Device-measured 24-hour movement behaviours and risk of incident cardiovascular disease



Rosemary Walmsley

Green Templeton College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

2022

For my grandmothers, who I only wish had the opportunities I have to pursue science and technology, and for my grandfathers, who saw the beginning but not the end.

Acknowledgements

Thank you to participants in UK Biobank and CAPTURE-24, without whom none of this would be possible. Thank you also to the teams of people who make these studies happen.

Thank you to my supervisors, Aiden Doherty and Derrick Bennett. Thank you for inducting me into this fascinating research field, and supporting me through the twists, turns, pandemics, and more.

Thank you to the Medical Research Council, for supporting this research and for supporting the wider training that so much enhanced this experience.

Thank you to the wonderful communities at the Big Data Institute and Nuffield Department of Population Health. Thank you especially to Scott Small, Ben Feakins, Hang Yuan, Aidan Acquah, Adam Sturge, Lynne Morley-Johnson, Karl Smith-Byrne, Rema Ramakrishnan, Angel Wong, Dani Kim, Diego Aguilar-Ramirez, Marie-Alice Frappat, Jenny Censin, Stefania Benonisdottir, Laura Wittemans, Mei Sum Chan, Huaidong Du, Khizr Nawab, and many more. An extra thank you to Shing Chan, for everything from help with Python muddles to philosophical discussions of machine-learning. Thank you to the students who went through our 'Data Challenge', for good times and research ideas.

Thank you to colleagues further afield who I have learned so much from. Too many

to name individually, but especially James Sheppard, Tessa Strain, Terence Dwyer, Mark Woodward, and Charlotte Lund Rasmussen. Thank you to the welcoming movement behaviour research community, especially the wonderful people I met at INTUE, ICAMPAM, and the Granada Workshop in 2019.

Thank you to all the teachers and mentors who have taught, encouraged, and supported me to get here. Thanks especially to Oyinlola Oyebode, without whom I would never have dared start this work, and to Richard Earl, who taught me so much and trusted me to use it.

Thank you to the many friends, old and new, who I am lucky to have in my life.

Thank you to my family, especially Mum, Dad, and Joanna: I am so grateful for your unwavering love and support.

Thank you to Achille, the best part of these years by far.

Abstract

More time in moderate-to-vigorous physical activity (MVPA) is associated with lower cardiovascular disease (CVD) risk. However, most of the 24-hour day is spent in other behaviours, including light physical activity, sedentary behaviour, and sleep. Device-based movement measurement in large prospective cohorts now enables detailed investigation of 24-hour movement behaviours. However, traditional approaches to classifying behaviours in device data, based on 'cut-points', can be crude and sometimes inaccurate. Furthermore, approaches to associate 24-hour movement behaviours with disease risk do not always account for how behaviours make up the 24-hour day. While biases are known to affect analyses, they are rarely assessed quantitatively. This work therefore aimed to investigate effects of device-measured 24-hour movement behaviours on risk of incident CVD, addressing challenges around behaviour classification, analysis of 24-hour movement behaviours, and interpretation.

This work developed machine-learning-based classification methods to classify free-living wrist-worn accelerometer data as sleep, sedentary behaviour, light physical activity behaviours, or moderate-to-vigorous physical activity behaviours. The classifier used a Random Forest with a Hidden Markov model for time smoothing. It was developed using free-living data from a study of 152 participants who had a camera- and diary-based ground-truth measured alongside accelerometry (the CAPTURE-24 study). In Leave-One-Participant-Out analysis, this method classified free-living

movement behaviours with a mean accuracy of 88% (87%, 89%) and a mean Cohen's kappa of 0.80 (0.79, 0.82). UK Biobank (UKB) is a population-based prospective cohort of middle-to-older aged adults, in which ∼100,000 participants wore a wrist-worn Axivity AX3 accelerometer for a 7-day period between 2013 and 2015. The machine-learning model was used to classify movement behaviours of UKB participants, and classification patterns by time of day and by participant characteristics showed good face validity.

UK Biobank participants' 24-hour movement behaviours were associated with risk of incident cardiovascular disease (ICD-10 codes I20-25, I60-69) using a multivariable-adjusted Cox regression model. To account for the compositional nature of 24-hour movement behaviour data, it was modelled using a Compositional Data Analysis approach. Over a median of 6.2 years' follow-up, there were 4,105 incident CVD events among 87,498 participants. Reallocating time from other behaviours to MVPA was associated with a lower risk of incident CVD. Reallocating time from sedentary behaviour to other behaviours was also associated with lower risk. For example, compared to the average movement behaviour composition in this cohort, reallocating 20 minutes/day to MVPA from all other behaviours proportionally was associated with a Hazard Ratio (HR) of 0.91 (0.90, 0.93), while reallocating 1 hour/day to sedentary behaviour from all other behaviours proportionally was associated with a HR of 1.05 (1.03, 1.07).

A range of approaches were used to quantitatively assess the impact of information bias, selection bias, reverse causality, and residual/unmeasured confounding on results. Adjusting for within-person variation in measurements modestly de-attenuated some results and little affected others. Although there were clear selection effects from UK Biobank into the accelerometry sub-study, this did not appear to substantially affect analytic results. Findings on reverse causality were mixed, with little differ-

ence by time period but different findings on sedentary behaviour by (cardiovascular) health status. There was some evidence residual/unmeasured confounding affected MVPA results, but it did not entirely explain findings.

This work contributes to emerging evidence on device-measured 24-hour movement behaviours and risk of incident cardiovascular disease. It both supports existing guidelines encouraging adults to engage in regular moderate-to-vigorous physical activity and limit sedentary time, replacing it with physical activity of any intensity, and supports the development of future guidelines and interventions targeting behaviours throughout the 24-hour day. This work also supports future research, particularly by developing and making available a machine-learning-based classifier with good performance, and by using emerging methods for associating movement behaviours with incident disease and for assessing bias.

Contents

List of Tables **xix**

List of Figures **xxii**

Abbreviations **1**

Contributions **3**

Outputs **4**

Funding and resources **7**

Impact of the COVID-19 pandemic **8**

1 Introduction **9**

1.1 Cardiovascular disease 9

1.2 Physical activity 11

1.3 Physical activity and cardiovascular disease 12

1.4 From physical activity to 24-hour movement behaviours 15

1.5 Measuring 24-hour movement behaviours 18

1.6 Challenges 22

1.6.1 Classifying behaviours 23

1.6.2 Associating 24-hour movement behaviours with disease 25

1.6.3 Interpreting observed associations 28

1.7 Thesis aims and objectives 29

1.8 Thesis outline 30

1.9 Summary 31

2 Literature review on device-measured 24-hour movement behaviours and risk of incident cardiovascular disease 32

2.1 Introduction 32

2.2 Methods 33

2.2.1 Scope 33

2.2.2 Strategy 34

2.3 Findings 35

2.3.1 Systematic search 35

2.3.2 Studies identified 36

2.3.3 Study design 37

2.3.3.1 Accelerometry in prospective cohorts 37
2.3.3.2 Approaches to classifying movement behaviours . . . 42
2.3.3.3 Analytic approaches 43
2.3.3.4 Approaches to assessing bias 44
2.3.4 Study findings 45
2.3.4.1 Sleep 45
2.3.4.2 Sedentary behaviour 45
2.3.4.3 Light physical activity 52
2.3.4.4 Moderate-to-vigorous physical activity 54
2.4 Discussion 56
2.4.1 Findings 56
2.4.2 Size of prospective cohorts with accelerometry 57
2.4.3 Classifying movement behaviours 58
2.4.4 Analytic approaches 59
2.4.5 Assessing bias 60
2.5 Summary 61

3 UK Biobank **62**

3.1 Background 62
3.2 Study design: UK Biobank 63

3.2.1 Recruitment 63
3.2.2 Baseline assessment 64
3.2.3 Follow-up for disease outcomes 64
3.2.4 Resource enhancements and repeat assessments 65
3.2.5 Ethical framework 65
3.2.6 Accessing UK Biobank data 66
3.3 Accelerometry in the UK Biobank study 66
3.3.1 Recruitment 67
3.3.2 Wear protocol 67
3.4 Data preparation 68
3.4.1 Accelerometer data 68
3.4.2 Participant data 68
3.4.3 Disease data 69
3.4.4 Code and data availability 72
3.5 Dataset description 73
3.5.1 Data quality 73
3.5.2 Cohort description 76
3.6 Summary 76

4 Classifying movement behaviours in accelerometer data with machine-

learning **78**

4.1 Introduction 78

4.1.1 Background 78

4.1.2 Aims 81

4.2 Methods 82

4.2.1 CAPTURE-24 study 82

4.2.2 Ethical framework 82

4.2.3 Data labelling 84

4.2.4 Data processing 86

4.2.5 Machine-learning methods 87

4.2.6 Evaluating performance 91

4.2.7 UK Biobank 93

4.2.8 Code and data availability 95

4.3 Results 95

4.3.1 Performance in labelled data 95

4.3.2 UK Biobank 98

4.4 Discussion 100

4.4.1 Overall findings 100

4.4.2 Comparison with previous work using CAPTURE-24 104

4.4.3 Comparison with wider literature 107

4.5 Summary 107

5 Device-measured 24-hour movement behaviours and risk of incident cardiovascular disease 109

5.1 Introduction 109

5.1.1 Background 109

5.1.2 Aims 113

5.2 Compositional Data Analysis: background and motivation 113

5.2.1 Challenges of compositional data 113

5.2.2 Compositional Data Analysis 114

5.2.3 A mathematical foundation for Compositional Data Analysis . 115

5.2.4 Statistical motivations for Compositional Data Analysis 118

5.2.5 Difficulties with a Compositional Data Analysis approach . . . 119

5.3 Methods 120

5.3.1 UK Biobank 120

5.3.2 Movement behaviour phenotypes from accelerometer data . . 121

5.3.3 Confounders 121

5.3.4 Compositional Data Analysis approach 123

5.3.5 Statistical analyses 125

5.3.6 Presentation of results 126
5.3.7 Sensitivity analyses 129
5.3.8 Code and data availability 129
5.4 Results 130
5.4.1 Study population 130
5.4.2 Descriptive characteristics of movement behaviour phenotypes 131
5.4.3 Cox regression analyses for incident cardiovascular disease . . 133
5.4.3.1 Modelling 133
5.4.3.2 Time reallocation 135
5.4.3.3 Sensitivity analyses 139
5.5 Discussion 147
5.5.1 Overall findings 147
5.5.2 Comparison with previous literature 147
5.6 Summary 149

6 Quantitative bias analysis of associations between 24-hour movement behaviours and incident cardiovascular disease 151

6.1 Introduction 151
6.1.1 Background 151
6.1.2 Aims 153

6.2 Information bias due to within-person variation 153
6.2.1 Background 153
6.2.2 Methods 158
6.2.3 Results 164
6.3 Selection bias 168
6.3.1 Background 168
6.3.2 Methods 170
6.3.3 Results 173
6.4 Reverse causality 176
6.4.1 Background 176
6.4.2 Methods 177
6.4.3 Results 178
6.5 Residual and unmeasured confounding 182
6.5.1 Background 182
6.5.2 Methods 182
6.5.3 Results 183
6.6 Discussion 186
6.6.1 Overall findings 186
6.6.2 Comparison with previous literature 188

6.7 Summary 190

7 Discussion **192**

7.1 Introduction 192

7.2 Findings 193

7.3 Strengths and limitations 196

7.3.1 Measurement of 24-hour movement behaviours 196

7.3.2 Associating 24-hour movement behaviours with incident cardiovascular disease 201

7.3.3 Assessment of bias 204

7.3.4 Code and data availability 210

7.4 Implications for population health 210

7.5 Implications for future research 213

7.6 Summary 219

A Supplementary Tables and Figures **220**

A.1 Supplementary Tables 220

A.2 Supplementary Figures 232

B Code and data availability **235**

B.1 Code 235

B.2 Data 236

C Development of an open-source R package for Compositional Data Analysis in epidemiology **238**

C.1 Introduction . . . 238

C.2 Package design . . . 240

C.2.1 Modelling . . . 240

C.2.2 Interpreting results . . . 240

C.2.3 Other functionality . . . 241

C.2.4 Data . . . 241

C.2.5 Testing . . . 242

C.3 Tutorial . . . 243

C.4 Accessing and using the package . . . 243

C.5 Evolution of the package . . . 244

C.6 Strengths and limitations . . . 245

C.7 Future development . . . 246

C.8 Demonstration: package vignette . . . 247

D Within-person variation in compositional data **273**

D.1 Within-person variation in compositional data . . . 273

D.2 Equivalence with perturbation on the simplex . . . 274

D.3 Consequences when working with isometric log-ratio coordinates . . . 276

References **278**

List of Tables

1.1 Comparison of self-reported and device-based measurement of 24-hour movement behaviours. 19

2.1 Inclusion and exclusion criteria for the literature review. 34

2.2 Search terms for literature search on device-measured movement behaviours and risk of incident cardiovascular disease. 35

2.3 Methods of primary studies in the literature review. 39

2.4 Results of primary studies in the literature review. 48

3.1 Comparison of the overall UK Biobank cohort and the sub-sample with accelerometer data by participant characteristics measured at baseline assessment. 74

3.2 Stability of selected UK Biobank variables between baseline assessment and first repeat assessment (2-7 year time gap). 75

4.1 Characteristics of participants in the CAPTURE-24 study. 96

4.2 Minute-wise confusion matrix for machine-learned classification of behaviours in accelerometer data from 152 CAPTURE-24 participants in Leave-One-Participant-Out Cross-Validation. 96

4.3 Movement behaviours of 87,498 UK Biobank participants by participant characteristics. 99

5.1 Movement behaviours of 87,498 UK Biobank participants by participant characteristics. 132

5.2 Coefficient of first isometric log-ratio pivot coordinate for each movement behaviour estimated using all Compositional Data Analysis Cox regression models. 135

6.1 Comparison of accelerometry study population and sub-sample with repeated accelerometer measurements. 164

6.2 Accelerometer measurements at each repeat and univariable attenuation factor. 166

6.3 Participant characteristics by Inverse Probability of Selection Weight in the analytic sample. 173

6.4 Summary of results of quantitative analyses of bias. 186

A.1 Derivation of variables relating to non-accelerometry participant characteristics. 221

A.2 Derivation of variables relating to disease outcomes in hospital inpatient data or death register data. 223

A.3 Features of accelerometry signal used for behaviour classification. . . 224

A.4 Assignment of fine-grained camera image annotations from the Compendium of Physical Activities to broad movement behaviour classes. 225

A.5 Odds Ratios from model for selection into accelerometer sub-cohort among 472,118 participants with no prior cardiovascular disease at baseline assessment. 230

A.6 Odds Ratios from model for selection into accelerometer sub-cohort among 455,921 participants who were alive and had no cardiovascular disease record by 2013-06-01. 231

List of Figures

1.1 Proposed mechanisms for effects of physical activity on cardiovascular disease risk. 14

1.2 Sleep, sedentary behaviour, light physical activity, and moderate-to-vigorous physical activity together make up the 24-hour day. 16

1.3 Research pipeline and challenges in using accelerometer data to understand effects of 24-hour movement behaviours on disease risk. 23

2.1 Flow diagram showing inclusion of studies in the literature review. . . 36

3.1 Timeline of the UK Biobank prospective cohort study. 64

4.1 CAPTURE-24 study design. 83

4.2 Processing of accelerometer data for machine-learning analysis. 87

4.3 An example of a decision tree to classify epochs using average acceleration vector magnitude (avm) and the 75th percentile of acceleration vector magnitude (75thp). 89

4.4 The structure of a Hidden Markov Model. 89

4.5 Participant-wise mean (a) precision and (b) recall for classification of behaviours from accelerometer data calculated in Leave-One-Participant-Out Cross-Validation. 97

4.6 Behaviour profiles for sleep, sedentary behaviour (SB), light physical activity behaviours (LIPA), and moderate-to-vigorous physical activity behaviours (MVPA) among 87,498 UK Biobank participants. 101

4.7 Behaviour profiles according to participant characteristics among up to 87,498 UK Biobank participants. 102

5.1 Illustration of the simplex for compositional data with three compositional parts embedded in three-dimensional real space. 111

5.2 Schematic diagram describing possible causal relationships between variables in the model. 122

5.3 Participant flow diagram for the analysis of 24-hour movement behaviours and incident cardiovascular disease in UK Biobank participants. 131

5.4 Distribution of movement behaviours in 87,498 UK Biobank participants. 134

5.5 Hazard Ratios for incident cardiovascular disease associated with balance between movement behaviours based on 4,105 events among 87,498 UK Biobank participants. 137

5.6 Hazard Ratios for incident cardiovascular disease associated with reallocating time to a given movement behaviour, from all other behaviours proportionally, based on 4,105 events in 87,498 UK Biobank participants. 138

5.7 Hazard Ratios for incident cardiovascular disease for all behaviour pairs estimated using multivariable-adjusted and minimally adjusted Cox regression models. 140

5.8 Hazard Ratios for incident cardiovascular disease for all behaviour pairs estimated using a multivariable-adjusted Cox regression model before and after stratification by BMI. 141

5.9 Hazard Ratios for all and fatal incident cardiovascular disease for all behaviour pairs estimated using a multivariable-adjusted Cox regression model. 142

5.10 Hazard Ratios for incident cardiovascular disease for all behaviour pairs estimated using a multivariable-adjusted Cox regression model for women and men. 143

5.11 Hazard Ratios for incident cardiovascular disease for all behaviour pairs estimated using a multivariable-adjusted Cox regression model for participants aged under 65 and participants aged over 65. 144

5.12 Hazard Ratios for incident cardiovascular disease for all behaviour pairs estimated using a multivariable-adjusted Cox regression model for all participants and in a sensitivity analysis excluding individuals with a zero value in any behaviour. 145

5.13 Hazard Ratios for incident cardiovascular disease estimated using a multivariable-adjusted Compositional Data Analysis Cox regression model and using a multivariable-adjusted linear isotemporal substitution Cox regression model. 146

6.1 Example of the measurement protocol for the repeated measurements study. 159

6.2 Multivariable regression calibration procedure. 161

6.3 Hazard Ratios for incident cardiovascular disease associated with reallocating time to a given movement behaviour, from all other behaviours proportionally, among 87,498 participants (4,105 events) with and without 'correction' for within-person variation using multivariable regression calibration. 167

6.4 Hazard Ratios for incident cardiovascular disease associated with reallocating time to a given movement behaviour, from all other behaviours proportionally, among 87,498 participants (4,105 events) with and without weighting by Inverse Probability of Selection from main study into accelerometer sub-cohort. 175

6.5 Hazard Ratios for incident cardiovascular disease associated with reallocating time to a given movement behaviour, from all other behaviours proportionally, before and after removing the first two years of follow-up (4,105 events among 87,498 participants and 2,947 events among 86,011 participants respectively). 179

6.6 Hazard Ratios for incident cardiovascular disease associated with reallocating time to a given movement behaviour, from all other behaviours proportionally, by period of follow-up (1,158 events in the first 2 years of follow-up, 1,371 events between 2 and 4 years of follow-up, and 1,576 events beyond 4 years of follow-up). 180

6.7 Hazard Ratios for incident cardiovascular disease associated with reallocating time to a given movement behaviour, from all other behaviours proportionally, among 64,048 participants without poor cardiovascular health (2,195 events) and 23,450 participants with poor cardiovascular health (1,910 events). 181

6.8 Hazard Ratios for incident cardiovascular disease and for non-activity related accidents associated with reallocating time to a given movement behaviour, from all other behaviours proportionally, among 87,498 participants (4,105 events) and 84,552 participants (1,375 events) respectively. 184

6.9 Hazard Ratios and corresponding E-values for incident cardiovascular disease associated with reallocating time to a given movement behaviour, from all other behaviours proportionally, among 87,498 participants (4,105 events). 185

7.1 Research and translation pipeline for 24-hour movement behaviour epidemiology. 212

A.1 Hazard Ratios for incident cardiovascular disease associated with reallocating time to a given movement behaviour, from all other behaviours proportionally, among 87,498 participants (4,105 events) with and without 'correction' for within-person variation using multivariable regression calibration, with adjustment for wear month in the calibration equation. 232

A.2 Hazard Ratios for incident cardiovascular disease associated with reallocating time to a given movement behaviour, from all other behaviours proportionally, among 87,498 participants (4,105 events) with and without 'correction' for within-person variation using multivariable regression calibration, precision-weighted based on number of valid repeat measurements. 233

A.3 Hazard Ratios for incident cardiovascular disease associated with reallocating time to a given movement behaviour, from all other behaviours proportionally, among 87,498 participants (4,105 events) with and without weighting by Inverse Probability of Selection from restricted main study cohort into accelerometer sub-cohort. 234

Abbreviations and conventions

ABC - Attitude, Behavior and Change study

BMI - Body Mass Index

BRHS - British Regional Heart Study

CAPTURE-24 - 'Comparing Annotated Pictures with Time-Use Diaries' Recording of Events over 24 hours' camera validation study

CART - Classification and Regression Tree

CHD - Coronary Heart Disease

CI - Confidence Interval

CoDA - Compositional Data Analysis

CVD - Cardiovascular disease

DLM - Daily life movement

EPIC-Norfolk - European Prospective Investigation into Cancer - Norfolk

GPS - Global Positioning System

HAI - Healthy Ageing Initiative

HR - Hazard Ratio

ICD - International Classification of Diseases

ilr - isometric log-ratio

IQR - Inter-quartile range

LIPA - Light physical activity (abbreviates 'light-intensity physical activity')

MEMS - Microelectromechanical systems

MET - Metabolic Equivalent of Task

MI - Myocardial infarction

MPA - Moderate physical activity

MrOS - Osteoporotic Fractures in Men study

MVPA - Moderate-to-vigorous physical activity

NHANES - National Health and Nutrition Examination Survey

NHS - National Health Service

OPACH - Objective Physical Activity and Cardiovascular Health

SB - Sedentary behaviour

SD - Standard deviation

SE - Standard error

TDI - Townsend Deprivation Index

US - United States

USA - United States of America

UK - United Kingdom

UKB - UK Biobank

VPA - Vigorous physical activity

WHI - Women's Health Initiative

Unless otherwise stated:

- The notation X (Y, Z) denotes an estimate followed by its 95% confidence interval.
- 'log' refers to the natural log (ln).

Contributions

The work in this thesis is my own work, and was supervised by Associate Professor Aiden Doherty and Associate Professor Derrick Bennett. I planned and conducted data preparation, data analysis, interpretation, and writing.

Some other people contributed to elements of this work:

- Shing Chan assisted with data preparation of the CAPTURE-24 data in Chapter 4, and provided guidance on many aspects of the work.
- My co-authors in [1] provided critical discussion in both the planning and interpretation of several of the analyses in this thesis.
- The annotation of the CAPTURE-24 dataset, as described in Chapter 4, was performed by trained annotators. This annotation occurred prior to this project, and was conducted independently of it.

Figures and tables in the text are my own, except where otherwise indicated.

Outputs

Much of the work in this thesis has been published as:

Walmsley R, Chan S, Smith-Byrne K, Ramakrishnan R, Woodward M, Rahimi K, Dwyer T, Bennett D, Doherty A. Reallocation of time between device-measured movement behaviours and risk of incident cardiovascular disease. *British Journal of Sports Medicine.* 2021 doi: 10.1136/bjsports-2021-104050

I have contributed to the following papers:

Migueles JH, Aadland E, Andersen LB, Brønd JC, Chastin SF, Hansen BH, Konstabel K, Kvalheim OM, McGregor DE, Rowlands AV, Sabia S, Van Hees VT, **Walmsley R**, Ortega FB. GRANADA consensus on analytical approaches to assess associations with accelerometer-determined physical behaviours (physical activity, sedentary behaviour and sleep) in epidemiological studies. *British Journal of Sports Medicine.* 2022;56:376-384.

Ramakrishnan R, Doherty A, Smith-Byrne K, Rahimi K, Bennett D, Woodward M, **Walmsley R**, Dwyer T. Accelerometer measured physical activity and the incidence of cardiovascular disease: Evidence from the UK Biobank cohort study. *PLoS medicine.* 2021 Jan 12;18(1):e1003487.

Nawab KA, Storey BC, Staplin N, **Walmsley R**, Haynes R, Sutherland S, Crosbie

S, Pugh CW, Harper CH, Landray MJ, Doherty A. Accelerometer-measured physical activity and functional behaviours among people on dialysis. *Clinical kidney journal.* 2021 Mar;14(3):950-8.

Farrahi V, Kangas M, **Walmsley R**, Niemelä M, Kiviniemi A, Puukka K, Collings PJ, Korpelainen R, Jämsä T. Compositional associations of sleep and activities within the 24-H cycle with Cardiometabolic health markers in adults. *Medicine and Science in Sports and Exercise.* 2021 Feb;53(2):324.

I have contributed to open-source software packages, as described in Appendix B [2–4].

I have also presented my work internally and externally:

Internal:

- British Heart Foundation Centre for Research Excellence annual symposium, November 2021
- Nuffield Department of Population Health annual symposium, University of Oxford, March 2021
- China Kadoorie Biobank Seminar, Nuffield Department of Population Health, University of Oxford, March 2021
- Nuffield Department of Population Health annual symposium, University of Oxford, January 2020

External:

- UK Biobank Scientific Conference, November 2021

- International Society for Behavioural Nutrition and Physical Activity Conference, June 2021
- British Heart Foundation Data Science Workshop, May 2021
- Invited talk, Cardiac Electro-Mechanics Research Group, King's College London, April 2021
- International Society for Behavioural Nutrition and Physical Activity Conference, June 2020
- Invited talk, Department of Computer Science, Newcastle University, March 2019

Funding and resources

This research was funded by a Medical Research Council Industrial Strategy Studentship (grant number MR/S502509/1).

Research in this thesis used data from the CAPTURE-24 study and UK Biobank. CAPTURE-24 data collection and annotation was supported by the UK Economic and Social Research Council (grant number ES/L011662/1). Use of UK Biobank was conducted under Application Number 59070.

Some computation used the Oxford Biomedical Research Computing facility, a joint development between the Wellcome Centre for Human Genetics and the Big Data Institute supported by Health Data Research UK and the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre. The views expressed are those of the author and not necessarily those of the UK National Health Service, the NIHR, or the Department of Health and Social Care.

Impact of the COVID-19 pandemic

As suggested by university guidance, this statement describes the impact of the COVID-19 pandemic on work for this thesis. Work on this thesis began in October 2018, with a scheduled finish date of March 2022. The impact of the COVID-19 pandemic in the UK began in March 2020. Following national guidance, I worked entirely from home for 20 months: March 2020 - September 2021; December 2021 - February 2022.

Fortunately, this project used existing data, so suffered little direct disruption. However, there were some impacts. I did not have access to secure computation in March - April 2020, delaying progress on core projects. Remote access to secure computation throughout the 20 month period of working from home was less efficient than office-based access. Contact with supervisors and other key colleagues was substantially reduced over an extended period. The pandemic exacerbated a longstanding health condition and delayed access to treatment.

Chapter 1: Introduction

1.1 Cardiovascular disease

Cardiovascular disease (CVD) is a broad term covering a range of diseases of the cardiovascular system [5]. In this thesis, cardiovascular disease primarily refers to ischaemic heart disease and cerebrovascular disease. Ischaemic heart disease describes an under-supply of oxygen to the heart muscle, usually caused by a reduction of blood flow due to atherosclerosis (build-up of lipid plaques) in arteries of the heart [6]. 'Coronary heart disease', an alternative terminology, may be considered interchangeable with 'ischaemic heart disease' for the purpose of this thesis [6]. Depending on its form and extent, reduction of blood flow to heart muscles can result in different conditions, including angina and myocardial infarction [5]. Cerebrovascular disease describes disruption of the blood supply to part of the brain by ischaemia or bleeding [7]. Strokes are a common form of cerebrovascular disease. Strokes are most frequently ischaemic [8, 9], i.e. caused by a clot or blockage preventing blood flow to part of the brain. Ischaemic strokes often result from atherosclerosis. Less commonly, strokes may be haemorrhagic, i.e. caused by a blood vessel inside the brain bursting.

Cardiovascular disease has a very high global burden, which continues to grow as the global population ages [10]. Ischaemic heart disease caused 9.14 (95% uncertainty

interval[1]: 8.40, 9.74) million deaths in 2019 [11] and stroke 6.55 (95% uncertainty interval: 6.00, 7.02) million [11, 12]. This made them the first and second most common causes of death respectively [12]. Much of this mortality is premature (i.e. occurs before age 70 years): CVD was estimated to be responsible for 6.2 million deaths occurring between the ages of 30 and 70 years in 2019 [10]. In high-income countries, advances in both prevention and treatment led to falling age-standardised CVD mortality rates over recent decades [5, 10, 13, 14]. In the United Kingdom (UK), age-standardised CVD mortality rates fell 70% between 1979 and 2013 [13]. However, the burden of CVD remains substantial [10, 13]. In the UK, ischaemic heart disease and cerebrovascular disease remain common causes of death: in 2018, they were listed as the underlying cause for 63,876 and 35,960 deaths respectively (out of 614,690 total deaths) [15]. In the same year, the last year for which comparable statistics are available, ischaemic heart disease was the leading cause of death for men, and the second most common (after dementia and Alzheimer disease) for women [15]. Declines in age-standardised CVD mortality rates in the UK appear to have stalled in the 2010s [16], a pattern that has been observed in several high-income countries [17, 18].

Cardiovascular disease prevention has a central role in reducing this high public health burden [10]. Behavioural risk factors for CVD include smoking, alcohol consumption, diet, and physical activity [10, 19, 20]. Conditions associated with high risk include high blood pressure, an adverse blood lipid profile, and atrial fibrillation [19]. Interventions target both behavioural risk factors and high-risk conditions [20]. For example, tobacco control policies are widely adopted globally [21]. UK guidelines recommend prescription of cholesterol- and blood-pressure-lowering medication for primary prevention of CVD in at-risk groups [20]. However, stalling declines in CVD

[1]The Global Burden of Disease Study uses uncertainty intervals, rather than confidence intervals. The intervals they use attempt to integrate both systematic and random error.

mortality in high-income settings highlight a need for new approaches to complement existing approaches [10]. Broadening and deepening the evidence base on a range of risk factors may support the design of new interventions to reduce CVD burden and further reduce CVD mortality rates [5]. Understanding behavioural risk factors, such as physical activity, that may be addressed outside of clinical encounters and hence enable 'upstream' public health prevention is an important component of this [5].

1.2 Physical activity

Physical activity is defined as bodily movement produced by the skeletal muscles which requires energy expenditure [22–24]. Physical activity is a *behaviour* integral to human life [25]. Physical activity should be distinguished from *exercise*, which is that subset of physical activity that is planned, structured, and repetitive, and that has as an objective improving physical fitness [23]. Exercise can be studied as a standalone exposure, akin to a medication, that might be prescribed by a clinician in particular doses for particular effect [26]. By contrast, physical activity is woven through daily life, occurring at work [27], for travel (e.g. walking to the shops, cycling to work), and in leisure time (including structured, intentional activities like sports, and activity in the course of other pursuits, like caring for children) [28].

Beyond the overall definition, physical activity can be further categorised in different ways [25]. Physical activity may occur at different *intensities*. Physical activity intensity can be quantified as the rate of energy expenditure, and can be measured in units of Metabolic Equivalent of Task (METs), where 1 MET is resting metabolic rate i.e. the rate of energy expenditure in quiet sitting. For example, physical activity in which energy expenditure is ≥ 3 times energy expenditure at rest ('moderate-to-vigorous physical activity') is often investigated. Physical activity may occur with

different frequencies and in different patterns [25]. As observed above, it may occur in different domains: at work, for transport, or in leisure-time [25, 29]. Other contextual factors, such as the social context [30], may vary. It may be performed with different purpose, intentionality, or affective state e.g. exercise to maintain physical fitness, or incidental physical activity while grocery shopping [31].

1.3 Physical activity and cardiovascular disease

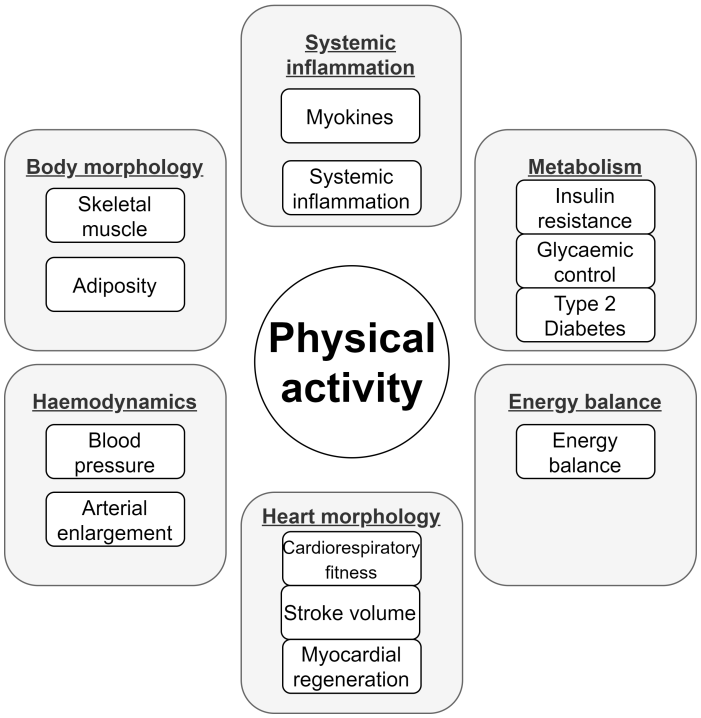
Physical activity (or lack thereof) has been extensively studied as a risk factor for and cause of CVD. The origins of physical activity epidemiology are often traced to the work of Jerry Morris and colleagues in the early 1950s [32]. Morris *et al.* [33] observed different coronary heart disease rates in different occupational groups: bus drivers compared to conductors; telephonists or clerks compared to postmen. Working through various possible explanations for this, they suggested that the most likely was the different physical activity demands of the different jobs [33]. However, notable questions remained: perhaps the disease process preceded the exposure, with men in poor cardiovascular health not selected into the more active occupations [34], or perhaps differences in body size between occupational groups partially drove results [32, 35]. Morris and colleagues therefore went on to study the health effects of physical activity in greater detail and in different types of study, in particular through setting up prospective cohort studies of British civil servants [36, 37]. Establishing exposure status prior to disease incidence minimises reporting bias and determines temporal order, supporting causal inference, and prospective studies also enable collection of data on confounders at the individual level. Large prospective cohort studies now routinely incorporate measurement of physical activity in questionnaires: examples include the Million Women Study [38], the China Kadoorie Biobank [39], and the Nurses' Health Study [40], among many others. Many studies have published find-

ings on the association between self-reported physical activity and CVD [41].

Higher levels of physical activity have consistently been associated with lower CVD risk. The epidemiological evidence on physical activity and CVD risk was recently reviewed for the 2018 Physical Activity Guidelines for Americans [42]. This work identified six systematic reviews associating physical activity with incident coronary heart disease, and five addressing incident stroke [41, 42]. All found higher levels of physical activity to be associated with lower disease risk [41, 42]. While observational evidence may be particularly vulnerable to some forms of bias (e.g. confounding), using randomised controlled trials to study effects of physical activity on incident CVD is very challenging, requiring an intervention able to change physical activity over a sustained period at large scale [43]. However, the prospective observational evidence is supported by several other sources of evidence, including observational studies of associations with cardiometabolic biomarkers [44, 45], detailed laboratory studies of exercise physiology [46–48], and randomised controlled trials of effects of activity interventions on aspects of cardiometabolic health [49]. Possible mechanisms for effects of physical activity on CVD risk include effects on blood pressure, glucose metabolism, inflammatory biomarkers, and energy balance (Figure 1.1) [46–49].

Associations between physical activity and CVD risk are relevant to population health: low levels of physical activity are very prevalent, so even modest effects of physical activity on CVD risk may result in substantial attributable disease burden [50, 51]. In 2012, Lee *et al.* [52] estimated that 6% of coronary heart disease globally was attributable to physical inactivity. They suggest this is likely an underestimate of the attributable burden, as several aspects of the analysis may bias towards the null (e.g. physical activity measured imprecisely; adjustment for potential mediators) [52]. With more than seven million deaths annually from coronary heart disease at that time [52], and more than nine million annually as of 2019 [10], this

Figure 1.1: Proposed mechanisms for effects of physical activity on cardiovascular disease risk [46–48].



represents a large annual death toll. Moreover, economic development, urbanisation, changes in transportation, and technological change are likely to decrease physical activity levels [51]. Assuming that this is borne out, the population-health relevance of physical activity may rise further.

Evidence on physical activity and disease influences public health policy and practice. In 1996, the United States Surgeon General released a Report on Physical Activity and Health, containing guidelines on the levels of moderate-to-vigorous physical activity people should engage in for health [53]. Since then public health agencies of many countries and international organisations have developed physical activity guidelines. The most recent World Health Organisation guidelines (henceforth, the 2020 WHO Guidelines) were published in November 2020. They recommend that, for good health, adults should engage in at least 150 to 300 minutes of moderate physical activity (e.g. brisk walking), 75 to 150 minutes of vigorous physical activity

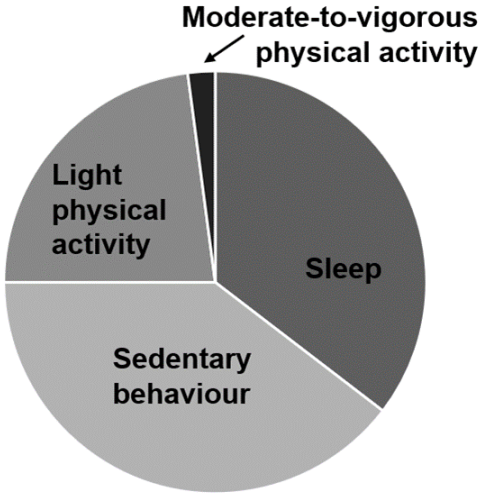
(e.g. running), or an equivalent combination each week [22]. Many public health and clinical interventions aim to change physical activity to improve health [54, 55].

1.4 From physical activity to 24-hour movement behaviours

Most people only spend a small proportion of their time in moderate-to-vigorous physical activity [56]. Time not spent in moderate-to-vigorous physical activity is divided between several other movement behaviours. These other behaviours include sleep, sedentary behaviour, and light physical activity. Moderate-to-vigorous physical activity is physical activity at $\geq$ 3.0 METs. Examples include may include walking the dog or cycling [28] (where reference MET values for a range of movement behaviours may be obtained from the Compendium of Physical Activities [28], a resource developed to systematically estimate and classify the energy cost of behaviours). There is some variation in how light physical activity is defined [57], but broadly it is physical activity occurring between 1.5 and 3.0 METs [22]. Examples include many tasks around the home or standing talking to someone [28]. Sedentary behaviour is waking behaviour with energy expenditure $\leq$ 1.5 METs in a sitting, lying or reclining posture with a low energy expenditure [58]. Examples include sitting while watching television or working at a computer [28]. Sleep is 'a reversible behavioral state of perceptual disengagement from and unresponsiveness to the environment' [59]. Figure 1.2 illustrates how these behaviours together make up the 24-hour day. These behaviours can be collectively described as '24-hour movement behaviours' [60].

There is a strong case to investigate health effects of all the movement behaviours which make up the 24-hour day, rather than just moderate-to-vigorous physical activity. Unsurprisingly, most evidence on health effects of movement behaviours is on

Figure 1.2: Sleep, sedentary behaviour, light physical activity, and moderate-to-vigorous physical activity together make up the 24-hour day.



moderate-to-vigorous physical activity, which has historically been easier to measure. However, there is no *a priori* reason to expect health effects of movement behaviours to relate only to moderate-to-vigorous physical activity: physiology and biomechanics vary substantially across different movement behaviours. In sedentary behaviour, most of the major muscles of the body are relaxed; static standing engages some key muscles, but still involves minimal energy expenditure; some forms of light physical activity involve higher levels of energy expenditure; while sleep involves very little energy expenditure and muscle activity, but performs a range of necessary restorative functions.

Better characterisation of the health effects of 24-hour movement behaviours may have important implications for public health and clinical practice. Potentially deleterious movement behaviour profiles are common, so even moderate relative risks could result in a substantial disease burden. Notably, in many societies high levels of sedentary

behaviour are prevalent [61, 62], and may be rising [62, 63]. Establishing health effects of behaviours beyond moderate-to-vigorous physical activity would also expand the range of behavioural targets for interventions and offer new opportunities to improve health [56, 64, 65]. Commenting on the recent Canadian 24-hour Movement Guidelines, Tremblay & Ross [66] note that evidence and recommendations covering a range of movement behaviours provides more options for those seeking to improve health, and allows people to modify behaviours in a way that fits with their preferences and constraints. Beyond simply expanding the range of behavioural targets, it has further been suggested that non-MVPA behaviours may represent better intervention targets than moderate-to-vigorous physical activity, although empirical evidence is lacking [64]. In particular, physical fitness, discretionary time and access to facilities, which are barriers to increasing MVPA, may be less relevant for other movement behaviours [64].

Initial evidence suggests that 24-hour movement behaviours beyond moderate-to-vigorous physical activity are associated with cardiovascular disease risk. Outside of moderate-to-vigorous physical activity, evidence is most developed for sedentary behaviour. Sedentary behaviour has attracted substantial research interest since the early 2000s [64, 67], and high levels of sedentary behaviour have been associated with higher CVD risk [22, 68]. Overall, the review conducted for the 2020 WHO Guidelines concluded that there was moderate certainty evidence for a dose-response association between sedentary behaviour and incident CVD [22]. Some evidence suggests higher levels of light physical activity may be associated with lower risk of CVD [69–72], although this finding is not consistent. Studies suggest a U-shaped association between sleep duration and CVD risk [73, 74].

However, substantial uncertainty remains about the associations between 24-hour movement behaviours and cardiovascular disease risk. The quantitative levels of

sedentary behaviour associated with higher risk remain unclear [68, 75]. The robustness of this association to adjustment for physical activity is also debated [68]. Evidence on light physical activity remains scarce and inconsistent [65, 69, 72, 76]. Again, the role of other behaviours in this association is unclear [77]. Evidence on sleep and CVD risk often remains disjoint from the study of other movement behaviours. Sleep is a neurobiological state rather than a state primarily defined by motion or posture [59], which presents challenges for measurement, analysis, and interpretation. Current evidence on device-measured movement behaviours and CVD will be reviewed in more detail in Chapter 2.

1.5 Measuring 24-hour movement behaviours

24-hour movement behaviours have been difficult to study historically, largely because of a lack of suitable measurement tools [68, 75, 78]. Until recently, most epidemiological evidence on movement behaviours used self-report to measure behaviours [78]. However, self-reported measures of movement behaviours have substantial disadvantages. Self-reported measures may be inaccurate and unreliable. Notably, recall and reporting bias may have a substantial impact e.g. participants may over-report time in physical activity because it is socially desirable. Meta-analysing studies comparing 'direct' measurements (of which accelerometry was the most common) with self-report, Prince *et al.* [79] found a mean correlation of 0.37 (standard deviation 0.25). While various factors may contribute to this low-to-moderate overall agreement, inaccuracy of self-report is likely important [79]. Studies also suggest error in self-reported measurements may be differential between groups of interest, making it more difficult to address analytically. For example, Sabia *et al.* [80] compared questionnaire- and accelerometer-assessed physical activity among 3,975 participants in the Whitehall II study, and found correlations between the two measurement meth-

ods were lower among participants of lower socio-economic status. Ferrari *et al.* [81] found questionnaire-based physical activity measurements were more accurate in men, in younger participants, and in those with a lower Body Mass Index. Importantly for this thesis, self-report is poorly suited to measuring some behaviours [78, 82]. For example, self-report may struggle to capture light physical activity, which often occurs in the course of daily life (e.g. self-care, tasks around the home) and may not be remembered [78]. Questionnaires sometimes measure proxies, rather than the behaviour itself [75]. An association between the proxy and disease may not reflect an effect of the behaviour. For example, television-watching is often used as a proxy for sedentary behaviour [83]. However, television-watching often co-occurs with snacking [75, 84]; shows a social gradient [75]; and may entail exposure to psychological stressors or advertising [75]. This makes it difficult to infer specific effects of sedentary behaviour from television-watching studies [75]. Importantly, these disadvantages of self-report substantially affect epidemiological inferences [85]. Ekelund *et al.* [85] conducted a harmonised meta-analysis to study the association between device-measured physical activity and mortality across several cohorts. They found that the maximal risk reduction observed using device-based measurement was more than twice that previously observed using self-reported measurements [85]. They attributed this difference to several of the factors described above, including more accurate and more comprehensive measurement [82]. Advantages and disadvantages of self-reported and device-based measurement are summarised in Table 1.1.

Table 1.1: Comparison of self-reported and device-based measurement of 24-hour movement behaviours.

	Self-reported measurement	**Device-based measurement**
Examples	International Physical Activity Questionnaire [86] Global Physical Activity Questionnaire [87]	Accelerometers (e.g. Axivity AX3, Actigraph GT3X+; may be worn at wrist, hip, thigh, or other placements) Other devices e.g. heart rate monitor, gyroscope (remain rare in practice)
Advantages	Historically easier to implement at scale	Now implemented at scale

	Large existing evidence base	Precise measurement of behaviour during measurement period Low risk of recall and reporting bias Complete: captures full range of 24-hour movement behaviours
Disadvantages	Some behaviours difficult to measure (e.g. light physical activity; non-intentional activity) Substantial recall and reporting bias, which may be differential Imprecise	Best processing methods and metrics still to be established Measurement error and within-person variation not currently understood Do not capture subjective aspects of behaviour (e.g. intentionality)

Device-based measurement is a promising alternative to self-report. The most widely-used device for measuring movement behaviours in epidemiological studies is the accelerometer. An accelerometer measures acceleration automatically, usually many times per second, and thus records information about the wearer's movement [88, 89]. While not free from measurement error [90], accelerometers are at lower risk of recall and reporting bias than self-report. They measure movement throughout the wear period, including behaviours that occur incidental to daily life and that may not be remembered [82, 91]. They can record behaviours that are difficult or impossible to measure by self-report, notably light physical activity [78].

Although accelerometers were used in epidemiological studies as early as the 1990s [92], early applications faced substantial challenges. Older accelerometers typically measured acceleration in a single axis of movement (e.g. vertical movement only) [88]. Due to memory limitations, they stored only aggregate metrics (e.g. minute-level summaries) [88]. Most early studies used accelerometers worn at the hip/waist, which were removed for sleep [89]. Removing accelerometers for sleep introduces burden for participants, and contributes to high levels of non-wear [89, 93, 94]. For example, the US National Health and Nutrition Examination Survey (NHANES) first used accelerometers in its 2003-2004 and 2005-2006 waves [89, 95]. These accelerometers were hip-worn and removed for sleep [95]. Only 25% of participants provided the requested seven valid days ($>$ 10 hours) of data; in some age groups up to 40% of the

sample had less than four valid days of data [89, 95]. Cost and logistical challenges meant that early accelerometer studies were also limited in scale [89]. Writing in 2016, Shephard & Tudor-Locke [96] concluded that most studies with device-based measurement had relatively small samples and insufficient follow-up to provide information on clinical outcomes. For example, the NHANES study is often cited as an example of early large-scale data collection, but even it had only $\sim 4,000$ participants aged over 40 years and with usable data [89, 97]. Evenson *et al.* [97] analysed the data after a median of 7 years of follow-up, at which time just 337 deaths had occurred (of which 107 were cardiovascular deaths), meaning it remains underpowered for robust epidemiological analysis.

Accelerometer technology has matured in recent years. Modern accelerometers, such as the Axivity AX3 and Actigraph GT3x, measure acceleration in three axes of movement. This provides more comprehensive measurement than the older uniaxial devices, particularly for irregular movement and non-hip placements [98]. Rather than piezo-electric sensors, they use the newer MEMS (microelectrical mechanical systems) sensors, which measure both static and dynamic acceleration [98]. They have lower cost, longer battery life, and more on-board memory than their predecessors. For example, the Axivity AX3 monitor advertises 14 days of battery life recording at 100 Hz (i.e. 100 readings per second) and 512 MB of on-board memory [99]. Supported by improvements in on-board memory and by demand for more granular data, manufacturers have developed devices providing raw data [89, 100]. The explosion in the quantity and granularity of data enables development of signal-processing methods with new capabilities [89]. Accelerometer placement has been another important area of innovation, with new placements (e.g. wrist) enabling low-burden, 24-hour wear [93]. 24-hour wear protocols enable studies of sleep alongside other movement behaviours, and, importantly, also typically have much higher wear-time compliance. In contrast to the earlier 2003-2004 and 2005-2006 waves of NHANES described above,

the 2011-2012 wave used a 24-hour protocol with a wrist-worn device and achieved much better wear-time compliance: between 70 and 80% of participants provided at least six days of data with at least 18 hours of wear time, and compliance was fairly consistent across age groups [89].

With improvements in technology, accelerometers are now commonly used to measure movement behaviours in large prospective cohort studies [101]. Most commonly, a 7-day wear period is used. Longer wear periods are desirable in order to achieve more reliable measurements of a person's typical behaviours [102]. However, this theoretical benefit of a longer wear period must be set against practical constraints, such as the need to charge the device, and diminishing analytical returns to longer wear [102]. Studies have shown that a 7-day wear period is feasible in practice [93], and can reliably measure a range of different device-based metrics in adults [103] and older adults [102].

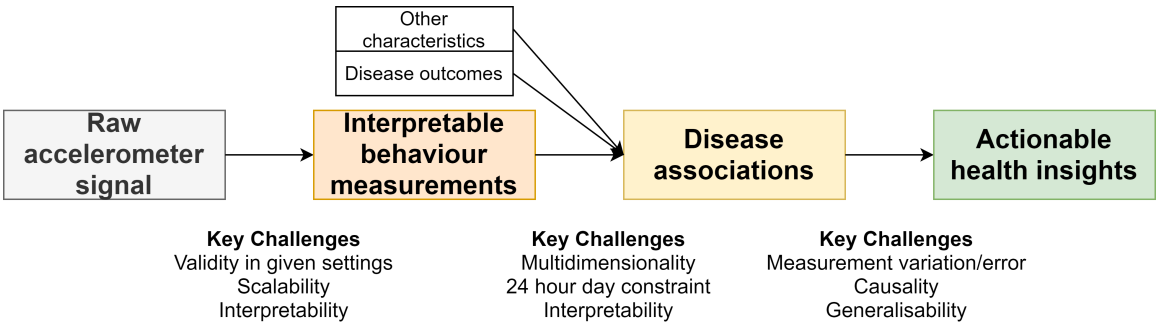
1.6 Challenges

24-hour accelerometry in large prospective cohort studies opens opportunities for understanding health effects of 24-hour movement behaviours, but challenges persist in how to best use this emerging data source.

Several stages are required to go from raw accelerometer data to addressing epidemiological questions. First, measurements of time in different movement behaviours are derived from the raw triaxial acceleration signal. Next, 24-hour movement behaviours are associated with disease. Finally, where associations are observed, these are interpreted in light of potential biases, such as information bias, selection bias, reverse causality, and residual and unmeasured confounding. Challenges at each of these stages limit our understanding of the relationship between 24-hour movement

behaviours and risk of incident cardiovascular disease. The following sections expand on each research stage, highlighting key challenges and emerging possibilities to address those challenges. Figure 1.3 illustrates these research stages.

Figure 1.3: Research pipeline and challenges in using accelerometer data to understand effects of 24-hour movement behaviours on disease risk.



1.6.1 Classifying behaviours

The dominant approach for deriving device-based measurements of movement behaviours from accelerometer data is to use cut-points. Under a cut-point approach, an epoch (time window) of accelerometer data is classified as a particular intensity of activity based on whether average acceleration or 'counts' (an alternative measure of movement produced by some devices) in that window exceed a threshold value (the cut-point). However, cut-point based behaviour classification has several important disadvantages. Notably, classification is based on a single aspect of the accelerometer signal [104]. Also, while device measurements of behaviours derived using cut-points may be interesting in their own right, these phenotypes sometimes correspond poorly to the movement behaviours they are intended to capture [105]. Measurements are also sensitive to processing choices [106]. For example, one study found that prevalence of meeting physical activity guidelines in a sample of children varied between 8 and 96% based on device-measured data using different accelerometer placement and

processing choices [107]. Different calibration studies have resulted in divergent estimates of appropriate cut-points, leaving conclusions potentially contingent on choice among these cut-points [89]. These challenges and inconsistencies make it difficult to translate between different studies using cut-point approaches and to translate them into public health guidelines and practice [106].

A promising method in the movement behaviour literature is machine-learning for behaviour classification. Most commonly, forms of supervised learning are used, whereby labelled training data is used to develop a classifier. A range of machine-learning-based approaches have been successfully applied to behaviour classification from accelerometer data. Some models classify behaviours directly [108], while others predict energy expenditure, which may be used to determine activity intensity [109, 110]. Farrahi *et al.* [108] identified studies using a range of machine-learning methods, including logistic regression, Artificial Neural Networks, K-Nearest Neighbours, Bayesian Networks, Support Vector Machines, and Random Forests.

However, the data used to develop classifiers often does not reflect the settings in which they will be applied. Studies developing classifiers have predominantly relied on data collected in laboratory or semi-structured settings [108]. The range of behaviours observable in free-living is much greater than in the laboratory or in semi-structured settings [111]. Movement in free-living is also more irregular and complex than in the laboratory: for example, walking over uneven ground is less regular than walking on a treadmill [104]. Therefore, classification methods developed using laboratory data may perform poorly when applied in free-living data [104, 112]. Moreover, in (semi-)structured settings, participants perform behaviours in a prescribed sequence, meaning temporal sequence is by design uninformative. However, in free-living data temporal sequence may be highly informative, and incorporating it into modelling may improve real-world performance [113, 114]. Reviewing the literature on machine-

learning-based behaviour classification, Narayanan *et al.* [113] note that studies where a laboratory-trained classifier is applied in free-living often show substantial drops in performance, but that including free-living data in the training data can mitigate this. A further challenge for machine-learning-based behaviour classification is that it has largely been studied as a standalone problem: associations of machine-learning-based phenotypes with disease have rarely been studied. One reason for this is models not being available for reuse: Clevenger *et al.* [115] found that of 162 studies developing physical activity intensity or energy expenditure prediction models for accelerometer data, 107 did not make their models available for reuse. Notably, often available models were those which could be most easily shared in text form, including cut-points and regression equations, not machine-learning-based methods [115].

1.6.2 Associating 24-hour movement behaviours with disease

'Traditional' statistical methods for analysing the association between movement behaviours and disease outcomes commonly involve categorising participants by level of sedentary behaviour, light physical activity or moderate-to-vigorous physical activity (e.g. by quartile) and associating these categorised variables with the outcome [69, 71, 97, 116–120]. Models treating the behaviour variable as a continuous term, in both linear and non-linear models, are also used [69–71, 117, 121].

However, these methods all fall within a 'single behaviour paradigm' i.e. they associate a single behavioural exposure variable with an outcome. They consider higher levels of one behaviour, holding fixed other characteristics of the individual (which may include time spent in other behaviours). This is conceptually problematic: higher levels of one behaviour entail lower levels of some other behaviours, meaning 24-hour movement behaviours are fundamentally interdependent [122]. Modelling behaviours as freely-varying individual variables does not allow for this interdependence [123].

One approach to address the multivariable, constrained nature of 24-hour movement behaviours is 'linear isotemporal substitution' [124][2]. Under a linear isotemporal substitution approach, all but one of the behaviour variables are included in the model and the model is also adjusted for the total time in all measured behaviours [124]. If total time in all measured behaviours is fixed (e.g. if the full 24-hour day is measured), it should be omitted, but a similar method can be followed [126]. The coefficient of a behaviour X is interpreted as the effect of substituting the left-out behaviour with behaviour X [124, 126]. This approach engages with the particular nature of movement behaviour data: coefficients are interpreted in terms of time substitutions, not unilateral differences in a single behaviour [124]. However, linear isotemporal substitution only partially accounts for the structure of 24-hour movement behaviour data [127, 128]. It continues to treat movement behaviour variables as if they come from multidimensional real space and can vary freely within it (e.g. exceeding the 24-hour day constraint), except at the final interpretation stage [129]. Assumptions can also be poorly suited to the distribution of movement behaviour data (e.g. linearity of associations on the standard scale) [129].

Recently, several authors have argued that the multidimensional, constrained nature of movement behaviour data means it is best understood as *compositional data* [122, 127]. Compositional data is multivariable, positive, real-valued data where relative, as opposed to absolute, values of the variables are relevant. Perhaps more intuitively, this means it is data that can be expressed as proportions of a whole without losing relevant information. In the case of movement behaviours, an individual divides their 24 hours each day between different intensities of activity, sedentary behaviour and sleep [130, 131]. As total time cannot be increased but only allocated differently

[2]The term 'isotemporal substitution' can be ambiguous: it is used to refer to the specific approach of Mekary *et al.* [124], but has also been used more generally to describe approaches where the effect of interest is substitution of time in one behaviour for time in other behaviours (with different underlying analytic approaches) [125]. To differentiate uses in this thesis, the approach of Mekary *et al.* [124] will be referred to as 'linear isotemporal substitution'.

between behaviours, only the relative values of the variables, not their absolute values, matter [127]. Traditional approaches are designed for univariable, unconstrained data carrying absolute information, meaning problems can arise when they are used to analyse compositional data. One challenge is identifying the effect of interest. When some variables in a composition are higher, other variables must be relatively lower: more of the day spent in physical activity entails less spent in other behaviours [131, 132]. Therefore, rather than the effect of 'more physical activity', the effect of interest may be the effect of more physical activity and less sedentary behaviour, more physical activity and less sleep, or more physical activity and less of all other behaviours [131, 132]. Relatedly, it is not obvious how to appropriately conceptualise and account for associations between variables making up the composition. Variables within the composition may have associations both with the outcome and with one another. Appropriately estimating the effect of interest therefore requires somehow accounting for all the variables making up the composition. The usual approach with non-compositional data would be to estimate associations with one behaviour while holding all others fixed, achieved statistically by adjusting for them within a regression analysis. However, in the compositional context this is impossible: as highlighted above, relatively more time in one behaviour entails relatively less time in one or several other behaviours [127]. Another challenge is to appropriately account for the particular statistical properties of compositional data. For example, the constraints on compositional data affect the distributions that are commonly seen [133, 134].

The most prominent approach designed to analyse compositional data is known as Compositional Data Analysis (CoDA). CoDA has been used extensively in movement behaviour epidemiology [127, 135], as well as in other fields including geology [136], the study of the microbiome [137, 138], and chemistry [139]. A CoDA approach aims to analyse and model the relative values of variables. This is typically achieved by using particular sets of log-transformed ratios of variables to perform the analysis. By

analysing 24-hour movement behaviour data with methods appropriate to its compositional structure, the challenges highlighted above may be addressed. Although CoDA is increasingly widely used in movement behaviour epidemiology, a systematic review in 2020 found large evidence gaps still remain [140]. In particular, most studies using CoDA have been cross-sectional, and there is a lack of evidence using this method to associate movement behaviours with incident disease outcomes [140, 141].

1.6.3 Interpreting observed associations

As for many exposures, the extent to which observational associations can be interpreted as causal effects of movement behaviours on disease outcomes is not always clear. Potential biases include information bias (in particular due to measurement error in the exposure), selection bias, reverse causality, and residual confounding. Measurement error is the difference between the true underlying value of a variable and a measurement of that variable. Some measurement error manifests as within-person variation, where the measured value fluctuates around the true average value from measurement to measurement. Measurement error may bias observed associations (known as 'information bias'). Selection bias can mean the cohort studied differs from the target population such that associations do not generalise, and can manifest as collider bias, whereby selection caused by both the exposure and the outcome induces an association between these variables which is not present in the overall population [142, 143]. Another source of bias is different causal explanations for observed associations, including reverse causality and confounding. Reverse causality, whereby disease which has not yet been detected clinically causes changes in movement behaviours, is an important possibility in the context of movement behaviours [144, 145]. Populations with chronic disease have been shown to have lower activity levels than disease-free populations [146], and, in particular, incipient CVD

may result in a loss of functional capacity and a preference for sitting [144]. Confounding, where a third factor causes both differences in movement behaviours and disease, may underlie observed associations. While confounders are typically adjusted for, important confounders may not be measured or may be measured poorly, leading to under-adjustment and residual confounding. In recent years, several methods have emerged which go beyond qualitative discussions to quantitatively address bias in observational studies [147].

