterranean-style diet for the primary and secondary prevention of

cardiovascular disease. Vol. 2019, Cochrane Database of Systematic Reviews. John Wiley and Sons Ltd; 2019. doi:10.1002/14651858.CD009825.pub3 pmid:30864165

425. Kaluza J, Stackelberg O, Harris HR, et al. Mediterranean Diet is Associated with Reduced Risk of Abdominal Aortic Aneurysm in Smokers: Results of Two Prospective Cohort Studies. *Eur J Vasc Endovasc Surg*. 2021;62(2):284–93. doi:10.1016/j.ejvs.2021.04.017 pmid:34144885

426. Bianchi VE, Rizzi L, Somaa F. The role of nutrition on Parkinson's disease: a systematic review [Internet]. Nutritional Neuroscience. Nutr Neurosci; 2022. doi:10.1080/1028415X.2022.2073107 pmid:35730414

427. Strikwerda AJ, Dommershuijsen LJ, Kamran Ikram M, Voortman T. Diet Quality and Risk of Parkinson's Disease: The Rotterdam Study. *Nutrients*. 2021;13(11). doi:10.3390/NU13113970 pmid:34836224

428. Muzerengi S, Rick C, Begaj I, et al. Coding accuracy for Parkinson's disease hospital admissions: implications for healthcare planning in the UK. *Public Health*. 2017;146:4–9. doi:10.1016/j.puhe.2016.12.024 pmid:28404473

429. Kunutsor SK, Laukkanen JA, Whitehouse MR, Blom AW. Adherence to a Mediterranean-style diet and incident fractures: pooled analysis of observational evidence. *Eur J Nutr*. 2018;57(4):1687–700. doi:10.1007/s00394-017-1432-0 pmid:28331983

430. Jennings A, Mulligan AA, Khaw KT, Luben RN, Welch AA. A Mediterranean Diet Is Positively Associated with Bone and Muscle Health in a Non-Mediterranean Region in 25,450 Men and Women from EPIC-Norfolk. *Nutrients*. 2020;12(4). doi:10.3390/NU12041154 pmid:32326165

431. Moreno-Montañés J, Gándara E, Gutierrez-Ruiz I, et al. Healthy Lifestyle Score and Incidence of Glaucoma: The Sun Project. *Nutrients*. 2022;14(4). doi:10.3390/NU14040779 pmid:35215429

432. Vergroesen JE, de Crom TOE, van Duijn CM, et al. MIND diet lowers risk of open-angle glaucoma: the Rotterdam Study. *Eur J Nutr*. 2022; doi:10.1007/s00394-022-03003-w pmid:36123555

433. García-Layana A, Ciufo G, Toledo E, et al. The Effect of a Mediterranean Diet on the Incidence of Cataract Surgery. *Nutrients*. 2017;9(5). doi:10.3390/NU9050453 pmid:28467363

434. Francisco SG, Smith KM, Aragonès G, et al. Dietary Patterns, Carbohydrates, and Age-Related Eye Diseases. *Nutrients*. 2020;12(9):1–19. doi:10.3390/NU12092862 pmid:32962100

435. Caldiroli L, Molinari P, Abinti M, et al. Can Mediterranean Diet Have a Positive Impact on Kidney Health? A Pending Answer to a Long-Time Question. *Nutrients*. 2022;14(20):4366. doi:10.3390/nu14204366 pmid:36297050

436. Khalatbari-Soltani S, Marques-Vidal P, Imamura F, Forouhi NG. Prospective association between adherence to the Mediterranean diet and hepatic steatosis: The Swiss CoLaus cohort study. *BMJ Open*. 2020;10(12). doi:10.1136/bmjopen-2020-040959 pmid:33371031

437. Ma J, Hennein R, Liu C, et al. Improved Diet Quality Associates With Reduction in Liver Fat, Particularly in Individuals With High Genetic Risk Scores for Nonalcoholic Fatty Liver Disease. *Gastroenterology*. 2018;155(1):107–17. doi:10.1053/j.gastro.2018.03.038 pmid:29604292

438. Guasch-Ferré M, Willett WC. The Mediterranean diet and health: a comprehensive overview. *J Intern Med*. 2021;290(3):549–66. doi:10.1111/JOIM.13333 pmid:34423871

439. Grammatikopoulou MG, Nigdelis MP, Theodoridis X, et al. How fragile are Mediterranean diet interventions? A research-on-research study of randomised controlled trials. *BMJ Nutr Prev Heal*. 2021;4(1):115–31. doi:10.1136/BMJNPH-2020-000188 pmid:34308119

440. Forster JL, Jeffery RW, VanNatta M, Pirie P. Hypertension prevention trial: do 24-h food records capture usual eating behavior in a dietary change study? *Am J Clin Nutr*. 1990;51(2):253–7. doi:10.1093/AJCN/51.2.253 pmid:2407098

441. Williams CM, Ashwell M, Prentice A, Hickson M, Stanner S. Nature of the evidence base and frameworks underpinning dietary recommendations for prevention of non-communicable diseases: A position paper from the Academy of Nutrition Sciences [Internet]. Vol. 126, British Journal of Nutrition. Br J Nutr; 2021. p. 1076–90. doi:10.1017/S0007114520005000 pmid:34515022

442. Shimonovich M, Pearce A, Thomson H, Keyes K, Katikireddi SV. Assessing causality in epidemiology: revisiting Bradford Hill to incorporate developments in causal thinking. *Eur J Epidemiol*. 2021;36(9):873–87. doi:10.1007/s10654-020-00703-7 pmid:33324996

443. Abbasi A, Sahlqvist AS, Lotta L, et al. A systematic review of biomarkers and risk of incident type 2 diabetes: An overview of epidemiological, prediction and aetiological research literature. *PLoS One*. 2016;11(10). doi:10.1371/journal.pone.0163721 pmid:27788146

444. Martín-Peláez S, Fito M, Castaner O. Mediterranean Diet Effects on Type 2 Diabetes Prevention, Disease Progression, and Related Mechanisms. A Review. *Nutrients*. 2020;12(8):1–15. doi:10.3390/NU12082236 pmid:32726990

445. Hernán MA. The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data. *Am J Public Health*. 2018;108(5):616–9. doi:10.2105/AJPH.2018.304337 pmid:29565659

446. Broadbent A. The C-word, the P-word, and realism in epidemiology. *Synthese*. 2021;198(10):2613–28. doi:10.1007/s11229-019-02169-x

447. Pfeiffer CM, Lacher DA, Schleicher RL, Johnson CL, Yetley EA. Challenges and Lessons Learned in Generating and Interpreting NHANES Nutritional Biomarker Data. *Adv Nutr An Int Rev J*. 2017;8(2):290–307. doi:10.3945/an.116.014076 pmid:28298273