1.7 Thesis aims and objectives

As outlined in Section 1.4, existing evidence suggests that 24-hour movement behaviours are associated with CVD risk. 24-hour movement behaviours may represent an important behavioural target for public health and clinical interventions in the future. However, while physical activity epidemiology is well-established, evidence on 24-hour movement behaviours remains limited. It is difficult to measure some behaviours using questionnaires; accelerometer measurement of movement behaviours in large epidemiological cohort studies makes it possible to study 24-hour movement behaviours with new detail and precision. The aim of this thesis is to investigate effects of device-measured 24-hour movement behaviours on risk of incident CVD. Three specific objectives support this overall aim:

1. Refine and apply state-of-the-art machine-learning-based behaviour classification methods to derive measurements of movement behaviours in large wrist-worn accelerometer datasets.

2. Study the association between movement behaviours and risk of incident cardiovascular disease, using methods appropriate to the compositional nature of movement behaviour data.

3. Quantitatively assess the impact of known sources of bias on estimates of the association between movement behaviours and risk of incident cardiovascular disease.

1.8 Thesis outline

The outline of the remainder of this thesis is:

Chapter 2 reviews existing evidence on the association between device-measured movement behaviours and CVD outcomes.

Chapter 3 introduces UK Biobank, with a focus on its wearable-device-based measurements of movement behaviours.

Chapter 4 refines state-of-the-art machine-learning-based behaviour classification methods and applies them to derive machine-learned measurements of movement behaviours from wrist-worn accelerometer data.

Chapter 5 studies the association between device-measured 24-hour movement behaviours and cardiovascular disease, using methods appropriate to the compositional nature of movement behaviour data.

Chapter 6 investigates the possible impact of several potential sources of bias on estimates of the association between device-measured movement behaviours and risk of incident cardiovascular disease.

Chapter 7 reviews the findings of this thesis, situates them in the broader context of current evidence and public health, and highlights implications for future research.

Appendix A contains supplementary tables and figures.

Appendix B describes availability of code and data.

Appendix C introduces `epicoda`, an open-source R package developed to facilitate use of Compositional Data Analysis in epidemiological applications.

Appendix D further develops the rationale for methods used in Chapter 6 to investigate within-person variation in compositional data.

1.9 Summary

Cardiovascular disease is important for population health. Physical activity is a well-established modifiable behavioural risk factor for cardiovascular disease. There is scope to broaden epidemiological studies to consider movement behaviours throughout the 24-hour day. 24-hour movement behaviours may represent a useful target for public health, but historical difficulties in studying them mean evidence remains limited. Accelerometer-based measurement has enabled investigation of 24-hour movement behaviours at scale for the first time. However, 24-hour movement behaviours still present some challenges for measurement, analysis, and interpretation. This thesis will investigate the relationship between 24-hour movement behaviours and risk of incident cardiovascular disease, addressing some of these challenges. The next chapter, Chapter 2, begins by reviewing the current literature on device-measured movement behaviours and risk of incident cardiovascular disease.

Chapter 2: Literature review on device-measured 24-hour movement behaviours and risk of incident cardiovascular disease

2.1 Introduction

Chapter 1 briefly described how movement behaviours have been associated with cardiovascular disease risk. It also described how and why device-based measurement is an important development in epidemiological research into movement behaviours, enabling more accurate and precise measurement, particularly of movement behaviours beyond moderate-to-vigorous physical activity. The aim of this thesis, to investigate effects of 24-hour movement behaviours on risk of incident cardiovascular disease, was described, alongside challenges in measurement, analysis, and interpretation that must be addressed. After the general introduction provided by Chapter 1, both to associations of movement behaviours with cardiovascular disease and to challenges of device-measured 24-hour movement behaviour data, it is useful to consider evidence on device-measured movement behaviours and risk of incident cardiovascular disease in more detail. Limiting considerations to studies using device-based measurement enables more in-depth consideration of individual studies and methodology. It was

also noted in Chapter 1 that measurements using self-report and device-based measurement are quite different [79], suggesting they should be interpreted separately. Considering evidence on device-measured movement behaviours and incident CVD provides important context for the work of this thesis and its interpretation, and it also helps to identify methodological strengths and weaknesses, to further inform this work.

This chapter aims to review and summarise the existing evidence on 24-hour movement behaviours and risk of incident cardiovascular disease (CVD) or cardiovascular disease mortality. It will consider evidence on individual movement behaviours, as well as evidence taking an integrated approach.

2.2 Methods

2.2.1 Scope

Studies were reviewed if they investigated the association between one or more device-measured movement behaviours and either incident cardiovascular disease events (including studies limited to some cardiovascular disease subtypes e.g. coronary heart disease) or cardiovascular disease mortality in adults. Studies using self-reported measurements of moderate-to-vigorous physical activity, light physical activity, sedentary behaviour, and sleep were beyond the scope of this review. Studies of phenotypes other than time in particular behaviours (e.g. overall energy expenditure, steps per day [148]) were also not reviewed. Although evidence on movement behaviours and cardiovascular disease can also come from studies associating movement behaviours with intermediate phenotypes or surrogate endpoints (e.g. cardiometabolic biomarkers), these were beyond the scope of this work. Only prospective studies (i.e. con-

sidering incident disease) were considered. Both primary studies and systematic reviews/meta-analyses were considered, although results had to be presented separately for studies using device-based measurement only. Exclusion criteria included restriction to a patient population (e.g. stroke patient cohort) and being written in a language other than English. Inclusion and exclusion criteria are described in Table 2.1.

Table 2.1: Inclusion and exclusion criteria for the literature review.

Inclusion criteria
Association between: One or more *device-measured* movement behaviours AND either incident cardiovascular disease (including subtypes) or cardiovascular disease mortality
Prospective design
Adult population (> 18 years)
Primary study OR systematic review/ meta-analysis with results disaggregated by measurement method
Exclusion criteria
Patient population
Non-English language (practical constraint)

2.2.2 Strategy

Although a formal systematic review process was not followed, the review was informed by a systematic search. This search was performed on MEDLINE on 2022-03-07. The search terms aimed to identify terms in the title or abstract indicating: device-based measurement, a CVD outcome, at least one movement behaviour exposure, and a prospective cohort design (included reference to 'incident' disease). Full search terms are shown in Table 2.2. Titles were screened to identify relevant studies, and abstracts screened where relevance could not be determined from the study title alone.

Table 2.2: Search terms for literature search on device-measured movement behaviours and risk of incident cardiovascular disease.

Search terms	
	((activity AND device) OR (acceleromet* OR actigraph* OR activity monitor OR pedometer OR inclinometer))
AND	(cardiovascular disease OR cardiovascular events OR cerebrovascular OR stroke OR heart disease OR infarction OR mortality)
AND	(physical activity OR sedentary OR sleep OR movement behav* OR physical behav* OR time us*)
AND	(prospective OR incident OR cohort)

First, results of the search process will be described. Methods of the included studies will then be reviewed, including their approaches to measure movement; to classify movement behaviours; to associate movement behaviours with cardiovascular disease or cardiovascular disease mortality; and to interpret results in light of potential biases. Findings across studies on sedentary behaviour, light physical activity, and moderate-to-vigorous physical activity will then be reviewed.

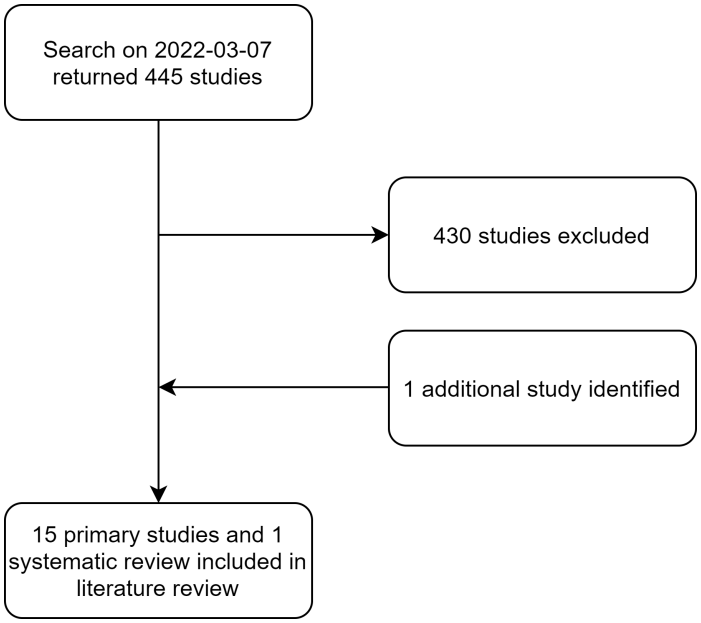
2.3 Findings

2.3.1 Systematic search

The systematic search returned 445 results, and one additional relevant study was identified ([149]). 15 primary studies and 1 systematic review (Qiu *et al.* [72]) were identified as falling within the review scope. Figure 2.1 illustrates selection of studies.

Almost all excluded studies were excluded as they did not fall within the scope of the review (Section 2.2.1). One study ([150]) was excluded as it analysed the same dataset as an included study ([151]), but drew additional distinctions between occupational and leisure-time moderate-to-vigorous physical activity. As results for moderate-to-vigorous physical activity in this cohort were reported in [151], this was considered

Figure 2.1: Flow diagram showing inclusion of studies in the literature review.



beyond the intended scope. Ramakrishnan *et al.* [118] and Walmsley *et al.* [1] were also excluded, as work contributing to this thesis was foundational to these studies.

2.3.2 Studies identified

Six cohorts were identified in which the association between some or all of the device-measured movement behaviours and incident cardiovascular disease events had been investigated. These were the Attitude, Behavior and Change study (ABC), the British Regional Heart Study (BRHS), the European Prospective Investigation into Cancer-Norfolk (EPIC-Norfolk), the Healthy Ageing Initiative (HAI), the Objective Physical Activity and Cardiovascular Health study (OPACH; sub-study of the Women's Health Initiative, WHI), and the Whitehall II study. Four cohorts were identified in which the association between some or all of the device-measured movement behaviours and cardiovascular disease mortality had been studied. These were ABC, National Health and Nutrition Examination Survey (NHANES), the OPACH study, and the

Osteoporotic Fractures in Men Study (MrOS). CVD mortality was typically studied as a secondary outcome (with all-cause mortality the primary outcome).

2.3.3 Study design

2.3.3.1 Accelerometry in prospective cohorts

Table 2.3 summarises methodology of the included studies, including implementation of accelerometry.

The earliest data collection in the identified studies was in the ABC study. This is a study of a nationally representative sample of Swedish adults. In 2001-2002, participants wore an Actigraph 7164 on the lower back in a 7-day waking-hours-only protocol [151]. Another important early example of large-scale accelerometer-based data collection was in the 2003-2004 and 2005-2006 waves of NHANES, a representative study of US adults [95, 97, 149]. In these waves, a hip-worn Actigraph 7164 and a 7-day, waking-hours-only wear protocol were used [97, 149]. EPIC-Norfolk is a study of middle-to-older aged UK adults [70]. Between 2004 and 2016, movement behaviours were measured using either a hip-worn Actigraph GT1M or a hip-worn Actigraph GT3X+ with a 7-day, waking-hours-only wear protocol [70]. MrOS is a study of older men in the USA, and between 2007 and 2009, movement behaviours were measured using an upper-arm-worn SenseWear Pro Armband with a 7-day, 24-hour wear protocol [120]. The BRHS is a study of men in the UK, which took accelerometer measurements between 2010 and 2012 (at which time participants were aged 71-92 years) [71]. They used a hip-worn Actigraph GT3X with a 7-day, waking-hours-only wear protocol [71]. The HAI is a study of adults aged 70 in Umeå, Sweden [121]. Between 2012 and 2017, accelerometer measurement was carried out using a hip-worn Actigraph GT3X+ with a 7-day, waking-hours-only wear protocol [121]. The OPACH

study is a sub-study of the Women's Health Initiative, a study of older women in the USA. Between 2012 and 2014, participants wore a hip-worn Actigraph GT3X+ in a 7-day, 24-hour wear protocol [69, 117, 119, 152]. Whitehall II is a study of people who were civil servants based in London when recruited in 1985-1988 (participants aged 60-83 years by the time of accelerometer wear) [76, 153]. It was the only study included in the literature review to use a wrist-worn accelerometer; specifically, a wrist-worn GENEActiv with a 9-day, 24-hour wear protocol [76, 153]. Accelerometer data collection took place in 2012-2013 [76, 153].

Table 2.3: Methods of primary studies in the literature review.

Cohort	Study	Population	n	Wear year	Device	Wear	Movement behaviours[a]	Classification approach	Analytic approach
CARDIOVASCULAR DISEASE									
Attitude, Behavior and Change study	Dohrn *et al.* [151]	representative sample of Swedish adults	1,176	2001-2002	lower-back-worn Actigraph 7164	7 day, waking-hours	SB, LIPA, MVPA	count-based cut-point	thirds of each behaviour individually
British Regional Heart Study	Jefferis *et al.* [71]	older men in the UK	1,181	2010-2012	hip-worn Actigraph GT3x	7 day, waking-hours	SB, LIPA, MVPA, bouted PA, sedentary breaks	count-based cut-point	quarters of each behaviour individually; linear, quadratic and spline approaches
EPIC-Norfolk	Dempsey *et al.* [70]	middle-to-older aged UK adults	5,585	2004-2016	hip-worn Actigraph GT1M/GT3X+	7 day, waking-hours	SB, LIPA, MVPA	count-based cut-point	splines
Healthy Ageing Initiative	Ballin *et al.* [121]	adults aged 70 in Umea, Sweden	3,343	2012-2017	hip-worn Actigraph GT3X+	7 day, waking-hours	SB, LIPA, MPA, MVPA	count-based cut-point	linear term
OPACH sub-study of WHI	LaCroix *et al.* [69]	older women in the US	5,861	2012-2014	hip-worn Actigraph GT3X+	7 day, 24-hour	LIPA, MVPA	count-based cut-point (study-specific)	quarters of LIPA, MVPA; splines
OPACH sub-study of WHI	Bellettiere *et al.* [117]	older women in the US	5,638	2012-2014	hip-worn Actigraph GT3X+	7 day, 24-hour	SB, pattern metrics	count-based cut-point (study-specific)	quarters of SB, unadjusted and adjusted for MVPA; splines
OPACH sub-study of WHI[b]	Nguyen *et al.* [152]	older women in the US	5,416	2012-2014	hip-worn Actigraph GT3X+	7 day, 24-hour	'daily life movement' (standing up while moving in a confined space)	machine-learning approach	quarters; splines

Cohort	Study	Population	n	Wear year	Device	Wear	Movement behaviours[a]	Classification approach	Analytic approach
Whitehall II	Yerramalla *et al.* [76]	civil servants based in London (when recruited in 1985 - 1988)	3,319	2012-2013	wrist-worn GENEActiv	9 day, 24-hour	the waking-day movement behaviour composition (MVPA, LIPA, SB)	raw-data cut-point	Compositional Data Analysis
Whitehall II	Yerramalla *et al.* [153]	civil servants based in London (when recruited in 1985 - 1988)	3,321	2012-2013	wrist-worn GENEActiv	9 day, 24-hour	SB, pattern metrics	raw-data cut-point	linear term; splines
CARDIOVASCULAR DISEASE MORTALITY									
Attitude, Behavior and Change study	Dohrn *et al.* [116]	Swedish adults aged 50+ years at follow-up	851	2001-2002	lower-back-worn Actigraph 7164	7 day, waking-hours	SB, LIPA, MVPA	count-based cut-point	thirds of each behaviour individually
Attitude, Behavior and Change study	Dohrn *et al.* [154]	Swedish adults aged 50+ years at follow-up	851	2001-2002	lower-back-worn Actigraph 7164	7 day, waking-hours	SB, LIPA, MVPA	count-based cut-point	linear isotemporal substitution
NHANES	Schmid *et al.* [149]	representative sample of US adults aged 50-85 years	3,702	2003-2004, 2005-2006	hip-worn Actigraph 7164	7 day, waking-hours	SB, PA	count-based cut-point	thirds of behaviours; linear isotemporal substitution
NHANES	Evenson *et al.* [97]	representative sample of US adults aged >40 years	3,809	2003-2004, 2005-2006	hip-worn Actigraph 7164	7 day, waking-hours	SB, sedentary bouts, LIPA, MPA, VPA, MVPA, moderate-to-vigorous bouts, meeting guidelines	count-based cut-point	categories of different behaviours

Cohort	Study	Population	n	Wear year	Device	Wear	Movement behaviours[a]	Classification approach	Analytic approach
OPACH sub-study of WHI	LaMonte *et al.* [119]	older women in the US	6,382	2012-2014	hip-worn Actigraph GT3X+	7 day, 24-hour	low light intensity physical activity, high light intensity physical activity, MVPA	count-based cut-point (study-specific)	thirds of each behaviour individually
Osteoporotic Fractures in Men study	Ensrud *et al.* [120]	older men in the US	2,918	2007-2009	upper-arm-worn SenseWear Pro Armband	7 day, 24-hour	waking time, SB, LIPA, MVPA	proprietary algorithm	quarters of each behaviour individually

[a] 'Sedentary time (ST)/sedentary behaviour (SB)' and 'LPA/LIPA' were considered interchangeable. The latter term was used throughout in each case.
[b] The study of Nguyen *et al.* [152] also includes analyses on CVD mortality.
Abbreviations used are: SB - sedentary behaviour; PA - physical activity; LIPA - light intensity physical activity; MVPA - moderate-to-vigorous physical activity; MPA - moderate physical activity; VPA - vigorous physical activity; DLM - daily life movement; OPACH - Objective Physical Activity and Cardiovascular Disease Health study; WHI - Women's Health Initiative; NHANES - National Health and Nutrition Examination Survey; EPIC - European Prospective Investigation into Cancer; CHD - coronary heart disease; CVD - cardiovascular disease; MI - myocardial infarction.

2.3.3.2 Approaches to classifying movement behaviours

The studies focussed on different movement behaviours (again, see Table 2.3 for a summary). Some studies were specifically designed to investigate a particular behaviour (e.g. sedentary behaviour [117]), whereas others considered several behaviours [70, 76]. Sedentary behaviour/time, light physical activity, and moderate-to-vigorous physical activity were most commonly studied. However, some studies drew further distinctions, including between moderate physical activity and vigorous physical activity [97], and between low-light intensity physical activity and high-light intensity physical activity [119]. None of the studies considered device-measured sleep.

Most of the identified studies used a cut-point-based approach to measure time in different behaviours. However, the cut-points used varied substantially between studies. The main reason for this was the highly divergent estimates of appropriate cut-points given by different calibration studies [89]. For example, using an Actigraph GT3X+, Dempsey *et al.* [70] identified sedentary time in the EPIC-Norfolk study with a cut-point of 100 counts per minute, whereas Bellettiere *et al.* [117] in OPACH used a cohort-specific cut-point of 18 counts per 15 seconds. A wide range of cut-points were used for the division between light physical activity and moderate-to-vigorous physical activity with an Actigraph worn at the hip: 2,020 counts per minute [97], 1952 counts per minute [121, 155], 1,040 counts per minute [71], and 760 counts per minute [97]. Notably, a study-specific calibration study was conducted in the OPACH study [156].

One study ([120]) used a proprietary algorithm to identify behaviours. One recent study identified 'daily life movement' using a machine-learning approach [152].

2.3.3.3 Analytic approaches

A wide range of analytic approaches were used in the studies (Table 2.3). The most common approach was the 'traditional' approach of categorising participants by their level of sedentary behaviour, light physical activity, or moderate-to-vigorous physical activity (e.g. by quartile) and then modelling the association between these categorised variables and CVD outcomes [69, 71, 97, 116, 117, 119, 120, 152]. Models using a linear term for the behaviour variable were also used, although less frequently. For example, this was used alongside other approaches by Jefferis *et al.* [71], LaCroix *et al.* [69], and Bellettiere *et al.* [117], and was the main approach in the study of Ballin *et al.* [121]. Splines were also widely used. For example, LaCroix *et al.* [69], Bellettiere *et al.* [117], Nguyen *et al.* [152], Jefferis *et al.* [71], and Yerramalla *et al.* [153] applied cubic regression splines modelling to better understand the dose-response association. Dempsey *et al.* [70] used a spline-based approach as their main approach. Studies broadly concurred that there are important interrelationships between the time spent in different behaviours, but there were big differences in how they approached those interrelationships. Many studies presented primary results from analyses unadjusted for other behaviours [69, 71, 117, 119–121, 151], although often also conducted additional analyses adjusted for some other behaviours [69, 71, 117]. Other studies considered their primary results to be those adjusted for other behaviours [97]. Nevertheless, whether or not they adjust for other behaviours, these approaches can be grouped as falling under the 'single behaviour paradigm' described in Chapter 1: they primarily consider higher levels of a single behaviour, and approach inclusion or exclusion of other behaviours in the model from the perspective of the behaviour of primary interest.

A few studies used approaches which explicitly move away from this 'single behaviour paradigm' and engage with the multidimensional, constrained structure of movement

behaviour data. Dohrn *et al.* [154] and Schmid *et al.* [149] applied a linear isotemporal substitution approach to mutually adjust waking behaviours and assess associations with replacing one behaviour with another. Yerramalla *et al.* [76] studied the movement behaviour composition within the day using a Compositional Data Analysis approach.

2.3.3.4 Approaches to assessing bias

Studies typically discussed several possible sources of bias, including within-person variation and error in measurements, selection bias, reverse causality, and residual confounding. However, addressing sources of bias quantitatively was rarer. One exception was exclusion of some period of follow-up and/or adjustment or exclusion based on disease status to address reverse causality, which was widely performed (although where CVD or CVD mortality was a secondary endpoint, often this analysis addressed another endpoint). However, studies were often limited by the overall length of follow-up or number of events, meaning the removed period was as short as 6 months in several cases [69, 117, 119]. Within-person variation and measurement error were mentioned by studies in two main respects: whether appropriate cut-points were used, and whether the wear period was sufficient to provide a stable measurement. For example, Dohrn *et al.* [116], Yerramalla *et al.* [76], and Evenson *et al.* [97] performed sensitivity analyses using different cut-points. LaCroix *et al.* [69] argued that using a cut-point calibrated within the study (and therefore in the population of interest) was a strength of their study. Yerramalla *et al.* [76] cited studies suggesting that 4 to 6 days of recording including weekend days provides a good measure of habitual physical activity. While many studies discussed aspects of selection into the cohort or generalisability of results, this was only addressed quantitatively in some analyses in the NHANES cohort (where weights relating to

probability of selection are available) [97, 149]. Although several studies acknowledged potential for residual confounding, this was not addressed quantitatively.

2.3.4 Study findings

2.3.4.1 Sleep

None of the reviewed studies associated device-measured sleep with risk of incident cardiovascular disease.

2.3.4.2 Sedentary behaviour

In analyses unadjusted for other movement behaviours, studies generally found associations between higher levels of sedentary behaviour and a higher risk of incident cardiovascular disease. In the British Regional Heart Study, a study of older men in the UK, Jefferis *et al.* [71] found a confounder-adjusted hazard ratio (HR) for CVD of 1.87 (1.03, 3.38) in the highest versus lowest quarter of SB. Modelling SB as a continuous variable, each additional 30 min/day in SB was associated with a HR of 1.09 (1.00, 1.19). In EPIC-Norfolk, Dempsey *et al.* [70] found that higher levels of sedentary time were associated with higher CVD risk (up to a HR of roughly 2 at around 12.5 hours/day sedentary). In the Healthy Ageing Initiative, a study of Swedish adults enrolled at age 70 years, Ballin *et al.* [121] found a HR of 1.37 (1.12, 1.68) for stroke or myocardial infarction per 1 hour/day increment of sedentary time. In OPACH, a study of older women in the US, Bellettiere *et al.* [117] found a HR for CVD of 1.69 (1.27, 2.26) in the highest versus lowest quarter of sedentary time. Results for CHD were similar but more pronounced (see Table 2.4). They found a linear dose-response association, with each additional hour per day of sedentary time asso-

ciated with a HR of 1.12 (1.05, 1.19) [117]. In Whitehall II, a study of former British civil servants now reaching older age, Yerramalla *et al.* [153] found a HR of 1.20 (1.06, 1.37) for cardiovascular disease per standard deviation of SB (100.2 min/day). One additional study found estimates which were directionally consistent but imprecise: in the ABC study, Dohrn *et al.* [151] found a HR for the highest versus lowest third of sedentary time of 1.41 (0.91, 2.20).

In studies which adjusted for other movement behaviours, SB mostly showed limited or null associations with incident CVD. However, there was notable inconsistency between studies. Associations reported above by Jefferis *et al.* [71] in the British Regional Heart Study were entirely attenuated by adjustment for MVPA and mobility disability (HR of 1.87 attenuated to 0.94 (0.45, 1.94)). Associations observed by Dempsey *et al.* [70] in EPIC-Norfolk were also almost fully attenuated by adjustment for MVPA and other covariates. The HR per SD of SB of 1.20 observed by Yerramalla *et al.* [153] in Whitehall II was attenuated to 1.02 (0.88, 1.19) by adjustment for reaching recommended levels of MVPA. In a separate study in the same cohort, Yerramalla *et al.* [76] studied the association between the waking-day movement behaviour composition (moderate-to-vigorous physical activity, light physical activity, sedentary behaviour) and cardiovascular disease using a Compositional Data Analysis approach. In that study, they found no evidence of lower CVD risk with higher light physical activity at the expense of SB [76]. In contrast, associations observed by Bellettiere *et al.* [117] in OPACH persisted after adjustment for MVPA, but were reduced in magnitude: for example, the HR of 1.69 for CVD in the highest quarter was attenuated to 1.53 (1.09, 2.14) after adjustment for MVPA, with a less clear dose-response association.

Associations with CVD mortality showed similar patterns, but estimates were often based on numbers of cardiovascular deaths too low to precisely estimate HRs, perhaps

because cardiovascular mortality was often a secondary outcome. Results reported here are unadjusted for other movement behaviours unless stated. In the ABC study, Dohrn *et al.* [116] found a HR for the highest versus lowest third of 7.61 (2.02, 28.65), although it was notable that this was based on only 24 cardiovascular deaths. Studying this association using a linear isotemporal substitution approach to model time substitutions and mutually adjust associations (see Chapter 1), Dohrn *et al.* [154] found that replacing 30 min/day of sedentary time with light physical activity was associated with a HR of 0.76 (0.63, 0.92) and replacing 10 min/day of sedentary time with MVPA was associated with a HR of 0.62 (0.42, 0.91). In MrOS, a study of older men in the US, Ensrud *et al.* [120] found a HR for the highest versus lowest quarter of sedentary behaviour of 1.71 (0.99, 2.97). Among participants in NHANES, Schmid *et al.* [149] found a HR comparing the highest third versus the lowest third of 3.56 (1.73, 7.32), which was attenuated to 1.47 (0.30, 7.19) by adjustment for physical activity. Using a linear isotemporal substitution approach, they found 30 min/day more light physical activity at the expense of sedentary time was associated with a HR of 0.86 (0.79, 0.93), and 30 min/day more MVPA at the expense of sedentary time was associated with a HR of 0.35 (0.14, 0.92) [149]. Evenson *et al.* [97] also studied the association between sedentary behaviour and cardiovascular disease mortality among NHANES participants, but after adjustment for other waking movement behaviours, did not find evidence that SB was associated with CVD mortality: the HR comparing the highest to lowest quarter was directionally consistent but imprecise (1.44 (0.71, 2.90)) and the trend was unclear (p for trend 0.52).

Table 2.4 summarises the results of different studies.

Table 2.4: Results of primary studies in the literature review.

Cohort	Study	Population	n	Follow-up (years)	Events	Mutual adjustment	Outcome	Findings			
								SB	LIPA	MVPA	Other
CARDIOVASCULAR DISEASE											
Attitude, Behavior and Change study	Dohrn *et al.* [151]	representative sample of Swedish adults	1,176	mean 14.4	139	No	CVD	T3 vs T1: 1.41 (0.91, 2.20)	T3 vs T1: 0.92 (0.60, 1.42)	T3 vs T1: 0.52 (0.33, 0.82)	
British Regional Heart Study	Jefferis *et al.* [71]	older men in the UK	1,181	median 4.9	122	Analyses without mutual adjustment	CVD	Per 30 min/day: 1.09 (1.00, 1.19)	Per 30 min/day: 0.94 (0.85, 1.04)	Per 10 min/day: 0.88 (0.81, 0.96)	
"	"	"	"	"	"	SB/LIPA analyses adjusted for MVPA; MVPA analyses for SB	"	*Per 30 min/day: 1.01 (0.90, 1.13)*	*Per 30 min/day: 0.99 (0.89, 1.11)*	*Per 10 min/day: 0.89 (0.80, 0.99)*	
EPIC-Norfolk	Dempsey *et al.* [70]	middle-to-older aged UK adults	5,585	mean 5.7	1,106	Analyses without mutual adjustment	CVD	Higher SB: higher risk	Higher LIPA: lower risk	Higher MVPA: lower risk	
"	"	"	"	"	"	MVPA analyses adjusted for LIPA; LIPA/SB analyses adjusted for MVPA	"	*Attenuated by covariate + MVPA adjustment*	*For LIPA > 3 hours/day, attenuated by covariate + MVPA adjustment*	*Persisted after covariate + LIPA adjustment*	Adjustment for covariates and movement behaviours in same model so cannot be distinguished
Healthy Ageing Initiative	Ballin *et al.* [121]	adults aged 70 in Umea, Sweden	3,343	mean 2.7	74	No	stroke and MI	Per 1 h/day: 1.37 (1.12, 1.68)	Per 30 min/day: 0.89 (0.79, 0.99)	Per 30 min/day: 0.58 (0.41, 0.83)	
OPACH sub-study of WHI	LaCroix *et al.* [69]	older women in the US	5,861	mean 3.5	143	Analyses without mutual adjustment	CHD		Q4 vs Q1: 0.58 (0.34, 0.99); Per 1 h/day: 0.80 (0.69, 0.93)	Q4 vs Q1: 0.54 (0.30, 0.96)	

Cohort	Study	Population	n	Follow-up (years)	Events	Mutual adjustment	Outcome	Findings[a, b]			
								SB	LIPA	MVPA	Other
"	"	"	"	"	"	Adjusted for MVPA	"		*Per 1 h/day: 0.86 (0.73, 1.00)*		
"	"	"	"	"	570	Analyses without mutual adjustment	CVD		Q4 vs Q1: 0.78 (0.60, 1.00); Per 1 h/day: 0.90 (0.83, 0.97)	Q4 vs Q1: 0.69 (0.53, 0.91)	
"	"	"	"	"	"	Adjusted for MVPA	"		*Per 1 h/day: 0.92 (0.85, 0.99)*		
OPACH sub-study of WHI	Bellettiere *et al.* [117]	older women in the US	5,638	mean 3.4	137	Analyses without mutual adjustment	CHD	Q4 vs Q1: 2.65 (1.34, 5.22)			
"	"	"	"	"	"	Adjusted for MVPA	"	*Q4 vs Q1: 1.68 (0.78, 3.60)*			
"	"	"	"	"	545	Analyses without mutual adjustment	CVD	Q4 vs Q1: 1.69 (1.27, 2.26); Per 1 h/day: 1.12 (1.05, 1.19)			
"	"	"	"	"	"	Adjusted for MVPA	"	*Q4 vs Q1: 1.53 (1.09, 2.14), less clear dose-response*			
OPACH sub-study of WHI	Nguyen *et al.* [152]	older women in the US	5,416	mean 6.0	616	No	major CVD				Q4 vs Q1 of 'daily life movement': 0.57 (0.45, 0.74)
"	"	"	"	"	268	"	CHD				Q4 vs Q1 of DLM: 0.57 (0.38, 0.84)
"	"	"	"	"	253	"	stroke				Q4 vs Q1 of DLM: 0.70 (0.47, 1.03)

Cohort	Study	Population	n	Follow-up (years)	Events	Mutual adjustment	Outcome	Findings[a, b]			
								SB	LIPA	MVPA	Other
Whitehall II	Yerramalla *et al.* [76]	civil servants based in London (when recruited in 1985-1988)	3,319	mean 6.2	299	Compositional Data Analysis approach	CVD	*Higher LIPA with lower SB: no evidence of lower risk. Higher MVPA with lower SB: lower risk.*	*Higher LIPA with lower SB: no evidence of lower risk. Higher MVPA with lower LIPA: lower risk.*	*Higher MVPA with lower SB/LIPA: lower risk.*	
Whitehall II	Yerramalla *et al.* [153]	civil servants based in London (when recruited in 1985-1988)	3,321	mean 6.2	299	Analyses without mutual adjustment	CVD	Per 100.2 min/day (1 SD): 1.20 (1.06, 1.37)			
"	"	"	"	"	"	Adjusted for meeting MVPA recommendation	"	*Per 100.2 min/day (1 SD): 1.02 (0.88, 1.19)*			
CARDIOVASCULAR DISEASE MORTALITY											
Attitude, Behavior and Change study	Dohrn *et al.* [116]	Swedish adults aged 50+ years at follow-up	851	mean 14.2	24	No	CVD mortality	T3 vs T1: 7.61 (2.02, 28.65)	T2 vs T1: 0.41 (0.17, 1.01) [No CVD deaths in LIPA T3]	T3 vs T1: 0.11 (0.03, 0.56)	
Attitude, Behavior and Change study	Dohrn *et al.* [154]	Swedish adults aged 50+ years at follow-up	851	mean 14.2	24	Linear isotemporal substitution	CVD mortality	*Per 30 min/day to LIPA: 0.76 (0.63, 0.92); 10 min/day to MVPA: 0.62 (0.42, 0.91)*	*Per 30 min/day from SB: 0.76 (0.63, 0.92)*	*Per 10 min/day from SB: 0.62 (0.42, 0.91)*	
NHANES	Schmid *et al.* [149]	representative sample of US adults aged 50-85 years	3,702	mean 6.4	196	Analyses without mutual adjustment	CVD mortality	T3 vs T1: 3.56 (1.73, 7.32)			T3 vs T1 of all PA (LIPA + MVPA): 0.27 (0.13, 0.57)
"	"	"	"	"	"	Mutual SB/PA adjustment	"	*T3 vs T1: 1.47 (0.30, 7.19)*			*T3 vs T1 of all PA (LIPA + MVPA): 0.35 (0.07, 1.86)*

Cohort	Study	Population	n	Follow-up (years)	Events	Mutual adjustment	Outcome	Findings[a, b]			
								SB	LIPA	MVPA	Other
"	"	"	"	"	"	Linear isotemporal substitution	"	*30 min/day to LIPA: 0.86 (0.79, 0.93). 30 min/day to MVPA: 0.35 (0.14, 0.92).*	*Per 30 min/day from SB: 0.86 (0.79, 0.93)*	*Per 30 min/day from SB: 0.35 (0.14, 0.92)*	
NHANES	Evenson *et al.* [97]	representative sample of US adults aged >40 years	3,809	mean 6.7	107	Other waking movement behaviours	CVD mortality	*Q4 vs Q1: 1.44 (0.71, 2.90), trend unclear (p 0.52)*	*Q4 vs Q1: 0.90 (0.44, 1.84), trend unclear (p 0.57)*	*Q4 vs Q1: 0.48 (0.23, 1.03) (p trend 0.001)*	
OPACH sub-study of WHI	LaMonte *et al.* [119]	older women in the US	6,382	mean 3.1	154	No	CVD mortality		T3 vs T1: low LIPA 0.63 (0.41, 0.99); high LIPA 0.30 (0.17, 0.51)	T3 vs T1: 0.42 (0.24, 0.75)	
OPACH sub-study of WHI	Nguyen *et al.* [152]	older women in the US	5,416	mean 6.1	331	No	CVD mortality				Q4 vs Q1 of DLM: 0.38 (0.26, 0.56)
Osteoporotic Fractures in Men study	Ensrud *et al.* [120]	older men in the US	2,918	mean 4.5	138	No	CVD mortality	Q4 vs Q1: 1.71 (0.99, 2.97)	Q1 vs Q4: 1.73 (0.90, 3.34)	Q1 vs Q4: 2.86 (1.50, 5.45)	

[a] Results for different cardiovascular outcomes are shown on different lines. Results without and with mutual adjustment for some other movement behaviours are also shown on different lines; results with mutual adjustment are italicised.

[b] Where studies presented estimates from several different models (e.g. different covariate sets), extracted results are based on those the study investigators presented as primary (e.g. in abstract). Generally, these were adjusted for non-movement-behaviour covariates.

Abbreviations used are: SB - sedentary behaviour; PA - physical activity; LIPA - light intensity physical activity; MVPA - moderate-to-vigorous physical activity; MPA - moderate physical activity; VPA - vigorous physical activity; DLM - daily life movement; OPACH - Objective Physical Activity and Cardiovascular Disease Health study; WHI - Women's Health Initiative; NHANES - National Health and Nutrition Examination Survey; EPIC - European Prospective Investigation into Cancer; CHD - coronary heart disease; CVD - cardiovascular disease; MI - myocardial infarction; SD - standard deviation; T3/T1 - third 3, third 1; Q4/Q1 - quarter 4, quarter 1.

2.3.4.3 Light physical activity

In analyses unadjusted for other movement behaviours, results on light physical activity were mixed. Several studies found evidence that higher levels of LIPA are associated with lower risk of incident CVD. In EPIC-Norfolk, Dempsey *et al.* [70] found higher levels of LIPA to be associated with lower risk of CVD. In the HAI, Ballin *et al.* [121] found a HR for stroke/MI of 0.89 (0.79, 0.99) per 30 min/day increment of light physical activity. In OPACH, LaCroix *et al.* [69] found HRs comparing the highest versus lowest quarter of light physical activity of 0.58 (0.34, 0.99) for CHD and 0.78 (0.60, 1.00) for CVD [69]. Using splines to study the dose-response association, they found a linear dose-response [69]. Each 1 hour/day increment in LIPA was associated with a HR of 0.80 (0.69, 0.93) for CHD and 0.90 (0.83, 0.97) for CVD [69]. Also in OPACH, Nguyen *et al.* [152] studied the association between 'daily life movement', a construct related to light physical activity, and CVD and found HRs for comparing the highest versus lowest quarter of 0.57 (0.45, 0.74) for major CVD, 0.57 (0.38, 0.84) for CHD, and 0.70 (0.47, 1.03) for stroke. In other studies, confidence intervals for estimates of association contained the null (although the fact most studies had small numbers of events means many of these estimates are imprecise to an extent that they do not exclude clinically meaningful associations). For example, in the ABC study (139 events), comparing the highest versus lowest third of LIPA Dohrn *et al.* [151] found a HR of 0.92 (0.60, 1.42). In the BRHS (122 events), comparing the highest versus lowest quarter of LIPA Jefferis *et al.* [71] found a HR of 0.67 (0.38, 1.20). Modelling LIPA as a continuous variable, they found that for each additional 30 min/day in LIPA, the HR for CVD was 0.94 (0.85, 1.04).

Results of analyses adjusted for other movement behaviours were inconsistent. In OPACH, LaCroix *et al.* [69] found that associations persisted after MVPA adjustment, with each 1 hour/day increment in LIPA associated with a MVPA-adjusted HR of 0.86

(0.73, 1.00) for CHD and 0.92 (0.85, 0.99) for CVD. In EPIC-Norfolk, Dempsey *et al.* [70] found adjustment for covariates and MVPA substantially attenuated associations, with significant associations persisting only for participants with the lowest levels of LIPA (< 3 hours/day). In the BRHS, after adjusting for MVPA, Jefferis *et al.* [71] found that for each additional 30 min/day in LIPA, the HR for CVD was 0.99 (0.89, 1.11). As noted above, in the Whitehall II study (299 events), which considered the waking-day movement behaviour composition, Yerramalla *et al.* [76] found no evidence of lower CVD risk with higher LIPA at the expense of SB [76].

Results on light physical activity and CVD mortality also varied between studies. Results reported here are unadjusted for other movement behaviours unless stated. In the MrOS study, Ensrud *et al.* [120] found a HR comparing the lowest versus highest quarter of LIPA of 1.73 (0.90, 3.34). In OPACH, LaMonte *et al.* [119] studied the association between low light intensity physical activity and high light intensity physical activity and CVD mortality. Comparing the highest versus lowest third of low light intensity physical activity, they found a HR of 0.63 (0.41, 0.99), and comparing the highest versus lowest third of high light intensity physical activity they found a HR of 0.30 (0.17, 0.51) [119]. Also in OPACH, comparing the highest versus lowest quarter of 'daily life movement', Nguyen *et al.* [152] found a HR of 0.38 (0.26, 0.56). In the ABC study, there were 24 CVD deaths overall and no CVD deaths in the highest third of LIPA, but the HR comparing the middle versus lowest third was 0.41 (0.17, 1.01) [116]. In linear isotemporal substitution analyses (i.e. considering time substitutions, with mutual adjustment), replacing 30 min/day of sedentary time with LIPA was associated with a HR of 0.76 (0.63, 0.92) (as also noted in the previous section) [154]. Also using a linear isotemporal substitution approach, in NHANES, Schmid *et al.* [149] found 30 min/day more LIPA at the expense of sedentary time was associated with a HR of 0.86 (0.79, 0.93). Evenson *et al.* [97] also studied participants in NHANES, but did not find evidence for an association between LIPA and CVD

mortality after adjustment for other waking movement behaviours.

Qiu *et al.* [72] meta-analysed studies on the association between objectively-measured LIPA and CVD mortality published up to 2020-04-05. Studies included in the meta-analysis for higher levels of LIPA were Dohrn *et al.* [116], Ensrud *et al.* [120], Evenson *et al.* [97], and LaMonte *et al.* [119] (two sets of results). All these studies used 'traditional' statistical approaches, and had a combined enrolment of 13,960 participants, with 423 cardiovascular deaths [72]. The pooled HR for CVD mortality per 30 min/day of LIPA was 0.80 (0.67, 0.96), with evidence for heterogeneity between studies [72]. In studies controlling for MVPA, this was attenuated to 0.92 (0.76, 1.12), and in studies controlling for sedentary time this was attenuated to 0.98 (0.92, 1.04) [72]. Separately, Qiu *et al.* [72] also considered LIPA replacing sedentary time, and found two studies which addressed this: Dohrn *et al.* [154] and Schmid *et al.* [149] (both following the linear isotemporal substitution methodology, see Chapter 1). Substituting 30 min/day of sedentary time with LIPA was associated with a HR for CVD of 0.84 (0.73, 0.96) [72].

2.3.4.4 Moderate-to-vigorous physical activity

Studies not adjusting for time in other movement behaviours consistently found associations between higher levels of MVPA and lower risk of incident CVD. In the ABC study, Dohrn *et al.* [151] found a HR comparing the highest versus lowest third of MVPA of 0.52 (0.33, 0.82). In the BRHS, Jefferis *et al.* [71] found a HR comparing the highest versus lowest quarter of MVPA of 0.34 (0.18, 0.62). Modelling MVPA as a continuous variable, they found that for each additional 10 min/day in MVPA, the HR for CVD was 0.88 (0.81, 0.96). In the HAI, Ballin *et al.* [121] found a HR for stroke or myocardial infarction of 0.58 (0.41, 0.83) per 30 min/day increment of MVPA. In EPIC-Norfolk, Dempsey *et al.* [70] found that higher levels of MVPA were

associated with lower risk of CVD. In OPACH, LaCroix *et al.* [69] found HRs comparing the highest versus lowest quarter of MVPA of 0.54 (0.30, 0.96) for CHD and 0.69 (0.53, 0.91) for CVD [69].

Strong associations between MVPA and CVD risk were also observed in studies adjusting for time in other movement behaviours. For example, in the BRHS Jefferis *et al.* [71] found that adjusting for SB had little impact on associations between MVPA and incident CVD. In EPIC-Norfolk, Dempsey *et al.* [70] found associations were attenuated only slightly by adjustment for covariates and LIPA. Studying the waking-day movement behaviour composition in Whitehall II, Yerramalla *et al.* [76] found that lower MVPA with higher SB or LIPA was associated with higher CVD risk.

Similarly, higher levels of moderate-to-vigorous physical activity were consistently associated with lower risk of CVD mortality. Results reported here are unadjusted for other movement behaviours unless stated. In OPACH, LaMonte *et al.* [119] found a HR in the highest versus lowest third of MVPA of 0.42 (0.24, 0.75). In the MrOS study, Ensrud *et al.* [120] found a HR comparing the lowest versus highest quarter of MVPA of 2.86 (1.50, 5.45). In the ABC study, the HR comparing the highest versus lowest third of MVPA was 0.11 (0.03, 0.56; again the very small number of CVD deaths in this study should be noted) [116]. As noted above, using a linear isotemporal substitution approach Dohrn *et al.* [154] found that replacing 10 min/day of sedentary time with MVPA was associated with a HR of 0.62 (0.42, 0.91) [154]. In NHANES, and as was also noted above, Schmid *et al.* [149] found that 30 min/day more MVPA at the expense of sedentary time was associated with a HR of 0.35 (0.14, 0.92) for CVD mortality [149]. Also in NHANES, Evenson *et al.* [97] found that, adjusted for other waking movement behaviours, levels of MVPA and moderate physical activity exceeding the lowest quarter were associated with lower risk: the

HR for the highest versus lowest quarter of MVPA was 0.48 (0.23, 1.03) (overall trend was statistically significant).

2.4 Discussion

2.4.1 Findings

Studies identified above consistently found that higher levels of moderate-to-vigorous physical activity were associated with lower risk of cardiovascular disease and cardiovascular death. Where this was examined, it seemed to persist after adjustment for other movement behaviours (e.g. sedentary behaviour). Several factors make drawing conclusions about magnitude and dose-response difficult. Magnitude of estimates varied widely between studies, and the limited size of most included studies means estimates were often notably imprecise. Quantitative meta-analysis of results may improve precision of estimates, but this is made more challenging by the diversity of approaches used throughout the research process [85]. Harmonising device-based measurements of physical activity is difficult, and generally requires the use of compatible devices and re-analysis of original study data [85].

In most studies, higher levels of sedentary behaviour were associated with higher risk of cardiovascular disease. Studies' primary analyses were often unadjusted for other movement behaviours; further adjusting for other movement behaviours (usually MVPA) almost entirely attenuated this association in some studies [70, 71], and partially attenuated it in others [117].

Results on light physical activity suggested higher levels of light physical activity were associated with lower CVD risk, although not all results were consistent and several studies found no significant association. Moreover, adjustment for other movement

behaviours often substantially attenuated observed associations [72].

None of the above studies considered device-measured sleep.

The limited number of studies using multi-behaviour or 24-hour approaches, considering several behaviours within a unified framework, makes it difficult to draw holistic conclusions about 24-hour movement behaviours from the existing evidence.

2.4.2 Size of prospective cohorts with accelerometry

Studies using device-based measurements of movement behaviours have been substantially limited by sample size. The earliest accelerometer-based measurement in any of the studies discussed was in the ABC study, which collected data in 2001-2002, in approximately 1,000 individuals (exact numbers differ between reports) [151]. Another early study was in NHANES 2003-2004 and 2005-2006 waves, which comprised a few thousand individuals [95, 97]. Many of the studies reviewed had limited numbers of events and discuss challenges with statistical power (e.g. ABC, 139 events [151], 24 cardiovascular deaths [116, 154]; HAI, 74 events [121]). Meta-analysing all available studies on device-measured LIPA and CVD mortality, Qiu *et al.* [72] reported only 423 deaths. Studies which are too small to precisely estimate effects threaten the validity of the epidemiologic literature: not only does it lead to imprecise estimates of effect, but it also means that when significant effects are observed, they are more likely to be 'false-positives' and to over-estimate true effects [157]. Really large-scale data collection using device-based measurements only occurred in the 2010s [101], meaning that follow-up time has only recently become sufficient to reliably study associations with incident disease.

2.4.3 Classifying movement behaviours

Notable challenges in deriving useful measurements from accelerometer data remain and affect the interpretation of the epidemiological literature.

Perhaps most notably, sleep was usually not recorded, and therefore not studied. Historically, accelerometers have often been worn at the hip and removed for sleep. Among the studies discussed, only OPACH, MrOS and Whitehall II used a 24-hour wear protocol [69, 76, 117, 119, 120, 153]. Although the accelerometer in OPACH was worn on a 24-hour basis, times in bed were still identified using a diary [117]. Although sleep periods were identified in Whitehall II using a validated algorithm, Yerramalla *et al.* [76] noted that as a very distinctive behaviour type, sleep presents conceptual difficulties for classification and analysis, and so they did not study sleep. None of the studies analysed the association between device-measured sleep and CVD risk, and most analyses of other behaviours were unadjusted for sleep. This is problematic as sleep is known to be associated both with other behaviours [158] and with cardiovascular disease risk [73, 74].

Even for behaviours that were measured, measurement challenges make it difficult to interpret some associations. Almost all studies used traditional 'cut-point' behaviour classification methods, whereby activity intensity is classified based on whether counts or another metric of acceleration falls above a certain threshold. As noted in Chapter 1, cut-point-based approaches have several known difficulties when classifying behaviours. Moreover, many studies noted that the cut-points used had often not been validated in the population of interest e.g. using cut-points developed using data from younger adults in an older adult population [76, 121]. Differences in measurement can have important consequences for epidemiological inferences. For example, a recent federated meta-analysis of six studies on the association between the movement

behaviour composition and all-cause mortality found results differed substantially between studies using hip-worn and wrist-worn devices and cut-point-based classification, with different inferences on light physical activity in particular [159]. They attributed this to overestimation of time in MVPA and misclassification of light physical activity as sedentary behaviour and vice versa using a cut-point in wrist-worn accelerometer data [159].

2.4.4 Analytic approaches

Differences in how studies approach interrelationships between behaviours make it difficult to compare and assess results between studies. When discussing whether or not to adjust results on sedentary behaviour for MVPA, Bellettiere *et al.* [117] observed that MVPA has in previous studies 'been viewed as a confounder, a mediator, an effect modifier, or a competing behavior', highlighting the inconsistency across the literature. A recent review on light physical activity identified treatment of MVPA as a key issue when interpreting studies of light physical activity [77]. For example, two studies ostensibly both studying light physical activity may be estimating something substantially different depending on whether they do or do not adjust for time in MVPA, and this is further complicated by the constraint imposed by the 24-hour day [132, 160]. Specifically, considering light physical activity time in a model without adjustment for time in other behaviours, associations are agnostic to which behaviour additional time in light physical activity is drawn from, but associations with the outcome may be confounded by MVPA. Adjusting for MVPA fixes time in MVPA, and so addresses confounding by MVPA. However, while the unadjusted model was agnostic to where additional time in light physical activity was drawn from, with time in MVPA fixed, additional time in light physical activity must be drawn from sedentary behaviour or sleep. The structure imposed by the 24-hour day constraint

means that adjustment, rather than simply controlling confounding, also changes the association being estimated [132, 160].

Approaches which explicitly acknowledge and account for the multidimensional, interdependent nature of 24-hour movement behaviours represent progress in this regard. Two studies did this through use of linear isotemporal substitution. However, as discussed in Chapter 1, linear isotemporal substitution approaches have their own limitations. The Compositional Data Analysis approach used by Yerramalla *et al.* [76] likely further addresses some of the nuances of movement behaviours as an exposure. However, as this literature review shows, the application of these methods in prospective studies of CVD or CVD mortality has been limited thus far.

2.4.5 Assessing bias

Studies commonly discussed potential biases, but addressing these biases quantitatively was much rarer (see Section 2.3.3.4). Within-person variation and measurement error were mentioned by studies principally to argue that the measurement approach used was appropriate (i.e. appropriate cut-points), and to argue that the measurement period was sufficiently long for a stable measurement. Most discussed generalisability of findings in their sample, but this discussion was usually brief and was not probed quantitatively. The potential for reverse causality to drive findings was widely acknowledged and discussed. To address this, many studies conducted analyses removing some period of follow-up and/or participants with known health conditions. However, as noted above, the removed period was as short as 6 months in several cases [69, 117, 119]. Residual confounding was not addressed quantitatively.

2.5 Summary

Existing literature suggests higher levels of moderate-to-vigorous physical activity are associated with a lower risk of incident cardiovascular disease. There is some evidence suggesting lower levels of sedentary behaviour, and possibly also higher levels of light physical activity, are associated with lower cardiovascular disease risk. The existing literature has many limitations. Most notably, most studies using device-based measurements of movement behaviours are relatively small in size. There are also challenges around how time in different behaviours is classified; how behaviours making up the 24-hour day are analysed alongside each other; and around interpretation of results. The next chapter, Chapter 3, will introduce a large prospective cohort study with device-based measurement of movement behaviours in approximately 100,000 participants, which subsequent chapters will use for analyses addressing some of these challenges.

Chapter 3: UK Biobank

3.1 Background

Chapter 2 noted that the relatively modest size of studies with device-based measurement of movement behaviours and longitudinal follow-up has substantially limited the analyses which were possible, particularly for incident disease. Largely, limitations in size were due to cost and difficulty associated with applying the technology at scale. However, in the last decade it has increasingly become feasible to deploy device-based measurement of movement behaviours in large prospective cohort studies (tens of thousands of participants) [101]. Incorporation of device-based measurement in large prospective cohort studies enables well-powered analyses associating movement behaviours with incident disease outcomes [93].

UK Biobank is a prospective cohort study which was established to enable research into the determinants of diseases of middle and older age [161]. Between 2006 and 2010, over 500,000 40 to 69 year old participants were recruited in England, Scotland and Wales [161, 162]. Extensive phenotypic information and biological samples were collected at the time of recruitment [161, 163]. Additional information is obtained by follow-up in health records and through invitations to participate in further data collection [162].

Between 2013 and 2015, UK Biobank invited participants to take part in a study collecting accelerometer-based measurements of movement behaviours [93]. Over 100,000 participants wore an accelerometer, making UK Biobank the largest prospective study to date with data on device-measured movement behaviours [93]. UK Biobank therefore offers an unparalleled opportunity to study how device-measured movement behaviours are associated with incident disease, and it will be used throughout this thesis. This chapter will describe the UK Biobank study, outline how data was prepared for use in this thesis, and describe key characteristics of the cohort.

3.2 Study design: UK Biobank

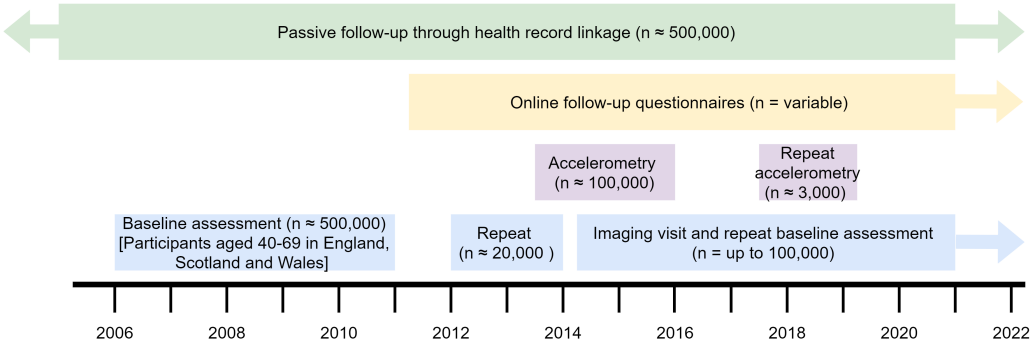
3.2.1 Recruitment

UK Biobank invitation letters were sent to all 40 to 69 year old adults registered with a National Health Service (NHS) general practitioner and living within approximately 25 miles of one of 22 UK Biobank assessment centres [161]. 5.5% of invited individuals joined the study (503,325 individuals from 9.2 million invitations) [161]. While this represents a low response rate relative to many studies, the goal was to recruit a large, diverse overall sample, and UK Biobank did not find following up on non-responses to be cost-effective in achieving this aim [164].

Figure 3.1 illustrates the timeline of data collection in the UK Biobank study. The initial study visit (the 'baseline assessment') took place between 2006 and 2010. Since then, subsets of participants have participated in repeats of the baseline assessment and in other 'resource enhancements', such as online follow-up questionnaires. Accelerometer measurement of movement behaviours, one of the 'resource enhancements', took place between 2013 and 2015, and was repeated for a subset

of participants in 2017 to 2018. Linkage to health-related datasets, such as hospital inpatient data, includes data from prior to recruitment (going back to the start of records) and is ongoing. Subsequent sections of the text elaborate on different study components.

Figure 3.1: Timeline of the UK Biobank prospective cohort study.



3.2.2 Baseline assessment

On recruitment, participants attended a UK Biobank assessment centre for an overall baseline assessment collecting information on many aspects of their health and its potential determinants. Participants completed a touchscreen questionnaire and were interviewed by a trained interviewer. They also gave blood, urine and saliva samples and several anthropometric measurements were made. Comprehensive descriptions of the data collection protocols are available elsewhere [161, 162, 165].

3.2.3 Follow-up for disease outcomes

Participants consented to ongoing linkage of their data with electronic health records. Hospital inpatient data and data from the death register are available for all participants. 45% of the cohort also have primary care data available for general use (planned to extend to the whole dataset in the future). Other linked health datasets,

such as COVID-19 testing data and Cancer Registry data, are also available. Only hospital inpatient data and the death register were used for analyses in this thesis.

3.2.4 Resource enhancements and repeat assessments

UK Biobank has been enhanced by measures derived from existing data and samples, notably genotype [166], exome sequencing data [167], and now some whole genome sequencing data [168].

UK Biobank participants with an email address on record have also been periodically recontacted to participate in further data collection. 20,000 participants attended a repeat visit in 2012-2013, in which the original protocol was repeated. Since 2014, participants have attended UK Biobank assessment centres for an imaging visit. At this visit, the original protocol was repeated and several additional imaging modalities were collected (e.g. brain and cardiac magnetic resonance imaging). UK Biobank aims to collect this data on approximately 100,000 participants; at the time of writing roughly half of this is available. Online questionnaires on diet, mental health and many other factors have been conducted. Over 100,000 participants agreed to wear an accelerometer (see Section 3.3).

3.2.5 Ethical framework

Research in this thesis using UK Biobank data is covered by the general ethical approval for UK Biobank, provided by the North West - Haydock Research Ethics Committee. It was most recently renewed on 2021-06-29 with reference 21/NW/0157 [169]. UK Biobank has an Ethics Advisory Committee (established 2018) and formerly had an Ethics and Governance Council to advise on ethical issues in its maintenance and development [169]. UK Biobank participants gave informed consent [161], and can

also withdraw this consent at any time [170].

3.2.6 Accessing UK Biobank data

UK Biobank data is available for use by *bona fide* researchers conducting health-related research in the public interest [171]. To access UK Biobank data an application must be submitted with details on the purpose of the research and the variables it requires. The research in this thesis was conducted under approved application 59070 [172].

Most data used in this thesis was downloaded and processed locally. Specifically, data used in this thesis was stored and processed on the University of Oxford Biomedical Research Computing system (a high-performance cluster computing system) [173] and on an Oxford Population Health department machine. Both systems follow information security and governance guidelines. Although this procedure was standard at the time of initial work for this thesis, UK Biobank has subsequently launched a Research Analysis Platform, allowing analyses to be performed without local download of data. This platform was used for some data processing in the latter stages of this thesis.

3.3 Accelerometry in the UK Biobank study

Accelerometer measurement in UK Biobank took place in a sub-study 3-10 years after initial recruitment, involving approximately one fifth of the original cohort (see Figure 3.1). A Physical Activity Expert Working group consisting of experts on physical activity measurement was convened to design and oversee the study [174].

3.3.1 Recruitment

Between February 2013 and December 2015, participants with an email address on record were emailed an invitation to wear an accelerometer [93]. Emailed participants were randomly selected, although the North West region was excluded for much of the project, due to concerns about use of this region for pilot studies resulting in undue participant burden [93]. Of 236,519 participants contacted, 106,053 responded consenting to participate in the study, a participation rate of 44.2% [93].

3.3.2 Wear protocol

Between June 2013 and December 2015, consenting participants were sent Axivity AX3 wrist-worn triaxial accelerometers. Participants were asked to wear the device on the dominant wrist for 7 days and then to return the device by post in a pre-paid envelope [93]. Participants were instructed to wear the device continuously and to carry out their normal activities while wearing the device [93]. Among the 106,053 consenting participants, 103,712 accelerometer datasets were obtained (97.8% datasets obtained; 0.1% consent withdrawn; 1.2% device not returned; 0.9% device unreadable) [93]. Note that small discrepancies with numbers of available datasets reported subsequently in this thesis relate to participant withdrawals.

The Axivity AX3 accelerometer was configured to record triaxial acceleration at 100 Hz, and with a dynamic range of ± 8 g [93]. Devices were programmed to start recording at 10am two working days after they were posted [93]. Accelerometer values were recorded in gravity units (denoted g), where 1 g is 9.8 m/s^2 i.e. acceleration due to gravity.

3.4 Data preparation

3.4.1 Accelerometer data

Accelerometer data was used to derive measurements of 24-hour movement behaviours (the exposure variables in this thesis). As improving measurement of movement behaviours using wrist-worn accelerometry is a key aim of this thesis, detailed discussion of accelerometer data preparation is deferred to Chapter 4. Briefly, accelerometer data were processed using a previously-developed pipeline (deployed on a high-performance computing cluster) and standard quality control criteria were applied. The machine-learning method developed in Chapter 4 was used to classify movement behaviours in accelerometer data.

3.4.2 Participant data

Tabular participant data was cleaned and prepared for use. Specifically, missing values, including stated non-responses such as 'Prefer Not to Answer' and 'Do Not Know', were first identified. Continuous variables were categorised where this was needed for analyses, and some categorical variables were also recoded as needed (e.g. where some categories had small numbers). Some variables were obtained from multiple underlying UK Biobank variables (e.g. overall fruit and vegetable consumption was obtained from fresh fruit, cooked vegetable, and raw/salad vegetable consumption). Supplementary Table A.1 gives UK Biobank field names and relevant details of the coding for all variables used. Code is also available (see Section 3.4.4).

Variables for subsequent analyses were chosen and prepared. Several variables may confound associations between movement behaviours and incident cardiovascular dis-

ease, and thus were required for epidemiological analyses (see Chapter 5). Some additional variables were used either for excluding participants from analyses, for descriptive analyses, or in assessing possible biases.

It should be observed that baseline assessment data was collected several years prior to accelerometer wear (see Figure 3.1). A small ($< 15\%$) proportion of participants in the UK Biobank cohort have attended a repeat assessment (see Figure 3.1). These participants therefore have covariate measurements on multiple occasions. Where participants have later measurements, some studies have used data from the latest assessment prior to accelerometer wear [175]. In this study, data from the baseline assessment was used for all participants. Relevant considerations were: consistency across participants; the proportion of participants with a second assessment before accelerometer wear being small; limited likely impact of additional covariate accuracy only in a small subset; and most key variables being relatively stable over an extended period (see Section 3.5.1).

3.4.3 Disease data

The outcome of interest in this thesis was the first cardiovascular disease event. This section will describe the health record data used; how cardiovascular disease was defined based on health record data; details of the definition; rationale for this definition; and suitability of the chosen approach. Supplementary Table A.2 summarises this information and code is available (see Section 3.4.4).

This thesis used hospital inpatient data and the death register. A hospital admission to an NHS hospital in England, Scotland or Wales generates records in a database. The hospital data used in this thesis contains only inpatient episodes (and not, for example, outpatient appointments). Different elements of the dataset record various

details of the admission, including dates, diagnoses, and procedures. For this thesis, diagnosis codes and dates were used. The death register contains data on deaths, including dates and causes of death.

Cardiovascular disease was defined as an ICD-10 code for ischaemic heart disease (I20-I25) or cerebrovascular disease (I60-I69) appearing either in hospital inpatient data or on the death register. For prevalent disease, the ICD-9 codes 410-414 and 430-438 were also considered (the ICD-9 coding scheme was used in a small proportion of historical records). In hospital data, both primary and secondary diagnoses were considered, and all codes appearing on the death register (i.e. both primary and contributory causes of death) were considered. Only the first appearance of these codes was used for analyses, either to exclude participants with prevalent disease at accelerometer wear, or to define incident disease during follow-up.

As health record data continues to accumulate, it is necessary to identify the final date of available data and censor participants accordingly. Participants' records were considered censored at the earliest of: their recorded date of death, the UK Biobank record censoring date for hospital inpatient data, and the UK Biobank record censoring date for the death register. The UK Biobank record censoring date is algorithmically determined based on numbers of records relative to previous months (with some modifications during the COVID-19 pandemic to account for disruptions to inpatient services) [176]. The censoring dates can differ between England, Scotland, and Wales. Participants were assigned to England, Scotland, or Wales based on their UK Biobank assessment centre. While participants may access care in a different country of the United Kingdom to the assessment centre they attended, the impact on analyses of either early or late censoring of a small number of participants was considered minimal. For analyses in this thesis, censoring dates were 2021-02-28 for participants in England and Scotland, and 2018-02-28 for participants in Wales.

The particular definition of cardiovascular disease used in this thesis was developed in consultation with collaborators (on [1] and [118]) with expertise in clinical cardiology and cardiovascular disease epidemiology and was previously used to study associations between accelerometer-measured activity and cardiovascular disease using more traditional approaches [118]. Retaining the same definition in this work enables comparability. Diseases falling under I20-25 and I60-69 also share aetiological similarities, so grouping them is biologically meaningful. For example, ischaemic heart disease and ischaemic stroke are both driven by atherosclerosis (hardening and narrowing of arteries due to deposition of lipid plaques). While haemorrhagic stroke and other cerebrovascular diseases may have less shared aetiology, some risk factors are shared. Strokes of different types may not be well-distinguished in health records: a substantial proportion of stroke records are of unspecified type [177].

Incorporation of procedural codes (OPCS) for procedures associated with these diseases was considered. However, in discussion with collaborators it was considered that inclusion of these codes may be largely redundant, as it would be unusual for patients to accrue a procedural code without a corresponding diagnostic code, and procedural codes may depend more closely on service access. Incorporation of primary care data was also considered, as some conditions falling under this disease definition (e.g. angina) may not lead to hospital admission. However, fewer than half of participants have primary care data available, limiting the value of this data source in this study. Furthermore, empirical evidence suggests a combination of hospital inpatient data and death register data is suitable for identifying incident cardiovascular disease cases for prospective analyses. Rannikmäe *et al.* [178] studied identification of stroke in different data sources in UK Biobank by analysing detailed medical information for participants in the Lothian region of Scotland. They estimated that Positive Predictive Value (also known as precision) for stroke in hospital inpatient data was 89% (82%, 94%). For most prospective epidemiological research Positive Predictive Value

is the primary metric for successful case definition [178]: high Positive Predictive Value means most positives are true positives, minimising bias in effect estimates. In terms of sensitivity (also known as recall), Rannikmäe *et al.* [178] found that of 225 cases identified as stroke based on detailed medical information, 39% were seen only in primary care data, and not in hospital or death records; similarly, using NHS England data Wood *et al.* [179] found that 30% of stroke/transient ischaemic attack cases and 8% of myocardial infarction cases identified across electronic health records appeared only in primary care data. This suggests a good overall level of sensitivity (although if the disease is rare and also power is not a major constraint, this is an important but secondary consideration, as false negatives make up only a very small proportion of controls).

3.4.4 Code and data availability

Much of the processing used the software package 'ukb_download_and_prep_template', which was developed and maintained in part for this project and is available on GitHub at `github.com/OxWearables/ukb_download_and_prep_template`. Detailed data preparation code can be found at `github.com/OxWearables/manuscript_ml_behaviours_cvd_2021` and `github.com/OxWearables/thesis_only_analyses`.

Appendix B provides a summary of code and data used throughout this thesis.

3.5 Dataset description

3.5.1 Data quality

Data in UK Biobank is clean and of good quality. Most variables collected at baseline assessment have little missing data, as shown in Table 3.1.

Some participants have repeated measurements data, enabling stability of UK Biobank variables to be assessed. Table 3.2 illustrates stability of some key variables, measured using Cohen's kappa for inter-rater reliability (the 'raters' are baseline assessment and the repeat of the baseline assessment in 2012-2013) and Fleiss-Cohen weighted kappa (ordered variables only; penalises movement between distant categories more than movement between adjacent categories). Most variables were stable even over this extended period, with Cohen's kappa values of 0.7 and above. Fruit and vegetable consumption was less stable than the other variables considered, with Cohen's kappa of just 0.33. More detailed examination showed that even in this case people most commonly moved between adjacent categories (55.5% of participants stayed in the same category, 39.5% of participants moved to an adjacent category and only 5.0% of participants moved more than one category). The higher Fleiss-Cohen weighted kappa values for ordered variables indicate that participants who moved between categories between baseline and repeat assessment moved more commonly to close than far categories.

Table 3.1: Comparison of the overall UK Biobank cohort and the sub-sample with accelerometer data by participant characteristics measured at baseline assessment.

	Overall UK Biobank cohort[a] (n=502,414)	**Analytic cohort for accelerometry analyses**[b] (n=87,489)
DEMOGRAPHIC		
Age at baseline assessment (years)		
Mean (SD)	57.0 (8.09)	56.3 (7.80)
Median [Q1, Q3]	58.3 [50.6, 63.7]	57.2 [50.1, 62.6]
Sex		
Female	273,329 (54%)	50,876 (58%)
Male	229,085 (46%)	36,613 (42%)
Ethnicity		
Asian	9,879 (2%)	756 (1%)
Black	8,058 (2%)	701 (1%)
Mixed/Other	9,084 (2%)	1,151 (1%)
White	472,617 (94%)	84,881 (97%)
Missing	2,776 (1%)	0 (0%)
SOCIOECONOMIC		
Townsend Deprivation Index		
Mean (SD)	-1.29 (3.09)	-1.76 (2.80)
Median [Q1, Q3]	-2.14 [-3.64, 0.55]	-2.47 [-3.83, -0.23]
Missing	623 (0%)	0 (0%)
Education		
School leaver	168,536 (34%)	19,533 (22%)
Further education	162,618 (32%)	29,058 (33%)
Higher education	161,131 (32%)	38,898 (44%)
Missing	10,129 (2%)	0 (0%)
BEHAVIOURAL		
Smoking status		
Never smoker	273,479 (54%)	50,886 (58%)
Ex-smoker	173,025 (34%)	30,712 (35%)
Current smoker	52,962 (11%)	5,891 (7%)
Missing	2,948 (1%)	0 (0%)
Alcohol consumption frequency		
Never	40,628 (8%)	4,745 (5%)
Less than 3 times/week	243,108 (48%)	39,756 (45%)
3 or more times/week	217,178 (43%)	42,988 (49%)
Missing	1,500 (0%)	0 (0%)
Daily fruit and vegetable consumption (servings/day)		
Less than 3	27,922 (6%)	3,594 (4%)
3-4.9	83,873 (17%)	14,289 (16%)
5-7.9	196,288 (39%)	36,988 (42%)
8 or more	182,591 (36%)	32,618 (37%)
Missing	11,740 (2%)	0 (0%)
BIOLOGICAL		
Body Mass Index (kg/m^2)		
Mean (SD)	27.4 (4.80)	26.6 (4.48)
Median [Q1, Q3]	26.7 [24.1, 29.9]	25.9 [23.5, 28.9]
Missing	3,104 (1%)	0 (0%)
Systolic blood pressure (mmHg)		
Mean (SD)	138 (18.7)	136 (18.1)
Median [Q1, Q3]	137 [125, 150]	135 [123, 148]
Missing	1,325 (0%)	49 (0%)

	Overall UK Biobank cohort[a] (n=502,414)	Analytic cohort for accelerometry analyses[b] (n=87,489)
HbA1c (mmol/mol)		
Mean (SD)	36.1 (6.78)	35.2 (5.34)
Median [Q1, Q3]	35.2 [32.8, 37.9]	34.7 [32.4, 37.1]
Missing	35,994 (7%)	5,243 (6%)
C Reactive Protein (mg/L)		
Mean (SD)	2.60 (4.36)	2.20 (3.86)
Median [Q1, Q3]	1.33 [0.66, 2.76]	1.12 [0.57, 2.33]
Missing	33,927 (7%)	4,784 (5%)
LDL cholesterol (mmol/L)		
Mean (SD)	3.56 (0.87)	3.60 (0.83)
Median [Q1, Q3]	3.52 [2.94, 4.12]	3.55 [3.02, 4.13]
Missing	33,789 (7%)	4,746 (5%)
HDL cholesterol (mmol/L)		
Mean (SD)	1.45 (0.38)	1.50 (0.38)
Median [Q1, Q3]	1.40 [1.17, 1.67]	1.46 [1.22, 1.73]
Missing	72,617 (14%)	11,623 (13%)
Triglycerides (mmol/L)		
Mean (SD)	1.75 (1.03)	1.64 (0.96)
Median [Q1, Q3]	1.48 [1.05, 2.15]	1.39 [0.99, 2.02]
Missing	33,281 (7%)	4,662 (5%)
HEALTH		
Self-reported overall health rating		
Poor	22,769 (5%)	1,761 (2%)
Fair	105,345 (21%)	12,595 (14%)
Good	288,970 (58%)	53,043 (61%)
Excellent	81,844 (16%)	19,958 (23%)
Missing	3,486 (1%)	132 (0%)
Self-reported medications for diabetes, cholesterol or blood pressure[c]		
Yes	138,223 (28%)	17,168 (20%)
No	364,191 (72%)	70,321 (80%)

[a] Total number of participants may differ slightly from earlier reports due to participant withdrawals.
[b] Cohort includes those participants included in analyses in Chapter 5. Cohort size differs by 9 participants due to participant withdrawals between conducting analyses in Chapter 5 and analyses in the present chapter (which were conducted later). Note that the lack of missing data for key covariates in the analytic sample is by design (see Chapter 5).
[c] Lack of missing data in medications variable is by design: non-responding participants were treated as having no relevant medications.
Q1, Q3 - quartile 1, quartile 3; SD - standard deviation.

Table 3.2: Stability of selected UK Biobank variables between baseline assessment and first repeat assessment (2-7 year time gap).

Variable	n with repeat	Cohen's Kappa	Fleiss-Cohen Weighted Kappa
Ethnicity	20,243	0.91 (0.90, 0.93)	
Qualifications	19,736	0.83 (0.83, 0.84)	0.90 (0.90, 0.91)
Body Mass Index category	20,257	0.71 (0.70, 0.72)	0.83 (0.82, 0.84)
Smoking status	20,243	0.88 (0.87, 0.88)	0.90 (0.90, 0.91)
Alcohol consumption	20,320	0.70 (0.69, 0.70)	0.75 (0.74, 0.76)
Fruit and vegetable consumption	19,993	0.33 (0.32, 0.34)	0.57 (0.56, 0.58)

3.5.2 Cohort description

Table 3.1 describes selected demographic, socioeconomic, behavioural, biological, and health characteristics of UK Biobank participants and of the sub-sample with accelerometer data. UK Biobank is not a representative sample of the UK population. For example, compared to the general population UK Biobank participants are (amongst other differences) more likely to be female, less likely to live in deprived areas, and less likely to experience a range of health conditions [180]. UK Biobank participants are also overwhelmingly white (94%) (comparable with the UK population of the same age in the 2001 Census but under-representing non-white participants relative to the 2011 Census) [180]. The sub-sample with accelerometer data is not representative of the overall cohort, with differential participation further accentuating many of the biases in the overall cohort. For example, as shown in Table 3.1, participants were even more likely to be female, more likely to be white, lived in less-deprived areas, and were generally in better health.

3.6 Summary

UK Biobank is a prospective cohort study, designed to investigate the determinants of diseases of middle and older age. It combines rich phenotypic, genetic, and disease data with a large scale which enables novel and reliable epidemiological research. The sub-sample of participants with accelerometer data represents the largest study of its kind to date. Over six years have now passed since this data was collected, meaning that large numbers of incident cardiovascular disease events have now accrued. It therefore offers an unprecedented opportunity to study how movement behaviours are associated with cardiovascular disease risk. The next chapter, Chapter 4, will

investigate how this accelerometer data can be best used for accurate measurements of time in different movement behaviours.

Chapter 4: Classifying movement behaviours in accelerometer data with machine-learning

The material on which this chapter is based has been published as part of an article in the ***British Journal of Sports Medicine*** *[1].*

4.1 Introduction

4.1.1 Background

Various methods have been used to classify movement behaviours from accelerometer data in epidemiological studies (see Chapter 1). Most commonly, cut-point-based methods are used. A cut-point is an acceleration threshold to distinguish different intensities of activity. However, cut-point-based methods have substantial disadvantages. They can only distinguish behaviours based on intensity, while properties of movement behaviours beyond intensity may be of great epidemiological relevance [58, 152]. Cut-point-based methods are prone to substantial misclassification, particularly when deployed in free-living settings and with accelerometers worn at the wrist [104, 105, 114, 181]. This misclassification materially impacts epidemiological findings: for example, as noted in Chapter 2, divergent results in a recent meta-analysis when

using cut-point-based methods to classify movement behaviours based on hip-worn compared to wrist-worn accelerometry may be explained by exposure measurement error [159]. In the literature reviewed on device-measured movement behaviours and incident cardiovascular disease (Chapter 2), almost all studies used cut-point-based methods to classify and study movement behaviours. Studies noted likely misclassification using cut-point-based methods and possible impacts of misclassification on epidemiological results [76, 97]. For example, Bellettiere *et al.* [117] identified improved measurement of sedentary behaviour as a research priority.

Machine-learning-based approaches can address some of the difficulties facing more traditional approaches. In particular, machine-learning methods can use a variety of features of the data to inform the classification [182]. They can capture complex relationships between features of the accelerometer data and behaviours, and these relationships may go beyond what researchers hypothesise based on biomechanical theory [182]. They are not constrained to only intensity-based classification [114, 182]. In the last decade, machine-learning-based approaches have been increasingly applied to movement behaviour classification from accelerometer data, and have shown very promising performance [108, 113].

However, large limitations remain in how machine-learning-based methods are applied (see also Chapter 1) [108, 113]. Most datasets used to develop classifiers are collected in structured settings, and they are often also relatively small [108, 113]. While extensively studied in methodological research, machine-learning-based classification methods are still rarely used to provide the exposure measurement in epidemiological studies. These challenges are discussed further in the paragraphs that follow.

Accelerometer data used to develop classification methods is often very different to accelerometer data collected in epidemiological studies. In particular, most behaviour classification methods are developed using laboratory-based data. It is therefore un-

known how well they well perform when used in data from epidemiological studies collected in free-living [104, 108, 113, 114, 181]. There is some evidence that methods developed in structured settings may perform poorly when used in free-living. For example, Bastian *et al.* [112] found very substantial drops in performance when using a classifier developed with laboratory data in semi-free-living conditions. In contrast, collecting ground-truth data in free-living settings helps to ensure classifiers will perform well when used in practice [104, 114, 181]. Reviewing the literature on behaviour classification using accelerometer data, Farrahi *et al.* [108] and Narayanan *et al.* [113] both identified the development of classifiers using free-living data as a key target for future studies. Moreover, most studies with ground-truth data are small-scale, principally due to the very high cost- and labour-intensiveness of collecting this data [183]. Small-scale studies for classifier development hinder the development of robust and performant classifiers: the range of behaviours represented is likely to be more limited, and the influence of random variation in the sample on the classifier is greater. The potential benefits of flexible, machine-learning-based classifiers are also generally more apparent in large samples.

Another key research need is the development of methods which are ready-to-use in large-scale epidemiological studies. Machine-learning-based behaviour classification has largely been studied as a standalone problem. Phenotypes derived using machine-learning methods have rarely been used in epidemiological studies, with traditional, cut-point-based methods remaining more common (see Chapter 2). For machine-learning-based phenotypes to be usable in epidemiological studies, several specific challenges must be addressed. Notably, the method should classify a suitable set of behaviours. In particular, the set of behaviours should be comprehensive, mutually exclusive, and health-relevant. Often, studies developing classifiers focus on circumscribed problems, where the behaviours available for classification are artificially limited. For example, in their ground-breaking algorithm development work, Ellis *et al.*

[184] classify behaviours as sitting, driving, standing, or walking/running. However, as Montoye *et al.* [111] note, were such a classifier to be used in an epidemiological study, it must classify all measured time as one of these four behaviours, although many behaviours do not fall into one of these categories. The classified behaviours should also be of interest for epidemiological research. Typically, this means they should have hypothesised associations with disease and be common in the population. For example, Willetts *et al.* [114] included 'bicycling' among the behaviours classified. Although representing an interesting classification problem, cycling is likely rare in a UK adult population in middle to older age, and its associations with health may not be expected to differ greatly from those of moderate-to-vigorous physical activity more generally. Readiness for application in epidemiological studies also highlights a need for end-to-end automation. Many methods for deriving time in different movement behaviours from accelerometer data include a 'human-in-the-loop' element of researcher judgement or verification [185]. This introduces a hard-to-quantify dependence on the researcher into behaviour measurement, and is infeasibly time-consuming and costly to apply in very large-scale epidemiological studies. Shareable, reusable methods also importantly enable between-study comparability: one success of cut-point-based methods is their ease of sharing, which facilitates comparing results. Finally, application-readiness emphasises the importance of holistically assessing aspects of validity which are most relevant for the particular application. Validity of measurements of movement behaviours is a complex, multi-faceted concept, which may not be well-captured by a single metric [186].

4.1.2 Aims

The aim of this work was to develop machine-learning-based methods to classify movement behaviours in free-living accelerometer data. Building on previous work,

it aimed to develop methods which were mature and ready for application in the UK Biobank.

4.2 Methods

4.2.1 CAPTURE-24 study

The CAPTURE-24 study is an accelerometer validation study of 152 adults aged 18-91 in Oxford, UK, in 2014-2015 [181, 187]. Participants were recruited via various routes including professional networks, online advertising, posters, social/sport clubs, emails to a list of research volunteers provided by a market research agency, and word-of-mouth [187]. Efforts were made to recruit a sample with broad variation across sex, age, and educational level [187].

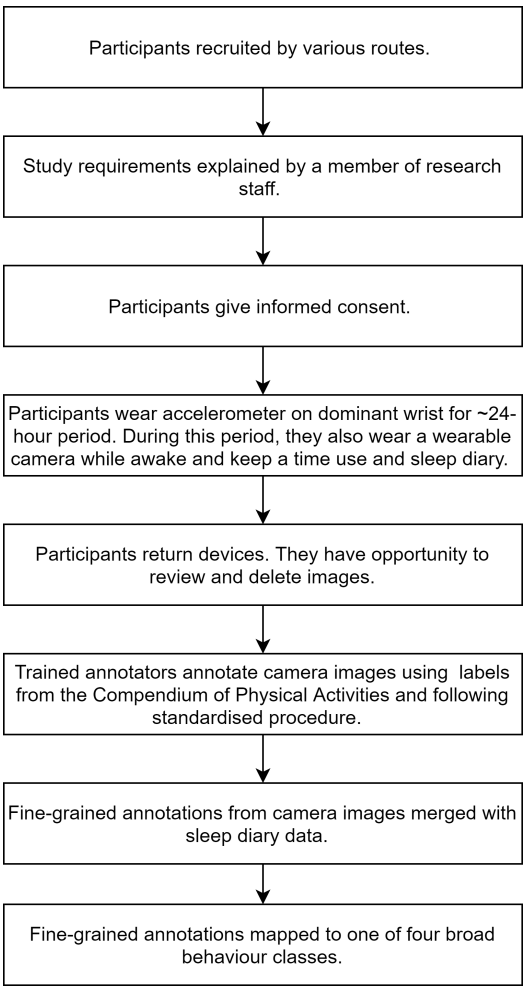
Consenting participants were asked to wear an Axivity AX3 wrist-worn accelerometer for 24 hours, wear a Vicon Autographer wearable camera while awake during that period, and keep a time use diary [114]. Participants were asked to wear the camera while going about their daily activities as normal; unlike in many similar studies, researchers did not prescribe particular behaviours. Study design is illustrated in Figure 4.1.

4.2.2 Ethical framework

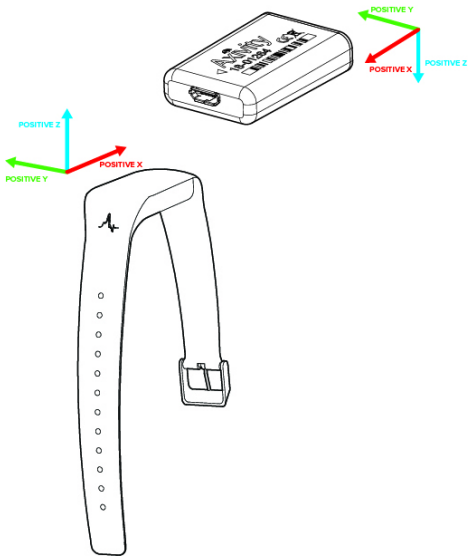
While wearable camera data is a rich and highly useful source of data for behaviour classification research, it also presents substantial ethical challenges [188]. Images are a rich and potentially intrusive modality. In passive data collection protocols, participants may forget they are wearing the device and may capture unwanted images.

Figure 4.1: CAPTURE-24 study design.

(a) Data collection steps in the CAPTURE-24 study.



(b) Diagram of the Axivity AX3 device. Image courtesy of Axivity Limited (`www.axivity.com`).



(c) Wearable camera.



Ethical considerations relate not only to direct study participants, but to people with whom they interact, including friends, family members, co-workers, and members of the public, who may not be aware they are being recorded. Kelly *et al.* [188] de-

veloped an ethical framework for automated wearable cameras in health behaviour research, which refined general ethical standards for image-based research to address the particular situation of automated wearable cameras. Their approach outlines how general ethical principles of respect for autonomy, beneficence, non-maleficence, and justice can be embedded in wearable camera studies [188]. For example, allowing participants to pause recording at any time, without questions from research staff, respects their autonomy [188]. Providing participants with guidance to verbally seek permission for recording from those they encounter in home and workplace settings while wearing the device respects the autonomy of third parties [188]. Instructing participants that they should remove the device if feeling threatened, uneasy, or that the device is attracting unwanted attention reduces risk of harm to participants (non-maleficence) [188].

The CAPTURE-24 study closely followed the ethical framework developed by Kelly *et al.* [188]. A member of the research team discussed the study requirements in detail with participants, and after this participants gave written informed consent [187]. Participants had an opportunity to review and delete images before the study team viewed them, and participants could not keep copies of their images [187]. CAPTURE-24 received ethical approval from the University of Oxford Inter-Divisional Research Ethics Committee (Ref SSD/CUREC1A/13-262).

4.2.3 Data labelling

Using camera images and time use diaries, trained annotators annotated accelerometer data with labels based on the Compendium of Physical Activities [28]. To do so, they used a specialised software designed for this purpose (available at `github.com/OxWearables/wearable-camera-desktop-browser`) [189]. Before they began annotation, annotators undertook training in annotation, in the software used, and in ethics.

They were only allowed to begin annotation once they attained a high level of inter-rater agreement with an expert (Cohen's kappa ≥ 0.80). They followed a detailed annotation protocol. They assigned 'events' to sequences of at least three consecutive images depicting the same behaviour (with some more generic labels available where a single dominating fine-grained behaviour was not available) [114]. Where the participant was doing multiple tasks at the same time, the primary behaviour was annotated, and it was suggested the movements of the participant's hands could be used to identify this [114]. Behaviours lasting less than three images would be grouped with the preceding event [114]. The annotation process is described in detail in the work of Willetts *et al.* [114].

Fine-grained annotations were mapped to sleep, sedentary behaviour, light physical activity behaviours, and moderate-to-vigorous physical activity behaviours. Describing intensity in METs (Metabolic Equivalent of Task), which measure energy expenditure relative to energy expenditure in quiet sitting, these behaviours were defined as:

1. Sleep: non-waking behaviour;
2. Sedentary behaviour: waking behaviour at ≤ 1.5 METs in a sitting, lying or reclining posture [58];
3. Light physical activity behaviours: waking behaviour at < 3 METs not meeting the sedentary behaviour definition;
4. Moderate-to-vigorous physical activity behaviours: all behaviour at ≥ 3 METs [28].

Examples of sedentary behaviour might include sitting working at a computer, or watching television; examples of light physical activity behaviours might include

cooking and some self-care activities; and examples of moderate-to-vigorous physical activity behaviours might include walking the dog or cycling [28]. Rule-based assignment of annotations to these classes was then reviewed by two people. This was found to be necessary as the fine-grained behavioural annotations were sometimes used in ways which did not appear to correspond to their assigned class (see Section 4.4.2). The review of annotation codes was conducted by two reviewers and a final set of labels agreed. No changes to the label classes were made after modelling results were seen. Supplementary Table A.4 lists fine-grained annotations assigned to each behaviour class.

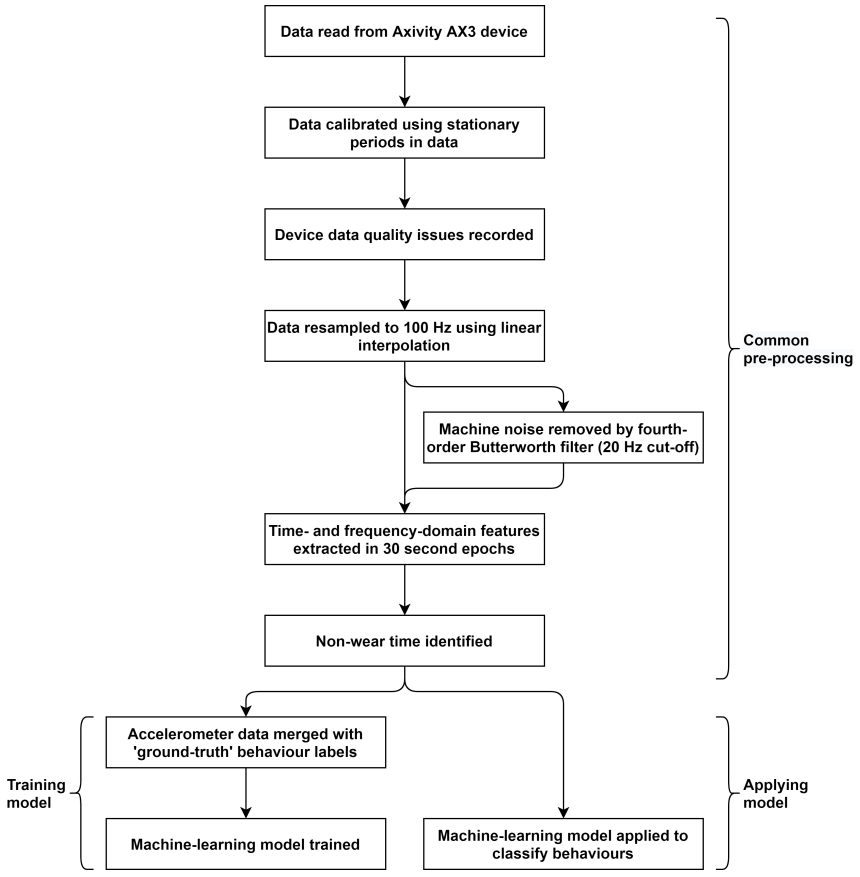
4.2.4 Data processing

The accelerometer data was processed to provide data on features of the acceleration signal in 30-second time windows (known as 'epochs'). Processing used 'biobankAccelerometerAnalysis' [4], which is a Python and Java-based tool to process accelerometer data in a scalable way (see also Section 4.2.8).

In brief, the steps to pre-process accelerometer data for subsequent machine-learning analysis are as follows (Figure 4.2). Accelerometer data is calibrated using a procedure based on stationary periods in the data which has been described previously [93, 190]. Device data quality issues are recorded; these include values exceeding the dynamic range of the device, recording errors, and 'interrupts' (gaps in the data). As the actual sampling rate can vary slightly, data is resampled to 100 Hz using linear interpolation (except interrupts lasting more than 1 second, which are treated as missing). The vector magnitude (the Euclidean norm of the x, y, and z axes) is calculated. For some subsequent features, it is filtered using a fourth-order Butterworth low-pass filter with a 20 Hz cut-off to remove machine noise. A range of time- and frequency- domain features are then extracted in each 30-second epoch of the data (see Section 4.2.5 for

a description of the features chosen for use in the model). Non-wear time is identified as periods of at least 60 minutes in which all three axes had a standard deviation of less than 13.0 mg. Ground-truth labels based on image data are merged with the epoch-level accelerometry data using timestamps. Where a 30-second accelerometer epoch contains more than one image label, the majority label is used. Epoch-level data is used for subsequent machine-learning analyses.

Figure 4.2: Processing of accelerometer data for machine-learning analysis.



4.2.5 Machine-learning methods

Using the labelled data, a balanced Random Forest (RF) model with a Hidden Markov model (HMM) for time smoothing was trained to classify sleep, sedentary

behaviour, light physical activity behaviours, and moderate-to-vigorous physical activity behaviours.

In each 30-second epoch of the acceleration signal, 50 features were used. These features were based on the features used by Willetts *et al.* [114]. They included time and frequency domain features e.g. mean and kurtosis of the acceleration vector magnitude, and power at different frequencies from the Fast Fourier Transform of acceleration vector magnitude. Willetts *et al.* [114] based the set of features they used on several previous studies [184, 191–193]. Due to concerns that the classifier was affected by device orientation relative to the wrist, the feature set in this work was limited to a subset of 50 rotation-invariant features. The features used are shown in Supplementary Table A.3.

Random Forests are based on decision trees. Decision trees assign class labels based on splits of the data using feature value thresholds (as shown in the example in Figure 4.3). They can be trained using the Classification and Regression Tree (CART) algorithm [194]. In a RF, many decision trees are used. Here, the RF used 100 trees. When training the trees (using the CART algorithm), randomness is introduced by (i) training each tree on a set of data points picked randomly (with replacement) and (ii) restricting the choice of splitting feature to a randomly picked subset of features at each split. To classify a data point using a RF, it is classified by each decision tree. Because they are trained on different subsets of data and use different features, different trees in the RF may classify data points differently. The overall classification given by the RF is the class that is assigned to the data point by the largest number of trees. The use of multiple randomly-differing instances to reduce variance on the output is the technique of bootstrap-aggregating or 'bagging' [195]. Due to a bug in the software used for model development, identified after submission of this thesis, in this case the subset of features available at each split was in fact the full feature set

(i.e. the model effectively consisted of a bagged decision trees).

For this application, a balanced RF was used. The fact some behaviours are much more common than others in the labelled data (e.g. sleep is much more common than moderate-to-vigorous physical activity behaviours) can cause a standard RF, which is trained by picking N examples at random with replacement, to favour assigning common labels at the expense of less common behaviours [114]. Using the balanced RF, if there were n_{rare} examples of the rarest behaviour, n_{rare} examples of each behaviour were picked with replacement to train each tree.

Figure 4.3: An example of a decision tree to classify epochs using average acceleration vector magnitude (avm) and the 75th percentile of acceleration vector magnitude (75thp).

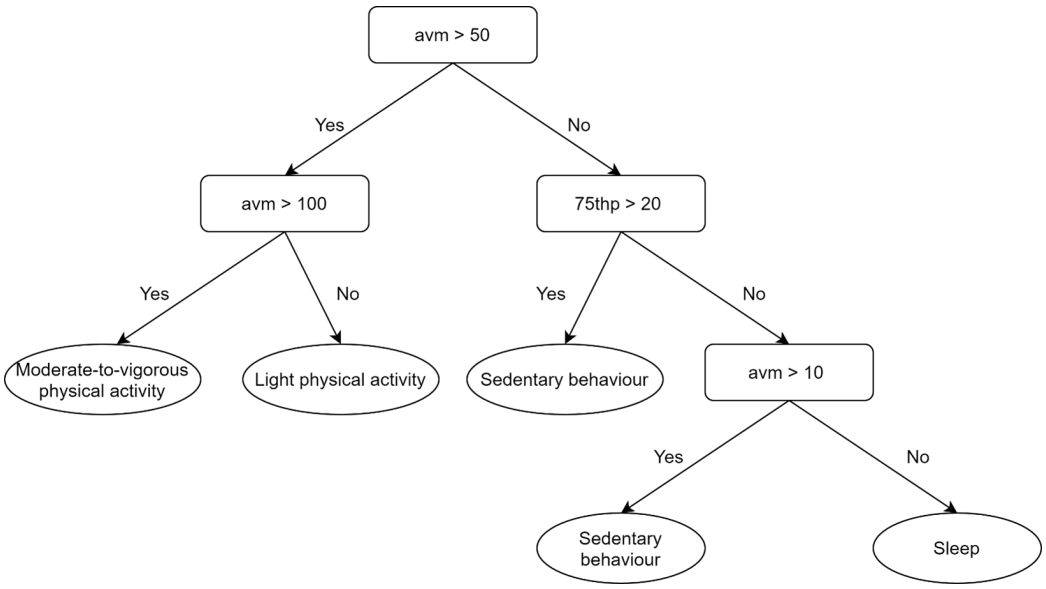
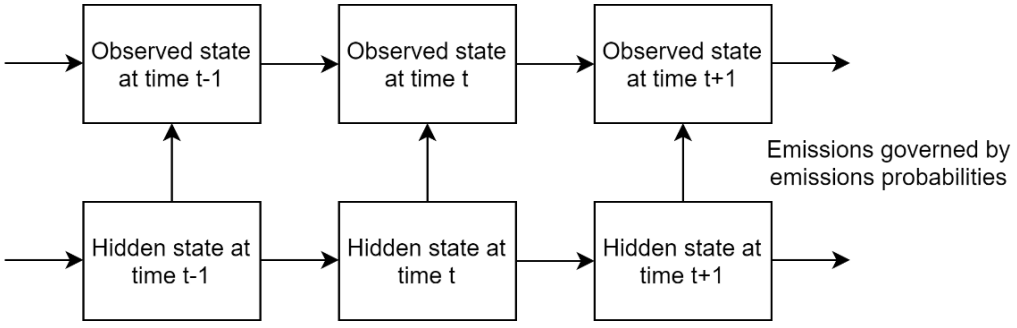


Figure 4.4: The structure of a Hidden Markov Model.



To provide time smoothing of predictions, a Hidden Markov model was used on top of the RF model. In the RF model, individual epochs are assumed to be drawn independently from the underlying distribution. In particular, the time structure in the data is not encoded in the model. In a HMM, there is a sequence of unobserved hidden states, which is assumed to have the Markov property (i.e. future states only depend on past states through present states). This sequence is governed by transition probabilities, which determine the probability of transitioning between each pair of states. There is a sequence of observed states, which depend probabilistically on the sequence of hidden states (described as 'emissions' from the sequence of hidden states; Figure 4.4). Here, the hidden states were the true behaviours, and the emissions were the RF-assigned labels. The Viterbi algorithm, the standard approach to this problem, was used to estimate the most likely true behaviour sequence given the observed sequence of RF-assigned labels [196]. Applying the Viterbi algorithm required estimates of transition probabilities between hidden states and emission probabilities of observed states from hidden states. Transition probabilities between behaviour pairs were estimated using the proportions of transitions that occurred between each behaviour pair in the labelled data. To estimate emission probabilities, epochs were first classified using out-of-bag predictions from the RF i.e. trees were used to classify data points on which they were not trained. This mimics use on unseen data, without requiring additional data. Emission probabilities were then estimated using the proportions of different pairs of true behaviour and RF out-of-bag estimate.

By using this HMM to estimate the most likely true behaviour sequence given the RF-assigned labels, a more plausible sequence of states was obtained. The HMM acts to re-label behaviours forming unrealistic sequences likely to be attributable to misclassification (e.g. short periods of sleep during moderate-to-vigorous physical activity behaviours). This has been shown to noticeably improve model performance on this classification problem [114].

4.2.6 Evaluating performance

Performance metrics were calculated using Leave-One-Participant-Out Cross-Validation. In Leave-One-Participant-Out Cross-Validation, for each participant, a model is trained on all other participants' data (i.e. with this participant's data left-out). The trained model is then used to label the left-out participant's data and evaluation metrics are calculated. This is repeated for all participants, and metrics are aggregated or calculated across all participants. Leave-One-Participant-Out Cross-Validation allows evaluation of the performance of the models on data not used in training, while retaining as much data as possible for use in training. Moreover, all available data may be used in training the final model used for classification. This makes it particularly relevant for human movement behaviour classification, where even in large studies the quantity of labelled data for training is more limited than in some other domains (e.g. image recognition) [197, 198]. Notably, it amends the more usual Leave-One-Out Cross-Validation to group data within participants. This is useful because variation at the participant level may be an important influence on classifier performance. If data from a particular participant is in both the training and test set for a cross-validation fold, this may act as a form of contamination between the training and test set and result in an over-estimation of performance in unseen data.

For model performance, several evaluation metrics were used. First, overall accuracy across all behaviours was reported. This was calculated within each participant, with the mean across participants reported. Accuracy is a simple, intuitive metric of model performance, describing the proportion of 30-second epochs that were correctly classified. Using mean per-participant accuracy, rather than aggregate accuracy over all data, prevents the result being dominated by performance on a few participants with larger amounts of data (important as there may be inter-individual differences

in classification performance). Second, Cohen's kappa, a metric of inter-rater reliability, was reported. Again, this was calculated within each participant, with the mean across participants reported. Cohen's kappa evaluates how much higher the agreement between two raters (here, annotator assigned 'ground-truth' label and model-assigned label) is than that which would be achieved by chance, given the proportions in each class. It has the advantage over accuracy that it takes into account the proportions in each class (in particular, in data where some classes are dominant, a classifier assigning solely to the dominant classes can achieve high accuracy but not high Cohen's kappa). Third, precision and recall for each behaviour were reported. Again, these were calculated for each participant, with the mean across participants reported. Precision for a given behaviour is the proportion of examples labelled by the model as that behaviour which are 'true' examples of that behaviour. Recall for a given behaviour is the proportion of 'true' examples of that behaviour labelled as that behaviour. Taking the mean across participants prevents performance being dominated by performance on participants with larger amounts of data. However, it also upweights the contribution of individuals with very small amounts of data for a given behaviour. Therefore, precision and recall were additionally calculated after excluding participants with up to 20 true minutes in the behaviour. Note also that all precision calculations must also exclude participants with no labelled time in a given behaviour, as precision is undefined in that case.

To compare the model with the traditional approach of cut-point-based classification, performance for classification of moderate-to-vigorous physical activity behaviours using this model was compared with using the cut-point in the labelled data. It was not possible to compare performance for sleep, sedentary behaviour, and light physical activity behaviours due to a lack of published, validated cut-points for these phenotypes in wrist-worn Axivity AX3 data.

To assess the performance of this model by age and sex subgroup, the performance metrics were recalculated in participant subgroups. Unfortunately, the data on demographic and participant characteristics available in the CAPTURE-24 study is limited to age and sex, so these analyses considered performance in men and women, and in participants older than the median age in the dataset (38 years). To further evaluate how model performance depended on age an additional model was trained using only data from participants aged 38 years or older.

4.2.7 UK Biobank

Having developed a model to classify movement behaviours in accelerometer data, it can be applied in large-scale epidemiological cohort studies. While the main goal of applying the model in UK Biobank data is to produce movement behaviour measurements for subsequent epidemiological studies, it also provides an important opportunity to further assess the developed model in context.

UK Biobank accelerometer data was processed using the 'biobankAccelerometerAnalysis' tool, and the machine-learning model was applied to classify movement behaviours based on signal features in 30-second epochs [4]. The data processing pipeline was as described above (Figure 4.2, Section 4.2.4; see also Doherty *et al.* [93]). Non-wear time was again identified as periods of at least 60 minutes in which all three axes had a standard deviation of less than 13.0 mg, and was imputed with the mean behaviour in the same minute of the day on other days.

As is standard in analyses using UK Biobank accelerometer data, data from participants with accelerometer data quality problems was excluded from subsequent analyses. To match other work with UK Biobank, exclusions used accelerometer variables available in tabular data (Category 1008 in the UK Biobank Data Showcase; derived

previously using 5-second epoch data) [93]. Specifically, participants were excluded if: their data could not be calibrated; they had more than 1% of epochs 'clipped' (i.e. beyond the dynamic range of the device) before or after calibration; they had poor wear time (less than 3 days data overall, or no data in at least one hour of the 24-hour cycle); or they had extreme values of overall average acceleration (overall average acceleration > 100 mg).

For maximum relevance to later analyses, the cohort included in descriptive tables and plots was limited to those who will be included in the analytic cohort in Chapter 5. Specifically, participants with prevalent cardiovascular disease and with missing data in key variables were excluded (see Chapter 5). Details on how UK Biobank data was processed were given in Chapter 3.

To summarise patterns of movement behaviours by participant characteristics, median and inter-quartile range of time in each movement behaviour within different participant groups were calculated. Face validity is 'the extent to which a measure looks like it will, or appears to, provide the desired information' [186]. To assess face validity, profiles of behaviours over the day were plotted. As human movement behaviours are patterned, most notably featuring sleeping overnight, these plots provide a visual check that model results are reasonable. Self-report questions in UK Biobank were also used to further explore model face validity. This approach has limitations, particularly as there is a time gap between the baseline assessment where the self-reported variables were collected and the device wear. Nonetheless, participants self-reporting they find it easier to get up in the morning might be expected to wake on average earlier than participants self-reporting that they find this difficult. Similarly, patterning may be expected by job characteristics and self-reported napping.

4.2.8 Code and data availability

Code and details of analyses are available at `github.com/OxWearables/manuscript_ml_behaviours_cvd_2021`. Most data processing and analysis was performed in Python 3.6.6 [199], with R 4.0.5 used for some additional analyses (mostly plotting) [200]. Development of machine-learning models and analyses were conducted using the 'biobankAccelerometerAnalysis' tool [4]. This project contributed to the maintenance and development of this tool. In particular, the machine-learning model developed in this chapter is now available through this tool. Data from the CAPTURE-24 study is openly available (see Appendix B, which provides a summary of code and data used throughout this thesis).

4.3 Results

4.3.1 Performance in labelled data

The available information on characteristics of participants in the CAPTURE-24 study is described in Table 4.1. The machine-learning method accurately classified movement behaviours in accelerometer data: when evaluated using Leave-One-Participant-Out Cross-Validation in 2,501 hours of free-living data from the CAPTURE-24 study, mean per participant accuracy was 88% (95% CI: 87, 89) and mean per-participant Cohen's kappa was 0.80 (95% CI: 0.79, 0.82). Mean per-participant precision and recall for each behaviour show most examples of all behaviours were correctly classified, with highest performance for sleep (Figure 4.5). As expected, mean precision and recall were slightly higher when removing individuals with few true examples of a behaviour (Figure 4.5). Misclassifications of behaviours

showed expected patterns (Table 4.2): for example, 94% of misclassified moderate-to-vigorous physical activity behaviour time was classified as light physical activity behaviours, with misclassifications to sedentary behaviour and sleep very rare.

Table 4.1: Characteristics of participants in the CAPTURE-24 study.

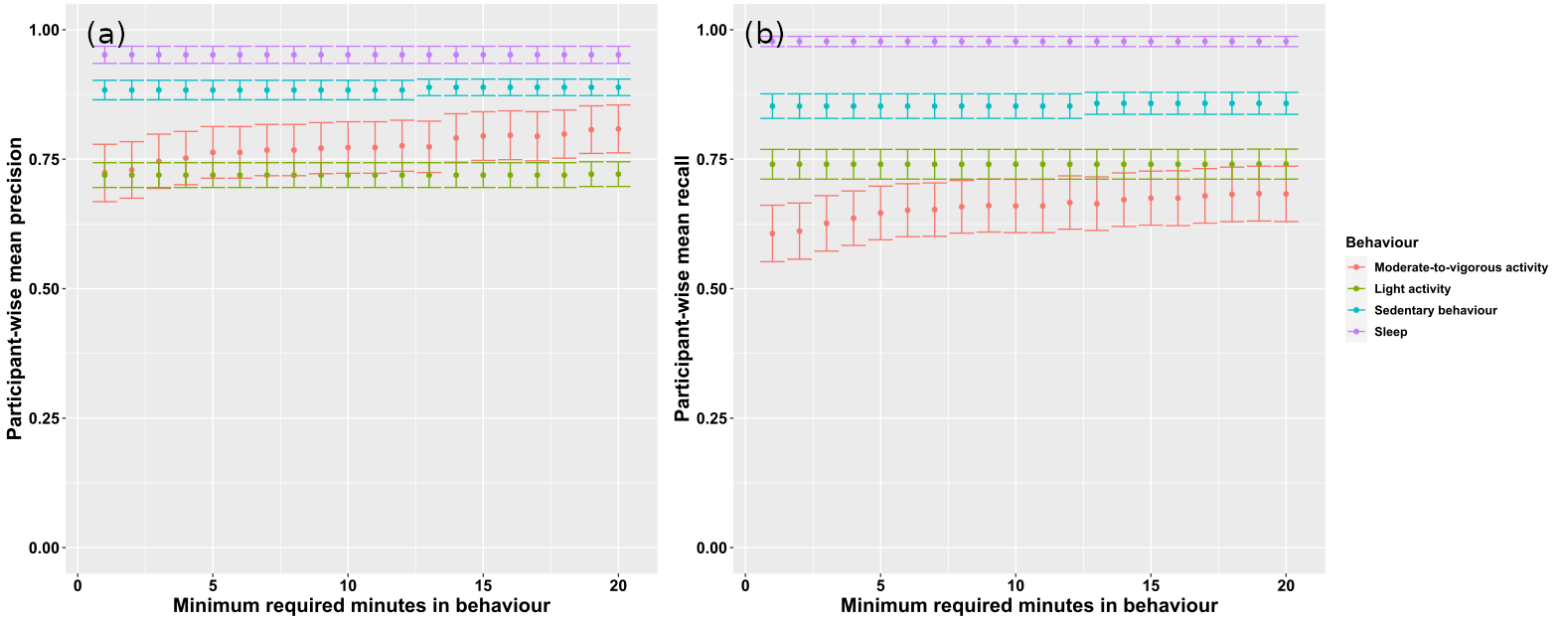
	n (%)
Overall	152
Age (years)	
18-29	39 (26)
30-39	41 (27)
40-49	24 (16)
50-59	21 (14)
60-69	17 (11)
70-79	7 (5)
80-91	2 (1)
Not recorded	1 (1)
Sex	
Female	99 (65)
Male	53 (35)

Table 4.2: Minute-wise confusion matrix for machine-learned classification of behaviours in accelerometer data from 152 CAPTURE-24 participants in Leave-One-Participant-Out Cross-Validation.

Model-assigned label:	**Sleep**	**Sedentary behaviour**	**Light physical activity behaviours**	**Moderate-to-vigorous physical activity behaviours**
'Ground truth':				
Sleep	51,347	980	215	0
Sedentary behaviour	2,322	53,052	5,717	87
Light physical activity behaviours	54	4,986	22,217	1,533
Moderate-to-vigorous physical activity behaviours	6	158	2,434	4,978

Performance was fairly consistent across age and sex groups. In the 72 participants aged 38 years or older, mean per-participant accuracy was 86% (85, 88) and mean per-participant Cohen's kappa was 0.79 (0.76, 0.82). Performance was also consistent across sex groups: mean per-participant accuracy was 88% (87, 89) in women and

Figure 4.5: Participant-wise mean (a) precision and (b) recall for classification of behaviours from accelerometer data calculated in Leave-One-Participant-Out Cross-Validation (with 95% confidence interval for the mean).



The x-axis gives the minimum required recorded annotator-labelled time in the behaviour for inclusion in the calculation. For precision, participants with no model-labelled time in the behaviour were also excluded as precision is undefined in this case.

87% (85, 89) in men; mean per-participant Cohen's kappa was 0.81 (0.79, 0.83) in women and 0.79 (0.77, 0.82) in men. The model trained only on participants aged 38 years or older showed similar performance in Leave-One-Participant-Out-Analysis: the mean per-participant accuracy was 86% (84, 88) and the mean per-participant Cohen's kappa was 0.79 (0.76, 0.81).

The machine-learning based method appeared to out-perform cut-point-based approaches for the same task in this dataset. While the different behaviours classified preclude an overall comparison, the machine-learning model identified moderate-to-vigorous physical activity behaviours with substantially higher precision (overall precision 0.75 vs 0.37) and similar recall (overall recall both 0.66) compared to using the standard 100 mg cut-point. The machine-learning based classification also had higher face validity in UK Biobank, with median 25 minutes/day in MVPA according to the model compared with 1.5 hours/day with the standard cut-point.

4.3.2 UK Biobank

The model-classified movement behaviours of UK Biobank participants are shown in Table 4.3. Patterns of movement behaviours by participant characteristics are plausible. For example, older participants spent less time in moderate-to-vigorous physical activity behaviours than younger participants. Participants with higher Body Mass Index (BMI) spent less time in light physical activity behaviours and moderate-to-vigorous physical activity behaviours than participants with lower BMI, and spent more time sedentary. Women spent more time in light physical activity behaviours than men, but spent less time in both moderate-to-vigorous physical activity behaviours and sedentary behaviour.

Table 4.3: Movement behaviours of 87,498 UK Biobank participants by participant characteristics. Movement behaviours are given as median (Inter-Quartile Range).

	n (%)	**Sleep** (hr/day)	**Sedentary behaviour** (hr/day)	**Light physical activity behaviours** (hr/day)	**Moderate-to-vigorous physical activity behaviours** (min/day)
Overall	87,498 (100)	8.6 (7.9,9.3)	9.2 (8.0,10.3)	5.5 (4.5,6.7)	25 (12,44)
Age at accelerometer wear (years)					
40-49	7,767 (9)	8.5 (7.8,9.1)	9.4 (8.1,10.6)	5.4 (4.3,6.6)	30 (16,50)
50-59	26,081 (30)	8.5 (7.8,9.2)	9.3 (8.1,10.5)	5.4 (4.4,6.7)	28 (14,47)
60-69	38,774 (44)	8.6 (8.0,9.3)	9.0 (7.9,10.2)	5.6 (4.6,6.7)	25 (12,43)
70-79	14,876 (17)	8.6 (7.9,9.4)	9.2 (8.1,10.3)	5.5 (4.5,6.7)	20 (9,36)
Sex					
Female	50,882 (58)	8.6 (8.0,9.3)	8.9 (7.8,10.0)	5.8 (4.8,7.0)	22 (10,38)
Male	36,616 (42)	8.4 (7.8,9.2)	9.6 (8.4,10.8)	5.1 (4.1,6.2)	31 (16,52)
Ethnicity					
Asian	756 (1)	8.4 (7.7,9.3)	9.3 (7.9,10.5)	5.6 (4.5,6.9)	19 (9,35)
Black	701 (1)	8.2 (7.3,9.0)	9.4 (8.0,10.6)	5.9 (4.8,7.2)	21 (10,35)
Mixed/Other	1,151 (1)	8.4 (7.6,9.1)	9.2 (8.0,10.5)	5.7 (4.5,7.0)	26 (13,44)
White	84,890 (97)	8.6 (7.9,9.3)	9.2 (8.0,10.3)	5.5 (4.5,6.7)	25 (12,44)
Townsend Deprivation Index (by quarter)					
Least deprived (Less than -3.8)	21,913 (25)	8.6 (7.9,9.3)	9.1 (7.9,10.3)	5.6 (4.6,6.7)	24 (12,43)
Second least deprived (-3.8 to -2.5)	21,839 (25)	8.6 (7.9,9.3)	9.1 (8.0,10.3)	5.6 (4.5,6.7)	24 (12,43)
Second most deprived (-2.5 to -0.2)	21,872 (25)	8.6 (7.9,9.3)	9.2 (8.0,10.3)	5.6 (4.5,6.7)	25 (12,44)
Most deprived (More than -0.2)	21,874 (25)	8.5 (7.8,9.2)	9.3 (8.1,10.5)	5.4 (4.3,6.6)	27 (13,47)
Education					
School leaver	19,535 (22)	8.7 (8.0,9.5)	8.9 (7.7,10.1)	5.7 (4.7,6.9)	20 (9,36)
Further education	29,061 (33)	8.6 (7.9,9.3)	9.1 (7.9,10.3)	5.6 (4.6,6.8)	23 (11,41)
Higher education	38,902 (44)	8.5 (7.8,9.1)	9.4 (8.2,10.5)	5.4 (4.4,6.5)	30 (16,50)
Smoking status					
Never smoker	50,888 (58)	8.6 (7.9,9.3)	9.2 (8.0,10.3)	5.5 (4.5,6.7)	26 (13,45)
Ex-smoker	30,717 (35)	8.5 (7.9,9.3)	9.2 (8.0,10.3)	5.6 (4.5,6.7)	25 (12,44)
Current smoker	5,893 (7)	8.6 (7.9,9.3)	9.4 (8.2,10.6)	5.4 (4.3,6.6)	21 (9,39)
Alcohol consumption					
Never drinker	4,745 (5)	8.6 (7.9,9.4)	9.1 (7.8,10.3)	5.6 (4.5,6.9)	20 (9,39)
Less than 3 times per week	39,760 (45)	8.6 (7.9,9.3)	9.2 (8.0,10.3)	5.5 (4.5,6.7)	23 (11,41)
3 or more times per week	42,993 (49)	8.5 (7.9,9.2)	9.2 (8.0,10.4)	5.5 (4.5,6.7)	28 (14,47)
Fruit and vegetable consumption (servings/day)					
Less than 3	3,595 (4)	8.6 (7.8,9.4)	9.7 (8.4,11.0)	5.0 (3.9,6.3)	21 (9,39)
3-4.9	14,293 (16)	8.6 (7.9,9.3)	9.4 (8.2,10.6)	5.3 (4.3,6.5)	24 (12,42)
5-7.9	36,991 (42)	8.6 (7.9,9.3)	9.2 (8.0,10.3)	5.5 (4.5,6.7)	26 (13,44)
8 or more	32,619 (37)	8.5 (7.9,9.2)	9.0 (7.8,10.1)	5.7 (4.7,6.9)	26 (13,45)
Body Mass Index (kg/m^2)					
Underweight (Less than 18.5)	511 (1)	8.5 (7.9,9.2)	8.5 (7.2,9.6)	6.3 (4.9,7.3)	34 (19,55)
Normal weight (18.5-24.9)	35,043 (40)	8.6 (8.0,9.2)	8.8 (7.7,10.0)	5.8 (4.7,6.9)	30 (16,50)
Overweight (25.0-29.9)	35,783 (41)	8.5 (7.9,9.3)	9.3 (8.1,10.4)	5.5 (4.4,6.6)	25 (12,43)
Obese (30 or more)	16,161 (18)	8.5 (7.8,9.3)	9.8 (8.5,10.9)	5.1 (4.1,6.3)	16 (7,31)

Patterns of model-classified movement behaviours over the day also showed face validity in UK Biobank. As shown in Figure 4.6, behavioural profiles of UK Biobank participants over the day are highly plausible: participants were likely to be asleep

overnight and unlikely to be asleep in the day; the probability they were sedentary was low overnight, but rose through the day, peaking in the evening; the probability they were engaged in light physical activity behaviours was very low overnight, rose quite rapidly in the morning, and got lower through the afternoon and evening; and participants were unlikely to be in moderate-to-vigorous physical activity behaviours at all times of day, but were most likely to be so in the middle hours of the day.

Considering UK Biobank participants' behaviours by self-reported sleep behaviours and job type also shows expected patterns, as shown in Figure 4.7. For example, night-shift workers were less likely to be classified as sleeping at night; sleep during daytime hours was more common among participants who self-reported napping; sleep times were shifted later on average in those who reported more difficulty getting up in the morning; those who self-reported a job involving walking/standing or a job involving manual/physical work were more likely to be classified as doing light physical activity behaviours during working hours; and patterns of moderate-to-vigorous physical activity behaviours were more common in conventional working hours for participants reporting a job involving manual/physical work, but less common at some other times of day.

4.4 Discussion

4.4.1 Overall findings

This work found that machine-learning methods trained using free-living data were able to accurately classify movement behaviours in accelerometer data. Performance was fairly consistent by age and sex.

The method developed can be applied to classify movement behaviours in UK Biobank.

Figure 4.6: Behaviour profiles for sleep, sedentary behaviour (SB), light physical activity behaviours (LIPA), and moderate-to-vigorous physical activity behaviours (MVPA) among 87,498 UK Biobank participants.