448. Sempionatto JR, Montiel VRV, Vargas E, Teymourian H, Wang J. Wearable and Mobile Sensors for Personalized Nutrition. *ACS Sensors*. 2021;6(5):1745–60. doi:10.1021/acssensors.1c00553 pmid:34008960

449. Shi Z, Li X, Shuai Y, Lu Y, Liu Q. The development of wearable technologies and their potential for measuring nutrient intake: Towards precision nutrition [Internet]. Nutrition Bulletin. John Wiley and Sons Inc; 2022. doi:10.1111/nbu.12581 pmid:36134894

450. McNairn M, Brito A, Dillard K, et al. Postprandial Dried Blood Spot–Based Nutritional Metabolomic Analysis Discriminates a High-Fat, High-Protein Meat-Based Diet from a High Carbohydrate Vegan Diet: A Randomized Controlled Crossover Trial. *J Acad Nutr Diet*. 2021;121(5):931-941.e2. doi:10.1016/j.jand.2020.10.024 pmid:33279463

Appendix

Appendix 2.1 Search strategies in Embase and Web of Science

- Embase search:

#1 “dietary pattern”.tw. OR “dietary patterns”.tw. OR “diet pattern”.tw. OR “diet patterns”.tw. OR “diet quality”.tw. OR “dietary quality”.tw. OR “food pattern”.tw. OR “food patterns”.tw. OR “diet score”.tw. OR “diet scores”.tw. OR “dietary score”.tw. OR “dietary scores”.tw. OR “diet index”.tw. OR “diet indices”.tw. OR “dietary index”.tw. OR “dietary indices”.tw. OR “eating index”.tw. OR “eating indices”.tw. OR “eating patterns”.tw. OR “eating pattern”.tw. OR “healthy diet”.tw. OR “healthy diets”.tw. OR “food score”.tw. OR “foods score”.tw. OR “diet diversity”.tw. OR “dietary diversity”.tw. OR “Mediterranean diet”.tw. OR "dietary approaches to stop hypertension".tw. OR “healthy eating index”.tw. OR “DASH”.tw. OR “HEI”.tw. OR “AHEI”.tw. OR “Nordic diet".tw.

#2 plasma.tw. OR serum.tw. OR circulating.tw. OR blood.tw. OR urin*.tw. OR excret*.tw.

#3 vitamin*.tw. OR mineral*.tw. OR ascorbate.tw. OR acid.tw. OR acids.tw. OR caroten*.tw. OR lycopene.tw. OR cryptoxanthin.tw. OR lutein.tw. or zeaxanthin.tw. OR folate.tw. OR tocopherol*.tw. OR polyphenol*.tw. OR phytochemical*.tw. OR nitrogen.tw. OR potassium.tw. OR sodium.tw.

#4 exp biomarkers/ OR biomarker*.tw. OR metabolomic*.tw. OR metabonomic*.tw. OR lipidomic*.tw. OR proteomic*.tw. OR omic*.tw. OR isotop*.tw. OR “metabolic profile”.tw. OR “metabolic profiles”.tw. OR “metabolite profile”.tw. OR “metabolite profiles”.tw. OR “metabolic signature”.tw. OR “metabolic signatures”.tw. OR “lipid signature”.tw. OR “lipid signatures”.tw. OR VOC*.tw. OR volatile.tw.

#5 exp microbiota/ OR exp gastrointestinal microbiome/ OR urin*.tw. OR plasma.tw. OR serum.tw. OR blood.tw. OR hair.tw. OR “adipose tissue”.tw. OR toenail*.tw. OR fingernail*.tw. OR “metabolic profile”.tw. OR “metabolic profiles”.tw. OR “metabolic signature”.tw. OR “metabolic signatures".tw. OR Microbiota.tw. OR “lipid signature”.tw. OR “lipid signatures”.tw. OR microbiome*.tw. OR microflora*.tw. OR microbiota*.tw. OR microbial.tw. OR gut flora*.tw. OR intestinal flora*.tw. OR intestine flora*.tw. OR fecal.tw. OR faecal.tw. OR faeces.tw. OR breath.tw.

#6 exp animal/ NOT exp human/

#7 #1 AND ((#2 AND #3) OR (#4 AND #5)) NOT #6

- Web of Science search:

#1 TS=(“dietary pattern” OR “dietary patterns” OR “diet pattern” OR “diet patterns” OR “diet quality” OR “dietary quality” OR “food pattern” OR “food patterns” OR “diet score” OR “diet scores” OR “dietary score” OR “dietary scores” OR “diet index” OR “diet indices” OR “dietary index” OR “dietary indices” OR “eating index” OR “eating indices” OR “eating patterns” OR “eating pattern” OR “healthy diet” OR “healthy diets” OR “food score” OR “foods score” OR “diet diversity” OR “dietary diversity” OR “Mediterranean diet” OR "dietary approaches to stop hypertension" OR “healthy eating index” OR “DASH” OR “HEI” OR “AHEI” OR “Nordic diet") OR TI=(“dietary pattern” OR “dietary patterns” OR “diet pattern” OR “diet patterns” OR “diet quality” OR “dietary quality” OR “food pattern” OR “food patterns” OR “diet score” OR “diet scores” OR “dietary score” OR “dietary scores” OR “diet index” OR “diet indices” OR “dietary index” OR “dietary indices” OR “eating index” OR “eating indices” OR “eating patterns” OR “eating pattern” OR “healthy diet” OR “healthy diets” OR “food score” OR “foods score” OR “diet diversity” OR “dietary diversity” OR “Mediterranean diet” OR "dietary approaches to stop hypertension" OR “healthy eating index” OR “DASH” OR “HEI” OR “AHEI” OR “Nordic diet")

#2 TS=(plasma OR serum OR circulating OR blood OR urin* OR excret*) OR TI=(plasma OR serum OR circulating OR blood OR urin* OR excret*)

#3 TS=(vitamin* OR mineral* OR ascorbate OR acid OR acids OR caroten* OR lycopene OR cryptoxanthin OR lutein or zeaxanthin OR folate OR tocopherol* OR polyphenol* OR phytochemical* OR nitrogen OR potassium OR sodium) OR TI=(vitamin* OR mineral* OR ascorbate OR acid OR acids OR caroten* OR lycopene OR cryptoxanthin OR lutein or zeaxanthin OR folate OR tocopherol* OR polyphenol* OR phytochemical* OR nitrogen OR potassium OR sodium)

#4 TS=(biomarker* OR metabolomic* OR metabonomic* OR lipidomic* OR proteomic* OR omic* OR isotop* OR “metabolic profile” OR “metabolic profiles” OR “metabolite profile” OR “metabolite profiles” OR “metabolic signature” OR “metabolic signatures” OR “lipid signature” OR “lipid signatures” OR VOC* OR volatile) OR TI=(biomarker* OR

metabolomic* OR metabonomic* OR lipidomic* OR proteomic* OR omic* OR isotop* OR “metabolic profile” OR “metabolic profiles” OR “metabolite profile” OR “metabolite profiles” OR “metabolic signature” OR “metabolic signatures” OR “lipid signature” OR “lipid signatures” OR VOC* OR volatile)