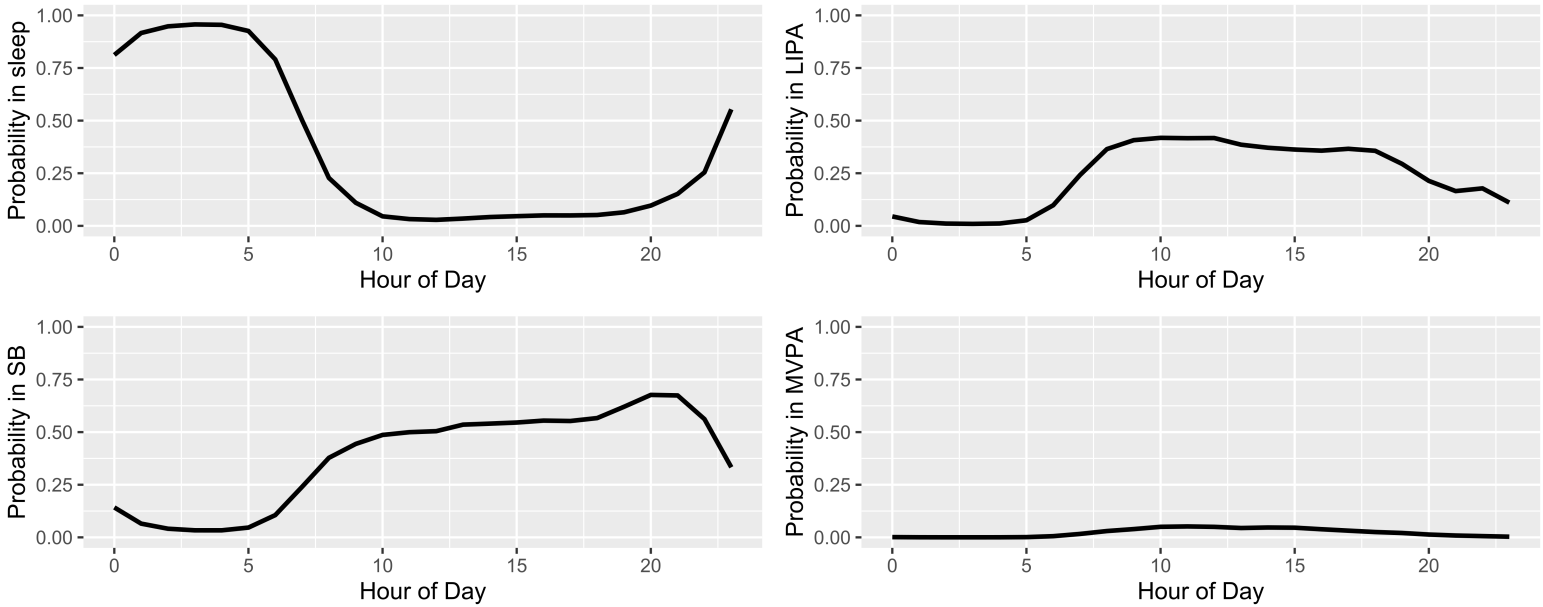
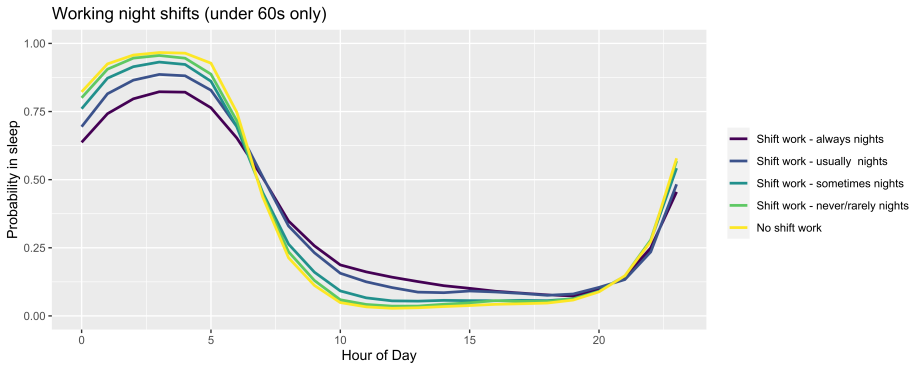


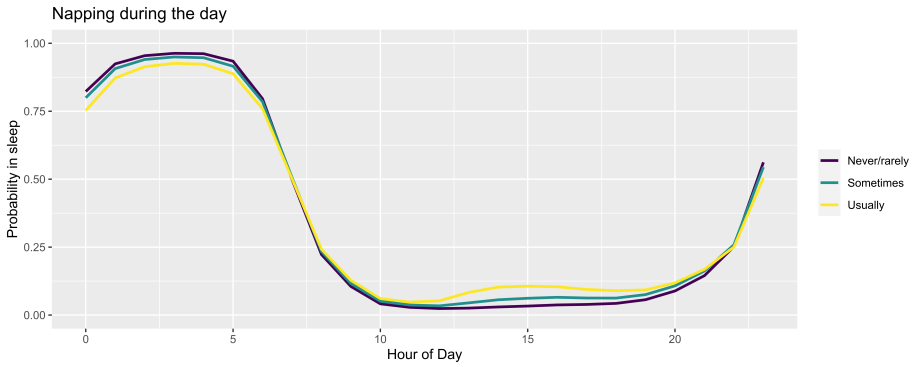
Figure 4.7: Behaviour profiles according to participant characteristics among up to 87,498 UK Biobank participants.

(a) Nighttime sleep was less common among participants reporting night-shift work, with a trend by frequency.

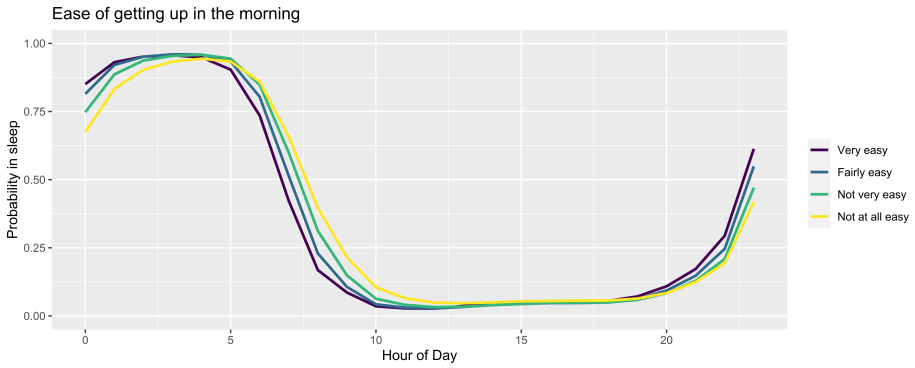


Reported in participants aged < 60 years to reduce probability of participants having left work before accelerometer wear.

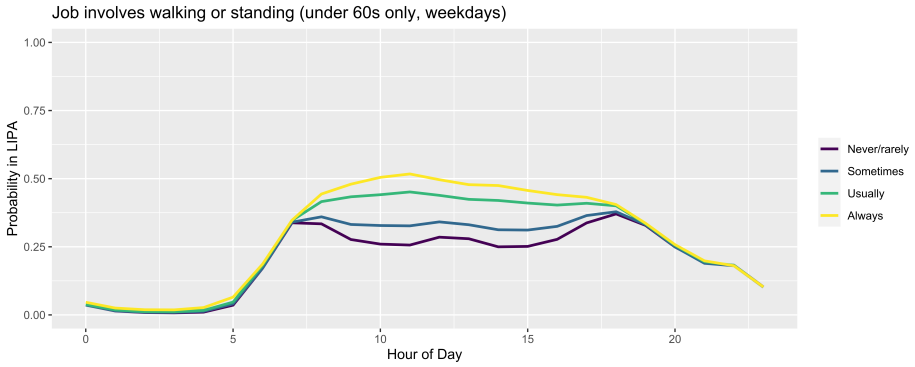
(b) Sleep during daytime hours was more common among participants reporting napping, with a trend by frequency.



(c) Sleep timing showed a trend by self-reported ease of getting up in the morning.

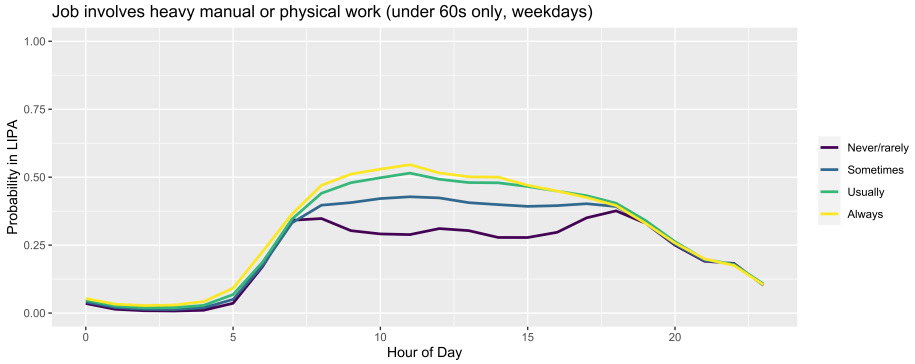


(d) Light physical activity behaviours during conventional working hours were more common for those reporting a job involving walking or standing, with a trend by frequency.



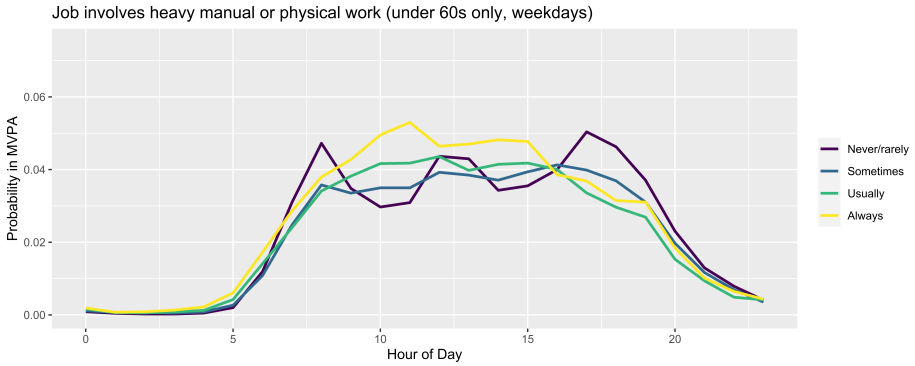
Reported in participants aged < 60 years to reduce probability of participants having left work before accelerometer wear; weekdays only.

(e) Light physical activity behaviours during conventional working hours were more common for those reporting a job involving heavy manual or physical work, with a trend by frequency.



Reported in participants aged < 60 years to reduce probability of participants having left work before accelerometer wear; weekdays only.

(f) Moderate-to-vigorous physical activity behaviours during conventional working hours were more common for those reporting a job involving heavy manual or physical work. Participants reporting a job never/rarely involving heavy manual or physical work were most likely to be in moderate-to-vigorous physical activity behaviours in the morning, around lunchtime, and in early evening.



Reported in participants aged < 60 years to reduce probability of participants having left work before accelerometer wear; weekdays only.

It classifies a set of behaviours which is highly relevant for epidemiological research. When applied to data from the UK Biobank, measured time in different movement behaviours showed high face validity, with expected patterns by time of day and by other participant characteristics.

4.4.2 Comparison with previous work using CAPTURE-24

Given similarities in the underlying methods in this work and the prior work of Doherty *et al.* [181] and Willetts *et al.* [114], it is useful to explore the ways in which this work builds on that earlier work. Willetts *et al.* [114] used data from the CAPTURE-24 study to develop a classifier of sleep, sitting or standing, vehicle time, mixed activity, walking, and bicycling, while Doherty *et al.* [181] developed a classifier of sleep, sedentary time, light tasks, walking, and moderate activity. As in this work, both these previous studies used a Random Forest model as their main classifier, with a Hidden Markov model additionally used for time-smoothing of the data [114, 181]. Nonetheless, several innovations made measurements of movement behaviours in the present study more suitable for subsequent epidemiological work.

Classification performance reported in this work improved on that reported in previous studies. Doherty *et al.* [181] reported an overall kappa score of 0.74 for a different (but closely related) set of behaviour classes in this dataset. Moreover, in this earlier study, performance varied substantially between different behaviours. For example, 41% of epochs labelled by the Doherty *et al.* [181] model as moderate activity were moderate activity in ground-truth labels, while 58% of epochs with a ground-truth label of moderate activity were classified as moderate activity. Poor performance on specific behaviour classes has been cited as an argument against more widespread use of that model [126]. In this respect, while the performance of the classifier developed here does vary between classes, this variation was less extreme and each single

class was well classified. Across all behaviours, the lowest precision was 73% (for light physical activity behaviours) and the lowest recall was 66% (for moderate-to-vigorous physical activity behaviours). The set of behaviours classified by Willetts *et al.* [114] was less closely related to the present work, but performance metrics in this work were comparable to those reported in that paper.

This work also developed a different set of movement behaviours for classification compared to previous studies. This set of behaviours was: sleep, sedentary behaviour, light physical activity behaviours, and moderate-to-vigorous physical activity behaviours. The set of behaviours developed was driven by the intended application in an epidemiological study of 24-hour movement behaviours. Based on the epidemiological literature, behaviour classes were developed which: were of epidemiological interest (e.g. had a hypothesised association with disease); were comprehensive and mutually exclusive; were sufficiently broad to constitute a tractable classification problem; and could be interpreted in the context of current epidemiological literature.

A detailed review of how fine-grained behavioural annotations were assigned to broad behaviour classes may have helped to improve classification performance. This review aimed to ensure fine-grained annotations were assigned to the most appropriate class. Most notably, an annotation associated with food preparation ('home activity;household chores;preparing meals/cooking/washing dishes;5035 kitchen activity general cooking/washing/dishes/cleaning up') appeared frequently in the data, and was associated with a MET value of 3.3 METs. While this MET value should indicate moderate-to-vigorous physical activity, this seems likely to over-estimate energy expenditure in most behaviours assigned this annotation, so this annotation was reassigned (to light physical activity behaviours) by our review of annotation codes. Another frequent example was 'home activity;miscellaneous;walking;5165 (generic) walking non-cleaning task such as closing windows lock door putting away items', which

was associated with a MET value of 3.5 METs. However, in the Compendium providing the reference value it was explicitly qualified as 'moderate effort' [28], whereas here it was considered more likely to have been used for light physical activity. Previously, these commonly-appearing annotations may have distorted classification of moderate-to-vigorous physical activity.

Concerns about sensitivity of early classifiers to orientation of the device and device axes were also addressed. Preliminary work for this thesis investigated face validity of the movement behaviour classification of Doherty *et al.* [181] in UK Biobank data. Some variables showed secular time trends in classified behaviours large enough to affect epidemiological analyses. These trends may have been caused by secular time trends in device orientation relative to the wrist in UK Biobank. If axes are oriented differently, some signal features differ. If there are trends in device orientation over the study, this could result in time trends in classified behaviours. Therefore, restricting the feature set used in this study to only the rotation-invariant features was investigated. This appeared to have little impact on performance in the labelled data, and derived phenotypes no longer showed marked secular time trends. There may also be further advantages to this restricted feature set, including lower risk of overfitting, additional robustness, and better generalisability to a range of devices (most research-grade devices on the market use a triaxial accelerometer, but axis labelling and orientation varies).

This work found similar performance in age and sex subgroups, a finding consistent with the earlier work of Willetts *et al.* [114]. In that study, they reported negligible change in performance when using a model trained in men on women and vice versa, and only small change in performance when using a model trained on younger participants on older participants [114].

4.4.3 Comparison with wider literature

Consistent with the performance of the classifier in this work, Narayanan *et al.* [113] found that 80% of machine-learning-based classifiers reported in the literature achieved an overall accuracy exceeding 85%. Studies achieving over 85% accuracy were much more common when evaluation was in a controlled setting; studies in free-living settings achieved lower overall accuracy in general. Performance in this work can therefore be interpreted as equalling, and indeed slightly exceeding, the standard set in the literature for accelerometer-based behaviour classification.

One important characteristic of this work compared to previous studies is the availability of the classifier for use in other epidemiological studies. As noted above, this remains relatively unusual for machine-learning based behaviour classifiers, but it is crucially important for enabling comparability between studies. Moreover, data classified using a version of this classifier has been returned to UK Biobank and will appear in the main Data Showcase, enabling this definition of movement behaviours to be used in future studies.

4.5 Summary

Machine-learning methods can accurately classify 24-hour movement behaviours in free-living wrist-worn accelerometer data. The methods developed in this work appear to be valid to derive measurements of time spent in sleep, sedentary behaviour, light physical activity behaviours, and moderate-to-vigorous physical activity behaviours from wrist-worn accelerometer data collected in UK Biobank, a large prospective cohort study. The next chapter, Chapter 5, will consider how these measurements can be used to better understand the associations between 24-hour movement behaviours

and incident cardiovascular disease in UK Biobank.

Chapter 5: Device-measured 24-hour movement behaviours and risk of incident cardiovascular disease

The material on which this chapter is based has been published as part of an article in the ***British Journal of Sports Medicine*** *[1].*

Subsequent to submission of this thesis, I observed a discrepancy between the published record censoring date in Wales and the actual distribution of diagnoses in the data. After I reported this to UK Biobank, they corrected the published record censoring date to 2016-03-31. Analyses in this chapter and Chapter 6 were therefore informed by an inaccurate censoring date for participants in Wales. Given the small proportion of participants affected, this is expected to affect results only minimally.

5.1 Introduction

5.1.1 Background

There is still some uncertainty about the associations between 24-hour movement behaviours and incident cardiovascular disease (see Chapters 1 and 2). Each person's day is divided between multiple different movement behaviours (sleep, seden-

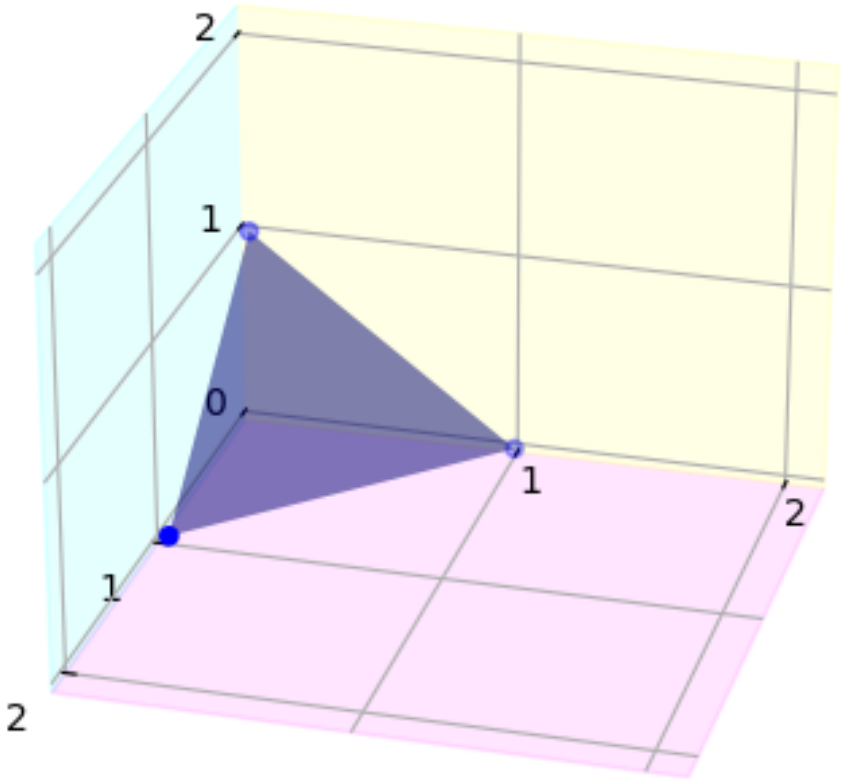
tary behaviour, light physical activity and moderate-to-vigorous physical activity), all of which may affect health. Historically, however, most studies have focussed on moderate-to-vigorous physical activity in isolation. Learning more about effects of 24-hour movement behaviours on cardiovascular disease risk may inform public health and clinical interventions, and open up new behavioural targets for cardiovascular disease prevention [68].

Studying 24-hour movement behaviours in large-scale epidemiological studies has only recently become possible. Before large-scale device-based measurement, movement behaviours were very difficult to measure. Chapter 3 described UK Biobank, a large prospective cohort study in which approximately 100,000 participants have worn an accelerometer to measure movement behaviours. Chapter 4 refined and applied methods to derive time in 24-hour movement behaviours from accelerometer data. Accurate and precise measurement of 24-hour movement behaviours presents new opportunities for epidemiological analysis, but in order to realise these opportunities analytic challenges must be addressed. 24-hour movement behaviour data has a complex structure: it is inherently multidimensional, and divides up the 24-hour day. The 24-hour day constrains how time in different behaviours can vary: increasing time in one behaviour reduces the amount of time available to spend on other behaviours.

The particular structure of 24-hour movement behaviour data can be conceptualised as *compositional data*. Compositional data is multivariable, positive, real-valued data where variables, sometimes known as 'parts', carry only relative information [201]. Perhaps more intuitively, compositional data might be thought of as data that can be expressed as proportions without losing relevant information. Note that sometimes variables carry both relative and absolute information, and may be considered as compositional or non-compositional depending on context [132]. For example, con-

sidering dietary composition, both proportion of total energy intake from different macronutrients and absolute energy intake from different macronutrients may be of interest. Also, although compositional data often arises where variables must have a fixed sum, as in this case variables cannot carry absolute information but only relative information, this is not necessary: relative information in the data may be relevant in types of data where the sum is not fixed [76]. One way to visualise compositional data is shown in Figure 5.1: expressed as proportions of a whole, compositional data can be viewed as a subset of real space, known as the *simplex*.

Figure 5.1: Illustration of the simplex for compositional data with three compositional parts embedded in three-dimensional real space.



With movement behaviours, it is impossible to unilaterally increase time in one behaviour: a person can only alter how they divide up time between different behaviours, not the total time each day. This makes the compositional framework appropriate: it

is relative time in different behaviours, not absolute time in any one behaviour, which can be changed.

As described in Chapter 2, most analyses of device-measured movement behaviours and incident cardiovascular disease have used 'single behaviour' epidemiological methods. That is, they have estimated the effect of higher levels of one behaviour, holding fixed other participant characteristics. Time spent in other behaviours may be adjusted for, but only by 'holding fixed' time spent in these other behaviours. This is conceptually problematic: in the context of compositional data, higher relative time in one behaviour entails lower relative time in other behaviours. Modelling behaviours as freely-varying separate variables also fails to account for their inherent interdependence. Moreover, statistical assumptions of traditional methods are often poorly suited to compositional data [127].

Compositional Data Analysis (CoDA) is a prominent approach to the analysis of compositional data. Rather than a specific method, CoDA describes a general approach based around the principle that relative, not absolute, values of compositional variables should be analysed. To analyse relative values of variables in a statistically rigorous way, CoDA methods typically use log-transformed ratios between variables ('log-ratios'). A CoDA approach can be motivated in different ways, drawing on mathematical and statistical considerations. The next section (Section 5.2) gives a detailed background and motivation.

However, while there is a substantial and rapidly growing evidence base associating 24-hour movement behaviours with cardiovascular risk factors using a Compositional Data Analysis approach, evidence on incident disease outcomes remains lacking. A systematic review published in 2020 identified just one prospective study for any incident outcome using a Compositional Data Analysis approach [140]. While more studies have since been published (e.g. [76, 159, 202, 203]), the literature remains limited.

The review in Chapter 2 identified only one study which had associated movement behaviours with incident CVD using a Compositional Data Analysis approach [76]. More evidence on how the movement behaviour composition of the 24-hour day is associated with incident disease has been identified as a key research need to inform future public health guidelines and interventions [75, 140, 204].

5.1.2 Aims

The aim of this work was to investigate the association between 24-hour movement behaviours and risk of incident cardiovascular disease. In particular, it aimed to do so using methods appropriate to the compositional nature of movement behaviour data (a Compositional Data Analysis approach).

5.2 Compositional Data Analysis: background and motivation

This section provides a more detailed background on the challenges compositional data presents for analysis. It also develops the motivation behind a Compositional Data Analysis approach.

5.2.1 Challenges of compositional data

The structure of compositional data, that is, multivariable, positive, real-valued data where variables carry only relative information, means it is not straightforward to analyse. In the context of epidemiological studies with a compositional exposure, such as 24-hour movement behaviours, a suitable analytic approach must address

several challenges (as introduced in Chapter 1). First, the effect of interest must be identified and appropriately quantified. This is important as unilateral changes in a single variable are impossible (some variables in a composition being relatively higher means others must be relatively lower) [131, 132]. For example, spending more time in physical activity means spending less time in some other behaviours, so the effect to be estimated may be the effect of more physical activity and less sedentary behaviour, more physical activity and less sleep, or more physical activity and less of all other behaviours in given proportions [131, 132]. Second, associations between variables making up the composition must be appropriately conceptualised and accounted for in analyses. Several variables in the composition may be associated in some sense with the outcome, and different variables in the composition are associated with one another. Estimating associations with one behaviour while holding all others fixed, the 'traditional' approach to this problem, is not possible in the compositional context where relatively more time in one behaviour entails relatively less time in one or several other behaviours [127]. Third, the analysis must be well-suited to the particular statistical properties of the compositional dataset being analysed. For example, the sample space of compositional data, the simplex, affects what distributions are possible and common [133, 134].

5.2.2 Compositional Data Analysis

Compositional Data Analysis describes a general approach to analysing compositional data, based around some fundamental principles. As described briefly above, CoDA methods typically use log-transformed ratios to analyse relative values in a statistically rigorous manner. Particular sets of log-ratios can be used for particular applications: for example, 'isometric log-ratio coordinates' form a non-redundant set and are particularly suitable for a wide range of applications (see Sections 5.2.3

and 5.3.4) [127, 201]. A Compositional Data Analysis approach can be applied in many different statistical analyses [205]. This thesis will focus on its application in regression modelling involving a compositional exposure, but will also make use of other definitions in a CoDA framework (such as the 'compositional mean', which is a measure of a central tendency suitable for compositional data [201]).

In this thesis, 'Compositional Data Analysis' is capitalised, although the term 'compositional data analysis' is often seen uncapitalised. Capitalisation is intended to denote that it is a proper name for the approach here and to differentiate between this approach and other possible approaches to the analysis of compositional data (e.g. ignoring compositionality; different transformation-based approaches) [134].

CoDA can be described and motivated in different ways. It is often now developed mathematically based on a geometry for the simplex [201]. A more applied statistical approach, which also resembles the earlier development of CoDA in the work of John Aitchison [201], focuses on the use of ratios to capture relative values and log-transformation of these ratios for better statistical properties. Both motivations are outlined here.

5.2.3 A mathematical foundation for Compositional Data Analysis

Mathematically, CoDA can be described via a geometry for the simplex, the sample space of compositional data (written Δ, shown in Figure 5.1) [201, 206]. There are operations of vector 'addition' (known as perturbation), mapping $\Delta \times \Delta \to \Delta$, and scalar 'multiplication' (known as powering), mapping $\mathbb{R} \times \Delta \to \Delta$. These operations satisfy the properties of a vector 'addition' or scalar 'multiplication' operation, but it should be noted that they are not the familiar operations from real space.

Perturbation is obtained by component-wise multiplication of the two vectors $(\mathbf{x}, \mathbf{y})$ and closing $(\mathcal{C})$ the result to 1 (or another constant) [206]:

$$(x_1, x_2, \ldots, x_D) \oplus (y_1, y_2, \ldots, y_D) = \mathcal{C}(x_1 y_1, x_2 y_2, \ldots, x_D y_D).$$

Powering is obtained by putting each element of a vector $(\mathbf{x})$ to the power of the scalar (α), and again closing $(\mathcal{C})$ the result to 1 (or another constant) [206]:

$$\alpha \otimes (x_1, x_2, \ldots, x_D) = \mathcal{C}(x_1^\alpha, x_2^\alpha, \ldots, x_D^\alpha).$$

In addition, an inner product can be defined as:

$$\langle (x_1, x_2, \ldots, x_D), (y_1, y_2, \ldots, y_D) \rangle = \sum_{i=1}^{D} \ln \left(\frac{x_i}{(x_1 x_2 \ldots x_D)^{\frac{1}{D}}} \right) \ln \left(\frac{y_i}{(y_1 y_2 \ldots y_D)^{\frac{1}{D}}} \right).$$

With these operations and inner product, the simplex is a Hilbert space. The inner product gives rise to a norm and so can define distances on the simplex.

This geometry is known as the Aitchison geometry on the simplex, and it satisfies desirable properties for an appropriate geometry for compositional data, including:

1. **Scale invariance.** As only the relative values of the variables, and not their absolute values, are informative, results should not change when the variables making up the composition are multiplied by the same scalar [201].

2. **Permutation invariance.** The same composition is described whichever order the compositional parts are presented in; results should not change if the order of the parts is changed [201].

3. **Sub-compositional coherence.** This consists of two aspects. First, sub-compositional dominance describes the property that the distance between two

sub-compositions should be no greater than the distance between the compositions they form part of [201]. Second, results should not change when a non-informative part of the composition is left out of analysis [201].

Being a finite-dimensional Hilbert space, the simplex equipped with this norm is isometrically isomorphic to standard Euclidean space. In practice, rather than working with the data directly in the simplex, data is transformed to Euclidean space so that standard techniques may be applied (equivalently, this can be considered as representing the data in coordinates with respect to an orthonormal basis of the simplex [206, 207]). Taking the log-transformation of ratios between parts maps the simplex into real space. By appropriate choice of the transformation, this mapping can be an isomorphism between the simplex for D parts and $(D-1)$-dimensional Euclidean space. As just noted, it is further possible to choose the transformation to be isometric [206]; that is, such that distances are preserved. This is the so-called *isometric log-ratio (ilr) transformation*. There are different possible isometric log-ratio transformations, corresponding to the choice of orthonormal basis in the simplex which is mapped to the standard basis for $\mathbb{R}^{(D-1)}$ (the different possible sets of isometric log-ratio coordinates can therefore be thought of as representing 'rotations' of one another). Non-distance-preserving isomorphisms, such as the so-called additive log-ratio transformations, may be used for applications such as regression modelling with a compositional exposure, but not for distance-sensitive applications. Isometric isomorphisms onto strict subspaces of real space, such as the so-called centred log-ratio transformation, may be used for distance-sensitive applications, but not for applications such as regression modelling where singularity may be a problem.

5.2.4 Statistical motivations for Compositional Data Analysis

An approach based on log-ratio transformation can also be motivated via more statistical considerations. Using ratios between variables allows information about relative values of variables to be captured, and information about absolute values dropped. By choosing an appropriate set of ratios, all of the relative information in the data can be captured without redundancy. Ratios of positive numbers take only positive values, and typically have a skewed distribution (all ratios $\frac{a}{b}$ where $a < b$ are less than 1, but for $b < a$, $\frac{a}{b}$ can take any value greater than 1). Empirically, log-transformation of ratios of compositional variables has generally been found to make distributions more normal and facilitate statistical analysis using standard methods.

In the particular context of regression modelling with a compositional exposure, a CoDA approach can answer the challenges outlined in Section 5.2.1, and so enable better answers to epidemiological questions. By design it is possible to adjust the model for all log-transformed ratios making up a non-redundant set, and so for the whole composition. By using log-transformed ratios, effects studied relate to differences in the relative values of compositional variables. Beyond directly interpreting coefficients of particular log-transformed ratios [159], interpretable effects are commonly obtained by considering different reallocations of times between parts of the composition [125]. Considering distributions and statistical properties, empirically models using a CoDA approach appear to fit the data well, and show some expected properties e.g. non-linear associations where moderate-to-vigorous physical activity levels are low [129].

5.2.5 Difficulties with a Compositional Data Analysis approach

There are some obstacles when applying a Compositional Data Analysis approach in the context of 24-hour movement behaviours, relating to treatment of zeroes, communication of results, and other modelling assumptions.

Zero values for parts of the composition (e.g. no measured time in moderate-to-vigorous physical activity behaviours) do not technically fall within the definition of compositional data, which refers to positive-valued data. Zero values also cannot be used in log-ratios (the log-ratio would be undefined), and so workaround approaches are used, with the appropriate approach depending on the source of the zeroes. 'Rounded' zeroes relate to measurement precision: even where no time in a given behaviour was observed, had the wear time been long enough, or time resolution of the measurement short enough, some time in the behaviour would be expected. If data contains rounded zeroes, they can be imputed as small positive values [208]. In contrast, 'true' zeroes occur where no matter the precision of the measurement, a zero for that part would be observed. For example, this may occur in movement behaviour research if someone is physically unable to take part in certain behaviours. If data contains true zeroes, participants with a true zero in a particular behaviour should be excluded from the main analysis and analysed separately.

Communicating models using a CoDA approach to the exposure can be challenging, as model coefficients relate to log-ratio transformed variables and hence are difficult to interpret directly. Nonetheless, several approaches have been developed; the approaches taken in this work will be outlined in Section 5.3.6. Also, while suggestive language is often used to communicate analyses using a Compositional Data Analysis approach (e.g. referring to 'reallocating time' [125]), it should be emphasised that

where data remains observational and based on movement behaviour measurement at a single timepoint, it cannot directly speak to within-person changes.

More generally, while Compositional Data Analysis is highly appropriate and useful, it does not address all the challenges of working with 24-hour movement behaviour data. The assumptions of statistical methods still apply when these methods are used in conjunction with a Compositional Data Analysis approach, and these assumptions may or may not be met according to the particular problem at hand. For example, if using a linear regression model, the assumption of linearity is still made; but now the assumption is of linearity in the transformed variables, rather than in the untransformed variables.

5.3 Methods

5.3.1 UK Biobank

This work used the UK Biobank dataset, described in detail in Chapter 3. UK Biobank is a population-based prospective cohort study of over 500,000 participants in England, Scotland and Wales. At baseline, participants attended an assessment involving a touchscreen questionnaire, biological sampling, an interview by a trained interviewer, and anthropometric measurements [161]. Between June 2013 and December 2015, participants with a valid email address were invited to wear an accelerometer. 106,053 consenting participants were sent an Axivity AX3 wrist-worn triaxial accelerometer to be worn on the dominant wrist for 7 days [93]. For the purpose of this work, readable accelerometer datasets were available for 103,683 participants. Protocols for UK Biobank data collection, particularly with respect to the accelerometry data, were described in detail in Chapter 3.

5.3.2 Movement behaviour phenotypes from accelerometer data

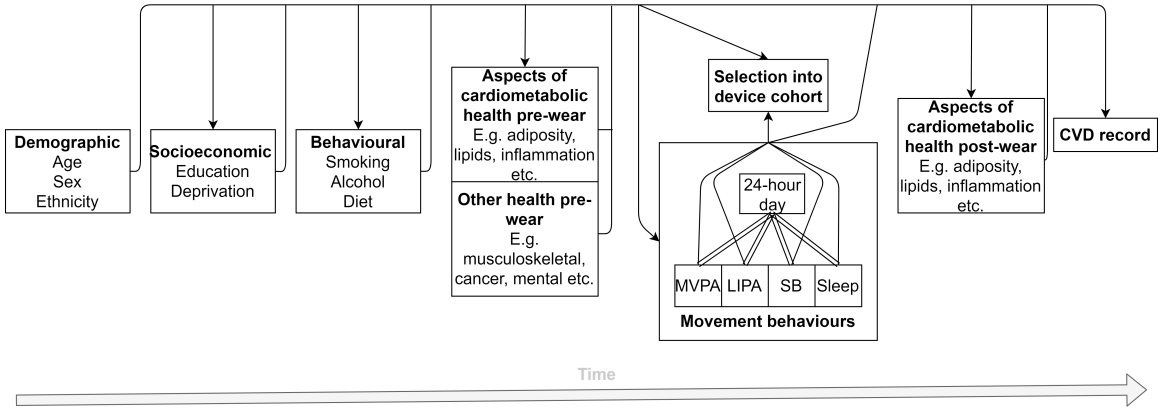
The machine-learning model developed in Chapter 4 was applied to classify accelerometer data of UK Biobank participants as sleep, sedentary behaviour, light physical activity behaviours, or moderate-to-vigorous physical activity behaviours. Data processing and data quality exclusions were for accelerometer data were described in detail in Chapter 4. Briefly, a standard accelerometer data processing pipeline was used to process accelerometer data (see Figure 4.2). Standard quality control criteria were applied to exclude participants with poor calibration, insufficient wear time, or implausible values [93]. The machine-learning model developed in Chapter 4 was then applied to classify behaviours.

5.3.3 Confounders

Preprocessing of variables used for model adjustment, including recategorisation where appropriate, was described in Chapter 3.

Analyses were adjusted for potential confounders. Variables were chosen based on existing literature, hypothesised causal relationships, and epidemiological theory. Statistical variable selection procedures were not used, as these procedures are most relevant where power considerations substantially constrain the number of variables which may be included. Variables were only adjusted for where they could be considered confounders. Variables which may mediate associations between movement behaviours and cardiovascular disease, such as cardiometabolic biomarkers, were not adjusted for. Assumptions about relationships between the variables in the model are illustrated in Figure 5.2.

Figure 5.2: Schematic diagram describing possible causal relationships between variables in the model. Notation for compositional variables is based on Arnold *et al.* [132].



Measures of body size/ adiposity require careful consideration in the context of 24-hour movement behaviours. High adiposity may be on the causal pathway between 24-hour movement behaviours and cardiovascular disease. Adiposity may also causally affect 24-hour movement behaviours [209]. With a single timepoint of measurement, this is difficult to address analytically: adjusting for adiposity may under-estimate the total effect of movement behaviours on cardiovascular disease risk (by excluding the indirect effect mediated by adiposity), but not adjusting for adiposity may over-estimate this effect (by not controlling confounding by adiposity). These considerations also affect some other variables describing aspects of cardiometabolic health. For example, glycaemic control may similarly act both as a confounder and as a mediator.

Age was used as the timescale for analyses [210]. Use of age as the timescale in a Cox proportional hazards regression model means that participants' timelines are aligned on age (with observation time therefore beginning at different points), and means that the risk set at each event time is participants at the same age as the participant who experienced an event at that time [210]. This is useful in a study such as the present study, where age is highly relevant to cardiovascular risk, but time-on-study

has limited relevance to cardiovascular risk [211] and calendar time is unlikely to be highly relevant over this limited calendar period (2013-2021). Using age as the timescale also means the parametric form of the association between age and risk does not need to be determined.

5.3.4 Compositional Data Analysis approach

As described in Section 5.1.1, a Compositional Data Analysis approach aims to analyse relative information in compositional data, and, in practice, this is typically achieved using log-ratio transformation. There are different log-ratio transformations which may be used in the context of regression analysis, including additive log-ratio transformation and isometric log-ratio transformation. As isometric log-ratio coordinates have elegant mathematical properties and hence are suitable for the widest range of analyses [201], they are typically used in movement behaviour research [127].

Isometric log-ratio coordinates can be set up in different ways. For example, the method of balances may be used to analyse particular behavioural trade-offs [135]. So-called 'pivot' coordinates, which sequentially consider one behaviour relative to remaining behaviours, are commonly used. The particular set of isometric log-ratio

'pivot' coordinates used in this work were:

$$z_1 = \sqrt{\frac{3}{4}} \ln \frac{sleep}{(SB.LIPA.MVPA)^{\frac{1}{3}}}$$
$$= \sqrt{\frac{1}{12}} \ln \frac{sleep}{SB} + \sqrt{\frac{1}{12}} \ln \frac{sleep}{LIPA} + \sqrt{\frac{1}{12}} \ln \frac{sleep}{MVPA}$$

$$z_2 = \sqrt{\frac{2}{3}} \ln \frac{SB}{(LIPA.MVPA)^{\frac{1}{2}}}$$
$$= \sqrt{\frac{1}{6}} \ln \frac{SB}{LIPA} + \sqrt{\frac{1}{6}} \ln \frac{SB}{MVPA}$$

$$z_3 = \sqrt{\frac{1}{2}} \ln \frac{LIPA}{MVPA}$$

As noted in Section 5.2.5, one difficulty when using a Compositional Data Analysis approach is how to approach zero values in the data, which cannot be used directly in log-ratios. In this work, where no time in a particular movement behaviour was observed, this was treated as a rounded zero (that is, considered to relate to measurement precision). They were therefore imputed using the log-ratio Expectation Maximisation algorithm, using the implementation in the 'zCompositions' R package [208]. Assuming multivariate normality of the data in coordinate space (i.e. multivariate normality of the transformed variables), this imputation method uses information from the covariance structure of the data to impute zero values with positive values [208, 212]. The initial covariance matrix was estimated using complete observations and the detection limit was set as the minimum non-zero value observed in the analytic dataset (0.0001 on the unitless scale in the full dataset, corresponding to 0.14 min/day). No participants recorded a zero value for sleep, sedentary behaviour, or light physical activity behaviours, while 1% recorded a zero value for moderate-to-vigorous physical activity behaviours.

5.3.5 Statistical analyses

Multivariable-adjusted Cox proportional hazards regression models were used to investigate the association between the movement behaviour composition and incident cardiovascular disease. The movement behaviour composition was modelled using isometric log-ratio coordinates (i.e. using a Compositional Data Analysis approach). As outlined above (Section 5.3.3), models used age as the timescale and were adjusted for potential confounders.

Appropriateness of adjustment with respect to the proportional hazards assumption was tested component-wise and globally using the Grambsch-Therneau test with the Kaplan-Meier transformation (at the 5% level). Plots of the Schoenfeld residuals were also examined. Where there was evidence of non-proportional hazards, notably for sex, adjustment was by stratification of the Cox model [213]. There was no evidence that other variables in the main analysis violated the proportional hazards assumption.

A minimally adjusted analysis used age as the timescale and was stratified by sex but had no further adjustment for potential confounders. To address potential sources of confounding, the main analysis used age as the timescale, was stratified by sex and adjusted for ethnicity (Asian, Black, Mixed/Other, White), education (school leaver, further education, higher education), deprivation (quarter of Townsend Deprivation Index in the study population), smoking status (current, ex-, never smoker), frequency of alcohol consumption (never, $<$ 3 times/week, 3+ times/week), fresh fruit and vegetable consumption ($<$ 3, 3-4.9, 5-7.9, 8+ servings/day), frequency of red and processed meat consumption ($<$ 1, 1-2.9, 3-4.9, 5+ times/week), and frequency of oily fish consumption ($<$ 1, 1, 2-4, $>$ 4 times/week).

As described in Section 5.3.3, adiposity may mediate associations between movement

behaviours and CVD. Therefore, the main analysis was not adjusted for BMI. However, BMI may also act as a confounder for associations between movement behaviours and CVD. Therefore, an additional analysis was performed adjusting for BMI in addition to the variables outlined above. As there was evidence that BMI violated the proportional hazards assumption, this adjustment was performed by categorising BMI and stratifying the Cox model by BMI category ($<$ 25, 25-30, 30+ kg/m^2).

To investigate associations with fatal cardiovascular events, a further multivariable-adjusted analysis was performed with fatal cardiovascular events as the outcome. To investigate consistency of associations between subgroups, separate analyses were also performed in women and men, and in those aged under 65 years and over 65 years at the time of accelerometer wear.

5.3.6 Presentation of results

As the coefficients in the model relate to the isometric log-ratio pivot coordinates, interpreting them is not straightforward. The first coordinate describes the balance between sleep and all other behaviours. Therefore, the coefficient of the first coordinate can be interpreted in terms of the effect of reallocating time to sleep from all other behaviours proportionally. If the coefficient of the first coordinate is greater than 0 (hazard ratio greater than 1) reallocating time to sleep from all other behaviours proportionally is associated with higher risk of cardiovascular disease. However, the second and third coordinates are harder to interpret analogously.

Therefore, to interpret the effect of reallocating time to each behaviour from all others proportionally, and following a standard approach in movement behaviour research, one model per behaviour was produced, with different first pivot coordinate [127]. Mathematically, the coordinates in these models represent rotations of the coordinate

system [205]. As such, they are in some sense equivalent, and overall aspects of model performance are identical [205]. However, they differ in the prioritisation of a particular movement behaviour for interpretation. Therefore, this approach can be used to present a set of model parameters interpretable in terms of reallocation to/from each behaviour (the 'first coordinate' from each of the models). It should be noted, however, that these coefficients do not parametrise a single model [205].

However, even using this approach, the magnitudes of the coefficients are hard to interpret. Therefore, model estimates of the hazard ratio at different compositions relative to the mean behaviour composition were also reported. In particular, results are described by pairwise time reallocation plots, which show the HR associated with reallocating time from one behaviour to another behaviour. A plot showing the HR associated with particular reallocations of time between behaviours (e.g. reallocating 1 hour/day to sedentary behaviour from all other behaviours proportionally) is also used. Reallocation results are obtained by using the model to estimate the outcome associated with differences in values of the compositional exposure variables relative to the mean behaviour composition (i.e. for different reallocations of time between behaviours) [125, 129, 159].

In detail, suppose we have a Cox regression model in the isometric log-ratio pivot coordinates laid out above:

$$\ln HR = \ln\left(\frac{h(t)}{h_0(t)}\right) = \beta_1 z_1 + \beta_2 z_2 + \beta_3 z_3 + \sum_{i=1}^{k} \gamma_i covariate_i$$

and that the value of the coordinates at the mean behaviour composition is $(\bar{z}_1, \bar{z}_2, \bar{z}_3)$. Note that this notation is not misleading, as the value of the coordinates at the compositional mean behaviour composition (defined as the closure of the component-wise geometric mean) is also the mean of the coordinate values.

Then, we consider a new behaviour composition, which corresponds to isometric log-ratio pivot coordinate values (z_1, z_2, z_3). In particular, when considering a pairwise time reallocation plot we would use a new composition where time in two behaviours remained at its value in the mean behaviour composition, a value was subtracted from time in one of the remaining behaviours, and that value added to the time in the other behaviour. The log hazard ratio, now relative to the mean behaviour composition, is then calculated as:

$$\ln HR = \beta_1(z_1 - \bar{z}_1) + \beta_2(z_2 - \bar{z}_2) + \beta_3(z_3 - \bar{z}_3)$$

The standard error on this can be calculated using the variance-covariance matrix of the coefficients $\beta = (\beta_1, \beta_2, \beta_3)$, denoted V. Writing $x = (z_1 - \bar{z}_1, z_2 - \bar{z}_2, z_3 - \bar{z}_3)^T$, then

$$SE = \sqrt{x^T V x}$$

This can be used to calculate the HR and an (approximate) 95% CI as:

$$\exp(\ln HR)(\exp(\ln HR - 1.96SE), \exp(\ln HR + 1.96SE))$$

While this exposition is similar to Dumuid *et al.* [125] in the context of linear regression and Chastin *et al.* [159], it can be shown that this approach is mathematically equivalent to that of McGregor *et al.* [214]. To ensure results were plotted within the range supported by the data, times plotted were truncated at the 5th and 95th centile for the behaviour in the pair for which this range was narrower.

5.3.7 Sensitivity analyses

Two sensitivity analyses were performed to address aspects of the analytical approach. One analysis considered the treatment of zero values. As noted in Section 5.3.4, a Compositional Data Analysis approach requires decisions on the treatment of zero values in analysis. Sensitivity of results to the method of treating zero values (imputation or exclusion) under the Compositional Data Analysis approach was examined by performing an analysis restricted to participants with non-zero values in all behaviour variables i.e. by treating zeroes as true zeroes, rather than rounded zeroes. A second analysis implemented a linear isotemporal substitution approach for comparability with previous literature. Under a linear isotemporal substitution approach, the total measurement time and time in all but one of the movement behaviours are included in the model (so the included variables are not perfectly multicollinear). As here all subjects had the same total measurement time (as non-wear time was imputed), a total time variable was not included (meaning the approach may be more properly called 'leave-one-out regression' than true linear isotemporal substitution [127]). Associations are modelled as linear (rather than linear in the log-ratios, as under a Compositional Data Analysis approach). The coefficient of each behaviour can be interpreted in terms of replacing time in the left-out behaviour with time in that behaviour. Linear isotemporal substitution has been widely used in movement behaviour epidemiology [215], but has been criticised for not addressing the fact movement behaviour data only conveys relative information [127].

5.3.8 Code and data availability

Code and details of analyses are available at `github.com/OxWearables/manuscript_ml_behaviours_cvd_2021`. Data preparation was performed in Python

3.6.6 [199] and R 4.0.5 [200], and used the 'ukb_download_and_prep_template' tool [3] for preparing covariate and outcome data. Statistical analysis was performed in R 4.0.5 [200] with 'zCompositions', 'survival', 'plyr', 'ggplot2', 'data.table', 'rlist', and 'gtools' [208, 216–221].

While writing functions to complete the analyses in this thesis, a standalone R package was developed to facilitate common epidemiological analyses using a compositional exposure. The package, known as 'epicoda', is available at `github.com/OxWearables/epicoda`. Its motivation and development are described further in Appendix C.

UK Biobank data is available to researchers who submit an application (`www.ukbiobank.ac.uk`).

See Appendix B for a summary of code and data used throughout this thesis.

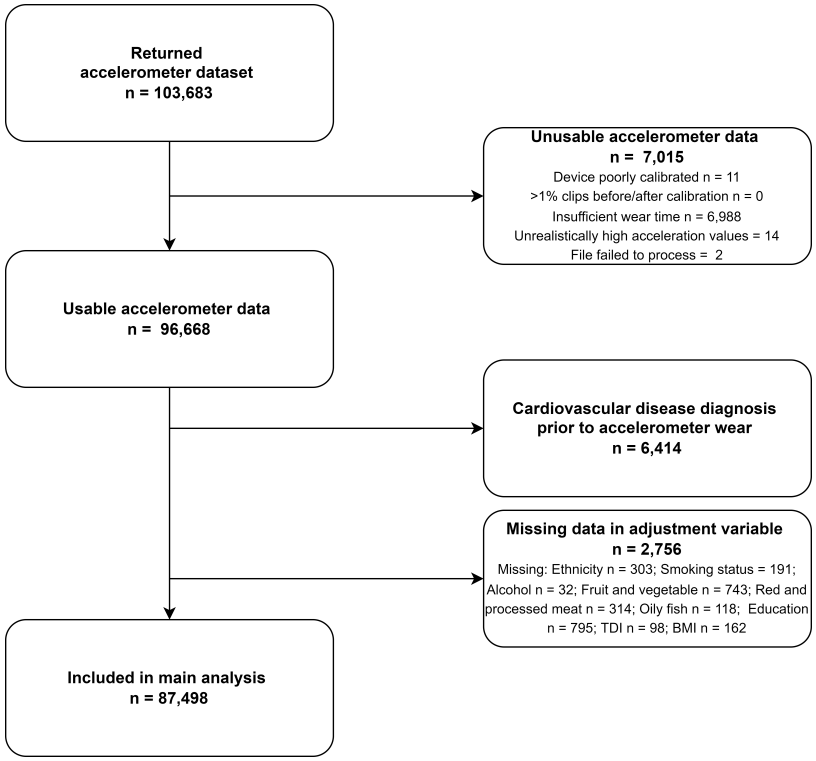
5.4 Results

5.4.1 Study population

Participant inclusion is described in Figure 5.3. 7,015 participants were excluded due to poor quality accelerometer data (mostly due to poor wear time). 6,414 participants were excluded due to a previous cardiovascular disease event (either an ischaemic heart disease or cerebrovascular disease record in hospital inpatient data before accelerometer wear or a self-reported heart attack or stroke at the baseline assessment). A further 2,756 participants were excluded due to missing data for variables used in the model.

Overall, 87,498 participants were included in the final analysis for incident cardiovascular disease. There were 524,919 person years of follow-up. Median follow-up was

Figure 5.3: Participant flow diagram for the analysis of 24-hour movement behaviours and incident cardiovascular disease in UK Biobank participants.



BMI, Body Mass Index; TDI, Townsend Deprivation Index.

6.2 years, with a maximum follow-up of 7.7 years. Over this follow-up period, there were 4,105 incident cardiovascular disease events.

5.4.2 Descriptive characteristics of movement behaviour phenotypes

Descriptive characteristics of movement behaviour phenotypes in the analytic sample were described in detail in Chapter 4; they are reproduced here in Table 5.1. In particular, older participants spent less time in moderate-to-vigorous physical activity behaviours than younger participants. Women spent more time in light physical activity behaviours than men, but spent less time in both moderate-to-vigorous physical

activity behaviours and sedentary behaviour. Participants with higher BMI spent less time in light physical activity behaviours and moderate-to-vigorous physical activity behaviours than participants with lower BMI, and spent more time sedentary.

Table 5.1: Movement behaviours of 87,498 UK Biobank participants by participant characteristics. Movement behaviours are given as median (Inter-Quartile Range).

	n (%)	**Sleep** (hr/day)	**Sedentary behaviour** (hr/day)	**Light physical activity behaviours** (hr/day)	**Moderate-to-vigorous physical activity behaviours** (min/day)
Overall	87,498 (100)	8.6 (7.9,9.3)	9.2 (8.0,10.3)	5.5 (4.5,6.7)	25 (12,44)
Age at accelerometer wear (years)					
40-49	7,767 (9)	8.5 (7.8,9.1)	9.4 (8.1,10.6)	5.4 (4.3,6.6)	30 (16,50)
50-59	26,081 (30)	8.5 (7.8,9.2)	9.3 (8.1,10.5)	5.4 (4.4,6.7)	28 (14,47)
60-69	38,774 (44)	8.6 (8.0,9.3)	9.0 (7.9,10.2)	5.6 (4.6,6.7)	25 (12,43)
70-79	14,876 (17)	8.6 (7.9,9.4)	9.2 (8.1,10.3)	5.5 (4.5,6.7)	20 (9,36)
Sex					
Female	50,882 (58)	8.6 (8.0,9.3)	8.9 (7.8,10.0)	5.8 (4.8,7.0)	22 (10,38)
Male	36,616 (42)	8.4 (7.8,9.2)	9.6 (8.4,10.8)	5.1 (4.1,6.2)	31 (16,52)
Ethnicity					
Asian	756 (1)	8.4 (7.7,9.3)	9.3 (7.9,10.5)	5.6 (4.5,6.9)	19 (9,35)
Black	701 (1)	8.2 (7.3,9.0)	9.4 (8.0,10.6)	5.9 (4.8,7.2)	21 (10,35)
Mixed/Other	1,151 (1)	8.4 (7.6,9.1)	9.2 (8.0,10.5)	5.7 (4.5,7.0)	26 (13,44)
White	84,890 (97)	8.6 (7.9,9.3)	9.2 (8.0,10.3)	5.5 (4.5,6.7)	25 (12,44)
Townsend Deprivation Index (by quarter)					
Least deprived (Less than -3.8)	21,913 (25)	8.6 (7.9,9.3)	9.1 (7.9,10.3)	5.6 (4.6,6.7)	24 (12,43)
Second least deprived (-3.8 to -2.5)	21,839 (25)	8.6 (7.9,9.3)	9.1 (8.0,10.3)	5.6 (4.5,6.7)	24 (12,43)
Second most deprived (-2.5 to -0.2)	21,872 (25)	8.6 (7.9,9.3)	9.2 (8.0,10.3)	5.6 (4.5,6.7)	25 (12,44)
Most deprived (More than -0.2)	21,874 (25)	8.5 (7.8,9.2)	9.3 (8.1,10.5)	5.4 (4.3,6.6)	27 (13,47)
Education					
School leaver	19,535 (22)	8.7 (8.0,9.5)	8.9 (7.7,10.1)	5.7 (4.7,6.9)	20 (9,36)
Further education	29,061 (33)	8.6 (7.9,9.3)	9.1 (7.9,10.3)	5.6 (4.6,6.8)	23 (11,41)
Higher education	38,902 (44)	8.5 (7.8,9.1)	9.4 (8.2,10.5)	5.4 (4.4,6.5)	30 (16,50)
Smoking status					
Never smoker	50,888 (58)	8.6 (7.9,9.3)	9.2 (8.0,10.3)	5.5 (4.5,6.7)	26 (13,45)
Ex-smoker	30,717 (35)	8.5 (7.9,9.3)	9.2 (8.0,10.3)	5.6 (4.5,6.7)	25 (12,44)
Current smoker	5,893 (7)	8.6 (7.9,9.3)	9.4 (8.2,10.6)	5.4 (4.3,6.6)	21 (9,39)
Alcohol consumption					
Never drinker	4,745 (5)	8.6 (7.9,9.4)	9.1 (7.8,10.3)	5.6 (4.5,6.9)	20 (9,39)
Less than 3 times per week	39,760 (45)	8.6 (7.9,9.3)	9.2 (8.0,10.3)	5.5 (4.5,6.7)	23 (11,41)
3 or more times per week	42,993 (49)	8.5 (7.9,9.2)	9.2 (8.0,10.4)	5.5 (4.5,6.7)	28 (14,47)
Fruit and vegetable consumption (servings/day)					
Less than 3	3,595 (4)	8.6 (7.8,9.4)	9.7 (8.4,11.0)	5.0 (3.9,6.3)	21 (9,39)
3-4.9	14,293 (16)	8.6 (7.9,9.3)	9.4 (8.2,10.6)	5.3 (4.3,6.5)	24 (12,42)
5-7.9	36,991 (42)	8.6 (7.9,9.3)	9.2 (8.0,10.3)	5.5 (4.5,6.7)	26 (13,44)
8 or more	32,619 (37)	8.5 (7.9,9.2)	9.0 (7.8,10.1)	5.7 (4.7,6.9)	26 (13,45)
Body Mass Index (kg/m^2)					
Underweight (Less than 18.5)	511 (1)	8.5 (7.9,9.2)	8.5 (7.2,9.6)	6.3 (4.9,7.3)	34 (19,55)
Normal weight (18.5-24.9)	35,043 (40)	8.6 (8.0,9.2)	8.8 (7.7,10.0)	5.8 (4.7,6.9)	30 (16,50)
Overweight (25.0-29.9)	35,783 (41)	8.5 (7.9,9.3)	9.3 (8.1,10.4)	5.5 (4.4,6.6)	25 (12,43)
Obese (30 or more)	16,161 (18)	8.5 (7.8,9.3)	9.8 (8.5,10.9)	5.1 (4.1,6.3)	16 (7,31)

The mean composition of movement behaviours (the daily movement behaviours of a hypothetical average individual; calculated as the 'compositional mean') was 8.8 hours/day sleep, 9.3 hours/day sedentary behaviour, 5.6 hours/day light physical activity behaviours and 21 minutes/day moderate-to-vigorous physical activity behaviours (MVPA; Figure 5.4).

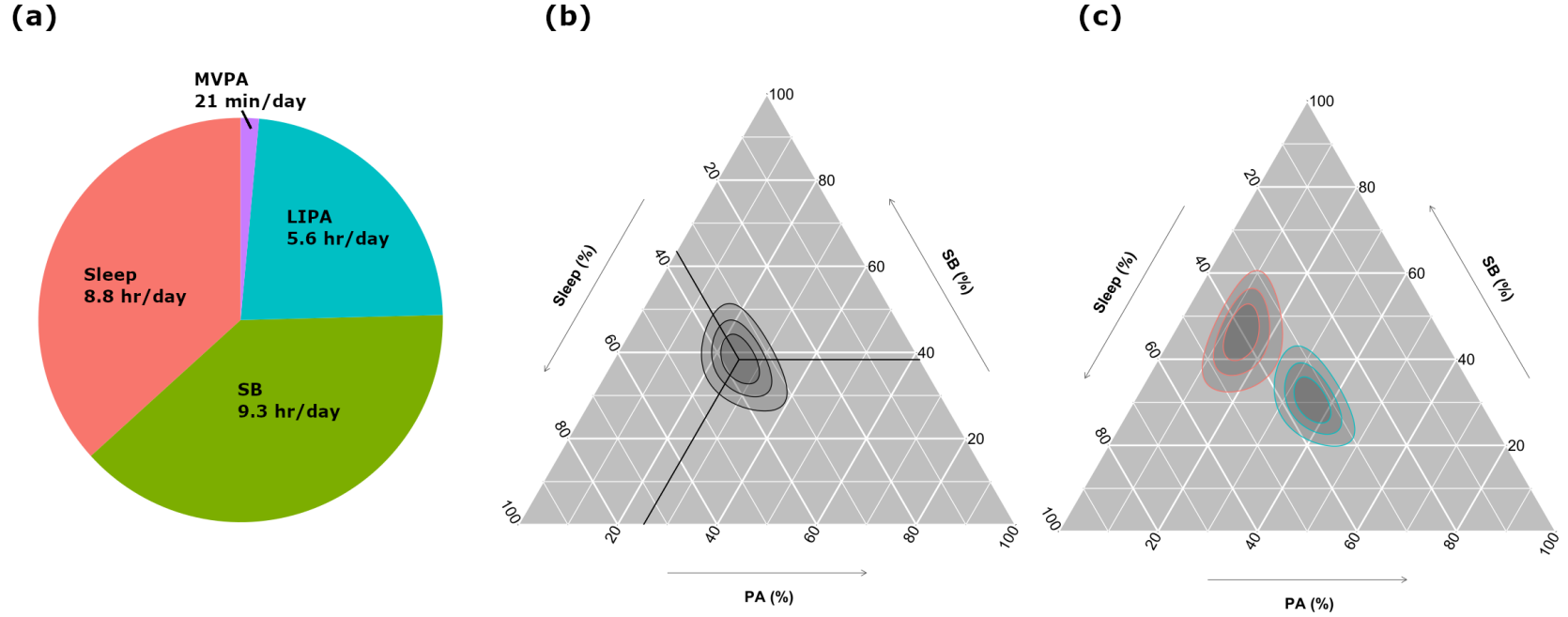
Time in physical activity behaviours and sedentary behaviour varied substantially among participants, while variation in sleep was more limited (Figure 5.4). The least and most active participants by average acceleration differed across all dimensions of behaviour (Figure 5.4).

5.4.3 Cox regression analyses for incident cardiovascular disease

5.4.3.1 Modelling

Table 5.2 shows the exponentiated model coefficients (hazard ratios) of the first pivot coordinate for each behaviour as derived from Compositional Data Analysis Cox regression models. As noted in Section 5.3.6, these exponentiated coefficients can be interpreted in terms of reallocating time from one behaviour to all other behaviours proportionally. Their magnitude is difficult to interpret directly, but they can be compared to 1 to indicate whether higher time in the named behaviour relative to all others proportionally was associated with higher or lower CVD risk. As shown in Table 5.2, there was evidence that a higher level of moderate-to-vigorous physical activity relative to all other behaviours proportionally was associated with lower risk of incident cardiovascular disease, and there was evidence that a higher level of sedentary behaviour relative to all other behaviours proportionally was associated with a higher risk of incident cardiovascular disease. Neither sleep nor light physical activity

Figure 5.4: Distribution of movement behaviours in 87,498 UK Biobank participants. (A) Mean movement behaviour composition among UK Biobank participants. (B) Movement behaviours of UK Biobank participants on a ternary plot, showing sleep, sedentary behaviour (SB) and physical activity behaviours (PA; combines light and moderate-to-vigorous physical activity behaviours). The crosshair marks the compositional mean. Concentric rings represent the 25, 50% and 75% predictive regions for the data. The behaviour composition at a point can be found by tracing out (parallel to the white lines and crosshair) from the point to the axes. (C) Ternary plot showing the behaviour distribution of the 5% most active (blue) and 5% least active (red) UK Biobank participants by average acceleration. Concentric rings represent the 25, 50% and 75% predictive regions for each group.



LIPA - light physical activity behaviours; MVPA - moderate-to-vigorous physical activity behaviours.

behaviours relative to all other behaviours proportionally were significantly associated with lower risk at the 5% level (both point estimates were below 1). As interpreting model coefficients for ilr-transformed variables is challenging, Section 5.4.3.2 considers the hazard ratios associated with particular allocations of time between behaviours.

Table 5.2: Coefficient of first isometric log-ratio pivot coordinate for each movement behaviour estimated using all Compositional Data Analysis Cox regression models.

	n	Events	First pivot coordinate: Sleep vs All other behaviours	First pivot coordinate: SB vs All other behaviours	First pivot coordinate: LIPA vs All other behaviours	First pivot coordinate: MVPA vs All other behaviours
MAIN ANALYSES						
A: Minimally adjusted	87,498	4,105	0.90 (0.77, 1.05)	1.37 (1.21, 1.56)	0.96 (0.88, 1.06)	0.84 (0.81, 0.86)
B: Main model	87,498	4,105	0.88 (0.75, 1.02)	1.38 (1.22, 1.57)	0.96 (0.88, 1.06)	0.86 (0.84, 0.88)
C: Further stratified by BMI	87,498	4,105	0.90 (0.77, 1.05)	1.29 (1.13, 1.46)	0.98 (0.89, 1.08)	0.88 (0.85, 0.90)
D: Fatal CVD	87,498	348	0.79 (0.48, 1.29)	2.13 (1.40, 3.22)	0.71 (0.52, 0.95)	0.85 (0.78, 0.92)
SUBGROUP ANALYSES						
E: Women only	50,882	1,654	0.78 (0.60, 1.02)	1.57 (1.28, 1.94)	0.95 (0.81, 1.11)	0.85 (0.82, 0.89)
F: Men only	36,616	2,451	0.93 (0.77, 1.13)	1.28 (1.09, 1.50)	0.97 (0.86, 1.09)	0.86 (0.83, 0.90)
G: Under 65s	51,180	1,425	0.98 (0.75, 1.28)	1.17 (0.95, 1.45)	1.01 (0.86, 1.18)	0.86 (0.82, 0.91)
H: Over 65s	36,318	2,680	0.83 (0.69, 1.00)	1.52 (1.30, 1.77)	0.93 (0.82, 1.04)	0.86 (0.83, 0.89)
SENSITIVITY ANALYSES						
I: Without zero values	86,696	4,017	0.86 (0.73, 1.01)	1.38 (1.21, 1.57)	0.99 (0.89, 1.09)	0.86 (0.83, 0.89)

Columns for coefficients report $\exp\hat{\beta}$ (95% CI). Model A (the minimally adjusted model) used age as the timescale and was stratified by sex. Model B was additionally adjusted for ethnicity, deprivation, education, smoking status, alcohol consumption, fresh fruit and vegetable consumption, red and processed meat consumption, and oily fish consumption. Model C was as Model B but additionally stratified by BMI (adjusted for BMI by stratifying the Cox model). Model D was as Model B but with fatal CVD, not incident CVD, as the outcome. Models E-I were as Model B but estimated only in the specified subgroups.

5.4.3.2 Time reallocation

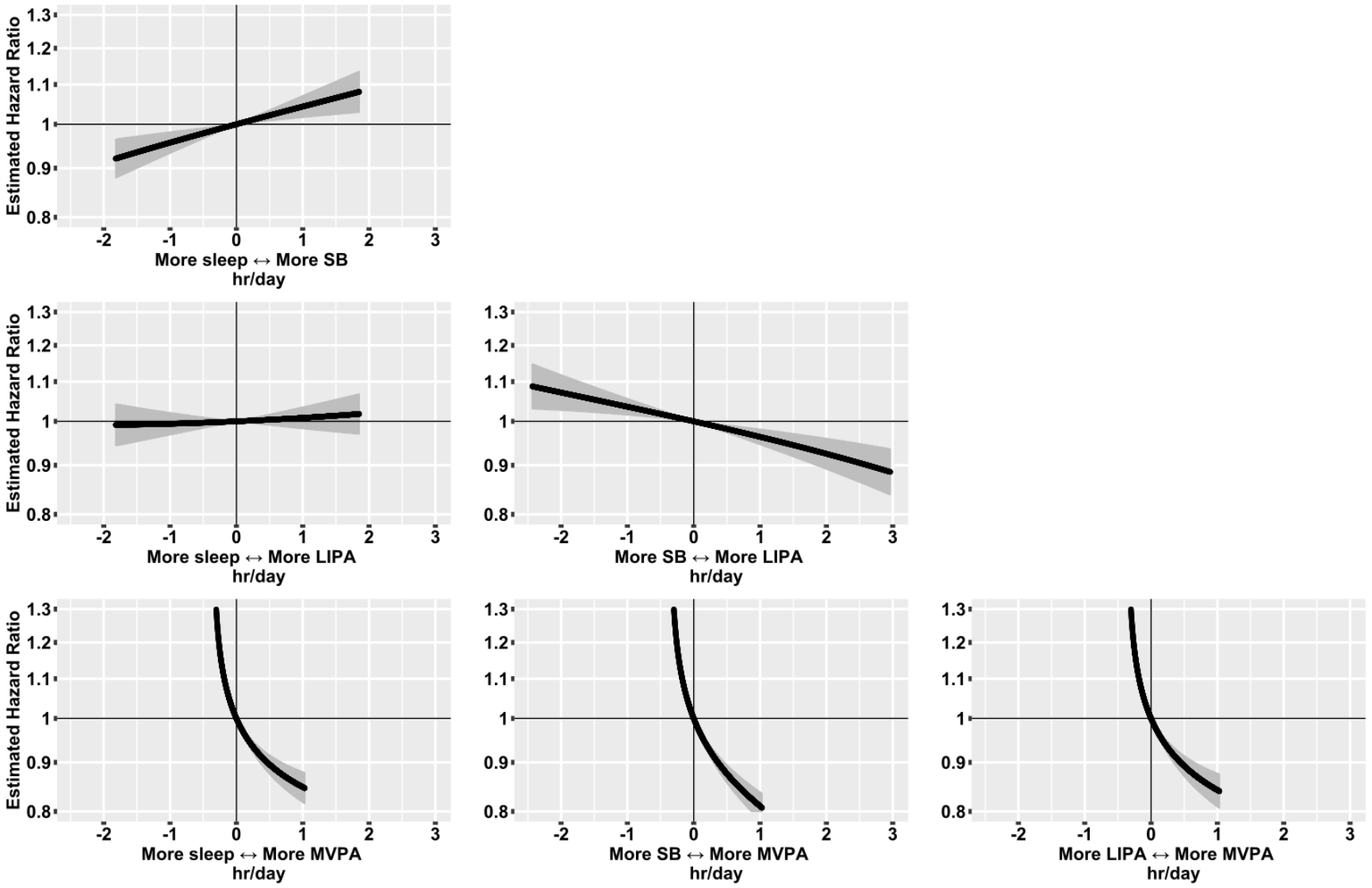
Pairwise reallocations of time between behaviours were considered. Reallocating time from sedentary behaviour to light physical activity behaviours was associated with a lower risk of cardiovascular disease (Figure 5.5): relative to the mean behaviour composition, the hazard ratio associated with reallocating 1 hour/day from light physical activity behaviours to sedentary behaviour was 1.04 (1.02, 1.06), while the HR

associated with reallocating 1 hour/day from sedentary behaviour to light physical activity behaviours was 0.96 (0.95, 0.98). Reallocating time from sedentary behaviour to MVPA was associated with more pronounced lower risk of cardiovascular disease (Figure 5.5): relative to the mean behaviour composition, the HR associated with reallocating 15 minutes/day from MVPA to sedentary behaviour was 1.19 (1.15, 1.22), while the HR associated with reallocating 15 minutes/day from sedentary behaviour to MVPA was 0.92 (0.91, 0.94). Reallocating time from light physical activity behaviours or sleep to MVPA, and reallocating time from sedentary behaviour to sleep were also associated with a lower risk of cardiovascular disease, while reallocating time from sleep to LIPA was not associated with cardiovascular disease risk (Figure 5.5).

Reallocations of time into a single behaviour from all other behaviours proportionally were also considered. Relative to the mean behaviour composition, reallocating 20 min/day to MVPA from all other behaviours proportionally was associated with 9% (7%, 10%) lower risk of CVD (Figure 5.6; 28% of the study population exceeded this level of MVPA). Reallocating 1 hour/day to sedentary behaviour, from all other behaviours proportionally, was associated with 5% (3%, 7%) higher risk of CVD (Figure 5.6; 26% of the study population exceeded this level of sedentary behaviour). Reallocating 1 hour/day to light physical activity behaviours, from sleep, sedentary behaviour and MVPA proportionally and reallocating 1 hour/day to sleep, from sedentary behaviour, light physical activity behaviours and MVPA proportionally, showed more modest and non-significant associations with lower risk of CVD (1% (-1%, 2%) and 2% (0%, 4%), respectively; Figure 5.6).

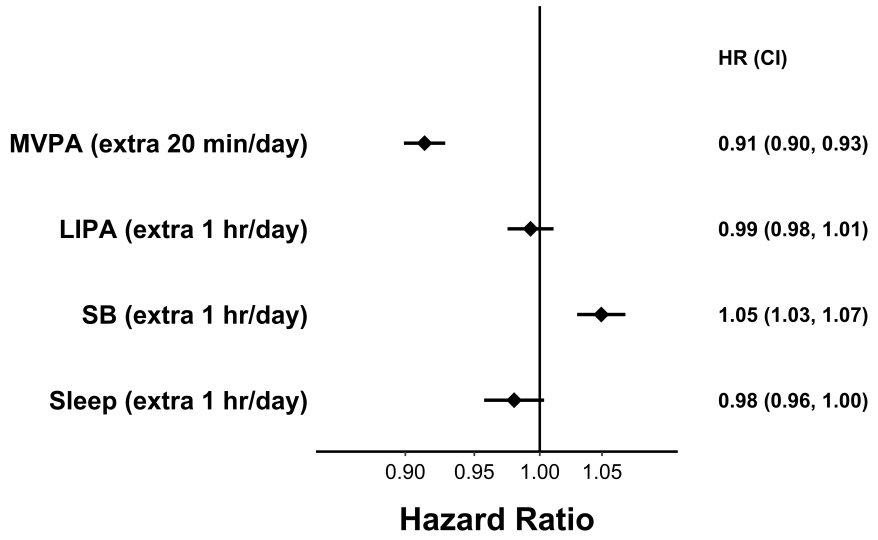
Results from the multivariable-adjusted model were very similar to those from a minimally adjusted model, with a slight attenuation for reallocations into moderate-to-vigorous physical activity behaviours (Figure 5.7, Table 5.2). Further adjustment

Figure 5.5: Hazard Ratios for incident cardiovascular disease associated with balance between movement behaviours based on 4,105 events among 87,498 UK Biobank participants.



All relative to the mean behaviour composition (8.8 hours/day sleep, 9.3 hours/day sedentary behaviour (SB), 5.6 hours/day light physical activity behaviours (LIPA), 0.35 hours/day (21 min/day) moderate-to-vigorous physical activity behaviours (MVPA)). Model used age as the timescale, was stratified by sex and was additionally adjusted for ethnicity, deprivation, education, smoking status, alcohol consumption, fresh fruit and vegetable consumption, red and processed meat consumption, and oily fish consumption. 95% CIs shown.

Figure 5.6: Hazard Ratios for incident cardiovascular disease associated with reallocating time to a given movement behaviour, from all other behaviours proportionally, based on 4,105 events among 87,498 UK Biobank participants.



All relative to the mean behaviour composition (8.8 hours/day sleep, 9.3 hours/day sedentary behaviour (SB), 5.6 hours/day light physical activity behaviours (LIPA), 0.35 hours/day (21 min/day) moderate-to-vigorous physical activity behaviours (MVPA)), and more time in named behaviour reallocated from all other behaviours proportionally. Model used age as the timescale, was stratified by sex and was additionally adjusted for ethnicity, deprivation, education, smoking status, alcohol consumption, fresh fruit and vegetable consumption, red and processed meat consumption, and oily fish consumption. 95% CIs shown.

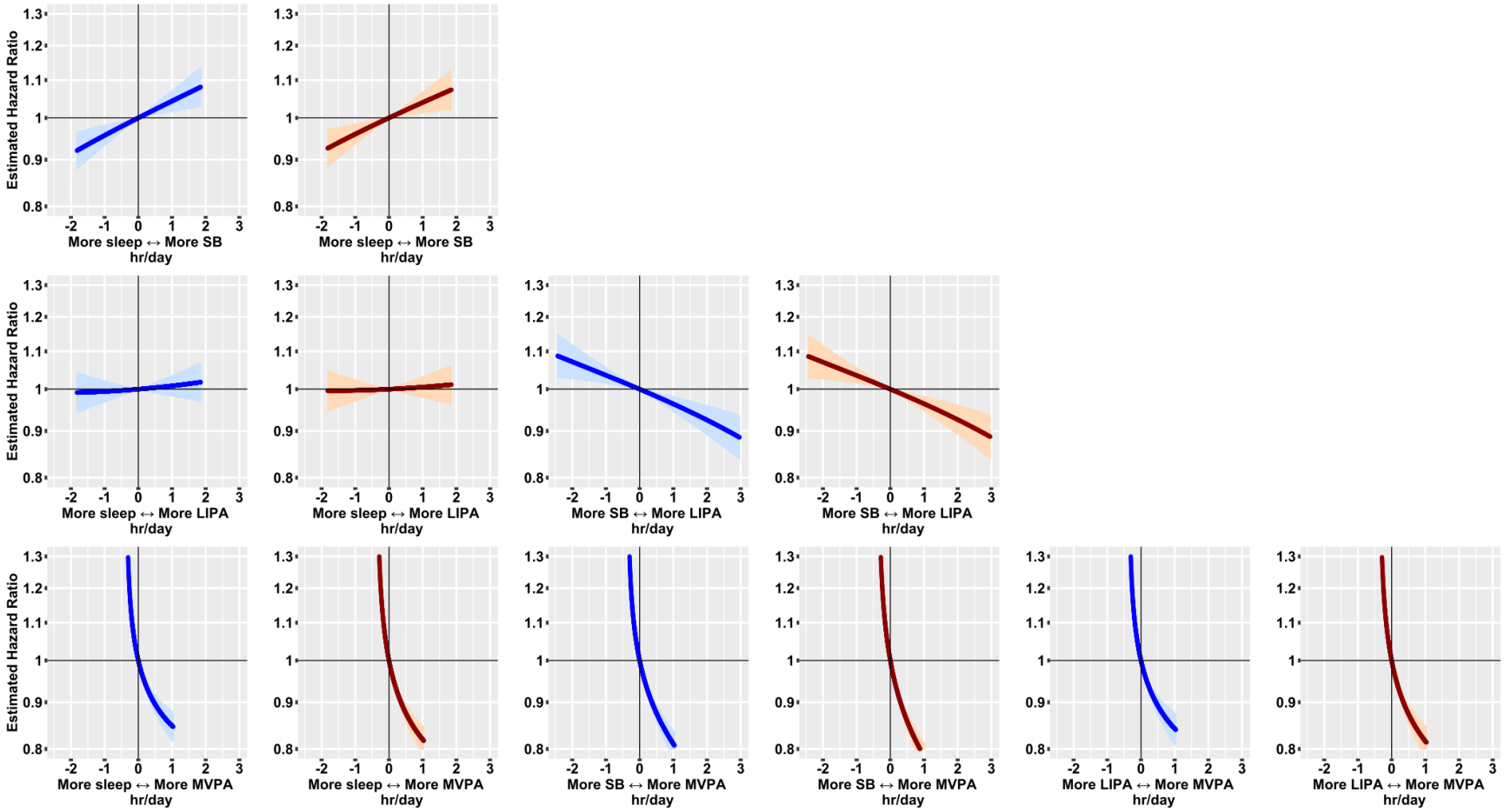
for BMI, using a model stratified by BMI category, resulted in some attenuation of the association between movement behaviours and incident CVD (Figure 5.8, Table 5.2). For example, the 9% (7%, 10%) lower risk of CVD associated with reallocating 20 min/day to MVPA relative to the mean behaviour composition was reduced to a 7% (6%, 9%) lower risk after stratification by BMI. Associations for fatal cardiovascular events were similar to those for all cardiovascular events, with reallocating time from sedentary behaviour to light physical activity behaviours and sleep appearing more strongly associated with fatal events (Figure 5.9, Table 5.2). Results for women and men separately were similar, with some associations appearing stronger for women (Figure 5.10, Table 5.2). Results for participants aged under and over 65 separately were similar, with some associations appearing stronger for older adults (Figure 5.11, Table 5.2).

5.4.3.3 Sensitivity analyses

Restricting the analysis to only those recording non-zero values in all behaviour variables (i.e. excluding the 1% of participants who registered no MVPA) did not materially impact results (Figure 5.12, Table 5.2).

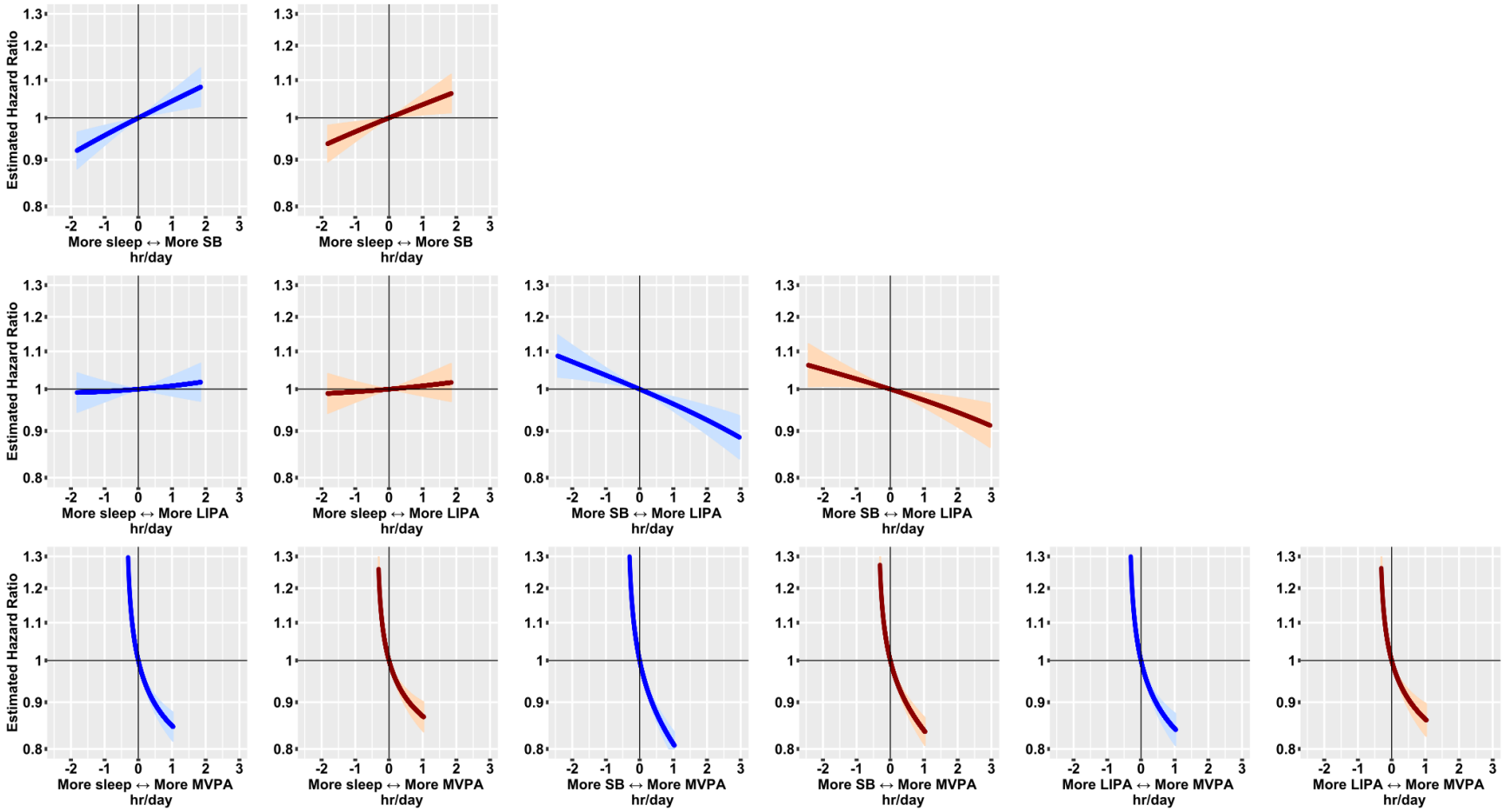
Results using a linear isotemporal substitution approach were broadly similar to results of the main analysis (Figure 5.13). However, there were some differences in shape of the associations observed, as may be expected as a consequence of the different modelling assumptions.

Figure 5.7: Hazard Ratios for incident cardiovascular disease for all behaviour pairs estimated using multivariable-adjusted (blue) and minimally adjusted (red) Cox regression models.



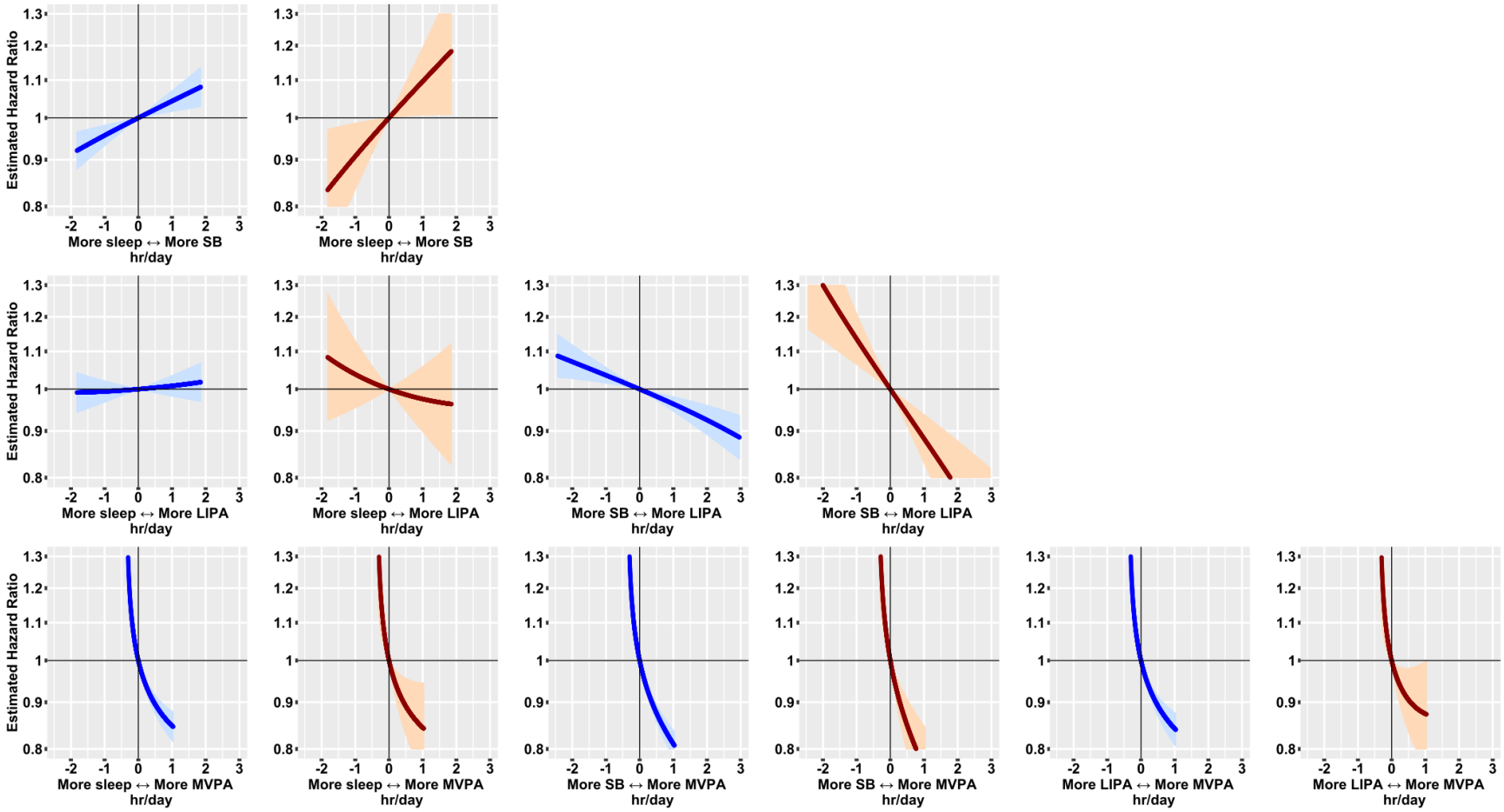
Models based on 4,105 events in 87,498 participants. All relative to the mean behaviour composition: 8.8 hours/day sleep, 9.3 hours/day sedentary behaviour, 5.6 hours/day light physical activity behaviours, 0.35 hours/day (21 min/day) moderate-to-vigorous physical activity behaviours. Multivariable-adjusted model used age as the timescale, was stratified by sex and was additionally adjusted for ethnicity, deprivation, education, smoking status, alcohol consumption, fresh fruit and vegetable consumption, red and processed meat consumption, and oily fish consumption. Minimally-adjusted model used age as the timescale and was stratified by sex.

Figure 5.8: Hazard Ratios for incident cardiovascular disease for all behaviour pairs estimated using a multivariable-adjusted Cox regression model before (blue) and after (red) stratification by BMI.



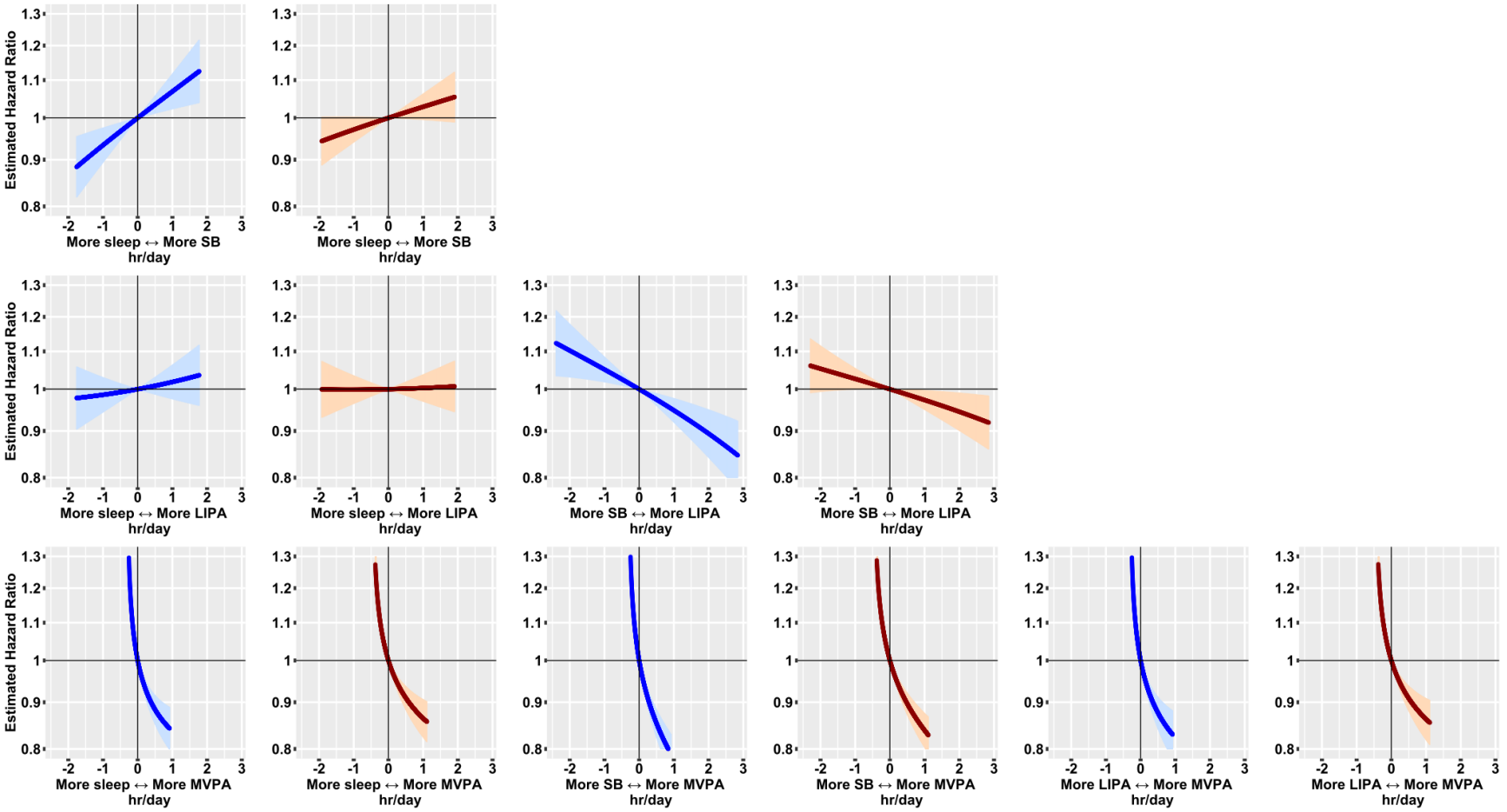
Models based on 4,105 events in 87,498 participants. All relative to the mean behaviour composition: 8.8 hours/day sleep, 9.3 hours/day sedentary behaviour, 5.6 hours/day light physical activity behaviours, 0.35 hours/day (21 min/day) moderate-to-vigorous physical activity behaviours. Models used age as the timescale, were stratified by sex and were additionally adjusted for ethnicity, deprivation, education, smoking status, alcohol consumption, fresh fruit and vegetable consumption, red and processed meat consumption, and oily fish consumption.

Figure 5.9: Hazard Ratios for all (blue) and fatal incident (red) cardiovascular disease for all behaviour pairs estimated using a multivariable-adjusted Cox regression model.



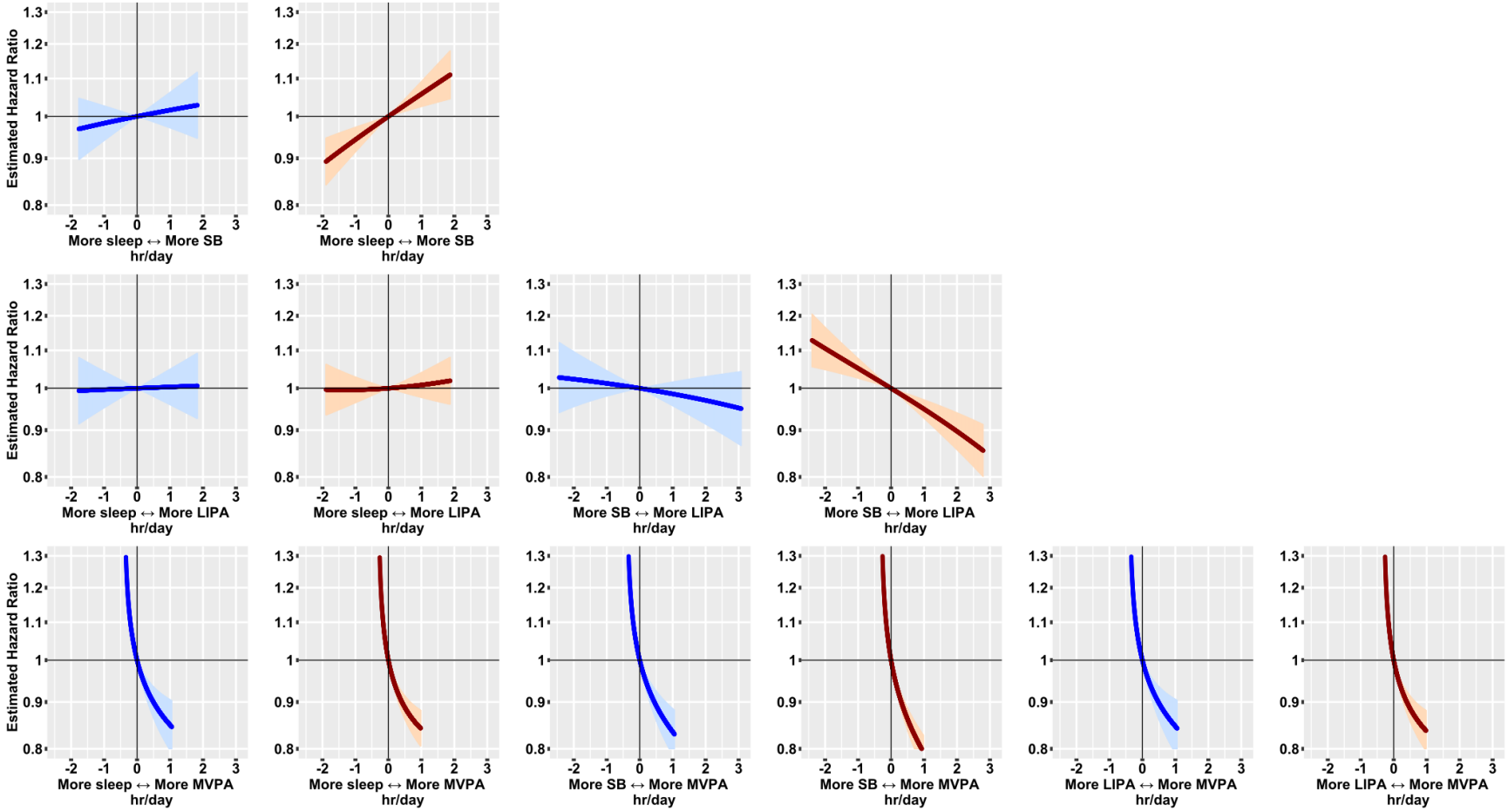
Models based on 4,105 events and 348 cardiovascular deaths in 87,498 participants. All relative to the mean behaviour composition: 8.8 hours/day sleep, 9.3 hours/day sedentary behaviour, 5.6 hours/day light physical activity behaviours, 0.35 hours/day (21 min/day) moderate-to-vigorous physical activity behaviours. Models used age as the timescale, were stratified by sex and were additionally adjusted for ethnicity, deprivation, education, smoking status, alcohol consumption, fresh fruit and vegetable consumption, red and processed meat consumption, and oily fish consumption.

Figure 5.10: Hazard Ratios for incident cardiovascular disease for all behaviour pairs estimated using a multivariable-adjusted Cox regression model for women (blue) and men (red).



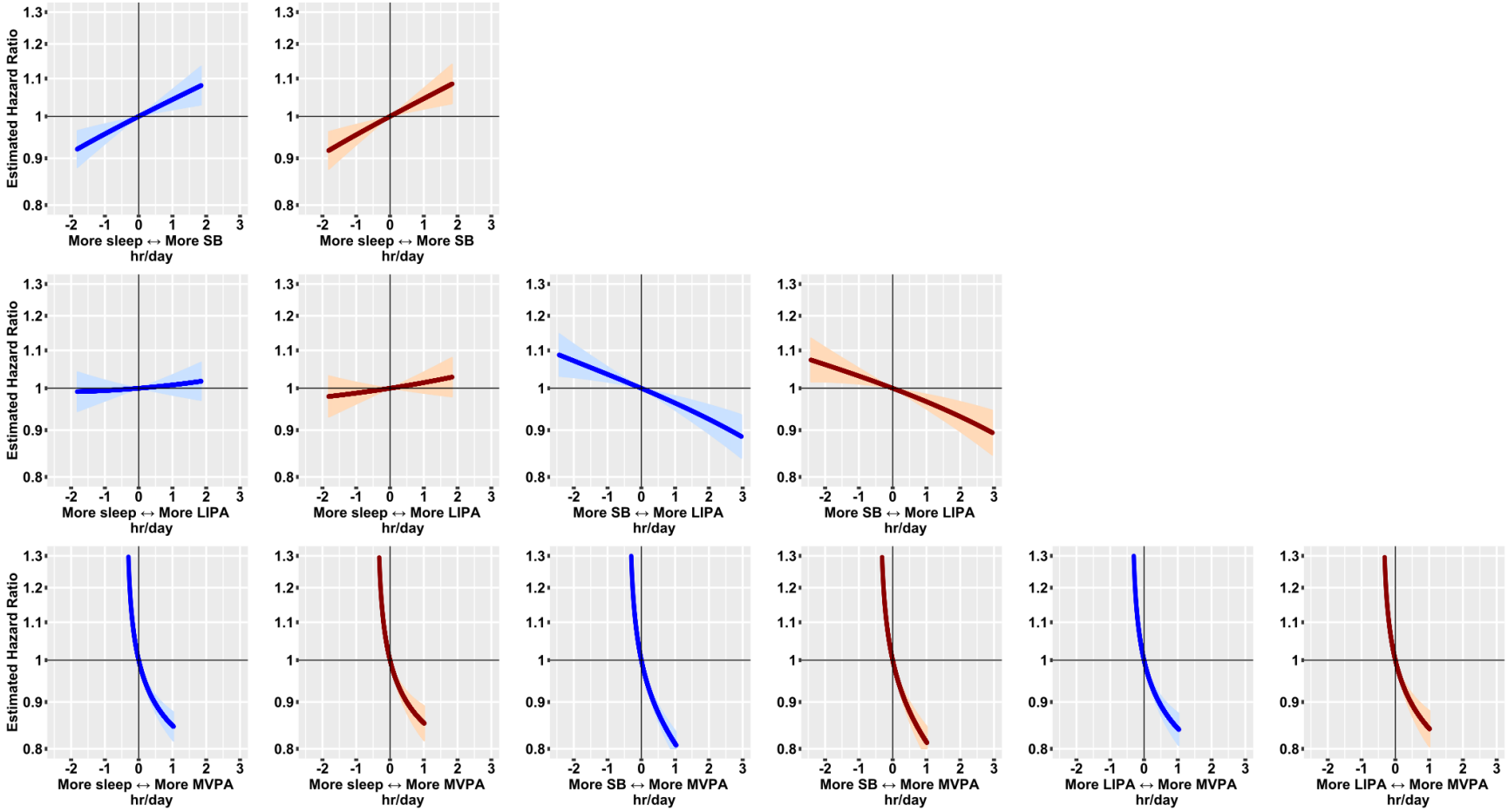
Models based on 1,654 events in 50,882 women and 2,451 events in 36,616 men. All relative to the mean behaviour composition in each case (women – 8.9 hours/day sleep, 9.0 hours/day sedentary behaviour, 5.9 hours/day light physical activity behaviours, 0.30 hours/day (18 min/day) moderate-to-vigorous physical activity behaviours; men - 8.7 hours/day sleep, 9.7 hours/day sedentary behaviour, 5.1 hours/day light physical activity behaviours, 0.45 hours/day (27 min/day) moderate-to-vigorous physical activity behaviours). Models used age as the timescale and were additionally adjusted for ethnicity, deprivation, education, smoking status, alcohol consumption, fresh fruit and vegetable consumption, red and processed meat consumption, and oily fish consumption.

Figure 5.11: Hazard Ratios for incident cardiovascular disease for all behaviour pairs estimated using a multivariable-adjusted Cox regression model for participants aged under 65 (blue) and participants aged over 65 (red).



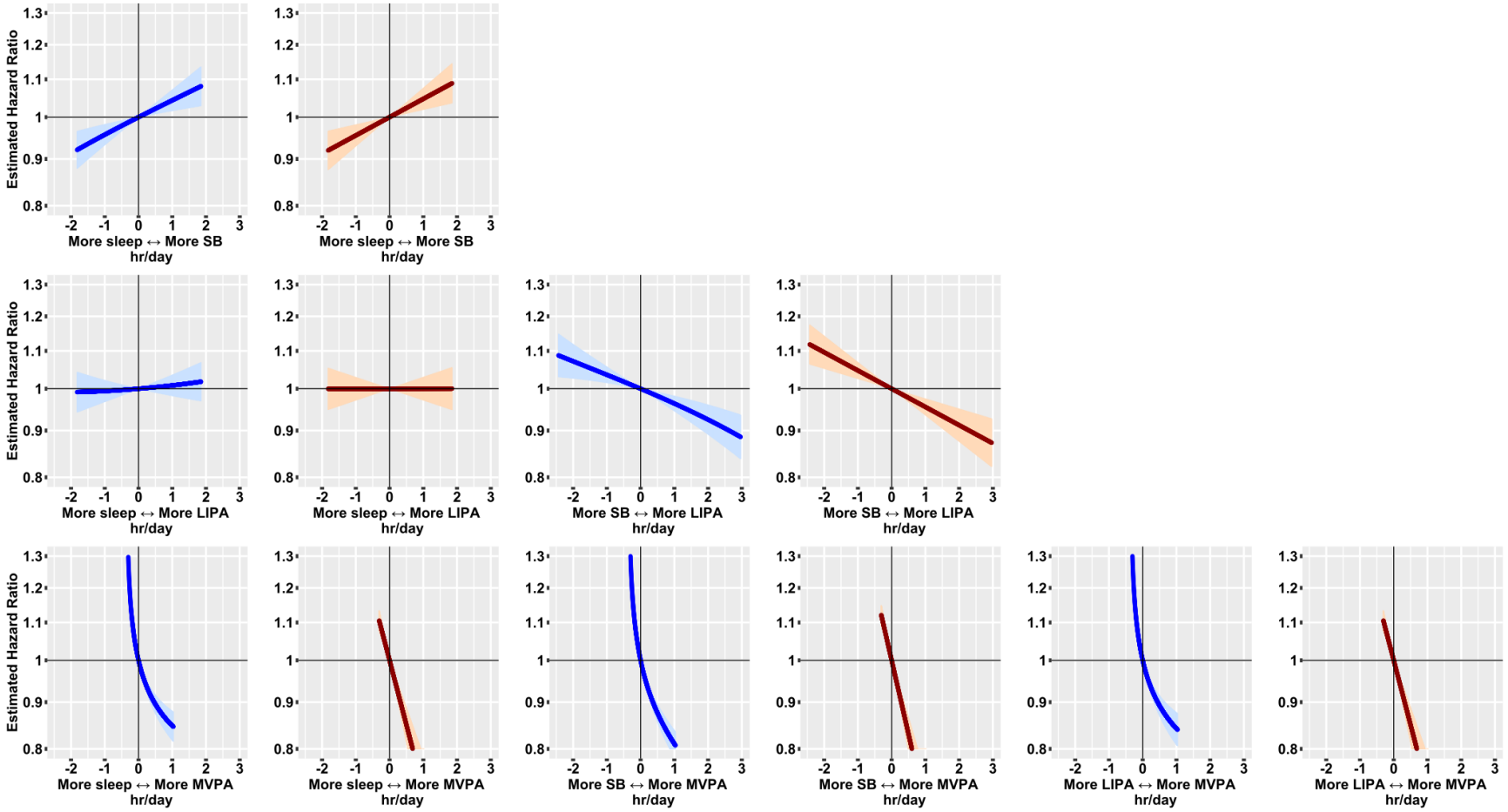
Models based on 1,425 events in 51,180 participants aged under 65 and 2,680 events in 36,318 participants aged over 65. All relative to the mean behaviour composition in each case (participants aged under 65 - 8.8 hours/day sleep, 9.3 hours/day sedentary behaviour, 5.5 hours/day light physical activity behaviours, 0.39 hours/day (23 min/day) moderate-to-vigorous physical activity behaviours; participants aged over 65 - 8.9 hours/day sleep, 9.2 hours/day sedentary behaviour, 5.6 hours/day light physical activity behaviours, 0.30 hours/day (18 min/day) moderate-to-vigorous physical activity behaviours). Models used age as the timescale, were stratified by sex and were additionally adjusted for ethnicity, deprivation, education, smoking status, alcohol consumption, fresh fruit and vegetable consumption, red and processed meat consumption, and oily fish consumption.

Figure 5.12: Hazard Ratios for incident cardiovascular disease for all behaviour pairs estimated using a multivariable-adjusted Cox regression model for all participants (blue) and in a sensitivity analysis excluding individuals with a zero value in any behaviour (red).



Main model based on 4,105 events in 87,498 participants. Model excluding individuals with zero values based on 4,017 events in 86,696 participants. All relative to the mean behaviour composition in each case (main analysis – 8.8 hours/day sleep, 9.3 hours/day sedentary behaviour, 5.6 hours/day light physical activity behaviours, 0.35 hours/day (21 min/day) moderate-to-vigorous physical activity behaviours; analysis excluding individuals with zero values – 8.8 hours/day sleep, 9.3 hours/day sedentary behaviour, 5.6 hours/day light physical activity behaviours, 0.37 hours/day (22 min/day) moderate-to-vigorous physical activity behaviours). Models used age as the timescale, were stratified by sex and were additionally adjusted for ethnicity, deprivation, education, smoking status, alcohol consumption, fresh fruit and vegetable consumption, red and processed meat consumption, and oily fish consumption.

Figure 5.13: Hazard Ratios for incident cardiovascular disease estimated using a multivariable-adjusted Compositional Data Analysis Cox regression model (blue) and using a multivariable-adjusted linear isotemporal substitution Cox regression model (red).



Models based on 4,105 events in 87,498 participants. All relative to the mean behaviour composition: 8.8 hours/day sleep, 9.3 hours/day sedentary behaviour, 5.6 hours/day light physical activity behaviours, 0.35 hours/day (21 min/day) moderate-to-vigorous physical activity behaviours. Models used age as the timescale, were stratified by sex and were additionally adjusted for ethnicity, deprivation, education, smoking status, alcohol consumption, fresh fruit and vegetable consumption, red and processed meat consumption, and oily fish consumption.

5.5 Discussion

5.5.1 Overall findings

This chapter examined how the allocation of time between behaviours was associated with incident cardiovascular disease events over a > 6 year follow-up period. Reallocating time to MVPA from sleep, sedentary behaviour or light physical activity behaviours or reallocating time from sedentary behaviour to light physical activity behaviours or sleep was associated with lower risk of incident cardiovascular disease. Per minute, the most pronounced differences in risk were seen for MVPA.

Associations were partly explained by BMI. As highlighted in Section 5.3.3, this is challenging to interpret, as it may represent mediation or confounding by BMI. Moreover, BMI was measured some years prior to accelerometer wear, meaning this may represent an under-adjustment for BMI (although BMI is relatively stable [222]; see also Chapter 3).

5.5.2 Comparison with previous literature

The associations found in this chapter are broadly consistent with the literature on device-measured movement behaviours and risk of incident cardiovascular disease as reviewed in Chapter 2.

Reallocation of time from any other behaviour into moderate-to-vigorous physical activity behaviours was associated quite steeply with a lower risk of incident cardiovascular disease. Other studies considering moderate-to-vigorous physical activity also typically found associations between higher moderate-to-vigorous physical activity levels and lower risk of incident cardiovascular disease. Where studied, this

typically persisted after adjustment for time in other movement behaviours.

This work also found that reallocation of time from sedentary behaviour into light physical activity behaviours or sleep was associated with a lower risk of incident cardiovascular disease. As noted in Chapter 2, most existing literature on device-measured sedentary behaviour and incident cardiovascular disease has found lower levels of sedentary behaviour to be associated with lower risk of incident cardiovascular disease. However, the literature was inconsistent, and in particular it was unclear to what extent these associations persisted after adjustment for time in moderate-to-vigorous physical activity (study findings varied). One way in which this work builds on these previous studies is via use of a Compositional Data Analysis approach to appropriately mutually adjust 24-hour movement behaviours. Therefore, this association is adjusted for time in other behaviours, notably including moderate-to-vigorous physical activity behaviours. Associations between sedentary behaviour and incident CVD in this work were of relatively modest strength. As most studies on device-measured 24-hour movement behaviours and incident cardiovascular disease have been relatively small in size, inconsistency in current literature may result at least in part from other studies being underpowered to detect the relevant effects.

Results relating to sleep cannot be compared to other studies, as studies in the literature review did not investigate device-measured overall sleep time: sleep was often not measured, and where it was measured difficulties integrating sleep within the same framework as the other behaviours were noted [76]. Sleep is very different to other behaviours: it is a neurobiological state, and is tightly regulated and patterned [76]. However, here sleep was considered to be best analysed alongside other behaviours: it is an integral component of the 24-hour day, and is associated with other behaviours. Results were suggestive, for example indicating that reallocating time from sedentary behaviour to sleep was associated with a lower risk of incident CVD. However, these

results should be considered as preliminary, in particular noting the distinctiveness of sleep; the need for better validation of sleep measurements (see Chapter 4); and the possibility of a fuller epidemiological treatment of sleep considering factors beyond duration (e.g. quality and timing).

Comparing this work to the only other study of movement behaviours and incident CVD using a Compositional Data Analysis approach, there were both commonalities and differences. Consistent with this work, Yerramalla *et al.* [76] found that higher levels of moderate-to-vigorous physical activity behaviours at the expense of sedentary behaviour and light physical activity were associated with lower risk of incident cardiovascular disease. However, in contrast to this work, they did not find evidence for associations between higher levels of light physical activity at the expense of sedentary behaviour and lower risk of incident cardiovascular disease. That earlier study had substantially fewer events than the present study (299 versus 4,105), meaning it may have been underpowered. In addition, it used a cut-point method to measure time in different movement behaviours using a wrist-worn accelerometer. The authors noted that this may lead to misclassification between sedentary behaviour and light physical activity [76]. It may also overclassify light physical activity as moderate-to-vigorous physical activity, as mean daily MVPA time was close to 1 hour/day, even in their older participant population [76].

5.6 Summary

This work found that reallocation of time from any other behaviour to machine-learned moderate-to-vigorous physical activity behaviours was associated with a lower risk of incident cardiovascular disease. Reallocation of time from machine-learned sedentary behaviour to other behaviours was also associated with a lower risk of inci-

dent cardiovascular disease. Using a suitable approach to analyse 24-hour movement behaviour data (Compositional Data Analysis) supported holistic consideration of the 24-hour day. To assess the robustness of these results, the next chapter, Chapter 6, will investigate the potential impact of sources of bias, including information bias, selection bias, reverse causality, and residual confounding.

Chapter 6: Quantitative bias analysis of associations between 24-hour movement behaviours and incident cardiovascular disease

Some of the analyses in this chapter have been published as part of an article in the ***British Journal of Sports Medicine*** *[1]. Other analyses in this chapter are currently unpublished, but in some cases may be included within future publications.*

6.1 Introduction

6.1.1 Background

The previous chapter showed that device-measured 24-hour movement behaviours were associated with risk of incident cardiovascular disease in UK Biobank. This chapter will consider the interpretation of these associations and assess the potential impact of bias. The exposure, average 24-hour movement behaviours, is measured with error, potentially resulting in information bias. Selection effects into the UK Biobank and the accelerometry sub-cohort mean estimates may be vulnerable to selection bias. Selection bias can manifest in different ways, including collider bias and

lack of generalisability. There may also be alternative causal explanations for observed associations, including reverse causality and residual and unmeasured confounding. Reverse causality occurs when disease causes changes in movement behaviours rather than *vice versa*. It is an important possibility as there are differences in movement behaviours between people with and without chronic disease [146], and the cardiovascular disease process may reduce functional capacity before a person presents to inpatient care [144]. Residual confounding occurs when confounding is not adequately adjusted for in analyses, and can occur when potential confounders were not measured or measured poorly. Each of these potential biases will be discussed in more detail in later sections.

In literature on device-measured movement behaviours and CVD, bias was commonly discussed, but rarely analysed quantitatively (Chapter 2). More widely in the literature, quantitative treatment of bias remains rare, despite the fact that in large studies systematic bias may be a bigger component of overall error than random sampling variation, which is routinely quantified [223]. More quantitative and comprehensive investigation of bias may support epidemiological inference: quantitative investigation clarifies the impact on inference, and considering several biases also supports overall inference and helps to prioritise the most pertinent biases for future work [224].

Quantitative bias analysis describes the process of quantitatively assessing the impact of bias on analytic results [147, 224]. It can inform interpretation of epidemiological evidence, guarding against both over-confidence and unwarranted dismissal of results [147]. Quantitative bias analysis is most useful under particular circumstances: the aim is understanding the causes of disease; there is only modest uncertainty due to sampling variation; a plausible shift in the effect size could meaningfully change inference; and there are plausible mechanisms for bias [223]. Movement behaviour epidemiology is among the sub-fields of epidemiology that meets these conditions.

Describing causal effects is often the ultimate goal, and policy relevance clear [22]. The large sample size and hence relatively precise estimates are a strength of the present work, minimising the impact of random variation (see Chapter 5). Associations are sufficiently modest that plausible shifts could change decision-making based on epidemiological evidence, especially when considering non-MVPA movement behaviours [225], and there are plausible mechanisms for bias.

6.1.2 Aims

The aim of this chapter is to investigate quantitatively the impact of several potential sources of bias on the estimates of associations between device-measured movement behaviours and risk of incident CVD derived in Chapter 5, in order to support the interpretation of earlier results and identify priorities for future studies. As different sources of bias require different approaches, this chapter addresses sources of bias in turn: specifically, it addresses information bias, selection bias, reverse causality, and residual and unmeasured confounding. This chapter interprets 'quantitative bias analysis' in a broad sense, and considers any analysis quantitatively informing assessment of bias as within scope.

6.2 Information bias due to within-person variation

6.2.1 Background

Measurement error is the difference between a measurement and the true underlying value of interest. Measurement error can arise by various mechanisms. For exam-

ple, in the context of movement behaviours it may result from fluctuations of time in different behaviours in a given week around their long-term average, from misclassification of behaviours by the classifier, or from the participant removing the device for certain behaviours (amongst other mechanisms). Consequently, measurement error can have different structure and properties. Measurement error may be random i.e. be randomly distributed with a mean of 0 [226]. It may be systematic i.e. show a systematic bias [226]. Error may occur at multiple levels: within-person and between-person [226]. One important property of measurement error is whether it is *differential* with respect to other key variables in the study, particularly the outcome [227]. Variables measured with differential error present particular difficulties for analysis [227]. Measurement error can also act additively or multiplicatively on the true value, and may follow different distributions [228]. Perhaps the simplest case is additive non-differential random error, known as *classical measurement error*. Specifically, when considering a single error-prone variable, the measured value X^* is the sum of the true value X and an error term ε: $X^* = X + \varepsilon$, where ε is a random variable with mean 0 and constant variance, and is independent of X and of the covariates and outcome in epidemiological analyses [229]. The classical measurement error model extends to the multivariable context: if $\mathbf{X}$ is a multivariable exposure $(X_1, X_2, \ldots, X_K)$, then $\mathbf{X}^* = \mathbf{X} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ is now a vector with mean zero and constant variance, and again is independent of $\mathbf{X}$, covariates, and the outcome [229].

While there may be measurement error in exposure, outcome, and confounders, this work focuses on the impact of measurement error affecting the movement behaviour exposure variables. Although device-based measurements of movement behaviours are often valued for being less error-prone than self-report [82], particularly as they are less vulnerable to recall and reporting bias, many forms of measurement error nonetheless affect device-based measurement of movement behaviours [90]. As was shown in Chapter 4, there is some misclassification of behaviours. Measurements may

be affected by participants wearing devices in different ways (e.g. placing on non-dominant wrist rather than the dominant as requested), by participants selectively removing the device, or by aspects of participant body size, shape, or mobility. Also, participant's behaviours in the week they were measured may differ from their long-term average. This list of examples is not exhaustive.

Not all these forms of measurement error are addressed in this work: in particular, the data available to characterise measurement error limits which types of error can be addressed. This thesis therefore specifically focuses on that component of measurement error in the exposure variables which manifests as within-person variation; that is, as variation in measurements made using the same method for the same person at different times. For epidemiological analyses which are concerned with effects of the 'usual' level of the exposure, but have only a single measurement with which to assess it, within-person variation is a component of measurement error (and typically an important one) [229]. Variation in time spent in different movement behaviours between weeks will manifest as within-person variation, as will random errors in measurement due to misclassification occurring in a (near-)random way.

Information bias describes bias arising from measurement error in any of the variables in a study. Measurement error can bias parameter estimates, lead to loss of power, and mask features of the data [228]. Depending on the magnitude and form measurement error takes, its impact can be substantial: for example, estimates of the effect of blood pressure on cardiovascular disease risk were revised up roughly two-fold by correcting for within-person variation [230].

With appropriate methods and data, exposure measurement error can be quantified and its impact on analyses assessed. Suitable data for quantification of measurement error can come from various types of study (for which various names are used in the literature; here the terminology of Keogh *et al.* [227] is used). A validation study is

a study which has a gold-standard reference measurement alongside the error-prone measurement [227]. Where this is not available, a calibration study is similar but the reference measurement is not a gold-standard but can be assumed to be subject only to random error [227]. Sometimes, no such reference is available, but instead a replicates study, whereby the error-prone measurement is repeated, is performed [227]. The data available determines what kinds of measurement error can be addressed: for example, a replicates study cannot be used to address systematic measurement error since replicates will all be equally affected (addressing systematic error requires a reference measurement).

The overall variance in measurements of a particular variable is made up of the within-person variance, σ_W^2 (i.e. within-person variation), and the between-person variance, σ_B^2 (i.e. differences between measurements for different individuals). The proportion of the overall variance in the measurement which is between-person variance, $\frac{\sigma_B^2}{\sigma_B^2+\sigma_W^2}$, can be used to characterise within-person variation. Here this quantity will be denoted λ. As it represents the proportion of variance *not* due to within-person variation, higher values of λ indicate measurements which are more stable within an individual. λ has various names, including 'attenuation factor' and 'regression dilution coefficient', and it can be estimated in various ways [231]. Similar principles apply in the context of multiple variables measured with within-person variation, although results can no longer be summarised so straightforwardly via a single interpretable number.

Perhaps the most widely-used approach to quantify the impact of measurement error on analyses is 'regression calibration' [227, 232]. Its motivating principle is that, rather than using the error-prone measured variable(s) to estimate the model, the expected/predicted true value of the variable(s) can be used instead [227]. The coefficient obtained using these expected values in the model can be interpreted as the

coefficient for the true value i.e. it is 'corrected' for measurement error. This method is closely related to simply dividing the coefficient calculated using the measured variable by λ (in the simple case of within-person variation in a single error-prone variable in the context of a linear model and with λ derived by regression, it is exactly equivalent [229]). Regression calibration has been used in epidemiological studies since the 1990s, particularly in nutritional epidemiology [227, 233].