#5 TS=(urin* OR plasma OR serum OR blood OR hair OR “adipose tissue” OR toenail* OR fingernail* OR “metabolic profile” OR “metabolic profiles” OR “metabolic signature” OR “metabolic signatures" OR Microbiota OR “lipid signature” OR “lipid signatures” OR microbiome* OR microflora* OR microbiota* OR microbial OR gut flora* OR intestinal flora* OR intestine flora* OR fecal OR faecal OR faeces OR breath) OR TI=(urin* OR plasma OR serum OR blood OR hair OR “adipose tissue” OR toenail* OR fingernail* OR “metabolic profile” OR “metabolic profiles” OR “metabolic signature” OR “metabolic signatures" OR Microbiota OR “lipid signature” OR “lipid signatures” OR microbiome* OR microflora* OR microbiota* OR microbial OR gut flora* OR intestinal flora* OR intestine flora* OR fecal OR faecal OR faeces OR breath)

#6 TS=(animal* NOT human*) OR TI=(animal* NOT human*)

#7 TI=("rat*" or "monkey*" or "rabbit*" or "cat*" or "dog*" or "primate*" or "mouse" or "mice" or "pig*")

#7 #1 AND ((#2 AND #3) OR (#4 AND #5)) NOT (#6 OR #7)

Appendix 3.1 Multi-country (subcohort) and country-specific nutritional biomarker scores of the Mediterranean Diet Score in the EPIC-InterAct subcohort: standardised coefficients*

Biomarker	Subc.	FR	IT	ES	UK-g	UK-h	NL	GE	SE	DE
n	10,569	453	1,537	3,083	865	187	1,235	1,686	892	1,710
α-carotene	0.07	-	-	0.04	-	-	0.11	0.06	0.2	0.11
β-carotene	-	-	0.11	0.02	-	-	-	-	-	-
β-cryptoxanthin	0.07	-	-	0.07	0.1	-	0.09	-	0.05	0.06
Lycopene	0.03	-	-	-	0.07	-	0.06	0.11	0.06	0.05
Lutein	0.07	0.14	0.07	0.05	-	0.22	-	0.07	-	0.07
Zeaxanthin	-0.02	0.02	-	-	-	-	-	-0.06	-	-
Vitamin C	0.07	-	0.07	0.06	0.08	0.1	-	0.06	-	-
C14:0	-	-	-	0.04	-	-	-	-	-0.08	-
C16:0	-0.08	-	-	0.04	-	-	-	-	-	-
C18:0	0.04	-	-	-	-	-	-	-	-0.07	-
C15:0	-	-	-0.1	-0.13	-0.11	-0.16	-	-	-0.14	-
C17:0	-0.02	-	-0.07	-0.05	-	-	-	-	-	-
C20:0	-0.01	-	-	-0.03	-	-	-	-	-0.09	-
C22:0	-	-	-	-0.06	-0.22	-	-	-0.06	-	-
C23:0	0.03	-	-	-	-0.06	-	-	-	-	-
C24:0	0.02	-	-	-	0.13	-	-	-	0.09	-
C18:3n-3	-	-	0.04	0.04	-	-	-	-	0.09	-0.07
C20:5n-3	-	-	-	-	-	-	-	-	-	-
C22:5n-3	0.02	-	-0.08	-0.06	-	-	-	-	-	-0.07
C22:6n-3	0.03	-	0.15	0.14	0.07	-	0.07	0.08	0.17	0.09
C18:2n-6c	-0.01	-	-	-	-	-	-	-	-	-
C18:3n-6	0.01	-	0.09	0.05	-	-	-	-	0.09	0.08
C20:2	-	-	-	0	-	-	-	-	-	-
C20:3n-6	0.02	-	-	0.03	-	-	-	-	-	-
C20:4n-6	-	-	-	-	-	-	-	-	-	-0.1
C22:4	0.00	-	-0.11	-0.09	-0.15	-	-	-0.1	-	-
C22:5n-6	-0.05	-	-	-	-	-	-	-	-0.1	-
C16:1	-0.14	-	-	-0.11	-	-	-	-	-	-
C17:1	-	-	-	-0.01	-	-	-0.06	0.04	-	-
C18:1n-9c	-	-	-	0.13	-0.04	-	-	-0.04	-	-
C20:1	0.06	-	-	-0.03	-	-	-	-	-	0.04
C24:1	-0.01	-	-	0.11	0.09	-	-	-	-	0.1
C18:1n-9t	0.05	-	-0.05	-	-	-	-0.07	-	-	-0.04
C18:2n-6t	0.07	-	-	0.01	-	-	-	-	0.06	-
Magnesium	-0.01	-	-	-0.04	-	0.13	-	-0.06	-	0.05
Calcium	0.00	-	-	-0.01	-	-	-	-	-	-
Vit. 25(OH) D_3	-0.02	-	-	-0.03	-	-	-	-	-	-
Ferritin	0.04	-	-0.04	0.03	-	-	-	0.08	-	0.06
Iron	0.00	-	-	0.03	-	-	-	-	-	-
Transferrin	0.04	-	-	0.06	-	-	-	-	-	0.06

Abbreviations: Subc. – subcohort (multi-country score); FR – France; IT – Italy; ES – Spain; UK-g – United Kingdom, general population; UK-h – United Kingdom, health-conscious participants (recruitment targeting a large proportion of vegetarians); NL – the Netherlands; DE – Denmark; SE – Sweden; DE – Germany, DK – Denmark; Vit. – vitamin

*Red colour highlights positive coefficients, and blue colour highlights negative coefficients with intensity proportional to value.

Appendix 3.2 Multi-country (subcohort) and country-specific nutritional biomarker scores of the alternative Healthy Eating Index-2010 in the EPIC-InterAct subcohort: standardised coefficients*

Biomarker	Subc.	FR	IT	ES	UK-g	UK-h	NL	GE	SE	DE
n	10,569	453	1,537	3,083	865	187	1,235	1,686	892	1,710
α-carotene	0.10	0.04	0.06	0.04	0.15	0.19	0.11	0.2	0.15	0.14
β-carotene	0.02	-	0.04	0.07	-	-	-	-0.14	-	-
β-cryptoxanthin	0.06	0.06	0.1	0.09	-	-	-	-	0.06	0.04
Lycopene	-	-	-	-0.01	-	-	-	-	0.1	-
Lutein	0.05	0.13	0.07	0.07	-	-	-	-	-	0.06
Zeaxanthin	-0.10	-0.11	-	-0.08	-	-	-	-0.09	-0.08	-0.09
Vitamin C	0.02	0.09	-	0.03	-	-	-	0.04	-	-0.05
C14:0	-	-	-	-	-	-	-	-	-	-
C16:0	-	-	-	-0.12	-	-	-	-	-	-0.01
C18:0	-0.09	-	-	-0.11	-	-	-	-	-	-
C15:0	-0.07	-	-	-0.06	-	-	-	-	-0.12	0.11
C17:0	0.04	-	-	0.04	-	-	0.13	0.06	-	-
C20:0	0.01	-	0.06	-0.03	-	-	-	-	-0.11	-
C22:0	-0.09	-	-	-0.08	-	-	-	-	-	-
C23:0	-0.09	0.08	-	-	-0.05	-	-	-	-	-
C24:0	0.02	-	-	0.16	-	-	0.16	0.09	0.12	0.03
C18:3n-3	-	-	-	-0.03	-	-	-	-0.02	-	-0.04
C20:5n-3	-0.02	-	-	-	-	-	-	-	-	-
C22:5n-3	-	-	-	-0.09	-0.09	-0.12	-0.06	-0.1	-	-0.03
C22:6n-3	0.05	-	0.12	0.03	0.13	-	0.21	0.14	0.16	-
C18:2n-6c	0.01	-	-	-0.18	-	-	0.13	-	-	-
C18:3n-6	0.00	-	-	0.04	-	-	0.07	0.06	0.08	0.09
C20:2	-	0.06	-	-0.01	-	0.1	-	-	-	0.06
C20:3n-6	0.05	-	0.08	-	-	-	-	0.02	-	-
C20:4n-6	-0.06	-	-	-0.11	-	-	-	-0.09	-	-0.08
C22:4	-	-	-	-	-	-	-	-	-	-
C22:5n-6	-0.05	-0.11	-0.13	-0.08	-	-	-0.07	-0.06	-	-0.01
C16:1	-	-	-	-0.09	-	-	-	-0.14	-	-0.09
C17:1	-0.07	-0.1	-	0.05	-	-	-	-	-0.06	0.01
C18:1n-9c	-0.08	-	-	-0.14	-0.17	-	-0.05	-0.06	-	-0.15
C20:1	0.09	-	-	-	-	-	-	-	-	-
C24:1	-0.03	-	-	-0.09	-	-	-0.13	-0.08	-	-
C18:1n-9t	0.09	-0.1	-	-0.01	-	-	-	0.05	-	0.03
C18:2n-6t	-0.03	-	-	0.01	-0.06	-	-	-	-	-
Magnesium	-0.01	-	-	0	-	-	-0.06	-0.04	-	0.02
Calcium	-	-	-	-0.01	-	-	-	-	-	0.01
Vit. 25(OH) D_3	-	-	-	0.01	-	-	0.04	-	-	-0.04
Ferritin	0.01	-	-	-	-	-	-0.04	0.05	-	-0.03
Iron	0.01	-	-	0.01	-	-	-	-	-	0.02
Transferrin	-	0.11	-	0.01	-	-	-	-	-	-

Abbreviations: Subc. – subcohort (multi-country score); FR – France; IT – Italy; ES – Spain; UK-g – United Kingdom, general population; UK-h – United Kingdom, health-conscious participants (recruitment targeting a large proportion of vegetarians); NL – the Netherlands; DE – Denmark; SE – Sweden; DE – Germany, DK – Denmark; Vit. – vitamin

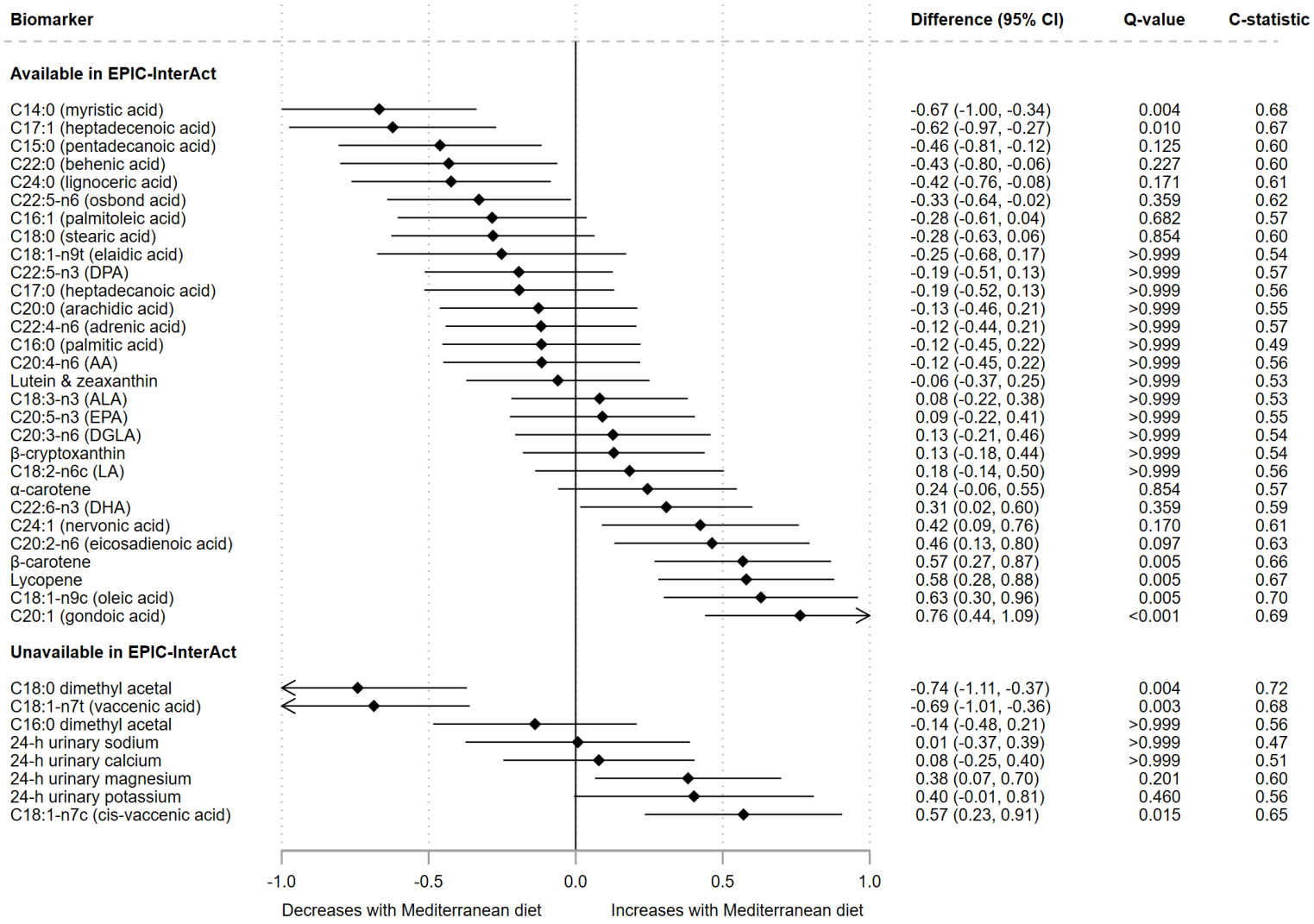
*Red colour highlights positive coefficients, and blue colour highlights negative coefficients with intensity proportional to value.