Compared to within-person variation in some other exposures, within-person variation in movement behaviour measurements has been little-studied [227]; evidence on device-measured movement behaviours is particularly scarce. A few studies have quantified within-person variation in device-measured movement behaviours [234–237]. However, these studies used hip-worn devices and most were short-term (week-to-week or within-year) [234–237]. Within-person variation in the medium term and for wrist-based measurements is poorly understood. There is little evidence on the impact of within-person variation in device-measured movement behaviours on analytic results.

Compositional data presents a particular and potentially challenging context for considering within-person variation. First, compositional data is inherently multivariable, and within-person variation in multivariable data is hard to assess qualitatively, as it can produce bias in any direction [229]. Moreover, in the compositional context, variation in relative time in one movement behaviour entails variation in relative time spent in at least one other movement behaviour. That is, the compositional structure of the data affects the structure of within-person variation. While particularities of measurement error in the compositional context have long been noted [238], analytic attempts to address measurement error or within-person variation in compositional data remain rare [239]. By developing an approach to within-person variation on the simplex, Appendix D shows that when working with isometric log-ratio coordinates,

standard approaches to multivariable within-person variation (such as regression calibration), can be applied in the compositional context.

6.2.2 Methods

Study participants

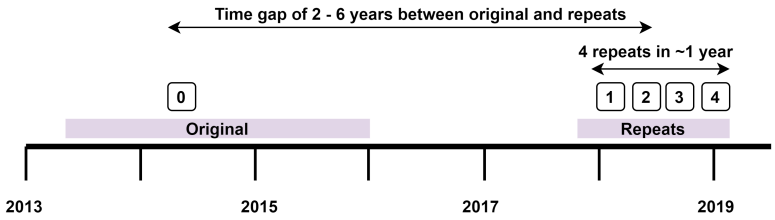
A subset of participants in the UK Biobank accelerometer study were invited to participate in repeated accelerometer wear. To ensure good coverage across sex and age groups, the study invited participants using a stratified sampling procedure. Participants with very good-quality data (including close to 7 days wear time, monitor returned promptly) in the original accelerometer study were stratified into eight groups based on sex and ten-year age bands, and invitations were sent to participants sampled randomly within these groups.

Measurement protocol

The measurement protocol for the repeated measurements study involved a further four accelerometer wear periods, which were equally spaced over a roughly one year period between November 2017 and January 2019. At each repeat participants were asked to complete the same wear protocol as in the original study. Therefore, a participant who completed the protocol perfectly and for whom there were no data quality issues would have accelerometer data on five occasions: one from the original study, plus four measurements spread approximately evenly over a yearlong period 2-6 years later. Figure 6.1 illustrates this.

Accelerometer data from the repeated measurements study data was processed using the same procedure as the original study (see Chapter 4). Due to a change in a

Figure 6.1: Example of the measurement protocol for the repeated measurements study.



dependency package, the machine-learning model was updated relative to Chapters 4 and 5, but the same procedure was used and values were very similar. At each repeat, the same quality control criteria were applied as in the original study (based on calibration, wear time, 'clips', and extreme values, again derived using 5-second epoch data; see Chapter 4). Participants were included in the analytic sample if they had good-quality data in the original accelerometer data collection and from at least one repeat. Only participants appearing in the analytic sample from Chapter 5 were included (i.e. participants with prior cardiovascular disease or missing covariate data were excluded).

Structure of the analytic dataset

This data takes the form of repeated measurements on a small subset of individuals in the study. However, the asymmetry in measurement timing (one at the original assessment; four within a one-year period 2-6 years later) is unusual in a repeated measurements study and creates a question over how to best analyse the data. For this study, the mean of good-quality measurements in the repeated measurements study was used to give a single follow-up value. For someone with four good-quality measurements this would be the mean of four repeated measurements; where only one to three of the measurements were available and of good quality the mean of these measurements was used.

This mean is no longer a true repeated measurement, as it has a longer overall measurement period and accounts for the component of variability due to time of year. Using the terminology of Keogh *et al.* [227] introduced above, with the original measurement and the mean of repeated measurements, the data structure resembles an internal calibration study more than true repeated measurements data. It should be noted throughout, however, that the fact the follow-up measurement also comes from repeated measurements data (and is not a gold-standard or reference measurement) means the model can only identify aspects of measurement error that manifest as within-person variation.

Descriptive characteristics of repeated measurements

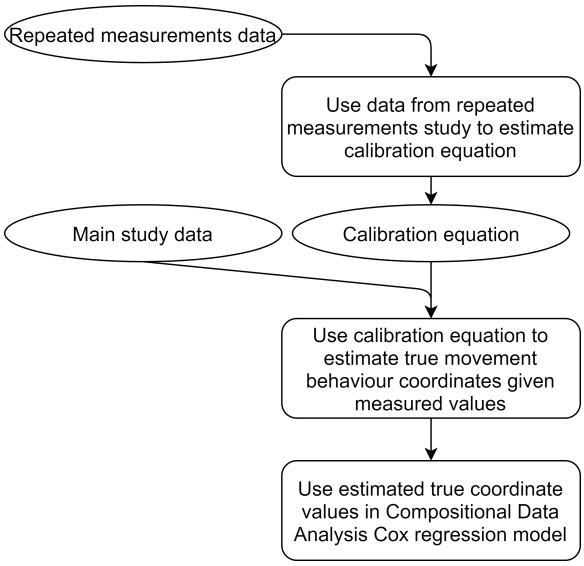
Descriptive characteristics of several different accelerometer metrics at each assessment were calculated. These metrics were: the simple summary metric of overall average acceleration, each movement behaviour individually, and the first pivot coordinate (Z_1) for each movement behaviour versus all others (as used in modelling in Chapter 5).

Univariable attenuation factors, λ, for each variable were also calculated, using the regression method attributed to Rosner *et al.* [231, 240, 241]. In detail, the mean of repeated measurements of the variable was regressed on the measurement of the variable in the original accelerometer study. This model was also adjusted for age at the time of accelerometer wear and sex (both assumed to be measured without error). The attenuation factor, λ, is obtained from this model as the coefficient of the original measurement [227].

Multivariable regression calibration

Multivariable regression calibration was used to assess the impact of within-person variation on analyses (Figure 6.2) [229]. The underlying principle of regression calibration is to replace the measured values in the model with the expected/predicted true values given the measured values. Available data is used to estimate a 'calibration equation' for the expected/predicted true values given the measured values.

Figure 6.2: Multivariable regression calibration procedure.



As noted above, the fact that 24-hour movement behaviour data is compositional data has consequences for the structure of within-person variation. It was argued above and in Appendix D that describing 24-hour movement behaviours using isometric log-ratio coordinates allows within-person variation to be modelled as additive, unconstrained within-person variation in the isometric log-ratio coordinates (Z_1, Z_2, Z_3).

The calibration equation was obtained by regressing each isometric log-ratio coordinate outcome variable (again, the mean of that variable across repeated measurements) on observed values of all three isometric log-ratio coordinates from the original accelerometer study. The model was also adjusted for age at accelerometer wear, sex,

ethnicity, deprivation, education, smoking status, alcohol consumption, fresh fruit and vegetable consumption, red and processed meat consumption, and oily fish consumption (i.e. the same variables as used in the Cox regression model for incident CVD in Chapter 5). For modelling purposes, the non-movement-behaviour variables were assumed to be measured without error.

The calibration equation was then applied in the full UK Biobank accelerometer cohort to predict the expected true value of the isometric log-ratio coordinates for each participant given observed values of the isometric log-ratio coordinates and covariates.

The Cox regression model was then estimated using the same methodology as in Chapter 5, but with the expected true values of the isometric log-ratio coordinates in place of their observed values. Using this approach, the hazard ratios in the Cox regression model are 'corrected' for within-person variation in the movement behaviour variables.

Following Keogh *et al.* [227], uncertainty on coefficient values was estimated using bootstrapping. While analytic formulae are available to estimate uncertainty, the bootstrap is often easier to implement in practice [227]. Specifically, 2,000 bootstrapped samples were drawn. A stratified sampling procedure was used: a bootstrap sample was drawn separately for participants with repeated measurements data and for those without, and the overall sample for the Cox model was formed from these two samples [227, 228]. This ensured the sub-sample with repeated measurements was of fixed size [228]. 95% confidence intervals were then calculated using the basic (also known as empirical or reverse percentile) bootstrap [242].

Modelling assumptions

A key assumption of regression calibration is that measurement error is non-differential [227, 228]. This is reasonable in this context, especially as exposure measurement preceded development of the outcome often by years. Nonetheless, it is possible to construct scenarios in which it may be violated e.g. if measurement accuracy depends on participant fitness, which is also associated with the outcome. A second assumption is that the form of the calibration equation, here a linear regression model, is a good model for the expected true values given the observed values and covariates [227]. In addition, as multivariable regression calibration is only approximately unbiased in the context of Cox regression (unlike the context of linear regression, in which it is unbiased), further assumptions are required for the approximation to be good. Specifically, the disease should be rare and either the log-hazard ratio for the exposure moderate or the measurement error variance small [227]. These conditions are met here. Finally, the parameters for the calibration equation estimated in the (non-representative) subset with repeated measurements data must be assumed to generalise to the whole accelerometer cohort.

Sensitivity analyses

Sensitivity analyses were conducted to assess whether results were affected by direct incorporation of month of measurement in the calibration model, or by precision weighting the calibration model according to the number of repeated measurements used to calculate the mean repeated measurement (the outcome in the model).

Code availability

Data processing and analyses were performed in Python 3.7.4 [243] and R 4.2.1 [200]. Code can be found at `github.com/OxWearables/repeatedMeasurementsProcessing` (accelerometer data processing), `github.com/OxWearables/thesis_additional_walmsley` (analysis), and `github.com/OxWearables/repeatedMeasurements` (functions to support analyses on within-person variation).

6.2.3 Results

5,504 participants (688 in each of the eight sampling subgroups) were contacted, of whom data was obtained from 3,197 participants, and 3,154 had valid data for at least one of the repeated measurements. 2,892 of these participants appeared in the analytic sample from Chapter 5 and so were included. Characteristics of the cohort with repeated measurements are shown in Table 6.1. The sampling strategy means that the cohort is more age- and sex-balanced than the overall accelerometer cohort; other participant characteristics were broadly similar.

Table 6.1: Comparison of accelerometry study population and sub-sample with repeated accelerometer measurements.

	Accelerometry study population (n=87,498)	**Repeated measurements sample** (n=2,892)
Age at accelerometer wear (years)		
Mean (SD)	62.0 (7.8)	58.9 (9.4)
Median [Q1, Q3]	63.0 [55.9, 68.2]	58.7 [50.3, 66.9]
Sex		
Female	50,882 (58%)	1,512 (52%)
Male	36,616 (42%)	1,380 (48%)
Ethnicity		
Asian	756 (1%)	21 (1%)
Black	701 (1%)	17 (1%)
Mixed/Other	1,151 (1%)	40 (1%)
White	84,890 (97%)	2,814 (97%)
Townsend Deprivation Index		
Mean (SD)	-1.76 (2.80)	-1.61 (2.84)
Median [Q1, Q3]	-2.47 [-3.83, -0.23]	-2.30 [-3.70, -0.05]

	Accelerometry study population (n=87,498)	**Repeated measurements sample** (n=2,892)
Education		
School leaver	19,535 (22%)	535 (18%)
Further education	29,061 (33%)	954 (33%)
Higher education	38,902 (44%)	1,403 (49%)
Smoking status		
Never smoker	50,888 (58%)	1,785 (62%)
Ex-smoker	30,717 (35%)	934 (32%)
Current smoker	5,893 (7%)	173 (6%)
Alcohol consumption frequency		
Never	4,745 (5%)	137 (5%)
Less than 3 times/week	39,760 (45%)	1,357 (47%)
3 or more times/week	42,993 (49%)	1,398 (48%)
Daily fruit and vegetable consumption (servings/day)		
Less than 3	3,595 (4%)	119 (4%)
3-4.9	14,293 (16%)	494 (17%)
5-7.9	36,991 (42%)	1,271 (44%)
8 or more	32,619 (37%)	1,008 (35%)
Body Mass Index (kg/m^2)		
Mean (SD)	26.6 (4.5)	26.4 (4.7)
Median [Q1, Q3]	25.9 [23.5, 28.9]	25.7 [23.3, 28.5]

Q1,Q3 - quartile 1, quartile 3; SD - standard deviation.

Median values of accelerometer metrics were fairly consistent across repeated measurements (Table 6.2). Some patterning was apparent, with slightly higher median time in moderate-to-vigorous physical activity behaviours and lower median time in sleep in spring/summer measurements. λ (lambda) values for the accelerometer metrics varied between 0.59 (pivot coordinate, light physical activity behaviours) and 0.71 (overall average acceleration; Table 6.2). The metrics with least within-person variation were overall average acceleration ($\lambda = 0.71$) and sedentary behaviour ($\lambda = 0.69$ absolute time, 0.70 pivot coordinate). Moderate-to-vigorous physical activity behaviours showed more within-person variation ($\lambda = 0.62$ absolute time, 0.62 pivot coordinate).

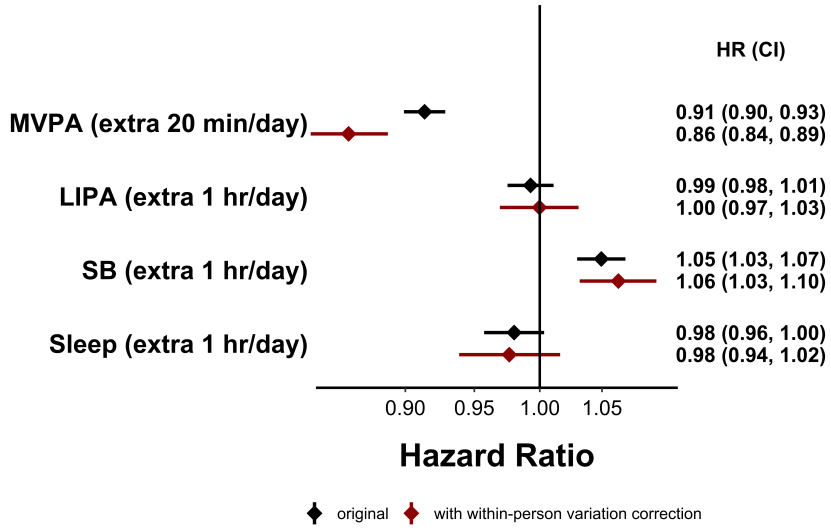
Table 6.2: Accelerometer measurements at each repeat and univariable attenuation factor.

Variable	Repeat 0[a] (Original)	Repeat 1[a] (Winter)	Repeat 2[a] (Spring)	Repeat 3[a] (Summer)	Repeat 4[a] (Autumn)	Lambda (SE)[b]
n	2,892	2,193	2,551	2,471	2,463	NA
Wear date	2014-11-14 (2014-05-27, 2015-05-29)	2018-02-08 (2018-01-08, 2018-03-01)	2018-05-05 (2018-04-13, 2018-05-24)	2018-08-03 (2018-07-13, 2018-08-24)	2018-11-01 (2018-10-12, 2018-11-26)	NA
Overall average acceleration (mg)	28.1 (23.5, 33.7)	26.1 (21.5, 31.9)	27.6 (22.8, 33.4)	27.4 (22.4, 32.9)	26.4 (21.5, 31.6)	0.71 (0.01)
ML moderate-to-vigorous physical activity (hr/day)	0.47 (0.24, 0.80)	0.40 (0.19, 0.72)	0.48 (0.23, 0.84)	0.48 (0.23, 0.83)	0.45 (0.20, 0.77)	0.62 (0.01)
ML light physical activity (hr/day)	5.4 (4.4, 6.6)	5.3 (4.2, 6.4)	5.5 (4.4, 6.6)	5.4 (4.4, 6.6)	5.3 (4.3, 6.4)	0.66 (0.01)
ML sedentary behaviour (hr/day)	9.3 (8.1, 10.5)	9.3 (8.1, 10.6)	9.2 (8.0, 10.5)	9.4 (8.1, 10.6)	9.3 (8.1, 10.6)	0.69 (0.01)
ML sleep (hr/day)	8.5 (7.9, 9.2)	8.7 (8.0, 9.5)	8.4 (7.8, 9.2)	8.4 (7.8, 9.1)	8.6 (8.0, 9.4)	0.66 (0.01)
1st pivot coordinate: MVPA vs other	-2.40 (-2.98, -1.92)	-2.55 (-3.20, -2.02)	-2.37 (-3.00, -1.88)	-2.40 (-3.02, -1.90)	-2.45 (-3.15, -1.95)	0.62 (0.01)
1st pivot coordinate: LIPA vs other	0.45 (0.18, 0.73)	0.47 (0.18, 0.75)	0.46 (0.19, 0.74)	0.45 (0.18, 0.72)	0.45 (0.17, 0.72)	0.59 (0.01)
1st pivot coordinate: SB vs other	1.06 (0.80, 1.33)	1.11 (0.84, 1.40)	1.04 (0.79, 1.32)	1.05 (0.81, 1.32)	1.08 (0.82, 1.36)	0.70 (0.01)
1st pivot coordinate: sleep vs other	0.95 (0.77, 1.18)	1.04 (0.82, 1.29)	0.94 (0.75, 1.18)	0.95 (0.75, 1.18)	0.99 (0.80, 1.25)	0.63 (0.01)

[a] Columns for each repeat show Median (Inter-Quartile Range) of the named variable at the given repeat.
[b] Lambda (SE) is the attenuation factor and its standard error, calculated by age- and sex-adjusted regression (see main text).

Assessing the impact of multivariable within-person variation on the main analyses, there was a fairly modest impact on analytic results (Figure 6.3). The magnitude of association for reallocating time from other behaviours into moderate-to-vigorous physical activity behaviours became more marked after accounting for within-person variation. Other results changed very modestly or not at all. All estimates were less precise after accounting for within-person variation. The multivariable structure of within-person variation in movement behaviour data was clear: unlike the univariable regression dilution coefficients, which would imply all associations should be de-attenuated by accounting for within-person variation, some associations were little affected by accounting for multivariable within-person variation.

Figure 6.3: Hazard Ratios for incident cardiovascular disease associated with reallocating time to a given movement behaviour, from all other behaviours proportionally, among 87,498 participants (4,105 events) with and without 'correction' for within-person variation using multivariable regression calibration.



All relative to the mean behaviour composition (8.8 hours/day sleep, 9.3 hours/day sedentary behaviour, 5.6 hours/day light physical activity behaviours, 0.35 hours/day (21 min/day) moderate-to-vigorous physical activity behaviours) and more time in named behaviour reallocated from all other behaviours proportionally. Calibration equation was adjusted for age, sex, ethnicity, deprivation, education, smoking status, alcohol consumption, fresh fruit and vegetable consumption, red and processed meat consumption, and oily fish consumption. Cox model used age as the timescale, was stratified by sex and was additionally adjusted for ethnicity, deprivation, education, smoking status, alcohol consumption, fresh fruit and vegetable consumption, red and processed meat consumption, and oily fish consumption.

Results were not materially changed by incorporation of month of year into the calibration equation (Supplementary Figure A.1), or by precision weighting the model based on the number of valid repeated measurements (Supplementary Figure A.2).

6.3 Selection bias

6.3.1 Background

Although UK Biobank invitations were sent to everyone registered with an NHS general practitioner, of the relevant age, and living near an assessment centre (over 9 million people), the overall response rate was low (5.5%) and there were selection effects [161]. For example, men, people from ethnic minorities, smokers, and people living in deprived areas are under-represented [180]. Participants are in better health than the general population of the same age (a 'healthy volunteer' effect): Fry *et al.* [180] found 46.2% lower all-cause mortality at age 70-74 for men in UK Biobank compared to the general population, and 55.5% lower all-cause mortality at age 70-74 in women. Selection bias following patterns seen in UK Biobank is common in prospective cohort studies, though a lack of resource to recruit non-responders may have accentuated it in UK Biobank [164].

As well as selection into the overall study, there are selection effects into the accelerometer sub-study of UK Biobank (shown descriptively in Table 3.1, Chapter 3). Tyrrell *et al.* [244] studied participation in optional components of UK Biobank, including the accelerometer sub-study, food frequency questionnaires, and mental health questionnaire. They found evidence for genetic predictors of participation in these optional components, and used Mendelian Randomisation to show causal effects of several traits (e.g. education, adiposity) on participation [244]. As with any longitudinal

study, attrition of participants is also a form of selection [245], although the passive nature of follow-up in UK Biobank and the low overall mortality rate means this is relatively modest.

Selection bias can have several analytic consequences: spurious associations can be induced where selection constitutes conditioning on a collider (a variable caused by both exposure and outcome), and findings may fail to generalise due to different effects in the selected sub-sample compared to the target population [143]. Munafò *et al.* [246] argue that collider bias may have a meaningful impact on associations even under modest selection. Others argue that, for aetiological analyses with appropriate adjustment, sample representativeness is not an important concern [163]. Selection can sometimes even support scientific inference, as a homogeneous sample reduces confounding [247]. This underlines the value of quantitative assessment of selection bias in practice in particular contexts.

Quantitatively assessing the impact of selection bias is difficult, as information on non-participants can be hard to obtain [142]. To assess the possible impact of selection bias on several associations, Batty *et al.* [248] compared estimates in UK Biobank with estimates for the same risk factors in more representative samples (Health Survey for England and Scottish Health Survey). Direct analytic approaches to selection bias are typically based on weighting of observations. Inverse Probability of Selection Weights aim to weight participants inversely to their probability of selection [245]. Weighting participants in model estimation inversely to their probability of selection means the cohort used to calculate the model more closely matches the distributions in the target population [245]. Many different approaches can be used to calculate appropriate probabilities to use in the weights [249]. The requirements for data to inform these probabilities is often the limiting factor: data on non-participants is usually required, although the particular form this takes can vary.

The structure of the UK Biobank accelerometry study, which is nested within the main UK Biobank cohort, offers a particular opportunity to investigate one selection process in detail: selection from the overall UK Biobank cohort into the accelerometry sub-study. There are known selection effects for the accelerometer sub-cohort versus the overall cohort. As there is baseline data on unselected participants from the main UK Biobank cohort, this is a more tractable problem than selection into the overall cohort. Mechanisms affecting selection from the target population into UK Biobank also affect selection within the UK Biobank (e.g. healthy volunteer effect, education), meaning analysing selection within UK Biobank can give broader insights. As previous analyses to quantify the impact of selection focussed on selection between the target population and the overall cohort [248, 250], this also complements existing literature.

6.3.2 Methods

To study the impact of selection within the UK Biobank on results in the accelerometer sub-cohort, an Inverse Probability of Selection Weighting approach was used, with logistic regression used to model probability of selection. First, selection probabilities from the overall cohort into the final analytic sample were modelled using logistic regression. All participants who had not experienced a hospital-recorded or self-reported cardiovascular event at the date of their baseline assessment were included in the model (i.e. all those participants who, at the time of baseline assessment, were theoretically eligible for inclusion in this study). The logistic regression model was adjusted for a broad set of variables which may be associated with selection, including several biomarkers and self-reported activity variables. These variables were: age (at baseline), sex, ethnicity, deprivation, education, smoking status, alcohol consumption, fruit and vegetable consumption, red and processed meat consumption, oily

fish consumption, Body Mass Index (BMI), systolic blood pressure (SBP), HbA1c, C Reactive protein (CRP), LDL cholesterol, HDL cholesterol, triglycerides, poor self-reported overall health, self-reporting taking medications for diabetes, cholesterol or blood pressure, MET minutes per week spent walking according to self-report, MET minutes per week spent in moderate physical activity according to self-report, MET minutes per week spent in vigorous physical activity according to self-report, and self-reported usual walking pace. Some variables with skewed distributions (based on visual inspection) were log-transformed: specifically, these variables were HbA1c, CRP, and triglycerides. MET minutes per week in different activities according to self-report were categorised for analyses (0, 1-500, 501-1,000, 1,001+ MET minutes/week). To avoid excluding participants due to missing data in variables used in the selection model, multiple imputation was used to impute missing values. Variables were assumed to be Missing At Random (i.e. missingness explained by measured variables only). Five datasets were imputed using the 'mice' R package [251, 252]. All variables in the selection model, including the indicator of presence in the analytic sample, were used to impute all others. Default methods were used i.e. predictive mean matching for numeric data and polytomous regression imputation for multi-level factors (which were all treated as unordered for this purpose), and the default five iterations were used, with convergence checked by inspection [253]. Each imputed dataset was then used to estimate a logistic regression model.

To obtain Inverse Probability of Selection Weights, the logistic regression model in each imputed dataset was used to estimate the log-odds of presence in the analytic sample for each participant. To obtain a single value per participant, the mean log-odds for that participant across models/imputed datasets was calculated. $\mathbb{P}(selection|C = c)$, for a participant with covariate values c, was obtained by inverting the logit transformation. Stabilised Inverse Probability of Selection Weights were

calculated as:

$$w_i = \frac{\mathbb{P}(selection)}{\mathbb{P}(selection|C = c)}$$

To avoid extreme weight values, weights were trimmed at the 99th percentile, with all weights above this assigned the same weight as the 99th percentile [254].

To verify the weights behaved as expected, a descriptive table on weights in the analytic sample was produced, and Odds Ratios from the pooled logistic regression model were examined.

A Cox proportional hazards model was then derived using these Inverse Probability of Selection Weights as sampling weights. The model followed exactly the methodology of the model in Chapter 5, except for use of weights and a robust estimator for the standard errors [217]. The robust estimator is a sandwich-type estimator incorporating the weights, as implemented in the 'survival' package [216, 217]. [It may be noted that, for the purpose of this model, weights are treated as known i.e. it does not account for uncertainty in the procedure for estimating the weights.]

A sensitivity analysis used only those participants who had not died or experienced a cardiovascular disease event by 2013-06-01, when accelerometer measurements began, to model selection weights i.e. it considered a cohort who were alive and eligible for inclusion in this study at the time of accelerometer data collection (rather than the cohort who were theoretically eligible for inclusion in this study at the time of baseline assessment, as in the main analysis).

Analyses were performed in R 4.2.1 [200]. Code can be found at `github.com/OxWearables/thesis_additional_walmsley`.

6.3.3 Results

Table 6.3 shows how participant characteristics varied by weighting for participants in the analytic sample. Higher weights were given to male participants, non-white participants, participants with less advanced qualifications, and participants in poorer health. Supplementary Table A.5 shows the Odds Ratios for selection associated with participant characteristics.

Table 6.3: Participant characteristics by Inverse Probability of Selection Weight in the analytic sample.

	Quarter 1 - Lowest weight (n=21,874)	**Quarter 2** (n=21,874)	**Quarter 3** (n=21,874)	**Quarter 4 - Highest weight** (n=21,875)	**Overall** (n=87,498)
Overall activity (mg)					
Mean (SD)	30.5 (8.3)	28.7 (8.0)	27.9 (8.0)	26.5 (8.1)	28.4 (8.2)
Median [Q1, Q3]	29.4 [24.9, 34.9]	27.7 [23.2, 32.9]	27.0 [22.4, 32.3]	25.6 [21.0, 30.8]	27.4 [22.8, 32.9]
Age at accelerometer wear (years)					
Mean (SD)	60.5 (7.6)	61.9 (7.8)	62.6 (7.8)	63.1 (7.9)	62.0 (7.8)
Median [Q1, Q3]	60.8 [54.3, 66.6]	62.8 [55.8, 68.0]	63.8 [56.6, 68.7]	64.7 [57.0, 69.4]	63.0 [55.9, 68.2]
Sex					
Female	15,132 (69%)	12,161 (56%)	12,102 (55%)	11,486 (53%)	50,882 (58%)
Male	6,742 (31%)	9,713 (44%)	9,772 (45%)	10,389 (47%)	36,616 (42%)
Ethnicity					
Asian	0 (0%)	3 (0%)	42 (0%)	711 (3%)	756 (1%)
Black	0 (0%)	4 (0%)	62 (0%)	635 (3%)	701 (1%)
Mixed/Other	22 (0%)	165 (1%)	372 (2%)	592 (3%)	1,151 (1%)
White	21,852 (100%)	21,702 (99%)	21,398 (98%)	19,937 (91%)	84,890 (97%)
Townsend Deprivation Index					
Mean (SD)	-2.50 (2.37)	-2.01 (2.63)	-1.76 (2.71)	-0.76 (3.14)	-1.76 (2.80)
Median [Q1, Q3]	-3.03 [-4.18, -1.34]	-2.68 [-3.94, -0.60]	-2.43 [-3.76, -0.28]	-1.49 [-3.21, 1.36]	-2.47 [-3.83, -0.23]
Education					
School leaver	2 (0%)	316 (1%)	4,888 (22%)	14,329 (66%)	19,535 (22%)
Further education	2,342 (11%)	9,139 (42%)	11,928 (55%)	5,652 (26%)	29,061 (33%)
Higher education	19,530 (89%)	12,419 (57%)	5,058 (23%)	1,894 (9%)	38,902 (44%)
Smoking status					
Never smoker	13,912 (64%)	13,735 (63%)	12,391 (57%)	10,850 (50%)	50,888 (58%)
Ex-smoker	7,901 (36%)	7,623 (35%)	7,913 (36%)	7,279 (33%)	30,717 (35%)
Current smoker	61 (0%)	516 (2%)	1,570 (7%)	3,746 (17%)	5,893 (7%)
Alcohol consumption frequency					
Never	295 (1%)	822 (4%)	1,185 (5%)	2,443 (11%)	4,745 (5%)
Less than 3 times/week	7,051 (32%)	9,521 (44%)	10,862 (50%)	12,326 (56%)	39,760 (45%)
3 or more times/week	14,528 (66%)	11,531 (53%)	9,827 (45%)	7,106 (32%)	42,993 (49%)
Daily fruit and vegetable consumption (servings/day)					
Less than 3	215 (1%)	510 (2%)	939 (4%)	1,931 (9%)	3,595 (4%)
3-4.9	2,808 (13%)	3,487 (16%)	3,854 (18%)	4,143 (19%)	14,293 (16%)
5-7.9	10,723 (49%)	9,534 (44%)	8,879 (41%)	7,855 (36%)	36,991 (42%)
8 or more	8,128 (37%)	8,343 (38%)	8,202 (37%)	7,946 (36%)	32,619 (37%)

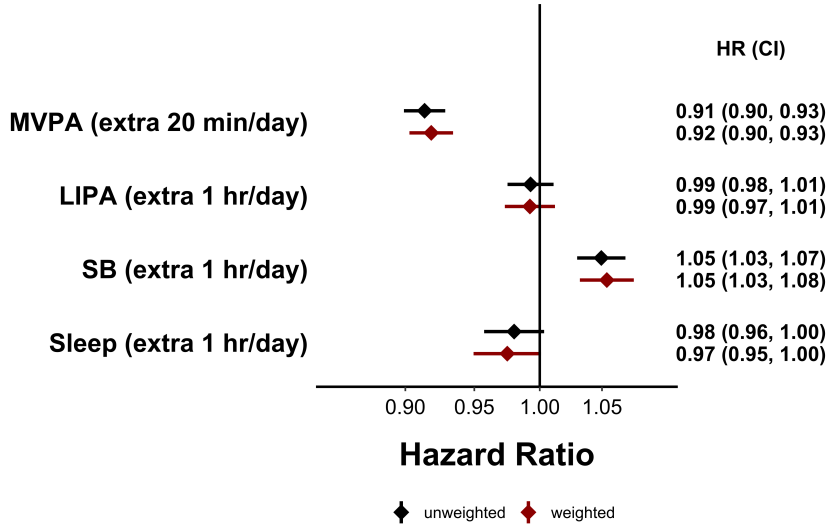
	Quarter 1 - Lowest weight (n=21,874)	Quarter 2 (n=21,874)	Quarter 3 (n=21,874)	Quarter 4 - Highest weight (n=21,875)	Overall (n=87,498)
Red and processed meat intake (times/week)					
Less than 1	3,179 (15%)	1,624 (7%)	1,242 (6%)	977 (4%)	7,023 (8%)
1-2.9	9,394 (43%)	8,478 (39%)	8,308 (38%)	7,619 (35%)	33,799 (39%)
3-4.9	5,547 (25%)	6,440 (29%)	6,694 (31%)	6,964 (32%)	25,645 (29%)
5 or more	3,754 (17%)	5,332 (24%)	5,630 (26%)	6,315 (29%)	21,031 (24%)
Oily fish intake (times/week)					
Less than 1	9,514 (43%)	8,992 (41%)	9,510 (43%)	10,201 (47%)	38,218 (44%)
1	9,216 (42%)	9,004 (41%)	8,341 (38%)	7,642 (35%)	34,203 (39%)
2-4	3,075 (14%)	3,713 (17%)	3,830 (18%)	3,773 (17%)	14,391 (16%)
More than 4	69 (0%)	165 (1%)	193 (1%)	259 (1%)	686 (1%)
Body Mass Index (kg/m^2)					
Mean (SD)	24.0 (2.9)	26.0 (3.6)	27.2 (4.3)	29.0 (5.3)	26.6 (4.5)
Median [Q1, Q3]	23.8 [22.0, 25.8]	25.7 [23.6, 28.2]	26.7 [24.2, 29.6]	28.2 [25.4, 31.7]	25.9 [23.5, 28.9]
Systolic blood pressure (mmHg)					
Mean (SD)	130 (16)	136 (18)	138 (18)	141 (18)	136 (18)
Median [Q1, Q3]	128 [118, 140]	134 [124, 147]	137 [126, 150]	140 [128, 153]	135 [123, 148]
Missing	11 (0%)	10 (0%)	8 (0%)	29 (0%)	58 (0%)
HbA1c (mmol/mol)					
Mean (SD)	33.8 (3.6)	34.7 (4.4)	35.4 (5.2)	37.0 (7.0)	35.2 (5.3)
Median [Q1, Q3]	33.9 [31.6, 36.0]	34.5 [32.2, 36.7]	35.0 [32.6, 37.3]	35.9 [33.4, 38.7]	34.7 [32.4, 37.1]
Missing	1,282 (6%)	1,246 (6%)	1,278 (6%)	1,446 (7%)	5,252 (6%)
C Reactive Protein (mg/L)					
Mean (SD)	1.22 (2.29)	1.96 (3.70)	2.36 (3.73)	3.28 (4.95)	2.20 (3.86)
Median [Q1, Q3]	0.68 [0.38, 1.28]	1.03 [0.56, 2.01]	1.30 [0.67, 2.57]	1.82 [0.91, 3.64]	1.12 [0.57, 2.33]
Missing	1,178 (5%)	1,169 (5%)	1,171 (5%)	1,275 (6%)	4,793 (5%)
LDL cholesterol (mmol/L)					
Mean (SD)	3.52 (0.79)	3.63 (0.83)	3.64 (0.83)	3.61 (0.88)	3.60 (0.83)
Median [Q1, Q3]	3.46 [2.96, 4.02]	3.58 [3.07, 4.15]	3.60 [3.05, 4.17]	3.57 [3.00, 4.18]	3.55 [3.02, 4.13]
Missing	1,166 (5%)	1,172 (5%)	1,155 (5%)	1,262 (6%)	4,755 (5%)
HDL cholesterol (mmol/L)					
Mean (SD)	1.65 (0.39)	1.51 (0.37)	1.46 (0.37)	1.38 (0.35)	1.50 (0.38)
Median [Q1, Q3]	1.62 [1.37, 1.89]	1.47 [1.25, 1.74]	1.41 [1.19, 1.68]	1.34 [1.12, 1.59]	1.46 [1.22, 1.73]
Missing	2,987 (14%)	2,866 (13%)	2,892 (13%)	2,887 (13%)	11,632 (13%)
Triglycerides (mmol/L)					
Mean (SD)	1.31 (0.71)	1.59 (0.89)	1.76 (1.00)	1.92 (1.09)	1.64 (0.96)
Median [Q1, Q3]	1.12 [0.84, 1.56]	1.37 [0.99, 1.94]	1.50 [1.07, 2.17]	1.65 [1.16, 2.38]	1.39 [0.99, 2.02]
Missing	1,145 (5%)	1,138 (5%)	1,136 (5%)	1,252 (6%)	4,671 (5%)
Poor self-reported overall health[a]					
Yes	24 (0%)	98 (0%)	299 (1%)	1,340 (6%)	1,761 (2%)
No	21,850 (100%)	21,776 (100%)	21,575 (99%)	20,535 (94%)	85,737 (98%)
Self-reported medications for diabetes, cholesterol or blood pressure[a]					
Yes	1,571 (7%)	3,587 (16%)	4,985 (23%)	7,025 (32%)	17,168 (20%)
No	20,303 (93%)	18,287 (84%)	16,889 (77%)	14,850 (68%)	70,330 (80%)
Self-reported weekly walking (MET minutes)					
No walking	285 (1%)	445 (2%)	601 (3%)	994 (5%)	2,325 (3%)
1-500	9,161 (42%)	8,210 (38%)	7,382 (34%)	6,546 (30%)	31,299 (36%)
501-1000	5,236 (24%)	4,542 (21%)	4,153 (19%)	3,387 (15%)	17,319 (20%)
1001+	4,808 (22%)	5,994 (27%)	6,532 (30%)	6,971 (32%)	24,305 (28%)
Missing	2,384 (11%)	2,683 (12%)	3,206 (15%)	3,977 (18%)	12,250 (14%)
Self-reported weekly moderate physical activity (MET minutes)					
No moderate activity	1,769 (8%)	2,437 (11%)	2,808 (13%)	3,641 (17%)	10,655 (12%)
1-500	9,991 (46%)	8,340 (38%)	7,238 (33%)	5,672 (26%)	31,242 (36%)
501-1000	4,235 (19%)	3,556 (16%)	3,300 (15%)	2,667 (12%)	13,758 (16%)
1001+	3,495 (16%)	4,858 (22%)	5,322 (24%)	5,918 (27%)	19,593 (22%)
Missing	2,384 (11%)	2,683 (12%)	3,206 (15%)	3,977 (18%)	12,250 (14%)
Self-reported weekly vigorous physical activity (MET minutes)					
No vigorous activity	4,338 (20%)	6,322 (29%)	7,404 (34%)	9,022 (41%)	27,086 (31%)
1-500	6,836 (31%)	5,824 (27%)	4,924 (23%)	3,648 (17%)	21,232 (24%)
501-1000	3,989 (18%)	2,840 (13%)	2,407 (11%)	1,734 (8%)	10,970 (13%)

	Quarter 1 - Lowest weight (n=21,874)	Quarter 2 (n=21,874)	Quarter 3 (n=21,874)	Quarter 4 - Highest weight (n=21,875)	Overall (n=87,498)
1001+	4,327 (20%)	4,205 (19%)	3,933 (18%)	3,494 (16%)	15,960 (18%)
Missing	2,384 (11%)	2,683 (12%)	3,206 (15%)	3,977 (18%)	12,250 (14%)
Self-reported usual walking pace					
Brisk pace	17,138 (78%)	11,324 (52%)	8,783 (40%)	4,924 (23%)	42,170 (48%)
Steady average pace	4,726 (22%)	10,382 (47%)	12,406 (57%)	14,181 (65%)	41,695 (48%)
Slow pace	0 (0%)	155 (1%)	665 (3%)	2,651 (12%)	3,471 (4%)
None of the above	0 (0%)	0 (0%)	0 (0%)	66 (0%)	66 (0%)
Missing	10 (0%)	13 (0%)	20 (0%)	53 (0%)	96 (0%)

[a] Lack of missing data in self-reported poor overall health and medications variables are by design: non-responding participants were treated as 'No'.
Q1, Q3 - quartile 1, quartile 3; SD - standard deviation.

Figure 6.4 shows the impact of weighting the Cox model by Inverse Probability of Selection. Associations observed were similar with and without weighting.

Figure 6.4: Hazard Ratios for incident cardiovascular disease associated with reallocating time to a given movement behaviour, from all other behaviours proportionally, among 87,498 participants (4,105 events) with and without weighting by Inverse Probability of Selection from main study into accelerometer sub-cohort.



All relative to the (unweighted) mean behaviour composition (8.8 hours/day sleep, 9.3 hours/day sedentary behaviour, 5.6 hours/day light physical activity behaviours, 0.35 hours/day (21 min/day) moderate-to-vigorous physical activity behaviours) and more time in named behaviour reallocated from all other behaviours proportionally. Models used age as the timescale, were stratified by sex and were additionally adjusted for ethnicity, deprivation, education, smoking status, alcohol consumption, fresh fruit and vegetable consumption, red and processed meat consumption, and oily fish consumption. Derivation of weights used the procedure described in the text.

In the sensitivity analysis using a more restricted cohort for calculation of weights, there were some differences in the Odds Ratios in the selection weighting model

(Supplementary Table A.6), but results of the weighted analysis were very similar (Supplementary Figure A.3).

6.4 Reverse causality

6.4.1 Background

Reverse causality has consistently been identified as an important possibility in epidemiological studies of 24-hour movement behaviours. Kujala [255] argued that it remains unclear whether physical activity causally increases longevity, highlighting the possibility of 'healthy exerciser bias', whereby physical activity participants are more likely to be healthy i.e. a form of reverse causality such that physical activity is a marker of underlying health, not causal in itself [255]. While most discussion has been in the context of mortality, similar considerations apply to incident disease [144]. The focus on mortality reflects authors' priorities and data availability: one of the longest-running studies using accelerometer measurement, in which reverse causality has been studied, is the NHANES study, for which follow-up is only for mortality and not for incident disease [256, 257].

Previous studies have addressed reverse causality in the context of device-measured movement behaviours, mostly with mortality as the outcome [256–258]. Established methods to address reverse causality include restriction to a healthy subgroup, exclusion of the early years of follow-up, and examining associations within particular periods of follow-up [145, 256–258]. Excluding early follow-up, a commonly-used approach in the literature review (Chapter 2), aims to exclude those cases most likely to have sub-clinical disease at time of movement behaviour measurement. CVD develops over an extended period; however, excluding early events reduces statistical power.

Often, 1-2 years are excluded, although for mortality some empirical results suggest excluding up to 5 years [256, 257]. This highlights the need for approaches which can assess reverse causality from different angles, such as examining associations across different periods of follow-up and by health status.

6.4.2 Methods

First, in order to exclude events where the disease process may have already been advanced at the time of movement behaviour measurement, the initial 2 years of follow-up (and so any events in this period) were excluded. By excluding early follow-up, this analysis aimed to exclude events where reverse causality is most likely to have played an important role.

Second, to examine the role of follow-up period in more detail, associations in different follow-up periods were then assessed quantitatively. The follow-up period was divided into 0-2 years follow-up, 2-4 years follow-up, and 4+ years follow-up. Interaction terms of the isometric log-ratio coordinates (i.e. the movement behaviour composition) with each follow-up period were included in the model. The final model used age as the timescale, and was adjusted for time on study. In detail, the dataset was prepared using Lexis expansion, whereby survival data is expanded along different timescales (in this case, time-on-study and age) [259, 260]. Participants' follow-up was then split at 2 and 4 years follow-up on the time-on-study timescale. A Cox model was then produced with age as the timescale and follow-up period (0-2, 2-4, 4+ years follow-up) treated as a time-varying covariate [261]. The interaction of follow-up period with the movement behaviour composition could then be investigated [262].

Third, to assess associations seen in groups with different health status, separate analyses were conducted in subgroups by (cardiovascular) health status. Statistical

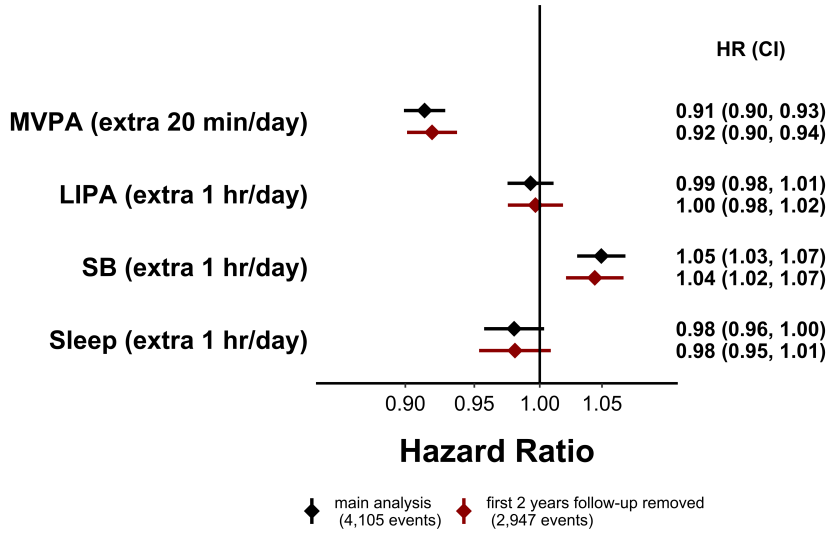
evidence for interaction was assessed using interaction terms between the ilr coordinates and (cardiovascular) health status in the main model. The group with poor (cardiovascular) health was defined as: those who self-reported poor overall health at baseline assessment; those who self-reported medications for diabetes, cholesterol or blood pressure at baseline assessment; or those who had a prior hospital admission for any condition of the circulatory system (I00–I99 as a primary diagnosis, e.g. admission for heart failure or aortic aneurysm). The group without poor (cardiovascular) health was the group not meeting these conditions. Supplementary Tables A.1 and A.2 describe variable preparation.

Analyses were performed in R 4.2.1 [200]. Code can be found at `github.com/OxWearables/thesis_additional_walmsley`.

6.4.3 Results

Removing the first 2 years of follow-up had a minimal impact on associations (Figure 6.5).

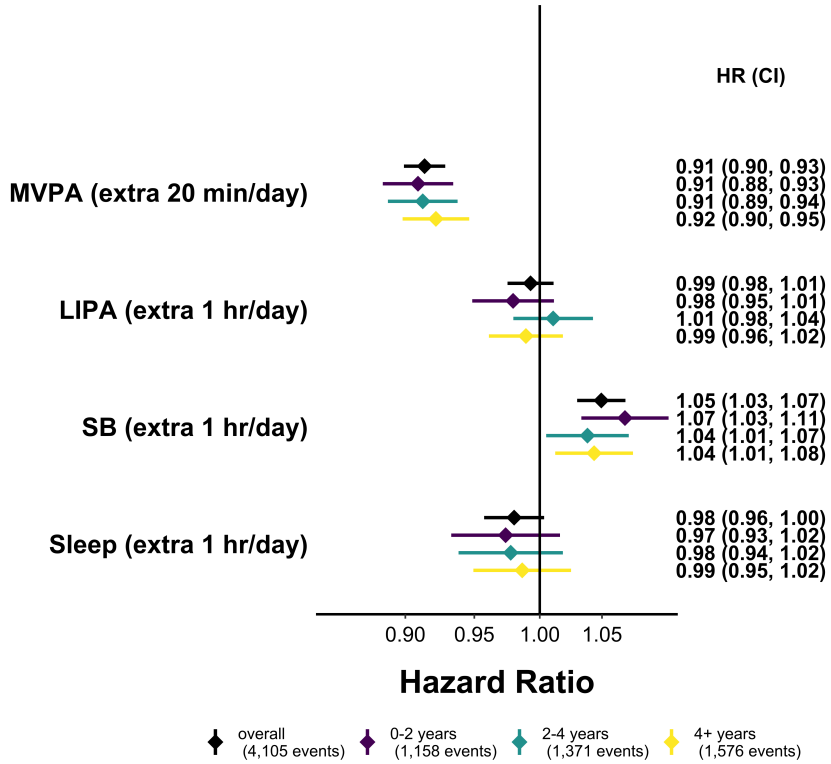
Figure 6.5: Hazard Ratios for incident cardiovascular disease associated with reallocating time to a given movement behaviour, from all other behaviours proportionally, before and after removing the first two years of follow-up (4,105 events among 87,498 participants and 2,947 events among 86,011 participants respectively).



All relative to the mean behaviour composition in the main analysis cohort: 8.8 hours/day sleep, 9.3 hours/day sedentary behaviour, 5.6 hours/day light physical activity behaviours, 0.35 hours/day (21 min/day) moderate-to-vigorous physical activity behaviours. More time in named behaviour reallocated from all other behaviours proportionally. Models used age as the timescale, were stratified by sex and were additionally adjusted for ethnicity, deprivation, education, smoking status, alcohol consumption, fresh fruit and vegetable consumption, red and processed meat consumption, and oily fish consumption.

1,158 events occurred in the first 2 years of follow-up, 1,371 events between 2 and 4 years of follow-up and 1,576 events beyond 4 years of follow-up. The HR by follow-up period for specific reallocations is shown in Figure 6.6. Visually, trends appear to be weak or absent. There was little statistical evidence for an interaction between time on study (0-2 years, 2-4 years, 4+ years) and the movement behaviour composition, either overall for the movement behaviour composition or considering the Wald p-value for the interaction between the first pivot coordinate for each behaviour and follow-up period.

Figure 6.6: Hazard Ratios for incident cardiovascular disease associated with reallocating time to a given movement behaviour, from all other behaviours proportionally, among 87,498 participants by period of follow-up (1,158 events in first 2 years of follow-up, 1,371 events between 2 and 4 years of follow-up, and 1,576 events beyond 4 years of follow-up).

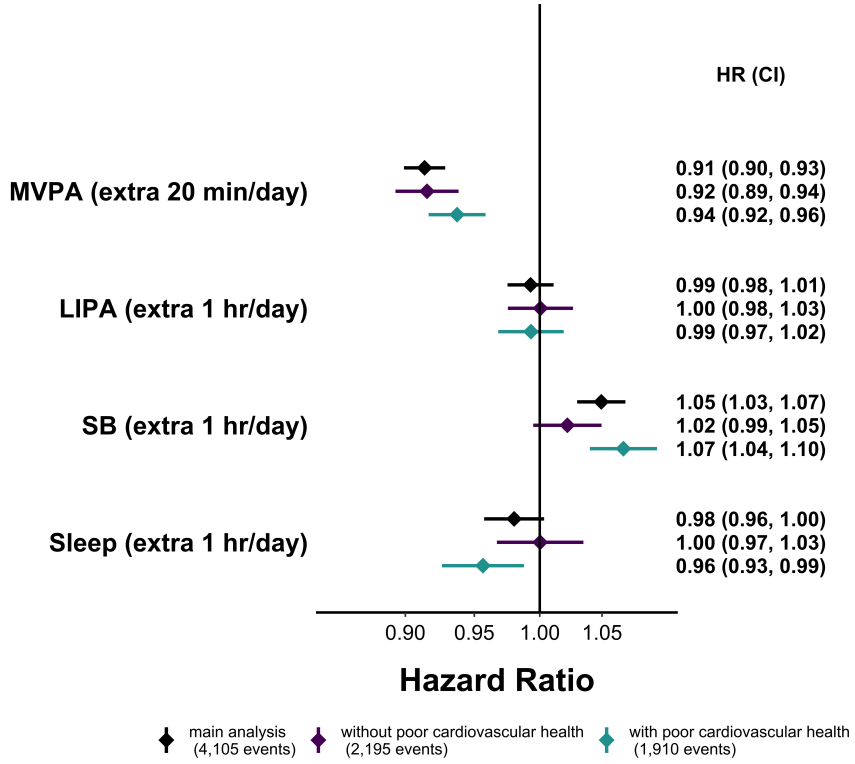


All relative to the mean behaviour composition (8.8 hours/day sleep, 9.3 hours/day sedentary behaviour, 5.6 hours/day light physical activity behaviours, 0.35 hours/day (21 min/day) moderate-to-vigorous physical activity behaviours) and more time in named behaviour reallocated from all other behaviours proportionally. Model used age as the timescale, was stratified by sex and was additionally adjusted for time period, ethnicity, deprivation, education, smoking status, alcohol consumption, fresh fruit and vegetable consumption, red and processed meat consumption, and oily fish consumption.

Among 23,450 participants identified as having poor (cardiovascular) health, either due to self-reported poor health, medication, or previous cardiovascular hospitalisation, there were 1,910 incident disease events. Among these participants, associations were fairly similar to those observed in the overall cohort (Figure 6.7). Among the 64,048 participants not identified as having poor (cardiovascular) health, there were 2,195 events. Associations for reallocating time into MVPA remained similar, but associations for reallocating time from sedentary behaviour to other behaviours were

substantially attenuated (Figure 6.7). There was not statistical evidence for interaction between (cardiovascular) health status and the overall movement behaviour composition (p-value for Likelihood Ratio Test 0.24), although interaction terms between health status and the first pivot coordinate for sleep and sedentary behaviour indicated some modest evidence (Wald p-values for interaction terms 0.08 and 0.06 respectively).

Figure 6.7: Hazard Ratios for incident cardiovascular disease associated with reallocating time to a given movement behaviour, from all other behaviours proportionally, among 64,048 participants without poor cardiovascular health (2,195 events) and 23,450 participants with poor cardiovascular health (1,910 events).



All relative to the mean behaviour composition in the main analysis cohort: 8.8 hours/day sleep, 9.3 hours/day sedentary behaviour, 5.6 hours/day light physical activity behaviours, 0.35 hours/day (21 min/day) moderate-to-vigorous physical activity behaviours. More time in named behaviour reallocated from all other behaviours proportionally. Models used age as the timescale, were stratified by sex and were additionally adjusted for ethnicity, deprivation, education, smoking status, alcohol consumption, fresh fruit and vegetable consumption, red and processed meat consumption, and oily fish consumption. Poor (cardiovascular) health was defined as self-reported poor overall health at baseline assessment; self-reported use of medications for diabetes, cholesterol or blood pressure at baseline assessment; or a prior hospital admission for any condition of the circulatory system (I00–I99 as a primary diagnosis).

6.5 Residual and unmeasured confounding

6.5.1 Background

Analyses in this thesis were adjusted for potential confounders, including sociodemographic and behavioural factors. However, residual and unmeasured confounding may bias results. Some potential confounders are unmeasured e.g. employment at the time of accelerometer wear. Some potential confounders, notably dietary variables, are measured poorly e.g. in Chapter 3, measurement of fruit and vegetable consumption was shown to be unstable.

Methods are available to quantify possible impacts of residual and unmeasured confounding. Negative control outcomes, where the exposure is associated with an outcome with a similar confounding structure but no hypothesised association, can be used to probe residual confounding (among other uses). For example, Hamer *et al.* [263] used non-activity-related accidental death as a negative control for CVD mortality in the Health Survey for England and Scottish Health Survey. Physical activity was associated with lower risk of non-activity-related accidental death although to a lesser extent than for CVD mortality, suggesting residual confounding may exaggerate associations between physical activity and CVD mortality [263]. Van Der Weele and colleagues recently introduced the E-Value to quantify the extent of residual and unmeasured confounding required to explain away a particular association [264, 265].

6.5.2 Methods

A negative control analysis was performed using a negative control outcome of accidents without a plausible mechanistic link to movement behaviours (accidents exclud-

ing falls [266], cycling accidents and intentional self-harm; see Supplementary Table A.2).

E-values were used to assess the minimum strength of an unmeasured confounder which would explain away the observed association. The E-value for the estimate quantifies the minimum strength of association that an unmeasured confounder would need with both exposure and outcome to explain away the observed association. The E-value for the 95% confidence interval quantifies the minimum strength of association an unmeasured confounder would need with both exposure and outcome to reduce the interval to overlap the null [264, 267]. As the exposure is continuous, here the risk ratio would apply to hypothetical groups with either the specified behaviour composition or the reference (the mean behaviour composition) [267].

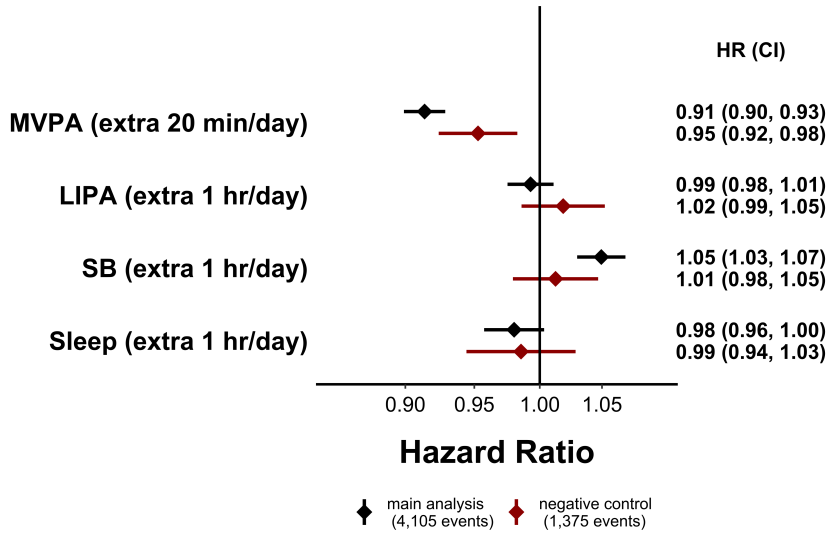
Analyses were performed in R 4.2.1 [200]. Code can be found at `github.com/OxWearables/thesis_additional_walmsley`.

6.5.3 Results

Most associations using the negative control outcome were near null (Figure 6.8). However, reallocations to moderate-to-vigorous physical activity behaviours were associated with the negative control outcome, although more modestly than with incident cardiovascular disease, suggesting residual confounding affects but does not entirely explain these results. Alternatively, this association may also be causal e.g. if low levels of moderate-to-vigorous physical activity result in lower muscular strength, raising the risk of presenting to hospital with an accident-related injury [263].

The E-values suggested that a substantial degree of unmeasured confounding would be required to reduce the observed associations to the null for MVPA and sedentary behaviour reallocated from other behaviours proportionally (Figure 6.9). For exam-

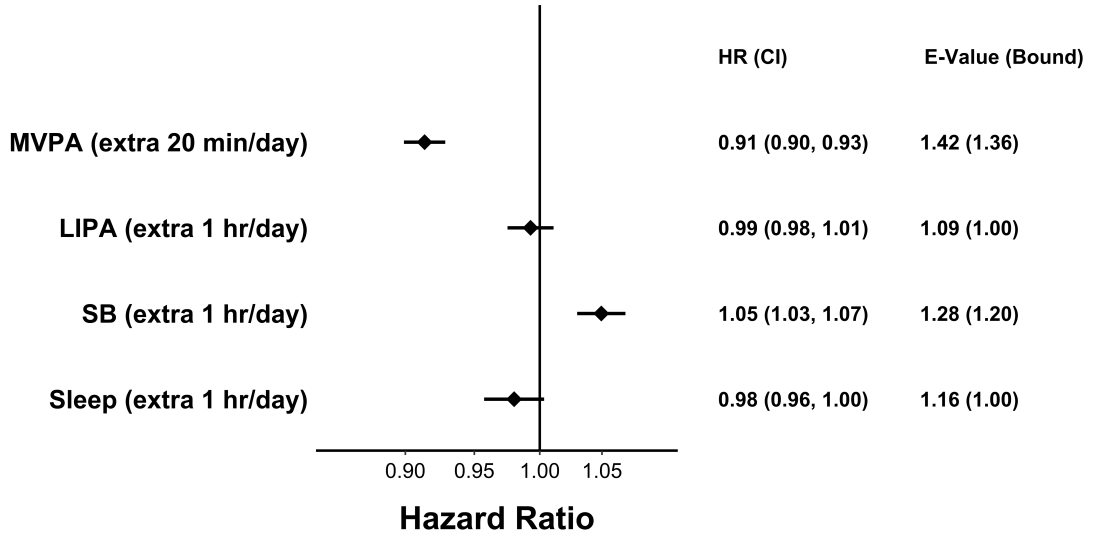
Figure 6.8: Hazard Ratios for incident cardiovascular disease and for non-activity related accidents associated with reallocating time to a given movement behaviour, from all other behaviours proportionally, among 87,498 participants (4,105 events) and 84,552 participants (1,375 events) respectively.



The analysis for the negative control outcome had a smaller number of participants as participants with prior accident were additionally excluded. All relative to the mean behaviour composition in the main analysis cohort (8.8 hours/day sleep, 9.3 hours/day sedentary behaviour, 5.6 hours/day light physical activity behaviours, 0.35 hours/day (21 min/day) moderate-to-vigorous physical activity behaviours) and more time in named behaviour reallocated from all other behaviours proportionally. Models used age as the timescale, were stratified by sex and were additionally adjusted for ethnicity, deprivation, education, smoking status, alcohol consumption, fresh fruit and vegetable consumption, red and processed meat consumption, and oily fish consumption.

ple, the E-value of 1.42 (for reallocating 20 minutes/day to MVPA, from all other behaviours proportionally) shows an unmeasured confounder would need to be associated with at least a 1.42-fold increase in risk for both exposure and outcome to explain away the observed association. For comparison, for the outcome this is similar in magnitude to the association found for current smoking versus never smoking in this data (HR 1.47 (1.31, 1.65)).