Appendix 3.3 Multi-country (subcohort) and country-specific nutritional biomarker scores of the DASH diet in the EPIC-InterAct subcohort: standardised coefficients*

Biomarker	Subc.	FR	IT	ES	UK-g	UK-h	NL	GE	SE	DE
n	10,562	453	1,537	3,076	865	187	1,235	1,686	892	1,710
α-carotene	0.12	-	-	0.08	0.11	-	0.09	0.2	0.22	0.16
β-carotene	-0.03	-	0.05	-	-	-	-	-0.14	-	0.02
β-cryptoxanthin	0.15	0.08	0.2	0.18	0.11	-	0.12	0.08	0.09	0.08
Lycopene	0.00	-	-	-	0.05	0.08	-	-	-	-
Lutein	0.03	0.13	-	0.06	-	0.18	-	0.08	-	0.04
Zeaxanthin	-0.04	-	-0.08	-0.06	-	-	-	-0.07	-	0.01
Vitamin C	0.05	-	-	0.05	0.09	0.14	-	0.06	-	0
C14:0	-0.01	-	-	-	-	-	-	-	-	-
C16:0	-0.08	-	-	-	-	-	-	-	-	0.06
C18:0	-0.10	0.06	-	-	-	-	-0.06	-	-	-
C15:0	-0.05	-	-	-	-	-0.15	-	-	-	-0.11
C17:0	0.06	-	-	-	-	-	0.07	0.05	0.06	0.15
C20:0	-0.01	-	-	-	-	-	-0.06	-	-	-0.05
C22:0	-0.08	-	-	-0.03	-0.13	-	-	-0.02	-	-0.08
C23:0	-0.02	-	-	-	-	-	-	-0.04	-	-0.06
C24:0	0.03	-	-	-	-	-	-	-	0.04	-
C18:3n-3	0.00	-	-	-	-	-	-	-	-	-
C20:5n-3	-0.06	-	-	-0.04	-	-	-	-	-	0.08
C22:5n-3	0.03	-	-	-	-	-	-	0.06	-	-
C22:6n-3	-0.08	-	-	-	-	-0.15	-	-0.1	-	-0.07
C18:2n-6c	-0.16	-	-	-0.07	-	-	-	-	-	-
C18:3n-6	0.04	-	-	-	-	0.09	-	0.05	0.12	0.07
C20:2	0.01	0.12	-	-	-	0.29	-	-0.02	-	0.05
C20:3n-6	-	-	-	-	-	-	-	-	-	-0.05
C20:4n-6	-0.08	-	-	-	-	-	-	-	-0.08	-0.06
C22:4	-0.04	-	-	-	-	-	-	-0.06	-	0.1
C22:5n-6	-0.06	-	-	-	-	-	-	-	-	-0.04
C16:1	0.03	-	-	-	-	-	0.11	-	-	0.13
C17:1	0.00	-	-	-	-0.04	-	-	-	-	0.02
C18:1n-9c	-0.13	-	-	-	-0.07	-	-0.11	-0.07	-	-0.14
C20:1	0.01	0.06	-	-	-	-	-	-	-	-
C24:1	0.02	-	-	-	-	-	-	-	-	0.12
C18:1n-9t	-0.01	-	-	-	-	-	-	-	-	-
C18:2n-6t	0.00	-0.06	-	-	-0.06	-0.12	-	-	-	0.05
Magnesium	-0.01	-	-	-	-	0.1	-	-0.05	-	0.02
Calcium	0.00	-	-	-	-	0.09	-	-	-	-
Vit. 25(OH) D_3	-0.01	-	-	-	-	-	-	-0.05	-	-
Ferritin	0.01	-	-	-	-	-	-	-	-	-
Iron	0.01	-	-	-	-	-	-	-	-	-
Transferrin	0.02	-	-	-	-	-	-	-	-	0.04

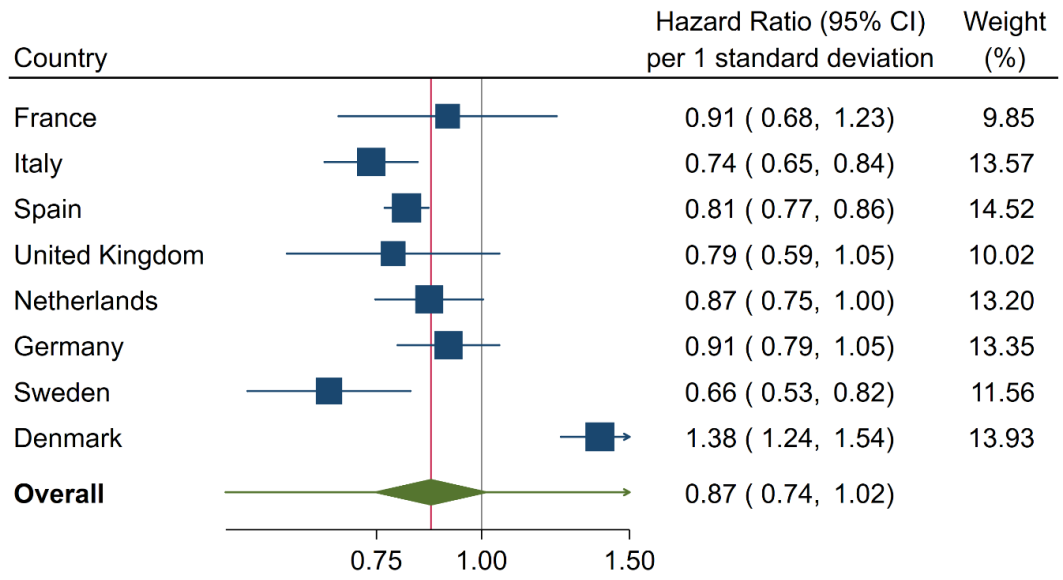
Abbreviations: Subc. – subcohort (multi-country score); FR – France; IT – Italy; ES – Spain; UK-g – United Kingdom, general population; UK-h – United Kingdom, health-conscious participants (recruitment targeting a large proportion of vegetarians); NL – the Netherlands; DE – Denmark; SE – Sweden; DE – Germany, DK – Denmark; Vit. – vitamin

*Red colour highlights positive coefficients, and blue colour highlights negative coefficients with intensity proportional to value.