Figure 6.9: Hazard Ratios and corresponding E-values for incident cardiovascular disease associated with reallocating time to a given movement behaviour, from all other behaviours proportionally, among 87,498 participants (4,105 events).



All relative to the mean behaviour composition (8.8 hours/day sleep, 9.3 hours/day sedentary behaviour, 5.6 hours/day light physical activity behaviours, 0.35 hours/day (21 min/day) moderate-to-vigorous physical activity behaviours) and more time in named behaviour reallocated from all other behaviours proportionally. Model used age as the timescale, was stratified by sex and was additionally adjusted for ethnicity, deprivation, education, smoking status, alcohol consumption, fresh fruit and vegetable consumption, red and processed meat consumption, and oily fish consumption.

6.6 Discussion

6.6.1 Overall findings

Overall, this work found that analytic results from Chapter 5 were fairly robust to potential biases, although some biases had a meaningful impact on associations. Several sources of bias were examined, including information bias, selection bias, and alternative causal mechanisms (reverse causality, residual confounding), using various quantitative bias analysis techniques. Biases acted in different directions. Table 6.4 provides a brief summary of results.

Table 6.4: Summary of results of quantitative analyses of bias.

Bias	Analysis	Findings
Information bias due to within-person variation	Multivariable regression calibration parametrised using repeated measurements	Modest de-attenuation of some results; little change in others
Selection bias	Inverse Probability of Selection Weighting (full UK Biobank to accelerometry cohort only)	Little change
Reverse causality	Excluding early follow-up	Little change
"	Associations by follow-up period	Little change
"	Splitting cohort by (cardiovascular) health status	MVPA results similar; SB results differed by (cardiovascular) health status
Residual confounding	Negative control outcome: accidents	Reallocation of other behaviours into MVPA modestly associated with negative control outcome; Other associations null
"	E-values	Substantial residual confounding needed to nullify associations for reallocations from other behaviours into MVPA and into SB

Device-measured movement behaviour variables show relatively limited within-person variation over an extended (3-5 year) period. Within-person variation was quite consistent between different movement behaviour metrics (0.59-0.71). The simplest metric, an overall acceleration-based summary metric, showed the least within-person

variation, although sedentary behaviour was comparable. Moderate-to-vigorous physical activity behaviours showed some of the highest levels of within-person variation among the metrics studied. Assuming appropriateness of the model, the impact of correction for within-person variation on analytic results was generally modest. The most substantial impact was seen for moderate-to-vigorous physical activity behaviours, where accounting for within-person variation de-attenuated the association by a factor of approximately 1.5 on the log-HR scale. While clinically meaningful, this remains more modest than is seen for self-reported physical activity [268]. However, although this is broadly reassuring, it should also be noted that this represents only one component of overall measurement error, and the way it is modelled imposes assumptions on its form. Results on selection of the accelerometer sub-cohort within the UK Biobank study showed that selection at this level likely had little impact on observed associations. This was in spite of evidence of substantial selection between the UK Biobank study and the accelerometer sub-study (as has been documented previously [244]). Indications on the impact of reverse causality were mixed. In particular, excluding early follow-up and analysing events by length of follow-up suggested little impact of reverse causality. However, this analysis was limited by the overall length of follow-up. In a subgroup without poor (cardiovascular) health, associations involving moderate-to-vigorous physical activity behaviours were similar to in the full cohort, but associations for reallocating time from sedentary behaviour to other behaviours showed some attenuation (this may be compared with results in younger and older adults in Chapter 5). This attenuation may suggest reverse causality, although an alternative explanation may be that associations in the overall cohort are driven by effects in those with poor health. Findings suggested residual and unmeasured confounding may affect results, but that their impact was likely modest. Specifically, the negative control analysis found an association between reallocation of time from other behaviours to moderate-to-vigorous physical activity behaviours

and the negative control outcome.

6.6.2 Comparison with previous literature

These results can be compared to previous studies. For within-person variation, an important caveat to any comparison is that earlier studies typically used hip-worn devices and focussed on short-term variation (week-to-week/within-year). Perhaps the most comparable previous work in terms of time period of variation is a study among 209 older women in the US (embedded in the Women's Health Study), which considered within-person variation in several hip-worn accelerometer metrics over a 2-3 year period [234]. They found λ values between 0.67 and 0.82 for different metrics. Differences with the present study may be due to the population, including the restriction to women, and particular properties of the hip-worn accelerometer metrics. In a within-year study of 914 adults wearing both hip- and thigh-worn devices in two wear periods 6-months apart, Saint-Maurice *et al.* [235] found λ values between 0.54 and 0.62 for different metrics. This appears somewhat lower than the present study, but again may be impacted by the population and particular measurement properties. Moreover, using two diametrically opposed points in the annual cycle may artificially inflate within-person variation (by disproportionately selecting very different points). In a small study (n = 87), Aadland & Ylvisåker [236] studied within-individual variation in physical activity and sedentary behaviour between consecutive weeks. They found crude λ values of between 0.52 and 0.84 for different accelerometer-based measurements, or between 0.70 and 0.80 after adjustment for wear time [236]. However, they noted that because weeks were consecutive, longer term sources of within-person variation, such as seasonality, may have been neglected [236]. Jaeschke *et al.* [237] studied within-individual variation in various measurements from a hip-worn Actigraph GT3X+ between consecutive weeks. As with the previous study, this study was

small (n = 50), and considered consecutive weeks. They found λ values of between 0.68 and 0.82 for different accelerometer-based measurements [237]. The results of the present work therefore complement existing literature by characterising within-person variation of wrist-worn accelerometer measurements over a longer (2-6 year) period. This timescale is more appropriate for assessing the impact of within-person variation on results of Chapter 5 (with $>$ 6 years follow-up) [269]. In addition, this study extended previous studies by not only characterising within-person variation in device-based measurements of movement behaviours but also quantitatively assessing its impact on epidemiological results. Previous studies have estimated the impact of within-person variation on analytic results in the context of self-reported physical activity [268, 270], where its impact was found to be much more substantial (e.g. Mahmood *et al.* [268] found an approximately fivefold de-attenuation of associations on the log-HR scale; Lim *et al.* [270] a nearly sevenfold de-attenuation).

The modest impact of selection into the accelerometer sub-cohort on exposure-outcome associations is consistent with, and complementary to, the emerging literature on the analytic impact of selection in UK Biobank. Batty *et al.* [248] compared associations of risk factors with cardiovascular mortality in UK Biobank with more representative samples (Health Surveys for England and Scottish Health Surveys), and found that associations in UK Biobank generalised well. Using a post-stratification approach to weight UK Biobank to the distribution of participant characteristics in the Health Survey for England, Stamatakis *et al.* [250] found most risk factors, including self-reported physical activity, showed consistent associations with and without representativeness correction (although there were exceptions, notably alcohol drinking as a risk factor for CVD mortality). However, this previous literature considered selection effects between the general population and UK Biobank; the present work complements this by considering selection effects between UK Biobank and its accelerometer sub-study.

The results on reverse causality are consistent with studies in the literature review (Chapter 2). Most studies which addressed reverse causality by excluding early follow-up and/or by adjustment or exclusion based on disease status found results consistent with their main analyses, although sometimes with some attenuation and with low event numbers and lack of power a persistent issue. More work is needed to determine how long an excluded early period is needed to compelling address reverse causality in the context of cardiovascular disease: the appropriate length of time is outcome-specific, and most empirical evidence with device-measured movement behaviours is in the context of mortality [256, 257]. However, these results are also consistent with findings on self-reported physical activity and incident cardiovascular disease, where associations have been found to be relatively robust to reverse causality and persist even at long (> 10 years) follow-up [271]. Differences in associations by health status should be investigated more thoroughly in the future, with pre-specified subgroup analyses and more thorough characterisation of measurement properties by subgroup.

There is little previous work quantitatively addressing residual and unmeasured confounding. A previous analysis of physical activity and accidental death as a negative control outcome found associations between self-reported physical activity and the negative control outcome that were in the same direction but more modest in magnitude than associations between physical activity and cardiovascular disease mortality [263], a finding similar to this work.

6.7 Summary

Analytic results from Chapter 5 were found to be fairly robust to a range of potential biases. There was some evidence that reverse causality and residual confounding may partially explain some associations, as well as some evidence that within-person

variation may mean some observed associations are under-estimated, particularly for reallocation of time into moderate-to-vigorous physical activity behaviours. The next and final chapter, Chapter 7, will discuss overall conclusions from this thesis, in light of the biases discussed in this chapter. Strengths, limitations, and implications of these results for future work will also be discussed.

Chapter 7: Discussion

7.1 Introduction

This thesis aimed to investigate the effects of device-measured 24-hour movement behaviours on risk of incident cardiovascular disease. To address this aim, it had three specific objectives:

1. Refine and apply state-of-the-art machine-learning-based behaviour classification methods to derive measurements of movement behaviours in large wrist-worn accelerometer datasets.

2. Study the association between movement behaviours and risk of incident cardiovascular disease, using methods appropriate to the compositional nature of movement behaviour data.

3. Quantitatively assess the impact of known sources of bias on estimates of the association between movement behaviours and risk of incident cardiovascular disease.

7.2 Findings

First, existing literature addressing device-measured movement behaviours and risk of incident cardiovascular disease and cardiovascular disease mortality was reviewed (Chapter 2). Six cohorts were identified in which associations between device-measured movement behaviours and incident CVD had been studied, and four cohorts were identified in which associations between device-measured movement behaviours and CVD mortality had been studied. Studies consistently found associations between higher levels of moderate-to-vigorous physical activity and lower risk of incident cardiovascular disease or CVD mortality, which persisted after adjustment for other movement behaviours where this was considered. Some, though not all, studies found associations between higher levels of light physical activity and lower risk of incident cardiovascular disease or CVD mortality; results on adjusting for time in other movement behaviours were mixed and in many cases entirely attenuated. Studies found higher levels of sedentary behaviour were associated with higher risk of incident cardiovascular disease or CVD mortality, but again results were mixed and in many cases entirely attenuated where studies considered adjusting for time in other movement behaviours. Sleep was not considered in previous epidemiological studies using device-based measurement of movement behaviours. Several gaps in the existing evidence were noted. Studies with device-based measurement were usually small, with most studies having only a few hundred events. Sleep was rarely measured, and measurement of other behaviours was limited by the use of threshold-sensitive cut-point-based methods to classify behaviours. No studies analysed 24-hour movement behaviours together, and studies differed in their approach to adjustment for other behaviours, making it hard to compare results. Many potential sources of bias were noted, but quantitative assessment of these biases was rarer.

The first aim of the thesis was to refine and apply state-of-the-art machine-learning-based behaviour classification methods to derive measurements of movement behaviours in large wrist-worn accelerometer datasets. Chapter 4 developed a machine-learning model to classify sleep, sedentary behaviour, light physical activity behaviours, and moderate-to-vigorous physical activity behaviours in free-living wrist-worn accelerometer data. The classifier used a Random Forest with Hidden Markov Model smoothing, and it was developed using data from the CAPTURE-24 study, a free-living study with wrist-worn accelerometer data and a camera- and diary-based ground-truth. In Leave-One-Participant-Out Cross-Validation, the machine-learning method was shown to achieve very good performance at classifying movement behaviours in free-living data (mean Cohen's kappa of 0.80 (0.79, 0.82)). This machine-learning-based classification also appeared to perform better than traditional cut-point-based methods, and showed good face validity when applied to data from UK Biobank participants. This work is consistent with previous studies showing good performance of machine-learning-based behaviour classification in accelerometer data [108, 113], and it builds on this earlier work by using free-living data to develop the models and by classifying a highly relevant set of behaviours for subsequent applications. Both the classifier and the methods are publicly available for reuse. Good performance of the classifier meant derived phenotypes were suitable for use in epidemiological analyses in subsequent chapters.

The second aim of the thesis was to study the association between movement behaviours and risk of incident cardiovascular disease, using methods appropriate to the compositional nature of movement behaviour data. Chapter 5 therefore investigated how 24-hour movement behaviours were associated with risk of incident cardiovascular disease in a large prospective cohort over more than six years of follow-up. To appropriately account for the constrained and structured nature of 24-hour movement behaviour data, a Compositional Data Analysis approach was used. Reallocating

time from other behaviours into moderate-to-vigorous physical activity behaviours was associated quite steeply with a lower risk of incident cardiovascular disease. Reallocating time from sedentary behaviour to other behaviours was also associated with a lower risk of incident cardiovascular disease. For example, compared to the average movement behaviour composition in this cohort, reallocating 20 minutes/-day to MVPA from all other behaviours proportionally was associated with a Hazard Ratio of 0.91 (0.90, 0.93), while reallocating 1 hour/day to sedentary behaviour from all other behaviours proportionally was associated with a HR of 1.05 (1.03, 1.07). These results are broadly consistent with existing literature [22]. However, they also address uncertainties in current literature [272]. By conceptualising and analysing movement behaviour data as compositional data, this work was able to appropriately study 24-hour movement behaviours together [140]. It was also able to incorporate behaviours (e.g. light physical activity, sleep) on which evidence remains relatively scarce [22, 65], and the large number of events supports inference.

The third aim of the thesis was to quantitatively assess the impact of known sources of bias on associations between movement behaviours and risk of incident cardiovascular disease. Chapter 6 investigated the extent to which associations between 24-hour movement behaviours and incident cardiovascular disease may be affected by information bias, selection bias, reverse causality, and residual confounding. Within-person variation in exposure measurements was found to attenuate some associations, predominantly those involving moderate-to-vigorous physical activity behaviours, and have little impact on other associations. Although there was clear evidence of selection from the overall UK Biobank into the accelerometer sub-cohort, the impact of selection bias on analytic results was likely modest. There was some evidence that both reverse causality and residual confounding may partially explain some associations, although in both cases results were mixed and suggested that their overall impact was likely to be modest.

7.3 Strengths and limitations

7.3.1 Measurement of 24-hour movement behaviours

Overall, good measurement of 24-hour movement behaviours was a key strength of this work. By measuring movement behaviours using wearable devices, rather than self-report, scope for recall and reporting bias was minimised. Moreover, UK Biobank used a low-burden, 24-hour wear protocol with a wrist-worn device (in contrast to earlier studies, where devices were often hip-worn and removed for sleep) [93]. This protocol could be deployed at large scale and with good wear-time compliance [89, 93], and allowed measurement and investigation of the full range of 24-hour movement behaviours. The machine-learning behaviour classification model developed in Chapter 4 supported precise, accurate classification of movement behaviours. As described above, when examined in Leave-One-Participant-Out Cross-Validation, it achieved very good classification performance in free-living data, and when applied to UK Biobank data, it showed good face validity. The favourable comparison, both in quantitative performance metrics and in face validity, between the machine-learning-based classifier and use of a cut-point (the traditional approach to behaviour classification) further demonstrates its potential usefulness and applicability.

Strengths of the final classifier, and hence of the measurement of 24-hour movement behaviours in this work, arise from a classifier development study with many important design features.

A major strength is the use of free-living data, which represents the variety and complexity of real-world movement, to develop the model. In contrast to classifiers developed using laboratory-based or semi-structured data, which encounter unseen behaviours when deployed (e.g. car travel, walking on uneven ground), this model

can learn patterns which generalise well to the contexts where it is deployed [112, 273]. Use of free-living data also facilitates incorporating information on temporal sequence into modelling, improving performance [113]. As well as benefits for model training, using free-living data has benefits for understanding model performance: assessing performance in free-living data likely gives a more accurate assessment of real-world performance than assessing performance in laboratory data only [112].

Another key strength is that the model was developed with the specific intention of being suitable for measuring 24-hour movement behaviours in order to study their associations with incident cardiovascular disease. Developing the model with this intended application had several benefits. The set of behaviours the model was developed to classify (sleep, sedentary behaviour, light physical activity behaviours, moderate-to-vigorous physical activity behaviours) are highly suitable for studying associations of 24-hour movement behaviours with disease risk: they are comprehensive and mutually exclusive, meaning any measured time period corresponds to one of these behaviours; they are known to be highly relevant to population health; they are the most commonly studied behavioural constructs in existing studies (see Chapter 2); and findings about these behaviours can be translated into interventions and guidelines [22, 141]. The additional work to characterise multiple aspects of model validity (a range of metrics, performance by subgroup, and checks of face validity in UK Biobank data) was also tailored to the intended application: while developers of machine-learning models often focus on single summary metrics of performance, a more broad-based evaluation of validity is needed to understand the extent to which measurements are suitable for particular applications [274]. In the context of movement behaviour epidemiology, measurement validity is notably complex and multifaceted, and required measurement properties can vary substantially [186].

A final strength is the reproducible, open-source approach taken to methods develop-

ment and methods sharing. The classifier and phenotypes developed here are available for reuse, supporting future studies and helping to ensure comparability across the literature [115]. The classifier can be accessed through the 'biobankAccelerometer-Analysis' tool [4], and derived movement behaviour variables based on this work have been returned to the UK Biobank resource. In addition to the availability of the model for reuse, the methods used to develop the model are open-source and the CAPTURE-24 dataset (accelerometer data and annotations) is openly available (see Appendix B) [275]. Availability of these methods and data will allow other researchers to apply, compare, criticise, and improve upon them.

However, movement behaviour measurement in this work had limitations. Accelerometer methods have general limitations, notably that they record movement of the attachment site and not cardiovascular effects of movement [156]. For example, climbing a flight of stairs may be negligible for an elite athlete, but involve substantial effort for an infirm person [156]. Population-specific classifiers or combination of multiple sensor modalities (e.g. heart rate sensor alongside an accelerometer) may help to address this in the future [156].

Development of the machine-learning model also had some limitations. Machine-learning methods are diverse, and open questions remain about optimal approaches for a given problem. This work focussed on further developing existing approaches to derive measurements suitable for use in epidemiological work, and the optimal modelling approach was largely beyond its scope. However, Random Forests are sometimes described as a 'traditional' machine-learning method, having been used widely for over 20 years [195]. In some domains, the 'deep-learning revolution' of the last decade has transformed performance [276], and different model structures, including deep networks, may improve behaviour classification performance. It is likely that performance gains will be possible, though modest: comprehensive benchmarking

of classification performance in CAPTURE-24 using different methods has recently been conducted, and while some improvements on the RF-HMM combination were seen, these were modest. Aspects of the modelling process may be refined in ways going beyond model architecture. In particular, methods to maximise performance with limited labelled data may add value: for example, self-supervised learning can make use of unlabelled data, of which there is a much larger quantity, to support modelling [197], and transfer learning can use learning from other problems in this problem [277]. Another key limitation of the classifier development is the fact that the model could not be tested in an external dataset from an independent study (due to a lack of studies like the CAPTURE-24 study). While the Leave-One-Participant-Out methodology ensures models are always tested on data on which they were not trained, machine-learning models are known to be vulnerable to drops in performance when the distribution of input data is different to the dataset in which they were trained [278]. This means they may perform less well when applied in new datasets or contexts; evaluation in an independent study as test set may help to quantify this performance drop and establish less biased estimates of performance in entirely unseen data. However, in the absence of an independent study, the face validity of the models when applied to data from UK Biobank is reassuring.

The data used for classifier development may also be refined further in the future. While indirect observation is a highly valid way of obtaining behavioural ground-truth labels in free-living, it has some flaws [273]. 'Ground-truth' labels for sleep came from a diary, which is not a 'gold-standard' sleep measurement. Ideally, sleep measurements should be further tested, developed, and refined using a true gold-standard for physiological sleep, such as polysomnography [279, 280]. The 'ground-truth' for other behaviours may also be refined further in the future. For example, when classifying behaviours based on images, it is easy for misclassifications at the light physical activity/moderate-to-vigorous physical activity borderline to occur: the

image annotator must infer activity intensity ($<$ or $\geq$ 3 METs) from images, and some behaviours occur at intensities both above and below this threshold (e.g. some household and garden tasks [281, 282]). However, while a less subjective ground-truth mechanism may seem desirable, there is a general trade-off between quality of the ground-truth labels and the range and volume of behaviours represented. For example, in a controlled laboratory setting behavioural intensity can be measured more accurately than in free-living (e.g. via indirect calorimetry [283]), but diversity of behaviours is more restricted; studies using video may have more accurate labels than studies using still-image cameras, but typically have shorter measurement periods [284]. For classifier development, the primary concern is that a wide range of realistic behaviours are represented; moderate noise in labels is secondary, especially as machine-learning models often cope well with noisy input data. The target phenotypes for classification may also be refined. For example, it may be informative to separate dynamic light physical activity from standing, or restrict light physical activity to between 1.5 and 3 METs (here, it included behaviours at $\leq$ 1.5 METs not in a sitting, lying, or reclining posture). The definition of light physical activity has been shown to impact research findings [57], highlighting the need for further epidemiological work with a variety of light-physical-activity-related constructs. While consistency and robustness of results across the age and sex subgroups studied is reassuring with regard to model robustness, the extent to which this could be investigated was limited by a lack of detailed participant data in the CAPTURE-24 study. Future studies may be able to more thoroughly study the extent to which other characteristics, such as body shape parameters (e.g. arm length), handedness, and other participant characteristics (e.g. frailty), affect performance. Despite efforts to recruit a diverse sample in terms of age and workplace [187], the CAPTURE-24 dataset is skewed towards younger participants and may be unrepresentative in other ways (again, a lack of detailed participant data makes this difficult to assess

directly). In itself, this is not necessarily problematic: given the aim was to develop a classifier able to classify free-living behaviours of UK Biobank participants, *participants* need not be representative of UK Biobank participants provided their *behaviours* and associated movements are sufficiently representative of those seen in UK Biobank (this is in contrast to, for example, risk prediction based on an individual's accelerometer data, where representativeness of participants might be the primary concern). However, participant non-representativeness may make behavioural non-representativeness more likely, and future ground-truth studies may seek to recruit a sample which is more closely matched to the population where the classifier will be used (for instance, a random sub-sample may be recruited within a larger epidemiological study, similarly to how gold-standard biomarkers are sometimes collected on a sub-sample within dietary studies [229]).

7.3.2 Associating 24-hour movement behaviours with incident cardiovascular disease

A notable strength of the analysis associating movement behaviours with cardiovascular disease was that 24-hour movement behaviours were considered alongside each other, as constituent parts of the 24-hour day [122, 141]. Conceptualising 24-hour movement behaviours together within the 24-hour day allows clarity about the effects being estimated [132]; as noted in Chapter 2, lack of clarity and substantial variation between studies on effects being estimated is an important limitation of existing literature. Moreover, using a Compositional Data Analysis approach, an analytic approach well-suited to the compositional structure of 24-hour movement behaviour data, enabled 24-hour movement behaviours to be appropriately modelled together [122, 127]. This means all results are mutually adjusted for time in other behaviours, and questions about independence of effects of different behaviours are

subsumed into considering 24-hour movement behaviours as necessarily making up a whole. Accurate, device-based measurement of 24-hour movement behaviours was also crucial to analyses associating movement behaviours with incident CVD: better exposure measurement supports a 24-hour perspective and minimises information bias in analyses. The large size of the UK Biobank accelerometer sample and consequent large number of disease events was also important to precisely answer questions about associations with incident disease. By deriving the outcome passively from electronic health records and register data, active follow-up is not required, allowing a bigger sample and minimising issues of loss-to-follow-up [161, 165]. Again, as in the context of measurement, the fact that all methods are open-source and available for reuse is a strength. In particular, the development of key analytic functions into a shareable, reusable, and documented R package is a useful secondary output of this work, discussed at greater length in Appendix C.

The most important limitations of the analysis related to the potential for bias, including from measurement, selection, reverse causality, and residual confounding. As these biases potentially present a substantial threat to inference based on these findings, Chapter 6 addressed these biases in more detail.

As well as the possibility for bias, this analysis had some other limitations. The statistical approach (a Compositional Data Analysis Cox regression) has assumptions and limitations. Specifically, associations between the transformed variables (the log-ratios) and the log hazard ratio are assumed to be linear. Further work using approaches more flexible with respect to functional form, such as using compositional splines, may help to understand shapes of associations in more detail [285]. In particular, sleep has previously been suggested to have U-shaped associations with incident disease outcomes [286], which may be incorporated in modelling. Moreover, communication of the model may be refined. First, while described in terms of reallocation,

it should be emphasised that all results are derived statistically across participants. Also, the graphs produced depend on the movement behaviour composition which they are produced relative to (the central line on the graph) and on the particular time reallocation considered, which means they may give only a partial summary of what is happening overall. One way to address this may be through directly interpreting coefficients. Another way is through allowing users to interactively explore model results. This work has been pioneered by Dumuid *et al.* in a work currently under review, in which I have been involved in the course of this project.

The conceptualisation of 24-hour movement behaviours may also be refined. Pattern and quality may be as important as total time for some behaviours. For example, clinical sleep assessments consider sleep efficiency, Wake-After-Sleep-Onset, and sleep timing; the concept of sedentary breaks/bouts was central to laboratory research on sedentary behaviour [287]. Moreover, in the future, it may be possible to consider intensity of physical activity in a more continuous way (avoiding the arbitrary 3 MET LIPA-MVPA threshold), for example by using a functional data approach to the intensity spectrum [288, 289]. It may also be possible to consider different aspects of the exposure (e.g. considering how energy expenditure is divided between behaviours [175, 290]), or to focus on functional behaviours (e.g. walking). All have advantages and disadvantages, and may give insights complementary to those in this work. Nonetheless, the fundamental issue is that human movement behaviour is complex, and it is difficult to summarise it in a low-dimensional way that captures key health-relevant aspects, is measurable, and is interpretable for end users of research. To this extent, consideration of time in the simple set of behaviour classes used here has proved its worth through its wide use.

It may also be possible to further refine the disease event definition in this work. As discussed in Chapter 3, defining incident disease using electronic health record data

is complex. In particular, use of primary care data may identify more cases in the future. However, as noted in Chapter 3, hospital and death record diagnoses have high Positive Predictive Value, which is the key concern for validity of inference [178]. The large number of events may also make it possible to refine the subtypes of disease studied in the future. For example, the role of 24-hour movement behaviours may differ between different disease subtypes e.g. risk factors are known to vary between different subtypes of cerebrovascular disease [291].

7.3.3 Assessment of bias

The quantitative examination of several sources of bias, conducted in Chapter 6, is a strength of this work and informs interpretation of key results. Although many biases are known to affect the epidemiological literature on movement behaviours, assessing likely magnitude has historically been more challenging [224, 255, 292, 293]. The literature review in Chapter 2 found that quantitative assessment of systematic bias in device-measured movement behaviour research was rare.

The particular analyses conducted also had strengths. A key strength of work on within-person variation in this thesis was the medium-term data on several device-based metrics: quantitative assessment of within-person variation in device-based measurements of movement behaviours is rare, and this lack of evidence is especially the case for wrist-worn devices and in the medium term. The sample with repeated measurements data was large enough to estimate λ coefficients with good precision (standard error of estimates $\sim$ 0.01). Notably, this work extended previous work with device-based measurements by not only quantifying within-person variation, but also quantitatively investigating its possible impact on results. Using a Compositional Data Analysis framework made it possible to address multivariable within-person variation in device-measured 24-hour movement behaviours. This is im-

portant as the impact of multivariable within-person variation can be more difficult to assess than the impact of univariable within-person variation: while univariable within-person variation in the exposure often manifests as regression dilution bias, attenuating associations, multivariable within-person variation may act in different and hard-to-predict directions [294].

The analysis on the impact of selection was able to exploit a particularity of UK Biobank study design to quantitatively address selection. Usually, non-observation of non-participants makes it difficult to address selection quantitatively [295]; here, selection from the overall cohort into the accelerometry sub-cohort could be studied quantitatively as non-participants had been observed at an earlier date. The detailed phenotyping at the baseline assessment meant that a variety of variables could be used to study selection.

A strength of the work on reverse causality is its quantitative assessment using several different approaches. Follow-up was longer than in some of the large prior studies using device-based measurement reported in the literature review, allowing associations by time period to be better addressed [69, 117, 119].

Addressing residual and unmeasured confounding through the use of emerging methods is also a strength of this work. Residual and unmeasured confounding has rarely been addressed in the literature. It is also challenging to address quantitatively (as it is by definition unmeasured).

Overall, investigating several biases, rather than focussing on a single bias, was useful both for scientific inference from these results and for designing future studies: it makes it possible to begin to assess their overall impact and relative importance.

However, these quantitative bias analyses can only be interpreted as indicative, as they encode numerous assumptions. The measurement error model used in this work

considered only within-person variation, and assumed it could be well-modelled by regression calibration with a linear calibration equation for the isometric log-ratio coordinates. The use of regression calibration to account for within-person variation in compositional data remains little-studied, and the approach may be refined in the future. As data to parametrise the measurement error came from repeated measurements, and no gold-standard or reference measurement was available, the error structure could not be probed in detail. Unlike for overall energy expenditure [296], there is no obvious gold-standard for 24-hour movement behaviours, but camera data (over a more extended period than in Chapter 4), detailed time use diary data, different accelerometer placements, and different sensor modalities may be exploited in the future to better understand measurement error. Nonetheless, within-person variation is certainly a major component of measurement error and quantitative bias analysis remains valuable even when only partially capturing the phenomenon under study. Other issues were not considered in modelling within-person variation. For example, confounders also show within-person variation, which could be modelled. However, limited within-person variability of most confounders (shown in Chapter 3), measurement of the main exposure by a different approach (making correlated errors less likely), and modest impact of confounding mean this is likely a secondary concern. Measurement error parameters must be assumed to be generalisable from the sub-sample with repeated measurements to the accelerometer sample. Finally, when assessing the analytic impact of measurement error, a range of methods are available, making different assumptions [232]. SIMEX, standing for 'Simulation Extrapolation', is based on adding further measurement error to a variable, quantifying the effect of this additional error on estimates, and extrapolating back from this to likely results had the variable been measured without error [227, 228]. SIMEX is potentially more flexible than regression calibration in the range of measurement error structures and epidemiological models it can be applied to. Other approaches include methods based

on Maximum Likelihood Estimation (i.e. incorporating the measurement error model in the overall Maximum Likelihood Estimation process for estimating the parameters of interest); Bayesian approaches; moment reconstruction and moment-adjusted imputation; and treating measurement error as an imputation problem [232]. Using different methods with different assumptions may help to establish the results of this thesis more robustly.

The work on selection bias had several particular limitations relating to methodology. Bradley & Nichols [249] found previously in UK Biobank that weighting-based approaches may not fully account for selection (albeit in the more challenging context of prevalence estimates). The Inverse Probability of Selection Weighting approach to selection bias assumes that measured factors are sufficient to address selection [245, 295]. It cannot address selection driven by factors unmeasured and uncorrelated with included variables [245, 295]. Nonetheless, the near-null apparent impact of weighting for selection based on measured factors makes a very large impact of unmeasured factors seem less plausible. Future studies may use more sophisticated modelling techniques to estimate selection weights; in particular, as the weights themselves do not need to be interpretable, this may be a context where more flexible and data-driven methods (e.g. penalised logistic regression with non-linear terms and more predictors) may be fruitfully applied [249]. Clearly, the major limitation of the work on selection bias in this thesis is that it only addresses selection effects within UK Biobank, and not from the general population into the UK Biobank cohort [246] or from attrition during the study (although passive follow-up means attrition here is very modest) [297]. The lack of an impact of selection within UK Biobank is reassuring with regard to the impact of selection more broadly, as drivers of selection may be shared (e.g. 'healthy volunteer' effects have been documented both into UK Biobank and from UK Biobank into its sub-studies [180, 244]). Moreover, selection between UK Biobank and its sub-studies is important in its own right and

little-studied. Nonetheless, future work should address selection into the study as well, perhaps by using more representative samples (e.g. Health Survey for England, Census microdata, UK Household Longitudinal Study) to design appropriate weights for UK Biobank [250]. Analytically addressing selection effects in UK Biobank is an active area of research, with at least two preprints on weighting in UK Biobank published in the first half of 2022, raising the prospect that in the future validated sampling weights will be available to use in sensitivity analyses addressing selection in UK Biobank [249, 298].

In the work on reverse causality, although this study had longer follow-up than previous comparable studies, total length of follow-up (maximum < 8 years) may still be a limitation: reverse causality in the context of cardiovascular disease may be better addressed with longer follow-up. Different study designs may also enhance treatment of reverse causality. Studies with device-based measurements of movement behaviours at multiple timepoints could help to understand behavioural trajectories, which may be useful in studying reverse causality: for example, Sabia *et al.* [299] considered trajectories of physical activity prior to dementia diagnosis to explore the role of reverse causality in associations between physical activity and dementia.

Analyses addressing residual confounding presented challenges. In particular, finding a suitable negative control outcome for 24-hour movement behaviours and cardiovascular disease was difficult: a causal link between movement behaviours and hospitalisation with accident is not entirely implausible. Assessing results using a different negative control outcome would be useful, but it is difficult to find a good negative control outcome given plausible associations of movement behaviours with a very wide range of outcomes [42, 300]. E-values, though a useful tool, should not be overinterpreted: they have been criticised for relying on unrealistic assumptions [301, 302] and for excessively automating consideration of confounders [303, 304]. One useful

avenue to further explore confounding in the future will be addressing associations between movement behaviours and incident cardiovascular disease in cohorts in different settings [39], where the confounding structure may be different: for example, in lower-income settings, people with lower socio-economic status may have higher physical activity levels (e.g. due to active transport), whereas in higher-income settings higher socio-economic status associates with higher physical activity levels [305].

Consistent with the overarching aim to roughly map the possible landscape of bias, each bias probed quantitatively in this work was addressed in limited detail. The quantification of bias could also be formalised further; some analyses did not result in a single 'bias-adjusted' estimate, often due to a lack of appropriate data to directly parametrise the bias model (e.g. for residual confounding). One option in these cases is 'probabilistic bias analysis': a plausible distribution of bias parameters is used to simulate the range of possible impact of bias [306–308]. A related limitation is that results of individual analyses were not integrated into a single overall bias analysis ('multiple bias modelling') [309]. Not integrating results makes it more difficult to assess the relative importance of different results and their overall impact: when biases act in different directions it can be unclear which will dominate [309]. In the future, many of the biases here may be addressed further using different studies and study designs; this could include cohorts with longer follow-up, cohorts in different settings, genetic epidemiology designs, 'econometric' causal inference designs, and trials. Triangulating results in this way is more robust: studies with very different designs are likely to be affected by different major biases [310]. However, it remains rare in movement behaviour epidemiology, perhaps because of difficulties in designing and executing appropriate studies [293].

7.3.4 Code and data availability

The availability of code used to produce results throughout this thesis, shared open-source via GitHub, is a strength across the different aims (see Appendix B). Work for this thesis has also contributed to open-source software packages, including software for accelerometer data processing [4], software for pre-processing UK Biobank data [3], and an R package to facilitate epidemiological analysis using a Compositional Data Analysis approach to the exposure (Appendix C). Adoption in epidemiology and health data research of industry-standard good practice in software development and code sharing (such as use of version control, automated testing, and packaging code [311]) helps to enable development of computational and statistical methods, and to ensure research is reproducible [312].

7.4 Implications for population health

Results in this thesis support existing public health guidelines on movement behaviours. Global guidelines from the World Health Organisation currently advocate at least 150 to 300 minutes of moderate physical activity (or vigorous equivalent) per week for adults, while noting that for the least active any activity has benefits [22]. The guidelines' recommendations on sedentary behaviour and light physical activity are more generic ('limit the amount of time spent being sedentary', and 'replace [sedentary behaviour] with more physical activity of any intensity'), and they do not currently incorporate sleep [22]. Supporting this, this work found that reallocating time from other movement behaviours into moderate-to-vigorous physical activity behaviours was associated quite steeply with a lower risk of incident cardiovascular disease, and also found that reallocating time from sedentary behaviour into other

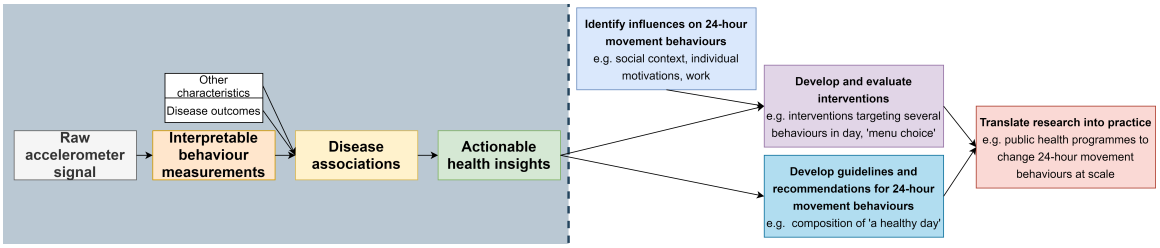
behaviours was associated with a lower risk of incident cardiovascular disease (albeit with a shallower gradient).

These results also provide evidence to support more detailed recommendations on 24-hour movement behaviours in the future. While 24-hour movement behaviours are widely acknowledged to have important potential as a behavioural target, evidence on their associations with incident disease has been relatively limited [30, 68, 75, 140, 204, 272, 313]. First, these results provide more evidence on sedentary behaviour and light physical activity: reallocating time from sedentary behaviour into other behaviours, notably including light physical activity behaviours, was associated with a lower risk of incident cardiovascular disease (and this was so while holding fixed time in other behaviours). For what may be the first time in the context of incident CVD, these results integrated sleep alongside other movement behaviours. While there are some difficulties with sleep measurement and analysis (see Chapters 4 and 5), this provides evidence towards future incorporation of sleep into guidelines. It also indicates, reassuringly, that accounting for sleep in analyses does not substantially alter conclusions about other behaviours. Analysing 24-hour movement behaviours together in an appropriate framework helps to fill evidence gaps on joint associations, and supports integration of evidence on individual behaviours into a coherent whole. However, open questions remain: modelling assumptions make it difficult to use these results to directly support particular quantitative thresholds of time in a given behaviour for benefit/risk, and so more flexible approaches which may better detect thresholds may be helpful in the future.

Beyond guidelines, a 24-hour approach to movement behaviours holds promise for interventions and policy [30, 66, 141]. It may broaden the range of contexts in which intervention is possible (e.g. intervention in workplaces where moderate-to-vigorous physical activity is not possible, but replacing sedentary behaviour with light physical

activity is feasible), and offer opportunities for collaborative intervention to prevent disease (e.g. collaboration between sleep and physical activity professionals) [314]. One possibility is to tailor interventions for an individual to focus on those aspects of the 24-hour day that are most tractable for them to change [66, 315]: for example, a relatively fit individual with a predominantly sedentary job may find incorporating a short period of moderate-to-vigorous physical activity the most convenient way to achieve healthier 24-hour movement behaviours [316], while a person with low fitness or lacking in confidence may find it easier to replace sedentary time with light physical activity [65]. However, development of 24-hour movement behaviour interventions is a new area of innovation, requiring an evidence base extending far beyond aetiology [314, 317]. This will include, for example, a better understanding of determinants of 24-hour movement behaviours, and of effective interventions to change 24-hour movement behaviours [317]. While research in this thesis, such as innovations in measurement of 24-hour movement behaviours, may support this, further methods and study designs are required. Figure 7.1 extends Figure 1.3 to consider how new evidence on 24-hour movement behaviours may eventually translate into changing population health, and the other kinds of evidence that will be required to achieve that goal.

Figure 7.1: Research and translation pipeline for 24-hour movement behaviour epidemiology.



7.5 Implications for future research

This work raises several broad considerations for future research.

As technological development in measurement (i.e. accelerometers deployable at scale) enabled analyses in this thesis, it is useful to consider how new technologies will further enable future studies. Devices to measure movement behaviours continue to evolve: the newest research-grade wearable devices are '9-axis' (triaxial gyroscope and triaxial magnetometer in addition to triaxial acceleration) [318], but at this early stage it remains unclear how much additional signal modalities will improve measurement. Device placement is another realm of innovation, with the thigh increasingly favoured as able to accurately capture postural information and achieve good 24-hour wear (a thigh-worn accelerometer is typically not removed during the wear period) [319, 320]. However, current protocols for collecting thigh-worn accelerometry data involve a researcher taping a device to a participant's thigh, which may limit scalability, increase cost, and potentially reduce acceptability to participants relative to wrist-worn accelerometry [321]. Use of multiple devices is another possibility, theoretically allowing advantages of several placements; however, behaviour classification studies have typically shown modest improvement from multi-device classification over single device classification [113, 322]. Other technological innovations may further broaden the range of tractable research questions. For example, devices worn for extended measurement periods may allow study of how movement behaviours vary over time (e.g. variations with weather; trajectories around a disease episode) [323]. However, the value of an extended measurement period for aetiological questions is less clear: results on measurement variation in Chapter 6 actually indicate relatively modest within-person variability for week-long measurements, and a longer measurement period would also raise difficulties (e.g. maintaining data quality). Wearable devices

deployed in future studies may also measure characteristics beyond movement that enhance aetiological research on movement behaviours. For example, heart rate monitors may indicate experienced movement intensity, while GPS devices may enable assessment of movement domain or context.

There are many opportunities for innovation in methods to go from raw device data to useful and usable measurements of 24-hour movement behaviours. Straightforwardly building on this work, one important question is the extent to which the classifier developed here will work in different studies, populations, and settings. Beyond external validation of the classifier, it will be important to understand if this classifier can be applied to studies using different brands of wrist-worn device (with the same underlying triaxial accelerometer technology); to populations differing from the present study in some characteristics (e.g. ethnicity, age); or to different settings (e.g. a low or middle income setting, perhaps with behaviours rarely seen in UK context, such as agricultural work).

There are also opportunities to innovate further in classifier development, in the hope of building classifiers that perform even better in the future. As described in Section 7.3, developments in machine-learning may be applied. An important domain of innovation is in the collection of free-living data suitable for use in classifier development. Collecting data with a source of ground-truth labels (e.g. image data) alongside accelerometry can be time-consuming, expensive, and complex. However, there are immediate opportunities to improve on current practice. Notably, several studies are already collecting this kind of data [184, 284, 324], but data is often not shareable or inter-operable, limiting its usefulness. Data sharing and development of common standards (e.g. labelling protocols) may help to get more use from existing data. As collecting ground-truth data is so resource-intensive, maximising value from individual data collections should also be a priority: data may be enhanced by col-

lecting more participant characteristics (including a full suite of commonly-studied variables, as well as more unusual variables with specific relevance e.g. handedness, balance tests) or more comprehensive movement behaviour 'ground-truth' data (e.g. polysomnography, other modalities like location, direct energy expenditure measurement). A planned data collection in the Oxford Biobank may achieve some of these aims [325].

As described in Section 7.3, another fruitful area for innovation may be in the phenotypes themselves which the methods aim to classify. Most straightforwardly, refinement of the set of behaviours may provide more information for subsequent epidemiological analyses (e.g. distinguishing static standing from dynamic light physical activity behaviours [57]). More generally, as noted above, there may be other ways to describe 24-hour movement behaviours which provide useful insights in subsequent research (e.g. a more granular summary, such as an activity or intensity 'spectrum') [128].

This work showed that a 24-hour approach provides a coherent framework for considering associations between movement behaviours and cardiovascular disease risk, and is feasible with appropriate measurement and analysis. While it is difficult to imagine future studies not routinely acknowledging and accounting for co-dependence of behaviours within the 24-hour day [66], this does not constrain future work to the particular analytic approaches used in this thesis. For example, as noted above, further insights may be achieved by relaxing assumptions in the Compositional Data Analysis model, perhaps by combining the compositional approach with approaches more flexible towards functional from e.g. isometric log-ratio coordinates may be used in polynomial or spline regression models. More generally, as was also noted above, there is an interplay between conceptualisation of the movement behaviour exposure and the analytic approach. Future work in movement behaviour epidemiology should

continue to work not only to show that movement behaviours are associated with cardiovascular disease risk, but also to identify *which aspects* of movement behaviours are most important: for example, analyses able to model pattern or treat intensity in a continuous manner may enhance future research.

Future work may also build on this work by investigating different disease outcomes; notably, analysing subtypes of cardiovascular disease separately may give insights into their shared and non-shared aetiology. Also, while the aim of analyses in this work was to understand the role of movement behaviours in cardiovascular disease aetiology, there are also prospects for analyses using device-measured movement behaviours with similar analytic approaches in risk prediction (with appropriate modification of the modelling framework to account for the different priorities of risk prediction and aetiological analysis [326]). For example, if device-measured 24-hour movement behaviours are predictive of cardiovascular disease risk, device-based movement behaviour measurements may be incorporated into risk scores used for clinical decision-making [327].

This work showed how quantitative assessment of the impact of different biases can help to assess the extent to which results may be interpreted causally, and noted avenues to improve this in future work (e.g. collection of data to better characterise measurement error). Nonetheless, quantitative analysis of bias often relies on extensive assumptions and is limited in what it can achieve, so other approaches are also needed in order to move beyond substantial concerns about bias and strengthen causal inference in movement behaviour epidemiology. Use of a more diverse range of studies and study designs will be crucial [43, 255].

Studying different observational cohorts will be one important source of insight: UK Biobank, like most cohort studies with device-measured movement behaviour data, is a socio-economically advantaged cohort in a high-income country. This determines

aspects of the confounding structure, and means most physical activity is likely for leisure or transport rather than work [29]. Whether similar results are found in different populations will therefore help to better understand what is driving them. For example, an ongoing collection of accelerometer data in the China Kadoorie Biobank includes a population where active agricultural work remains common.

Different study designs will help to establish (or otherwise) observational results [310]. Randomised controlled trials enable gold-standard causal inference, but are difficult to run in movement behaviour epidemiology [43]. For example, an adequately-powered trial of long-term effects of physical activity requires an intervention able to achieve substantial and sustained change in physical activity levels without high levels of dropout. Blinding is impractical, and a trial would also need to be large and lengthy to establish effects on incident CVD (although differences in intermediate biomarkers may be observed more quickly). At least one currently-running trial, the Women's Health Initiative Strong and Healthy Pragmatic Physical Activity Intervention Trial, aims to be large enough (49,331 participants randomised) to assess effects of a physical activity intervention on risk of incident cardiovascular disease [328]. However, it is currently unknown whether the physical activity intervention used (a combination of relevant materials, telephone reminders, and support, informed by behaviour change theory) will be sufficient to change physical activity substantially in the medium term [328]. In the absence of randomised controlled trials with disease outcomes, shorter-term trials which show effects on biomarkers of cardiovascular health may provide some evidence [43]. More generally, evidence on biological mechanisms can help to build the case for causality [329]. Biological mechanisms can be probed by various means, including epidemiological studies [44], laboratory studies on people [47], and animal studies [47]. The prospects here are perhaps better than for randomised controlled trials: there is already a wealth of evidence on the acute and long-term physiological changes induced by physical activity [46, 47]; an extensive if less mature

evidence base on the physiological response to sedentary behaviour [330, 331]; and detailed research into sleep biology [332]. Another promising prospect for improving causal inference in movement behaviour epidemiology is the leveraging of genetic epidemiology [292]. In particular, given suitable genetic instrument variables for movement behaviours, these can be used in Mendelian Randomisation analyses [181]. However, currently the variance in movement behaviour phenotypes explainable by known genetic variants is low. A previous study in the UK Biobank sub-cohort with device-measured movement behaviour data, albeit using different movement behaviour phenotypes, found 14 genome-wide significant loci associated with physical activity or sleep duration [181]. These loci together explained just 0.06% of variation in activity and 0.39% of variation in sleep duration [181]. However, there are prospects for this to be improved in the future with larger data collections: considering single nucleotide polymorphisms not reaching genome-wide significance, variance explained was substantially higher (13-18%, depending on the particular phenotype considered) [181]. Improvements to phenotype definition may also support future genetic work.

Finally, this thesis set out to directly support future research by making methods, code, and data as openly available as possible (see Appendices B and C). This may directly add value to future work: other researchers can use, develop, and criticise the methods developed and used here [333]. For example, future studies can use the classifier developed in Chapter 4 to use phenotypes which are directly comparable between studies. More generally, this thesis has demonstrated how emerging methods for measurement and analysis of movement behaviour data (accelerometer-based measurement; machine-learning for behaviour classification; Compositional Data Analysis for analysis of 24-hour movement behaviour data) can enhance epidemiological inference. The advance represented by increasing technological and analytic sophistication is widely acknowledged [78, 85, 204, 272, 313]. However, the research environment has

been relatively slow to evolve to support this, with complex data processing expertise and software sharing remaining relatively rare [334]. Adoption of computational best practices in movement behaviour epidemiology has the potential to substantially accelerate and strengthen research [335].

7.6 Summary

This thesis investigated associations between device-measured 24-hour movement behaviours and risk of incident cardiovascular disease. Analyses of 24-hour movement behaviours were made possible by the development of machine-learning methods able to accurately classify sleep, sedentary behaviour, light physical activity behaviours, and moderate-to-vigorous physical activity behaviours in free-living wrist-worn accelerometer data. Device-measured 24-hour movement behaviours were associated with risk of incident cardiovascular disease in UK Biobank. In particular, reallocating time from other behaviours into moderate-to-vigorous physical activity behaviours was associated with a lower risk of incident cardiovascular disease, and reallocating time from sedentary behaviour into other behaviours was also associated with a lower risk of incident cardiovascular disease. In analyses to quantitatively assess bias in these results, some meaningful differences were seen, but overall conclusions were fairly robust. Future work can build on this research by using different study designs and populations. Evidence from this work may inform development of public health guidelines and interventions addressing 24-hour movement behaviours.

Appendix A: Supplementary Tables and Figures

A.1 Supplementary Tables

Table A.1: Derivation of variables relating to non-accelerometry participant characteristics.

Characteristic	Source	Notes	UK Biobank field	Coding	Relevant chapters
Age	Baseline.		Derived from 90011, 34, 52.		3, 5, 6
Sex	Baseline.		31	Female, Male.	3, 5, 6
Ethnicity	Baseline.		Derived from 21000.	Asian and Asian British, Black and Black British, Mixed and Other, White. (Categorised as in Resource 100336; Mixed/Other includes Mixed, Chinese and Other due to small numbers).	3, 5, 6
Townsend Deprivation Index	Baseline.	Townsend Deprivation Index of address at time of UKB baseline assessment.	Derived from 189.	Within-cohort quarters used for some analyses.	3, 5, 6
Education	Baseline.		Derived from 6138.	School leaver, Further education (education beyond O-Levels/CSEs, excluding college/ university degree) and Higher education (college/university degree).	3, 5, 6
Smoking status	Baseline.		20116	Never smoker, Ex-smoker, Current smoker.	3, 5, 6
Frequency of alcohol consumption	Baseline.		Derived from 1558.	Never, less than 3 times/week, 3 or more times/week.	3, 5, 6
Daily servings of fresh fruit and vegetables.	Baseline.		Derived from 1289, 1299, 1309.	Less than one coded as 0.5, then sum of 1289, 1299, 1309. Categorised as less than 3, 3-4.9, 5-7.9, 8+ servings/ day.	3, 5, 6
Frequency of red and processed meat consumption.	Baseline.		Derived from 1369, 1379, 1389, 1349.	Less than one coded as 0.5, then sum of 1369, 1379, 1389, 1349. Categorised as less than 1, 1-2.9, 3-4.9, 5+ times/ week.	3, 5, 6
Frequency of oily fish consumption.	Baseline.		Derived from 1329.	Less than 1, 1, 2-4, more than 4 times/week.	3, 5, 6

Characteristic	Source	Notes	UK Biobank field	Coding	Relevant chapters
BMI	Baseline.		Derived from 21001.	For descriptive analyses: Underweight (less than 18.5 kg/m^2), Normal weight (18.5 -24.9 kg/m^2), Overweight (25 -29.9 kg/m^2), Obese (30+ kg/m^2). For BMI-stratified analysis, Underweight and Normal weight categories combined.	3, 5, 6
Systolic blood pressure	Baseline.		Derived from 4080 (automatic reading), and where this was missing from 93 (manual reading).	As each measurement was repeated, mean of measurements was used.	
C Reactive Protein	Baseline.		30710		3, 6
HbA1c	Baseline.		30750		3, 6
LDL cholesterol	Baseline.		30780		3, 6
HDL cholesterol	Baseline.		30760		3, 6
Triglycerides	Baseline.		30870		3, 6
Self-reported medications for diabetes, cholesterol or blood pressure	Baseline.		Derived from 6177, 6153.	Participants self-reporting use of insulin, cholesterol-lowering or blood-pressure-lowering medications coded as 'Yes'; other participants coded as 'No'.	3, 6
Self-reported overall health rating	Baseline.		2178	For some analyses split as 'Poor' vs 'Fair', 'Good' or 'Excellent'	3, 6
Prior self-reported cardiovascular disease	Baseline.	Self-reported heart attack or stroke.	Derived from 6150.		5, 6
MET minutes/week walking according to self-report	Baseline.		22037	Categorised as 0, 1-500, 501-1,000, 1,001+.	6
MET minutes/week moderate physical activity according to self-report	Baseline.		22038	Categorised as 0, 1-500, 501-1,000, 1,001+.	6
MET minutes/week vigorous physical activity according to self-report	Baseline.		22039	Categorised as 0, 1-500, 501-1,000, 1,001+.	6
Job involves standing/walking	Baseline.		Derived from 806, 6142.		4
Job involves heavy manual/physical work	Baseline.		Derived from 816, 6142.		4
Self-reported night shift work	Baseline.		Derived from 3426.		4
Self-reported napping during the day	Baseline.		1190		4
Self-reported ease of getting up in the morning	Baseline.		1170		4

Table A.2: Derivation of variables relating to disease outcomes in hospital inpatient data or death register data.

Outcome	Source	Notes	UK Biobank field	Relevant chapters
First cardiovascular disease event	Death register, hospital inpatient data.	First appearance of ICD-10 codes I20-25 (ischaemic heart disease) or I60-69 (cerebrovascular disease) in either hospital inpatient data or death register data.	Derived from Category 100093, Category 2000.	5, 6
Loss-to-follow-up	Death register.	Earliest of: date of latest study data available and date of non-cardiovascular disease death.	Derived from Category 100093.	5, 6
Prior hospital-recorded cardiovascular disease	Hospital inpatient data.	Appearance of ICD-10 codes I20-25 (ischaemic heart disease) or I60-69 (cerebrovascular disease) or ICD-9 codes 410-414 or 430-438 in hospital inpatient data prior to accelerometer wear.	Derived from Category 2000.	5, 6
Prior primary admission for disease of the circulatory system.	Hospital inpatient data.	Hospital admission with primary diagnosis of I00-I99 before accelerometer wear.	Derived from Category 2000.	6
First accident unrelated to movement behaviour.	Death register, hospital inpatient data.	First appearance of ICD-10 codes V00-09, V20-99, W20-W99, X00-X59.	Derived from Category 100093, Category 2000.	6
Prior hospital-recorded accident unrelated to movement behaviour.	Hospital inpatient data.	Appearance of ICD-10 codes V00-09, V20-99, W20-W99, X00-X59 in hospital inpatient data prior to accelerometer wear.	Derived from Category 2000.	6

Table A.3: Features of accelerometry signal used for behaviour classification.

Feature	Description
enmoTrunc	Euclidean Norm Minus One truncated below at 0
mean	Mean
sd	Standard Deviation
coefvariation	Coefficient of Variation
median	Median
min	Minimum
max	Maximum
25thp	25th percentile
75thp	75th percentile
autocorr	Autocorrelation
fmax	Frequency of signal with highest power
pmax	Maximal power of signal
fmaxband	Frequency of signal with highest power between 0.3 and 3 Hz
pmaxband	Maximal power of signal between 0.3 and 3 Hz
entropy	Entropy
fft1	Power at 1 Hz
fft2	Power at 2 Hz
fft3	Power at 3 Hz
fft4	Power at 4 Hz
fft5	Power at 5 Hz
fft6	Power at 6 Hz
fft7	Power at 7 Hz
fft8	Power at 8 Hz
fft9	Power at 9 Hz
fft10	Power at 10 Hz
fft11	Power at 11 Hz
fft12	Power at 12 Hz
MAD	Mean Amplitude Deviation
MPD	Mean Power Deviation
skew	Skew
kurt	Kurtosis
f1	Frequency of signal with highest power between 0.3 and 15 Hz
p1	Maximal power of signal between 0.3 and 15 Hz
f2	Frequency of signal with second highest power between 0.3 and 15 Hz
p2	Second highest power of signal between 0.3 and 15 Hz
f625	Frequency of signal with highest power between 0.6 and 2.5 Hz
p625	Maximal power of signal between 0.6 and 2.5 Hz
totalPower	Total power for frequencies between 0.3 and 15 Hz
vmfft1	Power at 1/30 Hz
vmfft2	Power at 2/30 Hz
vmfft3	Power at 3/30 Hz
vmfft4	Power at 4/30 Hz
vmfft5	Power at 5/30 Hz
vmfft6	Power at 6/30 Hz
vmfft7	Power at 7/30 Hz
vmfft8	Power at 8/30 Hz
vmfft9	Power at 9/30 Hz
vmfft10	Power at 10/30 Hz
vmfft11	Power at 11/30 Hz
vmfft12	Power at 12/30 Hz

Table A.4: Assignment of fine-grained camera image annotations from the Compendium of Physical Activities [28] to broad movement behaviour classes.