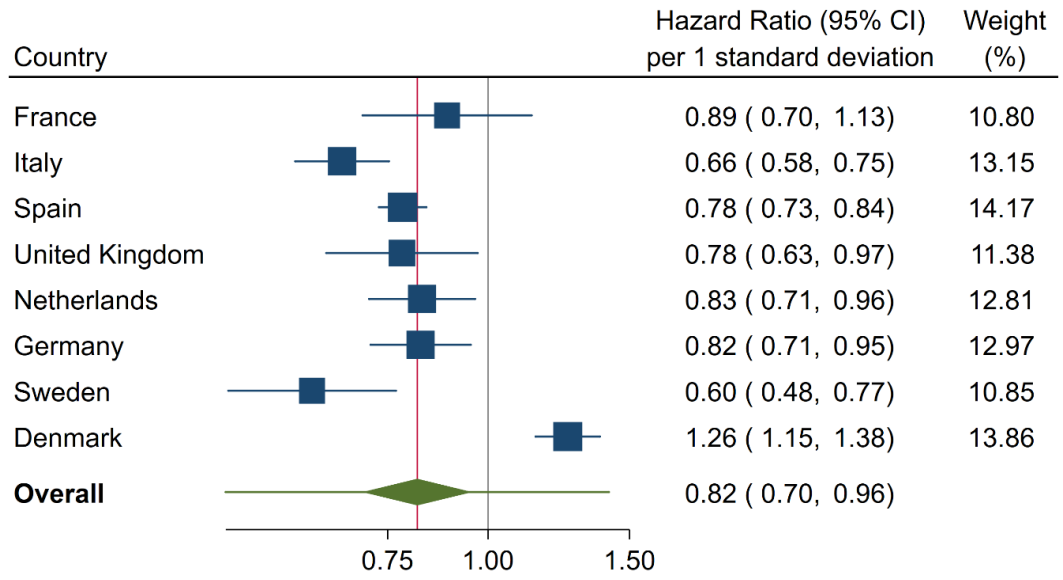


Appendix 4.1 Differences in standardised means of nutritional biomarkers between the Mediterranean and habitual diet groups in the MedLey trial at 6 months

Elastic net logistic regression, linear biomarker terms only



Logistic regression, variable selection based on minimising the Bayesian Information Criterion



Appendix 4.2 Secondary biomarker scores of the Mediterranean diet derived in the MedLey trial and incidence of type 2 diabetes in EPIC-InterAct: associations per 1 standard deviation

Green lines denote 95% prediction intervals. Associations were adjusted for: age (as timescale for effect modification by supplement use), sex, recruitment centre, prevalent cancer, cardiovascular disease, hypertension and hyperlipidaemia; familial history of type 2 diabetes, smoking status (never, former, current smoker), physical activity index (inactive, moderately inactive, moderately active, active), seasonality (sine and cosine function of the day of the year), fasting status (<3, 3-6, >6 hours), current use of vitamin or mineral supplements, marital status (single, married or cohabiting, divorced or separated, widowed), educational attainment (none, primary school, technical or professional school, secondary school, post-secondary school education), current employment, body mass index and waist circumference, and in women, menopausal status (pre-, peri-, postmenopausal, bilateral oophorectomy), current hormone replacement therapy use.

Appendix 4.3 Secondary biomarker scores of the Mediterranean diet derived in the MedLey trial and incidence of type 2 diabetes in Denmark and comparison with associations of individual biomarkers comprising the scores: associations per 1 standard deviation

Analysis	Biomarker scores, excluding biomarkers in turn		Biomarker scores, adjusting for biomarkers		Individual biomarkers	
	BIC-based*	Elastic net†	BIC-based	Elastic net	Denmark	EPIC-InterAct
Full biomarker score	1.26 (1.15-1.38)	1.38 (1.24-1.53)	1.26 (1.15-1.38)	1.38 (1.24-1.53)	-	-
ß-carotene	1.09 (1.06-1.12)	1.17 (1.12-1.23)	1.34 (1.22-1.48)	1.46 (1.31-1.63)	0.78 (0.70-0.87)	0.69 (0.60-0.79)
Lycopene	-	1.15 (1.10-1.20)	-	1.41 (1.26-1.57)	0.88 (0.79-0.98)	0.88 (0.81-0.96)
C24:0	0.99 (0.96-1.02)	1.03 (0.99-1.08)	1.14 (1.04-1.26)	1.19 (1.06-1.33)	0.52 (0.45-0.60)	0.76 (0.66-0.86)
C22:5n-6 (n6-DPA)	1.08 (1.05-1.12)	1.16 (1.10-1.22)	1.28 (1.13-1.43)	1.41 (1.24-1.60)	0.88 (0.76-1.01)	1.13 (1.03-1.25)
C17:1	-	1.11 (1.07-1.16)	-	1.37 (1.23-1.53)	0.80 (0.72-0.89)	0.92 (0.85-1.00)
C18:1n-9c	-	1.15 (1.10-1.20)	-	1.40 (1.26-1.57)	0.99 (0.90-1.10)	1.01 (0.94-1.09)
C20:1	1.04 (1.00-1.07)	1.10 (1.03-1.16)	1.06 (0.94-1.19)	1.19 (1.03-1.38)	1.29 (1.19-1.40)	0.96 (0.84-1.09)
C24:1	1.16 (1.12-1.20)	1.23 (1.17-1.29)	1.60 (1.43-1.79)	1.57 (1.40-1.77)	0.73 (0.66-0.80)	0.74 (0.70-0.78)
All biomarkers adjusted for	-	-	2.02 (0.98-4.16)	1.15 (0.49-2.70)	-	-

*Score formula: ß-carotene ×0.839 + C24:0×(-8.150) + C22:5n-6×(-3.426) + C20:1×5.348 + C24:1×7.495 - 3.859

†Score formula: ß-carotene ×0.453 + lycopene×0.323 + C24:0×(-4.220) + C22:5n-6×(-2.379) + C17:1×(-0.802) + C18:1n-9c×5.056 + C20:1×3.699 + C24:1×3.715 - 19.449

Multivariable adjusted model included the following covariates: age (as timescale for effect modification by supplement use), sex, recruitment centre, prevalent cancer, cardiovascular disease, hypertension and hyperlipidaemia; familial history of type 2 diabetes, smoking status (never, former, current smoker), physical activity index (inactive, moderately inactive, moderately active, active), seasonality (sine and cosine function of the day of the year), fasting status (<3, 3-6, >6 hours), current use of vitamin or mineral supplements, marital status (single, married or cohabiting, divorced or separated, widowed), educational attainment (none, primary school, technical or professional school, secondary school, post-secondary school education), current employment, body mass index and waist circumference, and in women, menopausal status (pre-, peri-, postmenopausal, bilateral oophorectomy), current hormone replacement therapy use.

Appendix 5.1 Associations of biomarker scores of Mediterranean diet and incident type 2 diabetes in EPIC-InterAct per 1 standard deviation by categories of use of dietary supplements and age at baseline*

Biomarker score and covariate category	Multiply imputed analysis			Complete-case analysis		
	n†	HR (95% CI)	$P_{interaction}$‡	n	HR (95% CI)	$P_{interaction}$‡
EPIC-InterAct score, MDS						
Dietary supplements						
Non-users	13,819	0.82 (0.78-0.86)		9,816	0.81 (0.76-0.86)	
Users	8,383	0.92 (0.86-0.97)	0.08	5,614	0.85 (0.79-0.91)	0.42
Age at baseline, years						
<45	4,234	0.86 (0.78-0.95)		3,174	0.86 (0.77-0.96)	
45-60	12,892	0.85 (0.81-0.90)		8,884	0.82 (0.77-0.88)	
>60	5,076	0.88 (0.83-0.94)	0.14	3,372	0.82 (0.76-0.90)	0.25
MedLey trial baseline score, MDS						
Dietary supplements						
Non-users	13,819	0.83 (0.79-0.87)		9,816	0.83 (0.79-0.88)	
Users	8,383	0.87 (0.82-0.93)	0.37	5,614	0.85 (0.79-0.92)	0.67
Age at baseline, years						
<45	4,234	0.84 (0.77-0.92)		3,174	0.85 (0.77-0.95)	
45-60	12,892	0.84 (0.80-0.89)		8,884	0.85 (0.80-0.90)	
>60	5,076	0.88 (0.81-0.95)	0.45	3,372	0.86 (0.78-0.95)	0.84
EPIC-Norfolk score, MDS-pyramid						
Dietary supplements						
Non-users	13,819	0.81 (0.76-0.85)		9,816	0.79 (0.74-0.85)	
Users	8,383	0.90 (0.84-0.96)	0.13	5,614	0.82 (0.75-0.89)	0.83
Age at baseline, years						
<45	4,234	0.80 (0.72-0.89)		3,174	0.80 (0.73-0.88)	
45-60	12,892	0.84 (0.79-0.89)		8,884	0.80 (0.75-0.86)	
>60	5,076	0.87 (0.81-0.94)	0.06	3,372	0.80 (0.73-0.88)	0.04

Abbreviations: EPIC – European Prospective Investigation into Cancer and Nutrition; HR – hazard ratio, CI – confidence interval; MDS – Mediterranean Diet Score; RCT – randomised controlled trial

*Models were adjusted for: age (as timescale), sex, recruitment centre, prevalent cancer, cardiovascular disease, hypertension and hyperlipidaemia; familial history of type 2 diabetes, smoking status (never, former, current smoker), physical activity index (inactive, moderately inactive, moderately active, active), seasonality (sine and cosine function of the day of the year), fasting status (<3, 3-6, >6 hours), current use of vitamin or mineral supplements, marital status (single, married or cohabiting, divorced or separated, widowed), educational attainment (none, primary school, technical or professional school, secondary school, post-secondary school education), current employment, body mass index and waist circumference, and in women, menopausal status (pre-, peri-, postmenopausal, bilateral oophorectomy), current hormone replacement therapy use.

†Numbers of participants by use of dietary supplements in multiply imputed analysis are mid-point values between the smallest and largest values in the imputation datasets.

‡Interaction *P* values for age are based on continuous-by-continuous interaction terms between age and biomarker scores.

Appendix 5.2 Associations of biomarker scores of Mediterranean diet with incidence of type 2 diabetes in EPIC-InterAct: sensitivity analyses

Model*	HR (95% CI) per 1 SD		
	EPIC-InterAct score, MDS	MedLey trial baseline score, MDS	EPIC-Norfolk score, MDS-pyramid
Main result	0.82 (0.76-0.89)	0.83 (0.77-0.89)	0.79 (0.72-0.87)
First 7 years of follow-up	0.80 (0.74-0.87)	0.77 (0.71-0.84)	0.77 (0.69-0.87)
> 7 years of follow-up	0.87 (0.80-0.94)	0.91 (0.82-1.00)	0.84 (0.75-0.95)
Excluding cases in first 2 years of follow-up	0.83 (0.77-0.89)	0.82 (0.77-0.88)	0.80 (0.73-0.88)
Excluding participants with HbA1c > 48mmol/mol	0.83 (0.77-0.90)	0.83 (0.77-0.89)	0.80 (0.72-0.89)
Excluding participants with prevalent cancer, MI or stroke	0.83 (0.77-0.89)	0.83 (0.77-0.89)	0.79 (0.72-0.87)
Excluding outliers in nutritional biomarkers	0.82 (0.76-0.88)	0.83 (0.77-0.89)	0.79 (0.73-0.87)
Biomarker score from a single elastic net regression	0.82 (0.76-0.88)	0.82 (0.76-0.88)	0.82 (0.74-0.90)
Biomarker score calculated using unpenalized coefficients	0.83 (0.77-0.89)	0.83 (0.77-0.89)	0.80 (0.72-0.88)
Biomarker selection with 1 SE higher lambda values†	0.74 (0.67-0.82)	no biomarkers selected	0.76 (0.69-0.84)
Biomarker score unadjusted at derivation stage	0.75 (0.67-0.84)	see Main result	0.76 (0.69-0.84)
Additional adjustment for alcohol	0.83 (0.77-0.90)	0.83 (0.77-0.90)	0.80 (0.73-0.88)
Additional adjustment for meat, olive oil and alcohol	0.83 (0.76-0.90)	0.83 (0.76-0.90)	0.80 (0.73-0.88)

Abbreviations: EPIC – European Prospective Investigation into Cancer and Nutrition; HR – hazard ratio; CI – confidence interval; MDS – Mediterranean Diet Score; MI – myocardial infarction; RCT – randomised controlled trial; SE – standard error

*Multivariable adjusted model included the following covariates: age (as timescale), sex, recruitment centre, prevalent cancer, cardiovascular disease, hypertension and hyperlipidaemia; familial history of type 2 diabetes, smoking status (never, former, current smoker), physical activity index (inactive, moderately inactive, moderately active, active), seasonality (sine and cosine function of the day of the year), fasting status (<3, 3-6, >6 hours), current use of vitamin or mineral supplements, marital status (single, married or cohabiting, divorced or separated, widowed), educational attainment (none, primary school, technical or professional school, secondary school, post-secondary school education), current employment, body mass index and waist circumference, and in women, menopausal status (pre-, peri-, postmenopausal, bilateral oophorectomy), current hormone replacement therapy use. 22,202 participants with non-missing biomarker score data were included in the analysis. Missing covariate data were imputed.

†In comparison to lambda values selected by cross-validation as yielding the lowest prediction error. Increasing lambda values leads to more parsimonious models but may result in no predictors being selected.

Appendix 5.3 Associations of observational biomarker scores of Mediterranean diet with incidence of type 2 diabetes in EPIC-InterAct: effect of adjustment for component biomarkers

Model*	HR (95% CI) per 1 SD†		
	EPIC-InterAct score, MDS	MedLey trial baseline score, MDS	EPIC-Norfolk score, MDS-pyramid
Main result‡	0.82 (0.76-0.89)	0.83 (0.77-0.89)	0.79 (0.72-0.87)
Additional adjustment for:			
a-carotene	0.92 (0.86-0.98)	-	0.88 (0.80-0.97)
ß-cryptoxanthin	0.87 (0.81-0.93)	-	0.84 (0.76-0.93)
Lutein & zeaxanthin	0.90 (0.84-0.96)	0.88 (0.82-0.95)	0.88 (0.80-0.96)
Lycopene	0.86 (0.81-0.91)	-	0.82 (0.75-0.89)
C14:0	0.82 (0.76-0.89)	-	0.80 (0.73-0.88)
C15:0	0.80 (0.74-0.86)	-	-
C16:0	0.79 (0.73-0.86)	-	-
C16:1	0.82 (0.76-0.89)	-	0.87 (0.80-0.95)
C17:0	0.81 (0.75-0.87)	-	0.83 (0.76-0.91)
C17:1	0.83 (0.77-0.88)	-	-
C18:1n-9c	0.83 (0.76-0.89)	0.70 (0.62-0.78)	-
C18:1n-9t	0.80 (0.74-0.87)	-	-
C18:2n-6	-	0.77 (0.72-0.83)	-
C18:3n-3	0.82 (0.76-0.89)	-	-
C20:1	0.82 (0.76-0.89)	-	-
C20:2	0.82 (0.76-0.88)	-	0.79 (0.72-0.87)
C20:3n-6	0.85 (0.79-0.92)	-	-
C20:4n-6	-		0.79 (0.72-0.87)
C20:5n-3	0.77 (0.70-0.85)	-	0.77 (0.69-0.86)
C22:0	0.82 (0.76-0.88)	-	-
C22:4n-6	0.83 (0.77-0.90)	-	-
C22:5n-3	-		0.78 (0.71-0.87)
C22:5n-6	0.84 (0.79-0.90)	-	-
C22:6n-3	0.81 (0.75-0.88)	0.82 (0.75-0.89)	0.78 (0.70-0.87)
C24:0	0.84 (0.78-0.91)	-	-
C24:1	0.91 (0.84-0.98)	-	-
All biomarkers from a given score	0.56 (0.36-0.87)	0.60 (0.40-0.90)	0.67 (0.49-0.91)

Abbreviations: CI – confidence interval; EPIC – European Prospective Investigation into Cancer and Nutrition; HR – hazard ratio; MDS – Mediterranean Diet Score

*Biomarkers were entered into the models as linear and squared terms. Only squared terms which were statistically significant when adjusting for individual biomarkers were used in the analysis adjusting for all biomarkers.

†Hyphen denotes biomarkers which were not components of a given score.

‡Main result was adjusted for the following covariates: age (as timescale), sex, recruitment centre, prevalent cancer, cardiovascular disease, hypertension and hyperlipidaemia; familial history of type 2 diabetes, smoking status (never, former, current smoker), physical activity index (inactive, moderately inactive, moderately active, active), seasonality (sine and cosine function of the day of the year), fasting status (<3, 3-6, >6 hours), current use of vitamin or mineral supplements, marital status (single, married or cohabiting, divorced or separated, widowed), educational attainment (none, primary school, technical or professional school, secondary school, post-secondary

school education), current employment, body mass index and waist circumference, and in women, menopausal status (pre-, peri-, postmenopausal, bilateral oophorectomy), current hormone replacement therapy use. 22,202 participants with non-missing biomarker score data were included in the analysis. Missing covariate data were imputed.

Appendix 6.1 Nutritional and metabolomic biomarkers correlated (absolute r ≥ 0.10) with the Mediterranean diet pyramid score* after Bonferroni correction in the EPIC-Norfolk study

Analyte†	Group or metabolism	Pearson correlation coefficient
Nutritional biomarkers (n = 6,068-6,709)		
α-carotene	Carotenoid	0.17
β-carotene	Carotenoid	0.14
β-cryptoxanthin	Carotenoid	0.19
lycopene	Carotenoid	0.16
lutein	Carotenoid	0.14
zeaxanthin	Carotenoid	0.13
C20:5n-3	Fatty acid	0.10
C22:6n-3	Fatty acid	0.17
vitamin C	Ascorbate and aldarate	0.15
Metabolomics (n = 10,544)		
indolepropionate	Tryptophan	0.12
tryptophan betaine	Tryptophan	0.15
dop -propyl-2-furanpropanoateamine 4-sulfate	Tyrosine	0.10
N-methylproline	Arginine and proline	0.13
glycerate	Glycolysis, gluconeogenesis	0.10
oxalate (ethanedioate)	Ascorbate and aldarate	0.12
threonate	Ascorbate and aldarate	0.14
docosahexaenoylcholine	Acyl choline	0.10
1-docosahexaenoyl-GPC (22:6)	Lysophospholipid	0.12
1-(1-enyl-palmitoyl)-GPC (P-16:0)	Lysoplasmalogen	0.12
1-(1-enyl-palmitoyl)-GPE (P-16:0)	Lysoplasmalogen	0.12
1-palmitoyl-2-docosahexaenoyl-GPC (16:0/22:6)	Phosphatidylcholine	0.13
1-stearoyl-2-docosahexaenoyl-GPC (18:0/22:6)	Phosphatidylcholine	0.13
docosahexaenoate (DHA; 22:6n3)	Fatty acid	0.13
4-allylphenol sulfate	Food component/plant	0.15
ergothioneine	Food component/plant	0.19
methyl glucopyranoside (α + β)	Food component/plant	0.12
stachydrine	Food component/plant	0.16
X - 09789	Unknown	0.11
X - 11315	Unknown	0.17
X - 11847	Unknown	0.15
X - 11849	Unknown	0.15
X - 11858	Unknown	0.11
X - 12212	Unknown	0.11
X - 12306	Unknown	0.10
X - 17145	Unknown	0.16
X - 21442	Unknown	0.11
X - 21752	Unknown	0.10
X - 23644	Unknown	0.14
X - 24431	Unknown	0.11
X - 24693	Unknown	0.15
X - 24738	Unknown	0.14

*The Mediterranean diet pyramid score was residual-adjusted for age, energy, sex, menopausal status, smoking status, educational attainment, marital status, Townsend index, current employment, physical activity, use of hormone replacement therapy, use of antihypertensive and lipid-lowering medication, use of any dietary supplements, use of fish oil, seasonality (sine and cosine function of the day of the year), time since last meal prior to blood draw, disease history (cardiovascular, cerebrovascular, diabetes, cancer), familial history of diabetes and myocardial infarction, and adiposity (body mass index and waist circumference).

†Unannotated metabolites are referred to by a unique numeric identifier with an "X" prefix.