SLEEP
7030 sleeping
SEDENTARY BEHAVIOUR
occupation;office and administrative support;11580 office/computer work general
occupation;office and administrative support;11580 office wok/computer work general
home activity;miscellaneous;sitting;9060 sitting/lying reading or without observable/identifiable activities
leisure;miscellaneous;sitting;0010 sitting / lying dialysis
transportation;private transportation;16010 driving automobile or light truck (not a semi)
home activity;miscellaneous;sitting;7010 sitting/lying and watching television with TV on as the primary activity
home activity;miscellaneous;sitting;11580 office/computer work general
home activity;miscellaneous;sitting;9055 sitting/lying talking in person/using a mobile phone/smartphone/tablet or talking on the phone/computer (skype chatting)
home activity;miscellaneous;sitting;11580 office work such as writing and typing (with or without eating at the same time)
home activity;eating;13030 eating sitting alone or with someone
home activity;miscellaneous;sitting;9030 sitting desk entertainment/hobby (with or without eating at the same time)
home activity;miscellaneous;sitting;5080 sitting non-desk work (with or without eating at the same time)
occupation;interruption;sitting;11585 sitting meeting/talking to colleagues with or without eating
leisure;miscellaneous;sitting;9060 (generic) sitting/lying reading or without observable/identifiable activities
occupation;interruption;11585 sitting meeting/talking to colleages with or without eating
home activity;leisure;activities for maintenance of a household;miscellaneous;9100 retreat/family reunion activities involving sitting eating relaxing talking with more than one person
home activity;miscellaneous;sitting;9060 sitting/lying reading or without observable activities
leisure;eating;social;13030 eating sitting indoor/outdoor
home activity;leisure;activties for maintenance of a household;miscellaneous;9100 retreat/family reunion activities involving sitting eating relaxing talking with more than one person
occupation;public admin/education/health;education;9065 students/attending seminars or talks
home activity;miscellaneous;sitting;9030 sitting desk work (with or without eating at the same time)
occupation;interruption;sitting;9060 sitting without observable/identifiable activities
leisure;miscellaneous;sitting;9055 sitting talking to person/using the phone
transportation;private transportation;16015 riding in a car or truck
leisure;eating;13030 eating sitting indoor/outdoor
occupation;interruption;13030 eating sitting
home activity;miscellaneous;sitting;9045 sitting playing traditional video game computer game
home activity;miscellaneous;sitting;9030 sitting desk work (with or without eating at the same time)
occupation;interruption;sitting;13030 eating sitting
home activity;leisure;individual activities;9075 sitting arts and crafts carving/wood weaving/spinning wool
home activity;household chores;washing/ironing/mending clothes;5080 knitting sewing sitting
occupation;interruption;sitting;9055 sitting using a mobile phone/smartphone/tablet or talking on the phone/computer (skype meeting etc.)
leisure;miscellaneous;sitting;21000 sitting meeting
leisure;miscellaneous;sitting;21005 (generic) sitting light office writing typing work
occupation;office and administrative support;11580 office work/computer work general
home activity;miscellaneous;sitting;7010 lying and watching television with TV on as the primary activity
leisure;miscellaneous;sitting;9060 (generic) sitting/lying reading or without observable activities
transportation;waiting;7021 sitting
occupation;interruption;11585 sitting meeting/talking to colleagues with or without eating
leisure;religious activities;20000 sitting in church in service attending a ceremony sitting quietly
transportation;private transportation;16030 motor scooter motorcycle
occupation;interruption;9055 sitting using a mobile phone/smartphone/tablet or talking on the phone/-computer (skype meeting etc.)
leisure;miscellaneous;21005 (generic) sitting light office writing typing work
home activity;miscellaneous;sitting;7021 sitting without observable actiivties
home activity;leisure;activties for maintenance of a household;with children;5170 sitting playing with child(ren)

home activity;leisure;activities for maintenance of a household;with children;5170 sitting playing with child(ren)
occupation;interruption;9060 (generic) sitting without observable activities
home activity;child/elderly/pet care;child care;5185 child care sitting/kneeling
home activity;miscellaneous;sitting;21010 sitting non-desk work (with or without eating at the same time)
leisure;eating;not-social;13030 eating sitting indoor/outdoor
leisure;miscellaneous;sitting;5080 sitting non-desk work (with or without eating at the same time)
home activity;leisure;activties for maintenance of a household;9100 retreat/family reunion activities involving sitting eating relaxing talking with more than one person
leisure;miscellaneous;sitting;21016 sitting child care only active periods
home activity;leisure;activties for maintenance of a household;5170 sitting playing with child(ren)
home activity;lawn and garden;gardening service;8055 driving tractor
occupation;interruption;9060 (generic) sitting without observable/identifiable activities
home activity;self care;13036 taking medication
leisure;miscellaneous;21000 sitting meeting or talking with others
home activity;self care;13046 having hair or nails done by someone else sitting
home activity;household chores;washing/ironing/mending clothes;5080 knitting sewing wrapping presents sitting
home activity;miscellaneous;sitting;9060 sitting reading or using a mobile phone/smartphone/tablet or talking on the phone/computer (skype chatting)
leisure;miscellaneous;21010 sitting non-desk work (with or without eating at the same time)
occupation;interruption;9060 sitting using a mobile phone/smartphone/tablet or talking on the phone/-computer (skype meeting etc.)
leisure;religious activities;20005 sitting in church talking or singing attending a ceremony sitting active
home activity;child/elderly/pet care;child care;5185 child care sitting/kneeling occasional lifting
leisure;miscellaneous;21016 sitting child care only active periods
home activity;miscellaneous;sitting;7021 sitting without observable activities
home activity;leisure;activities for maintenance of a household;with animals;5190 sitting playing with animals active periods
transportation;public transportation;16016 riding in a bus or train
leisure;sports;water activities;18012 boating power passenger

LIGHT PHYSICAL ACTIVITY BEHAVIOURS

home activity;household chores;preparing meals/cooking/washing dishes;5035 kitchen activity general cooking/washing/dishes/cleaning up
home activity;miscellaneous;walking;17150 walking household without observable loads
home activity;miscellaneous;walking;5165 (generic) walking non-cleaning task such as closing windows lock door putting away items
leisure;miscellaneous;walking;21070 (generic) walking and occasional standing (no more than two consecutive images)
transportation;walking;17161 walking not as the single means of transports e.g.from house to transports or vice versa/from car to places or vice versa/between transports
leisure;miscellaneous;walking;5060 shopping miscellaneous
occupation;interruption;11791 walking on job in office or lab area
home activity;miscellaneous;standing;9050 standing talking in person/on the phone/computer (skype chatting) or using a mobile phone/smartphone/tablet
home activity;household chores;washing/ironing/mending clothes;5090 folding or hanging clothes/put clothes in or out of washer or dryer/packing suitcase limited walking
home activity;miscellaneous;standing;9070 standing reading or without observable/identifiable activities
occupation;interruption;walking;11791 walking on job in office or lab area
occupation;manufacturing;11115 chef
home activity;self care;13040 (generic) self care such as grooming/washing hands/shaving/brushing teeth/putting on make-up not eliminating and bathing (not necessary in the toilet)
occupation;office and administrative support;11600 (generic) standing tasks such as store clerk/libarian/-packing boxes/repair heavy parts
leisure;miscellaneous;5060 shopping miscellaneous
home activity;household chores;washing/ironing/mending clothes;5070 ironing
leisure;miscellaneous;standing;9050 standing talking in person/using a phone/smartphone/tablet
home activity;miscellaneous;walking;5147 walking moving away light items (pens/papers/keys not included)
occupation;miscellaneous;11475 (generic) manual labour
occupation;interruption;standing;9050 standing talking in person/using a phone/smartphone/tablet
occupation;personal services;11413 kitchen maid
home activity;household chores;grocery shopping;5060 shopping
home activity;home repair;indoor;6126 home repair miscellaneous
leisure;miscellaneous;standing;9070 standing reading or without observable/identifiable activities
transportation;waiting;7040 standing in a line

home activity;child/elderly/pet care;child care;5185 child care sitting/kneeling occasional lifting
home activity;miscellaneous;standing;9071 (generic) standing miscellaneous
leisure;miscellaneous;standing;9071 (generic) standing miscellaneous
occupation;interruption;standing;9070 standing reading or without observable/identifiable activities
home activity;miscellaneous;standing;9020 standing writing/drawing/painting
home activity;leisure;individual activities;10074 playing musical instruments
occupation;interruption;5041 kitchen activity in the working place
home activity;miscellaneous;walking;17140 using crutches or frame
occupation;interruption;9050 standing talking in persone/using a phone/smartphone/tablet
home activity;leisure;individual activities;9085 standing arts and crafts/sand painting/carving weaving
home activity;miscellaneous;5025 (generic) multiple household tasks all at once including standing/lifting/sitting
home activity;miscellaneous;standing;9070 standing reading or without obvious activities
home activity;household chores;washing/ironing/mending clothes;5095 putting away /gathering clothes involving walking
home activity;lawn and garden;lawn care service;8135 planting potting transplanting seedlings or plants
occupation;interruption;standing;9071 (generic) standing miscellaneous
transportation;walking;9050 standing miscellaneous (talking to others etc.)
occupation;interruption;miscellaneous;5041 kitchen activity in the working place
home activity;household chores;house cleaning;miscellaneous;5100 making bed/changing linens
occupation;interruption;standing;9015 standing scanning documents
home activity;household chores;house cleaning;furniture;5032 dusting or polishing furniture
home activity;eating;13035 eating standing alone or with others
occupation;miscellaneous;11870 working in scene shop theatre actor backstage employee
occupation;office and administrative support;11600 (generic) standing tasks such as store clerk/librarian/-packing boxes/repair heavy parts
home activity;miscellaneous;standing;9050 standing talking in person on the phone/computer (skype chatting) or using a mobileo phone/smartphone/tablet
home activity;household chores;house cleaning;floors;5010 cleaning sweeping carpet or floors
leisure;sports;ball games;15090 bowling
home activity;self care;13020 dressing/undressing
occupation;interruption;9070 standing reading or without observable/identifiable activities
leisure;music playing;10074 playing musical instruments
home activity;home repair;indoor;6205 sharpening tools
home activity;home repair;indoor;6124 hammering nails
home activity;child/elderly/pet care;child care;5186 child care standing occasional lifting
occupation;interruption;13035 eating standing
home activity;child/elderly/pet care;pet care;5197 household animal care aside from feeding
leisure;recreation;outdoor;5171 standing playing with child(ren)
occupation;interruption;miscellaneous;13009 toilet break
home activity;household chores;house cleaning;miscellaneous;5148 watering plants
occupation;miscellaneous;11475 (generic) manual or unskilled labour
occupation;interruption;standing;9020 standing writing/drawing/painting
leisure;eating;social;13035 eating standing indoor/outdoor
home activity;household chores;house cleaning;floors;5131 scrubbing floors on hands and knees scrubbing bathroom bathtub
occupation;public admin/education/health;education;09071 teaching standing
home activity;household chores;preparing meals/cooking/washing dishes;5051 serving food/setting table implied walking and standing
home activity;household chores;preparing meals/cooking/washing dishes;5035 cleaning up table after meal implied walking (e.g. leaving from eating table to the kitchen)
leisure;eating;not-social;5060 buying foods or drinks as a takeaway
leisure;miscellaneous;21070 (generic) walking/standing combination indoor
home activity;self care;13009 toilet eliminating or squatting
home activity;self care;13045 hairstyling standing
home activity;lawn and garden;gardening service;8230 watering garden
occupation;interruption;13009 toilet break
leisure;miscellaneous;standing;9070 standing reading or without obvious activities
home activity;leisure;activities for maintenance of a household;miscellaneous;9101 retreat/family reunion activities playing games with more than one person
occupation;interruption;9015 standing scanning documents
home activity;child/elderly/pet care;child care;5183 standing holding child
occupation;interruption;9070 standing reading or without obvious activities
leisure;miscellaneous;standing;9020 standing writing/drawing/painting
leisure;religious activities;20039 walking/standing combination for religious purposes usher
leisure;recreation;indoor;9020 drawing writing painting standing
leisure;miscellaneous;21017 standing child care only active periods
home activity;child/elderly/pet care;pet care;5053 feeding household animals

leisure;eating;13035 eating standing indoor/outdoor
leisure;miscellaneous;9071 (generic) standing miscellaneous indoor or outdoor
leisure;sports;conditioning;2115 upper body exercise arm ergometer
occupation;interruption;standing;9050 standing talking in persone/using a phone/smartphone/tablet
home activity;self care;13000 getting ready for bed standing
leisure;eating;social;5060 buying foods or drinks as a takeaway
home activity;home repair;indoor;5160 standing light effort tasks
transportation;walking;9071 standing miscellaneous (talking to others etc.)
occupation;interruption;standing;13035 eating standing
occupation;interruption;9020 standing writing/drawing/painting
home activity;lawn and garden;gardening service;8220 walking applying fertilizer or seeding a lawn push applicator
leisure;eating;5060 buying foods or drinks as a takeaway
leisure;religious activities;20030 standing talking in church
leisure;eating;not-social;13035 eating standing indoor/outdoor
occupation;public admin/education/health;education;9071 teaching standing
home activity;household chores;washing/ironing/mending clothes;5082 sewing with a machine

MODERATE-TO-VIGOROUS PHYSICAL ACTIVITY BEHAVIOURS

transportation;private transportation;1010 bicycling
transportation;walking;17165 walking the dog
occupation;interruption;17133 walking upstairs
leisure;miscellaneous;walking;17031 loading /unloading a car implied walking
home activity;child/elderly/pet care;child care;5181 walking and carrying child
home activity;lawn and garden;gardening service;8050 digging spading filling garden compositing
leisure;sports;miscellaneous;17082 hiking or walking at a normal pace through fields and hillsides
home activity;lawn and garden;tree and shrub service;8025 clearing light brush thinning garden
home activity;lawn and garden;lawn care service;8165 raking lawn
occupation;interruption;walking;11795 walking on job and carrying light objects such as boxes or pushing trolleys
leisure;sports;water activities;18070 canoeing/rowing
home activity;lawn and garden;lawn care service;8080 laying crushed rock
leisure;sports;gymnasium and athletics;athletics;12150 running
leisure;sports;ball games;15235 football or baseball playing catch
home activity;miscellaneous;standing;5146 standing packing/unpacking household items occasional lifting
home activity;lawn and garden;gardening service;8245 gardening/picking up fruits vegetables flowers
leisure;miscellaneous;walking;17105 pushing a wheelchair non-occupational
leisure;sports;ball games;15680 tennis doubles
home activity;lawn and garden;gardening service;8192 shovelling dirt or mud
leisure;sports;miscellaneous;17082 hiking or walking at a normal pace through fields and hillsides
leisure;dancing;3010 ballet modern or jazz general rehearsal or class
leisure;sports;conditioning;2060 health club exercise
leisure;sports;ball games;15690 tennis singles
home activity;lawn and garden;lawn care service;8095 mowing lawn
occupation;public admin/education/health;health;11615 nursing patient care
occupation;miscellaneous;11615 (generic) standing lifting items continuously with limited walking
leisure;miscellaneous;17031 loading /unloading a car implied walking
leisure;sports;conditioning;2019 bicycling stationary RPM/Spin bike class
home activity;leisure;activties for maintenance of a household;with animals;5192 walking/running playing with animals active periods
home activity;lawn and garden;gardening service;8192 shoveling dirt or mud
home activity;household chores;house cleaning;furniture;5020 cleaning heavy such as car/windows/garage
leisure;sports;conditioning;2048 elliptical trainer
occupation;interruption;11795 walking on job and carrying light objects such as boxes or pushing trolleys
leisure;sports;conditioning;2050 resistance training
leisure;sports;conditioning;2070 rowing stationary ergometer
leisure;recreation;outdoor;5175 walking/running playing with child(ren)
leisure;sports;conditioning;2010 bicycling stationary
leisure;sports;conditioning;2120 water aerobics water calisthenics water exercise
home activity;home repair;outdoor;6020 automobile body work
occupation;agriculture/forestry/fishing;11192 taking care of animals
leisure;miscellaneous;standing;21017 standing child care only active periods
leisure;sports;conditioning;2065 stair-treadmill ergometer general
home activity;household chores;washing/ironing/mending clothes;5092 washing clothes by hand (with or without hanging wash)
leisure;miscellaneous;walking;17133 walking upstairs
transportation;walking;12150 running

occupation;interruption;walking;17070 walking downstairs
home activity;household chores;house cleaning;floors;5140 sweeping garage sidewalk or outside of house
occupation;interruption;walking;17133 walking upstairs
occupation;agriculture/forestry/fishing;11540 shovelling digging ditches
occupation;construction;11050 carrying heavy loads
transportation;walking;17250 walking as the single means to a destination not to work or class
leisure;miscellaneous;walking;17070 descending stairs
home activity;miscellaneous;walking;5121 walking with moving and lifting loads such as bikes and furniture
transportation;walking;17270 walking as the single means to work or class (not from)

Table A.5: Odds Ratios for selection into accelerometer sub-cohort among 472,118 participants with no prior cardiovascular disease at baseline assessment.

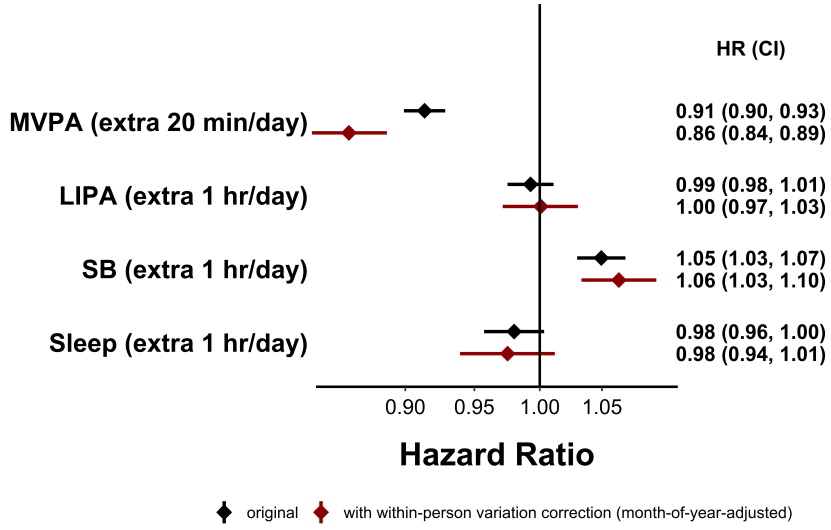
Variable	OR (CI)
Intercept	0.95 (0.76, 1.20)
Baseline age (per ten years)	1.02 (1.01, 1.03)
Sex (Ref: Female)	
Male	0.90 (0.88, 0.92)
Ethnicity (Ref: White)	
Asian	0.46 (0.43, 0.50)
Black	0.55 (0.51, 0.60)
Mixed/Other	0.70 (0.66, 0.75)
Townsend Deprivation Index	0.97 (0.97, 0.97)
Education (Ref: Higher education)	
School leaver	0.51 (0.50, 0.52)
Further education	0.76 (0.75, 0.77)
Smoking status (Ref: Never smoker)	
Ex-smoker	1.05 (1.03, 1.07)
Current smoker	0.69 (0.67, 0.71)
Alcohol consumption frequency (Ref: Less than 3 times/week)	
Never	0.85 (0.82, 0.88)
3 or more times/week	1.10 (1.09, 1.12)
Daily fruit and vegetable consumption (Ref: 5-7.9 servings/day)	
Less than 3 servings/day	0.85 (0.82, 0.88)
3-4.9 servings/day	0.95 (0.93, 0.97)
8 or more servings/day	0.97 (0.95, 0.98)
Red and processed meat consumption (Ref: 1-2.9 times/week)	
Less than 1 time/week	1.09 (1.06, 1.13)
3-4.9 times/week	0.96 (0.94, 0.98)
5 or more times/week	0.94 (0.93, 0.96)
Oily fish consumption (Ref: Less than 1 time/week)	
1 time/week	0.97 (0.96, 0.99)
2-4 times/week	0.91 (0.89, 0.93)
More than 4 times/week	0.81 (0.75, 0.88)
BMI	0.99 (0.98, 0.99)
Systolic blood pressure (per 10 mmHg)	0.98 (0.97, 0.98)
log HbA1c	0.84 (0.79, 0.90)
log C Reactive Protein	0.97 (0.96, 0.98)
LDL cholesterol	1.00 (0.99, 1.01)
HDL cholesterol	1.03 (1.00, 1.06)
log Triglycerides	0.98 (0.96, 1.00)
Self-reported poor overall health	0.83 (0.78, 0.87)
Self-reported medications for diabetes, cholesterol or blood pressure	0.93 (0.91, 0.95)
Self-reported weekly walking (Ref: None)	
1-500 MET minutes	1.06 (1.01, 1.12)
501-1,000 MET minutes	1.06 (1.01, 1.12)
1,001+ MET minutes	1.00 (0.95, 1.05)
Self-reported weekly moderate physical activity (Ref: None)	
1-500 MET minutes	1.03 (1.00, 1.05)
501-1,000 MET minutes	1.01 (0.97, 1.04)
1,001+ MET minutes	0.93 (0.90, 0.96)
Self-reported weekly vigorous physical activity (Ref: None)	
1-500 MET minutes	1.09 (1.06, 1.12)
501-1,000 MET minutes	1.13 (1.10, 1.16)
1,001+ MET minutes	1.08 (1.05, 1.11)
Self-reported usual walking pace (Ref: Steady average pace)	
Brisk pace	1.20 (1.18, 1.22)
Slow pace	0.78 (0.75, 0.81)
None of the above	0.44 (0.34, 0.57)

Table A.6: Odds Ratios for selection into accelerometer sub-cohort among 455,921 participants who were alive and had no cardiovascular disease record by 2013-06-01.

Variable	OR (CI)
Intercept	0.82 (0.65, 1.04)
Baseline age (per ten years)	1.04 (1.03, 1.05)
Sex (Ref: Female)	
Male	0.91 (0.90, 0.93)
Ethnicity (Ref: White)	
Asian	0.46 (0.43, 0.50)
Black	0.54 (0.50, 0.58)
Mixed/Other	0.70 (0.65, 0.74)
Townsend Deprivation Index	0.97 (0.97, 0.98)
Education (Ref: Higher education)	
School leaver	0.51 (0.50, 0.52)
Further education	0.76 (0.75, 0.77)
Smoking status (Ref: Never smoker)	
Ex-smoker	1.06 (1.04, 1.08)
Current smoker	0.70 (0.68, 0.72)
Alcohol consumption frequency (Ref: Less than 3 times/week)	
Never	0.86 (0.83, 0.89)
3 or more times/week	1.10 (1.09, 1.12)
Daily fruit and vegetable consumption (Ref: 5-7.9 servings/day)	
Less than 3 servings/day	0.86 (0.82, 0.89)
3-4.9 servings/day	0.96 (0.93, 0.98)
8 or more servings/day	0.97 (0.95, 0.98)
Red and processed meat consumption (Ref: 1-2.9 times/week)	
Less than 1 time/week	1.09 (1.06, 1.13)
3-4.9 times/week	0.96 (0.94, 0.98)
5 or more times/week	0.95 (0.93, 0.97)
Oily fish consumption (Ref: Less than 1 time/week)	
1 time/week	0.97 (0.96, 0.99)
2-4 times/week	0.91 (0.89, 0.93)
More than 4 times/week	0.82 (0.75, 0.89)
BMI	0.99 (0.98, 0.99)
Systolic blood pressure (per 10 mmHg)	0.98 (0.98, 0.98)
log HbA1c	0.86 (0.81, 0.91)
log C Reactive Protein	0.97 (0.97, 0.98)
LDL cholesterol	1.00 (0.99, 1.01)
HDL cholesterol	1.02 (0.99, 1.05)
log Triglycerides	0.98 (0.96, 1.00)
Self-reported poor overall health	0.88 (0.83, 0.93)
Self-reported medications for diabetes, cholesterol or blood pressure	0.95 (0.93, 0.97)
Self-reported weekly walking (Ref: None)	
1-500 MET minutes	1.07 (1.02, 1.12)
501-1,000 MET minutes	1.06 (1.00, 1.13)
1,001+ MET minutes	1.00 (0.94, 1.06)
Self-reported weekly moderate physical activity (Ref: None)	
1-500 MET minutes	1.03 (1.00, 1.06)
501-1,000 MET minutes	1.01 (0.98, 1.04)
1,001+ MET minutes	0.93 (0.90, 0.97)
Self-reported weekly vigorous physical activity (Ref: None)	
1-500 MET minutes	1.10 (1.06, 1.15)
501-1,000 MET minutes	1.12 (1.08, 1.17)
1,001+ MET minutes	1.10 (1.05, 1.16)
Self-reported usual walking pace (Ref: Steady average pace)	
Brisk pace	1.19 (1.17, 1.21)
Slow pace	0.80 (0.77, 0.83)
None of the above	0.44 (0.34, 0.56)

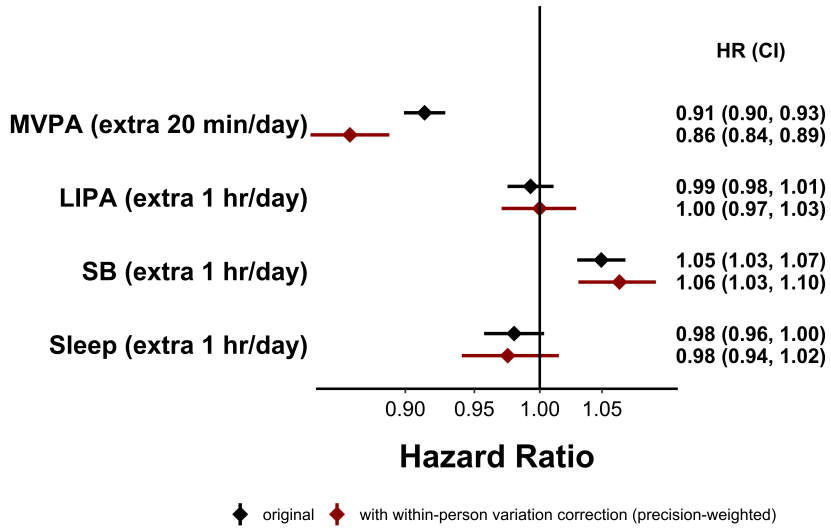
A.2 Supplementary Figures

Figure A.1: Hazard Ratios for incident cardiovascular disease associated with reallocating time to a given movement behaviour, from all other behaviours proportionally, among 87,498 participants (4,105 events) with and without 'correction' for within-person variation using multivariable regression calibration, with adjustment for wear month in the calibration equation.



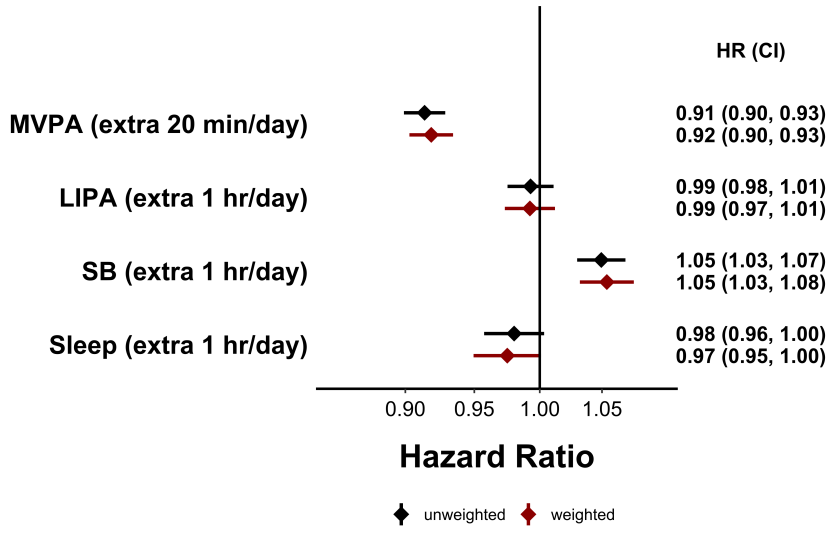
All relative to the mean behaviour composition (8.8 hours/day sleep, 9.3 hours/day sedentary behaviour, 5.6 hours/day light physical activity behaviours, 0.35 hours/day (21 min/day) moderate-to-vigorous physical activity behaviours) and more time in named behaviour reallocated from all other behaviours proportionally. Calibration equation was adjusted for age, sex, ethnicity, deprivation, education, smoking status, alcohol consumption, fresh fruit and vegetable consumption, red and processed meat consumption, oily fish consumption, and wear month. Cox model used age as the timescale, was stratified by sex and was additionally adjusted for ethnicity, deprivation, education, smoking status, alcohol consumption, fresh fruit and vegetable consumption, red and processed meat consumption, and oily fish consumption.

Figure A.2: Hazard Ratios for incident cardiovascular disease associated with reallocating time to a given movement behaviour, from all other behaviours proportionally, among 87,498 participants (4,105 events) with and without 'correction' for within-person variation using multivariable regression calibration, precision-weighted based on number of valid repeat measurements.



All relative to the mean behaviour composition (8.8 hours/day sleep, 9.3 hours/day sedentary behaviour, 5.6 hours/day light physical activity behaviours, 0.35 hours/day (21 min/day) moderate-to-vigorous physical activity behaviours) and more time in named behaviour reallocated from all other behaviours proportionally. Calibration equation was adjusted for age, sex, ethnicity, deprivation, education, smoking status, alcohol consumption, fresh fruit and vegetable consumption, red and processed meat consumption, and oily fish consumption, and was precision-weighted based on number of valid repeat measurements. Cox model used age as the timescale, was stratified by sex and was additionally adjusted for ethnicity, deprivation, education, smoking status, alcohol consumption, fresh fruit and vegetable consumption, red and processed meat consumption, and oily fish consumption.

Figure A.3: Hazard Ratios for incident cardiovascular disease associated with reallocating time to a given movement behaviour, from all other behaviours proportionally, among 87,498 participants (4,105 events) with and without weighting by Inverse Probability of Selection from restricted main study cohort into accelerometer sub-cohort.



Restricted main study cohort was all participants who were alive and had not had a cardiovascular disease record by 2013-06-01. All relative to the (unweighted) mean behaviour composition (8.8 hours/day sleep, 9.3 hours/day sedentary behaviour, 5.6 hours/day light physical activity behaviours, 0.35 hours/day (21 min/day) moderate-to-vigorous physical activity behaviours) and more time in named behaviour reallocated from all other behaviours proportionally. Models used age as the timescale, were stratified by sex and were additionally adjusted for ethnicity, deprivation, education, smoking status, alcohol consumption, fresh fruit and vegetable consumption, red and processed meat consumption, and oily fish consumption. Derivation of weights used the procedure described in the text.

Appendix B: Code and data availability

B.1 Code

Analytic code for Chapters 4 and 5 is available at `github.com/OxWearables/manuscript_ml_behaviours_cvd_2021`. This repository contains code and related information for: development of the machine-learning model to classify movement behaviours in Chapter 4; processing of accelerometer data in UK Biobank, including movement behaviour classification; preparation of other variables in UK Biobank; and analyses associating movement behaviours with incident cardiovascular disease in Chapter 5.

Additional analytic code is available at `github.com/OxWearables/thesis_additional_walmsley`. This code primarily relates to results in Chapter 3 (UK Biobank dataset description) and Chapter 6 (Quantitative Bias Analysis). Some accelerometer processing code used in Chapter 6 is also available at `github.com/OxWearables/repeatedMeasurementsProcessing`.

The analytic pipelines for these results are supported by several standalone software tools. This project has contributed to the development, improvement, or maintenance of all of these software tools. These tools are:

1. **biobankAccelerometerAnalysis** This is a tool to extract meaningful health information from large accelerometer datasets. It includes functionality for development and deployment of machine-learning models for movement behaviour classification. It was used extensively in analyses in Chapter 4. This project contributed to *ad hoc* maintenance of the tool, and the machine-learning model developed in Chapter 4 is now available to access through this tool. It can be accessed at `github.com/OxWearables/biobankAccelerometerAnalysis`;

2. **ukb_download_and_prep_template** This tool provides functionality and documentation for pre-processing UK Biobank data, and was used for this purpose in the thesis. This thesis contributed to the development and maintenance of this tool. However, its workflows are expected to be superseded by new cloud-based workflows using the new Research Analysis Platform. It can be accessed at `github.com/OxWearables/ukb_download_and_prep_template`;

3. **epicoda** This bespoke R package to support common epidemiological analyses using compositional exposure variables was developed through this project, and is described in detail in Appendix C. It can be accessed at `github.com/OxWearables/epicoda`;

4. **repeatedMeasurements** Functions to analyse repeated measurements data used in Chapter 6 are stored separately as part of a skeleton (i.e. in-development) R package. At the time of writing, this package is not mature, but can be accessed at `github.com/OxWearables/repeatedMeasurements`.

B.2 Data

Data from the CAPTURE-24 study are available at `10.5287/bodleian:NGx0JOMP5`.

UK Biobank data are available to researchers who submit an application (`www.ukbiobank.ac.uk`).

Appendix C: Development of an open-source R package for Compositional Data Analysis in epidemiology

C.1 Introduction

At the start of this project, there already existed several resources for facilitating Compositional Data Analysis in R (as well as in other statistical packages). For example, the 'compositions' package is a longstanding package for working with compositional data [336], and 'robCompositions' is a package useful both for general work with compositional data and specifically for the application of robust methods [337]. However, historically Compositional Data Analysis has not been widely used in epidemiology: the work of Chastin *et al.* [122] in 2015 is generally considered to have introduced Compositional Data Analysis to the movement behaviour epidemiology literature. Prior to that, Compositional Data Analysis had been more commonly used in physical sciences, such as geology, where the most common types of analyses differ substantially from in epidemiology. Software packages which are well-adapted to these fields, while extremely useful, may therefore not optimally reflect the demands of epidemiological analyses. For example, in epidemiology perhaps the most common types of analyses are regression models, including models which may be rare in other fields (e.g. Cox regression for survival analyses). Moreover, there is a great need to

communicate results in a way that is accessible and translatable to a broad audience. Compositional Data Analysis has rapidly gained traction in movement behaviour epidemiology, but there is some lag between the introduction of novel methods and the surrounding software ecosystem becoming mature. This was the context in which I began developing a package to facilitate Compositional Data Analysis in epidemiological research. It is worth noting that, in the intervening years, wider recognition of this gap means that the situation has evolved, with more options now available. For example, there is a package for linear regression with a compositional exposure known as 'deltacomp' [338], and many online resources such as the OpenCoDa suite of code are also available [339]. Nonetheless, there is some variation in functionality between these different packages; there is value to having a range of tools for a particular task; the software development effort represents a useful component of the overall thesis; and the process itself is instructive. Therefore, this appendix describes this software development effort.

The aim was to produce a reusable package for common analysis types in epidemiological research with a compositional exposure. This was done in tandem with the analyses in this thesis i.e. a software tool was developed and also used to produce the analyses in Chapter 5. Developing the package served multiple purposes: most immediately, it facilitated this work by making code easily transportable between projects and modular; second, it was useful in teaching; and third, it is now available for general use. It also represented an opportunity to build software development skills, which has had benefits for research code throughout the project and is also usable in future research.

The key aim of the package was to facilitate epidemiological analyses using compositional exposure variables. To achieve this, the package focuses on user-friendly functions for performing linear, logistic, and Cox regression analyses with a compo-

sitional exposure, as well as functions for presenting results from such analyses in a way that is easy to interpret for a range of practitioners. A secondary aim of the package was to facilitate dissemination of Compositional Data Analysis methods in epidemiological research by being designed for use as a teaching tool.

C.2 Package design

C.2.1 Modelling

The model contains a wrapper for linear, logistic and Cox regression with a compositional exposure (the 'comp_model' function). To accomplish this, the model performs on-the-fly isometric log-ratio transformation and sets up a model with the transformed variables. The 'comp_model' function allows for covariate adjustment. In the case of the Cox model, it also allows for model stratification and different timescales.

C.2.2 Interpreting results

Several functions exist for ease of interpreting results. There is a function for getting estimates from the model for different combinations of variables in the composition relative to the compositional mean. The function 'plot_transfers' enables plotting of reallocation plots centred on the compositional mean. There is also a function to illustrate different reallocations on a forest plot.

There is also a function to automate rotation through different first pivot coordinates to derive and tabulate the coefficient of the first coordinate with each different behaviour in the first pivot coordinate position. This enables some interpretation of the coefficients (see Chapter 5 for more detail on this).

C.2.3 Other functionality

The package contains several utility functions for performing tasks which need to be done throughout the package. For example, there are utility functions for changing the scale of the data outputs to different commonly-used units.

There are also functions to enable on-the-fly zero imputation in the context of other functions. This is potentially convenient, and it can also be useful in pedagogical contexts to abstract away from more advanced questions around zero imputation. However, this is not be the recommended way to use the package, as it may obscure an important analytic decision. More recently, the package has therefore moved towards prompting users to actively consider appropriate zero treatment in their own data.

For legacy reasons, the package contains some functions for plotting on a ternary plot. These may be retired as they do not build greatly on the existing available packages (notably 'ggtern') and are not particularly coherent with the purposes of the rest of the package.

C.2.4 Data

Data can help to test a package, demonstrate its use, and further develop functionality. However, for information governance and ethics reasons, many datasets used in health research cannot be shared. Therefore, the package contains a simulated dataset ('simdata').

This dataset includes:

1. Compositional variables (vigorous activity, moderate activity, light activity,

sedentary behaviour, and sleep).

2. Age groups and sex.
3. Body Mass Index (an outcome for linear regression).
4. A binary disease outcome.
5. An event and time-to-event outcome.

The simulation was originally loosely based on real data. However, its features are only partially realistic. For example, age group and sex are not correlated with the compositional variables. While data may be simulated more realistically in the future, for most practical testing and teaching purposes, this data has been sufficient.

C.2.5 Testing

At a relatively late stage of package development, a suite of tests were developed to verify that key aspects of the package worked as expected. The package also has Continuous Integration, whereby tests are run on multiple operating systems when a pull request is submitted to the package repository on GitHub.

Unit tests have been written testing key aspects of functionality of the package. Some tests implement comparisons of the package with other packages implementing related functionality. For example, isometric log-ratio transformation has been tested against the 'robCompositions' package [214]. The 'deltacomp' package has been used to test elements of the functionality relating to linear regression [338]. Other *ad hoc* tests have been performed of logistic and Cox regression functionality, using custom code from Duncan McGregor (logistic and Cox regression) [214, 339] and Louise Millard (Cox regression) [126].

However, test coverage is far from complete. In particular, testing of plotting functions in R presents a particular challenge: it is difficult to automate visual inspection of images [340], whereas testing exact identity of plot objects can be excessively sensitive (e.g. to changes to the plotting environment). As package functions become more flexible, the number of scenarios to test expands in a combinatorial fashion. Moreover, it is recommended to test common failure cases (e.g. null or missing values in arguments), which has not yet been implemented. In future development of the package the role of tests should become more central still (see Section C.7).

C.3 Tutorial

A tutorial in the form of an R vignette (i.e. a tutorial with integrated code) is also included within the package. This is reproduced at the end of this Appendix to briefly demonstrate the package's functionality.

C.4 Accessing and using the package

The package can be accessed on GitHub at 'github.com/OxWearables/epicoda'. It can be installed directly from an active R session by running:

```
remotes::install_github("activityMonitoring/epicoda")
```

[Note that this requires the 'remotes' package to be installed.]

Historically, the primary route to access R packages has been through CRAN. To make a package available on CRAN, the developer submits it to CRAN and it undergoes a range of checks. There are advantages to CRAN submission, such as verification that

the package meets certain standards and ease of use for users who may not be familiar with GitHub. However, there are also substantial disadvantages to CRAN submission, including substantial additional time required to prepare the CRAN submission and a slower update cycle. Moreover, the development of both GitHub and the 'remotes' package mean that installation via GitHub requires at most one additional line of code compared to CRAN installation, so need not represent a substantial additional overhead. Standard CRAN tests are also run as part of the Continuous Integration cycle. Therefore, the benefits of CRAN submission were not sufficient to justify pursuing it at this time (although this may be revisited should the package become more stable or widely used).

The package has been used in teaching three cohorts of the University of Oxford Centre for Doctoral Training in Health Data Science. This has shown that the package is usable for users who are generally conversant with R but have little-to-no familiarity with Compositional Data Analysis. It has also highlighted areas where the package could be clarified or improved and helped to identify bugs.

C.5 Evolution of the package

The package has evolved in various ways over the course of its development. A key driver of this has been the evolution of the package from a tool used principally by me in my own research, to a pedagogical tool and a tool which potentially has a wider audience. Changes have included:

1. The development of more extensive documentation. In particular, the package needed to be made self-explanatory. The use of a vignette, that is, a tutorial with code, has substantially supported this. Teaching experience suggests the package is now reasonably self-explanatory, at least for basic use, but it will be

further refined as and when the package is used more widely.

2. Making the package more robust to different uses and users. In particular, in early 2021 a series of changes were made to make the package more robust and, in particular, to minimise reliance on users consistently specifying arguments across different functions by deriving these in an automated fashion where possible. The extent of reliance on this has been reduced, but not eliminated, and more checks at an earlier stage would be useful (e.g. to throw errors if arguments are specified with the wrong type).

3. Some changes to particular defaults and design choices as my understanding of methods and use cases has evolved.

C.6 Strengths and limitations

'epicoda' is functional for use in everyday research. For example, the Compositional Data Analysis work in this thesis was conducted using this package. It has a level of flexibility suitable for a range of common use cases. A key strength of the package is its extensive documentation, which facilitates use by a wide range of users familiar with basic R. It is simple to install. Once installed, help can be accessed in the usual way ('?my_function'). The vignette described above also facilitates 'quickstart' use.

There are also several ways the package could be improved. Before wider use of the package, further testing and code review should be performed. Development by a single analyst brings with it a high risk of unidentified errors; there is evidence that code review can identify 55-60% of defects, integration testing 45% of defects, and unit testing 25% of defects, making the combination of these tools particularly helpful [333, 341]. Also, the package could probably be made simpler and more co-

herent for users. It may be possible to refine the outputs returned to users. For example, functions sometimes return outputs which may not be obvious to interpret (e.g. data frames containing the results of intermediate calculations). For historical reasons, some functionality is included within the package which may not be used going forwards. The under-the-hood implementation could also be improved. Limited attention was paid to speed of processing. While this rarely represents a major block in use of the package even in large data, there may be some relatively simple ways to improve processing speed. In retrospect, the implementation of different model types was *ad hoc*. A more coherent design principle may organise the implementation of linear and logistic regression under a more generic generalised linear model framework. As well as being statistically elegant, this would allow a wider range of regression analyses, such as Poisson regression. It may also be possible to make the under-the-hood implementation more sophisticated and coherent through the use of structures provided by R, such as S3 objects. Finally, it is an ongoing process to find the right balance between facilitating easy analyses by abstracting away from mathematical detail, and ensuring the analyst still makes key decisions (for example, see the discussion of zeroes in Section C.2.3 above).

C.7 Future development

Key priorities for future development are:

1. More extensive testing with different user groups, to better understand how usable and intuitive the package is and to unearth bugs.

2. Full independent code review for analytic correctness. As described in Section C.6, code review is an important means of ensuring quality in software development, and can discover a large proportion of bugs [333, 341].

3. More extensive unit testing. While unit testing is partially implemented, improving test coverage may help to further identify bugs and make it easier to reliably extend the package.

4. Making code clearer and more performant. There may be opportunities to improve the design of code and ensure code runs as efficiently as possible.

5. Rationalising what is included in the package for coherence. For example, in the future it may make the package more coherent to drop functionality on plotting prediction regions on ternary plots (Section C.2.3).

6. Implementing further functionality where this has a clear rationale. For example, this may include incorporating a broader range of generalised linear models or functions for using splines in the models.

A final key priority, which would be required for it to be useful to a wider audience, is to establish robust mechanisms for its ongoing maintenance and improvement.

C.8 Demonstration: package vignette

This vignette, a tutorial on how to use the package which comes with the package, demonstrates some of the package's functionality.

epicoda: Introduction and Examples

Contents

1 Loading the package **1**

2 A basic analysis **2**

2.1 Getting started with compositional data . . . 2

2.1.1 Loading simulated data . . . 2

2.1.2 What is compositional data? . . . 2

2.2 Exploratory data analysis . . . 3

2.2.1 The compositional mean . . . 3

2.2.2 Other descriptive statistics . . . 5

2.2.3 Plotting ternary diagrams . . . 5

2.3 Regression modelling . . . 8

2.3.1 Setting up regression models . . . 8

2.3.2 Understanding the model output . . . 9

2.3.3 Presenting the results: tabulate pivot-coordinate coefficients . . . 13

2.3.4 Presenting the results: substitution plots . . . 14

2.3.5 Presenting the results: estimates at particular compositions . . . 15

3 Additional information **19**

3.1 Mechanics of Compositional Data Analysis: Log-ratio transformations . . . 19

3.2 Presenting results via full model predictions . . . 20

3.3 Customising plots . . . 21

3.4 Regression modelling: further examples . . . 24

3.5 Getting the numbers out: estimates for particular compositions . . . 25

References **25**

1 Loading the package

If you have installed the package following the instructions on GitHub, it can be loaded by:

```
library(epicoda)
```

2 A basic analysis

This vignette illustrates some of the things this package can be used to do in an epidemiological analysis with a compositional exposure. More information about any function can be found by running `?function_name` (as usual).

Please note this package is shared as-is and without extensive review/testing. Use is at your own risk (https://github.com/activityMonitoring/epicoda). Bug reports are welcome.

It is highly recommended to deal with missing data both in compositional variables and in covariate or outcome data prior to using the package: functions were designed assuming complete data and may not behave predictably where data is not complete.

2.1 Getting started with compositional data

2.1.1 Loading simulated data

`simdata` is a (simulated) compositional dataset in the package. It has 5 columns which make up a hypothetical movement behaviour composition: "vigorous" (activity), "moderate" (activity), "light" (activity), "sedentary" (behaviour), "sleep." It also contains sex, age group and outcome columns. As the data is simulated, all results are hypothetical.

```
head(simdata)
#>   X    sex        agegroup follow_up_time event BMI disease vigorous moderate
#> 1 1 Female   Second oldest             24  TRUE  29       0   0.0016     0.35
#> 2 2   Male Second youngest             23 FALSE  27       0   0.0044     1.13
#> 3 3 Female          Middle             21 FALSE  31       0   0.0473     1.49
#> 4 4   Male          Middle             17 FALSE  33       0   0.0203     1.82
#> 5 5 Female   Second oldest             14 FALSE  35       0   0.0023     0.97
#> 6 6 Female          Middle             14 FALSE  23       0   0.0213     2.03
#>   light sedentary sleep
#> 1   2.5      11.4   9.8
#> 2   5.5       8.9   8.5
#> 3   2.9      10.1   9.4
#> 4   4.2       9.9   8.1
#> 5   3.7      11.0   8.3
#> 6   4.4       6.0  11.6
```

2.1.2 What is compositional data?

Data is compositional when it is positive-valued and the *relative* (not absolute) values of the variables are relevant to the problem being studied (Filzmoser, Hron, and Templ 2018). Another way to describe this is that compositional data is data that can be expressed as proportions of a whole without losing information.

Here, the compositional variables sum to the same total (24- the number of hours in the day), which is a common situation. As this means it is not possible to simply increase time in one behaviour - it is only possible to spend time in one behaviour rather than in another behaviour - it is the relative values that are relevant.

```
head(apply(simdata[, c("vigorous", "moderate", "light", "sedentary",
    "sleep")], 1, sum))
#> [1] 24 24 24 24 24 24
```

Other examples of compositional data don't have a fixed sum. For example, consider the vote counts in an election for different parties in different constituencies: although the overall number of votes cast might differ between constituencies, it is the proportion of the vote going to each party, and not the total number of votes going to that party in a given constituency, that is relevant.

2.1.2.1 Notes

- The variables in compositional data are often known as *parts*.
- In Compositional Data Analysis, expressing the data as proportions of a whole is known as taking the *closure*, written and defined as:

$$\mathcal{C}(part_1, part_2, ..., part_D) = \left(\frac{part_1}{\sum_{i=1}^{D} part_i}, \frac{part_2}{\sum_{i=1}^{D} part_i}, ..., \frac{part_D}{\sum_{i=1}^{D} part_i}\right) \quad (1)$$

2.2 Exploratory data analysis

2.2.1 The compositional mean

The compositional mean is a frequently-used summary statistic for compositional data (Filzmoser, Hron, and Templ 2018). Like the standard (multivariate) mean it is a measure of location, or central tendency. Unlike the standard mean, it interacts in an appropriate way with compositional data.

To calculate the compositional mean of columns `"vigorous"`, `"moderate"`, `"light"`, `"sedentary"`, `"sleep"` using `epicoda`, we can run:

```
comp_mean(data = simdata, # this is the dataset
          comp_labels = c("vigorous", "moderate", "light", "sedentary", "sleep") # this
           ↪  is the labels of the compositional columns,
          # which we specified above
          )
#> Error in process_zeroes(dCompOnly, comp_labels, rounded_zeroes, det_limit): There are
 ↪  zeroes in the data. rounded_zeroes is TRUE implying zeroes should be imputed, but
 ↪  det_limit has not been set.
#>  The det_limit must be set to perform zero imputation. See the help files or vignette
 ↪  for more information.
#>  NOTE: This represents a change to the default behaviour before 2021-11-26.
 ↪  Previously, the det_limit was guessed based on the data if it was not specified, but
 ↪  this might have unfortunate consequences in some datasets.
```

Unfortunately, it's thrown an error, with a message about zeroes. Compositional data analysis can't automatically include zero values. For more information about this issue, see (Palarea-Albaladejo and Martín-Fernández 2015) and (Lund Rasmussen et al. 2020). There are a couple of options:

- They can be dropped. This is the right option if they're so called "true zeroes," and not related to a limit in the precision with which we can measure.

- They can be imputed. This is the right option if they're so-called "rounded zeroes" i.e. they're related to a limit in the precision with which we can measure. For example, with movement behaviour data, it might be that if we recorded data for long enough, we would eventually see some vigorous activity, but our measurement period wasn't long enough. It might also be that our measurement device can't record values lower than a certain level.

For convenience, many functions in this package let you specify what to do with zeroes as part of the function, and will then perform this preprocessing step as part of the process. They have two arguments to do this: `rounded_zeroes` (default: `TRUE`) and `det_limit` (default: `NULL`).

- Setting `rounded_zeroes = TRUE` and `det_limit` to a numeric value on the same scale as the data imputes zeroes with that detection limit. Under the hood, it uses the `lrEM` algorithm from the `zCompositions` package. Note that currently the same detection limit must apply to all compositional columns. [Note that if the input data scale and the output scale (`units` argument), are different, it should match *the input data scale*.]
- Setting `rounded_zeroes = FALSE` will drop any rows with zeroes (i.e. use the "true zeroes" approach above)
- The default in the function is to run with `rounded_zeroes = TRUE` and `det_limit = NULL`. If there are no zeroes in the data, either as they've already been processed or because there just are no zero values, this will run fine: it doesn't need to do anything to the data. However, if there are zeroes it will throw an error message.

These convenience capabilities should be used with care. You can deal with zero values (i.e. perform zero imputation or drop them) prior to analysis. This has several advantages:

- It can speed up code as these calculations are done once, rather than repeatedly.
- You can inspect the data post-processing of zeroes. For example, you can check that zero imputation has worked as expected. If dropping zeroes, you will know how many rows have been dropped.
- You can do more complex zero imputation than implemented here (e.g. different detection limits for different parts; appropriate methods in the case of data with no complete columns).
- You can use different zero imputation methods (see the `zCompositions` package).

- If you are using compositions with non-constant sum, automated `det_limit` rescaling used in the package will fail. Again, these cases can be dealt with by pre-processing the zeroes using `zCompositions`.

As well as `comp_mean`, some other functions in the package also have `rounded_zeroes` and `det_limit` arguments: they also use calculations which need zeroes to be imputed or excluded. For all of them, it works the same as for `comp_mean`.

Note: Previously, functions in this package "guessed" the detection limit as the minimum non-zero value observed in the data if it wasn't set. This behaviour has been removed as this may have unfortunate consequences in some datasets and should be an explicit analytic decision.

So we can now find the mean again, this time imputing the zeroes with a specified detection limit. In the simulated data, the shortest period of a behaviour we could detect was 0.0012 hr/day (due to the length of an epoch for the accelerometer used to measure the data on which the simulated data is based).

```
comp_mean(data = simdata, # this is the dataset
          comp_labels = c("vigorous", "moderate", "light", "sedentary", "sleep"), # this
           ↪  is the labels of the compositional columns,
          # which we specified above
          rounded_zeroes = TRUE, # this option specifies that we'll treat the zeroes
          # as rounded zeroes i.e. we'll impute them
```

```
          det_limit = 0.0012, # this is the smallest value observable in the data
          )
#> Zeroes were imputed with detection limit
#>  5e-05  (on the unitless scale) using zCompositions::lrEM
#>   vigorous moderate light sedentary sleep
#> 1  0.00037    0.063  0.15      0.42  0.37
```

Even though we started out with sums which added to 24, the mean seems to add to 1. This is because we didn't specify the units - and so the package didn't know what units to return it to us in. But it could still return it as a sum to 1, without loss of information, as per the definition of compositional data.

To change, this we can specify the output units:

```
comp_mean(data = simdata,
          comp_labels = c("vigorous", "moderate", "light", "sedentary", "sleep"),
          rounded_zeroes = TRUE,
          det_limit = 0.0012,
          units = "hr/day" # this is the units. There are pre-specified options "hr/day",
          # "hr/wk", "min/day", "min/wk" and "unitless".
          # If you set units = "specified", you can also specify your own units using
          # specified = c("my_units_name", sum of a composition in these units)
          )
#> Zeroes were imputed with detection limit
#>  5e-05  (on the unitless scale) using zCompositions::lrEM
#>   vigorous moderate light sedentary sleep
#> 1   0.0088      1.5   3.5        10   8.8
```

2.2.2 Other descriptive statistics

While the compositional mean is a fairly clear compositional equivalent to the mean for non-compositional data, other standard descriptive statistics (such as the variance), do not have straightforward compositional equivalents for all purposes for which they are used (there may be several alternatives, depending on the purpose of the analysis) (Filzmoser, Hron, and Templ 2018). These are beyond the current scope of this package.

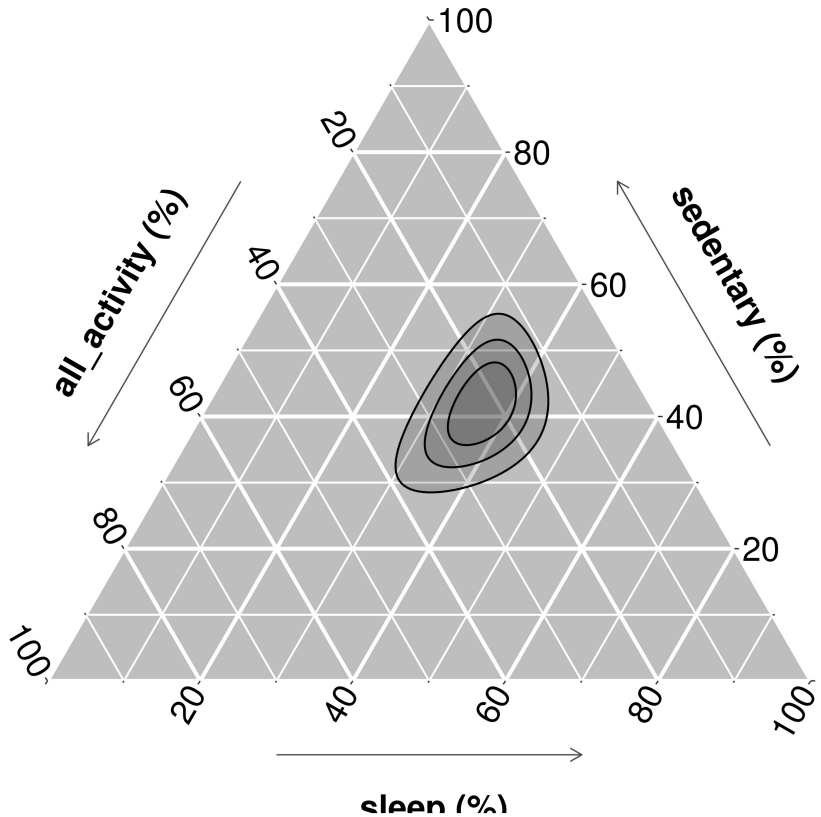
2.2.3 Plotting ternary diagrams

Ternary plots provide a way to plot compositional data with three parts. The R package `ggtern` is an extension of `ggplot2` to carry out this plotting. In this package, there is a wrapper for `ggtern` functions to implement common use cases: plotting the data density, and plotting prediction regions for the data, both overall and within subgroups. (Note that while the "geom_confidence_region" from `ggtern` is used, the regions produced seem to be more properly described as prediction regions)

Note (18.03.2021): A conflict between `ggtern` and `ggplot2` means using `ggtern` can stop axis labels appearing on `ggplot2` graphs. Pending a more permanent solution, these functions remain in the package, but `ggtern` is not automatically loaded and should be used with care.

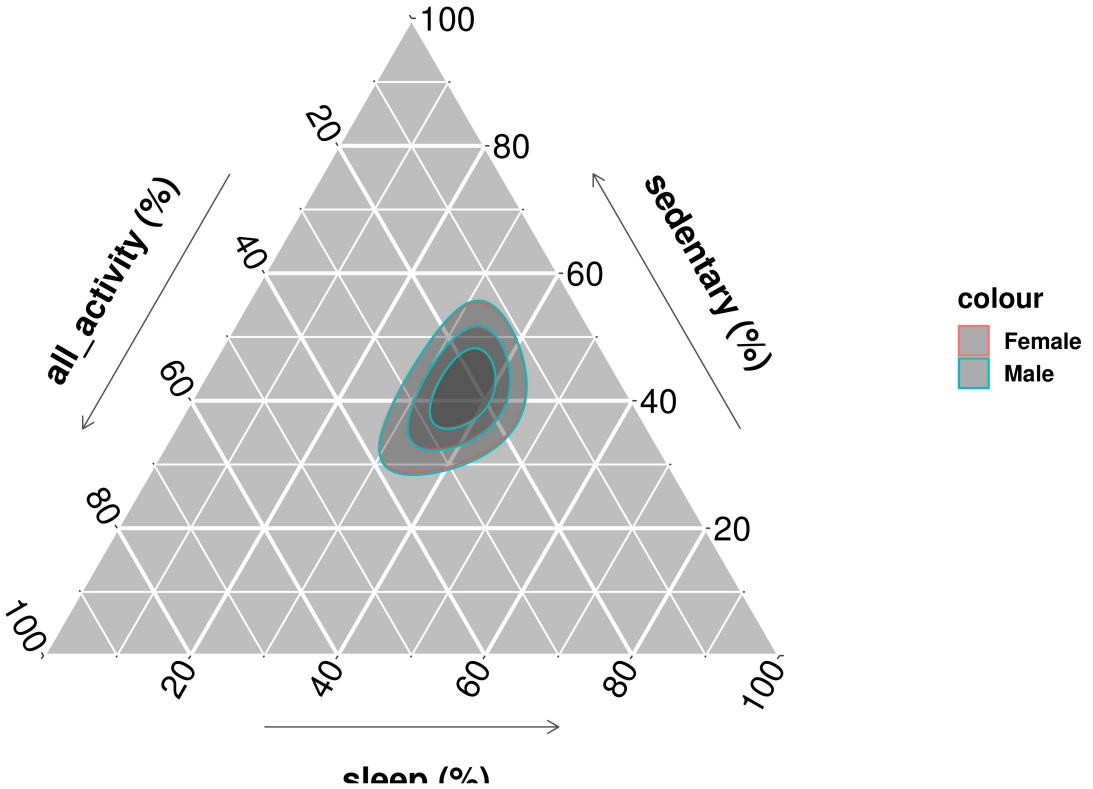
Frequently, data has more than three compositional parts (as here). There are two alternatives in that case: using three parts of the existing parts (looking at the subcomposition formed by these parts), or summing parts to consider a different composition (where that is informative). In the simulated data, `"vigorous"`, `"moderate"` and `"light"` represent three different intensities of activity, so it is meaningful to combine them:

```r
simdata$all_activity <- simdata$vigorous + simdata$moderate +
    simdata$light
plot_prediction_region_ternary(data = simdata, parts_to_plot = c("all_activity",
    "sedentary", "sleep"), probs = c(0.25, 0.5, 0.75)  # This will plot 25, 50 and 75%
     ↪  prediction regions (default is 0.5, 0.9, 0.95)
)
#> Registered S3 methods overwritten by 'ggtern':
#>   method           from
#>   grid.draw.ggplot ggplot2
#>   plot.ggplot      ggplot2
#>   print.ggplot     ggplot2
```



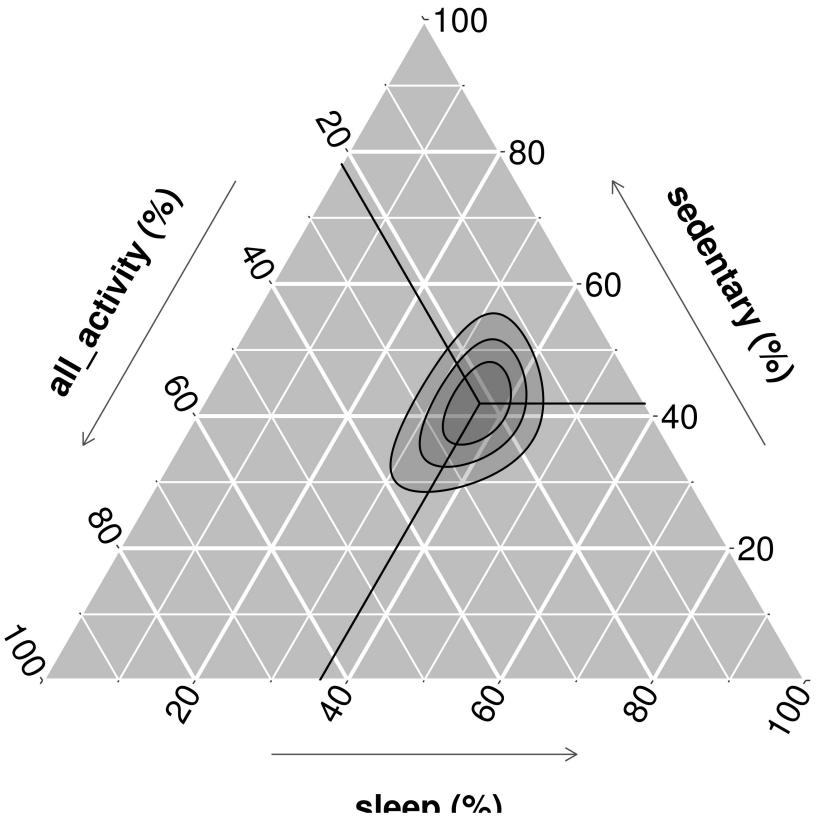
Another frequent use case is considering this plot separated by subgroups. For example, to look at different sexes (which in fact don't differ much in the simulated data):

```r
plot_prediction_region_ternary(data = simdata, parts_to_plot = c("all_activity",
    "sedentary", "sleep"), groups = "sex", probs = c(0.25, 0.5,
    0.75)  # This will plot 25, 50 and 75% prediction regions (default is 0.5, 0.9, 0.95)
)
```



The `mark_points` argument can be used to annotate particular points using crosshairs. For example, to mark the compositional mean:

```
points_to_mark <- as.data.frame(comp_mean(data = simdata, comp_labels = c("all_activity",
    "sedentary", "sleep")))
plot_prediction_region_ternary(data = simdata, parts_to_plot = c("all_activity",
    "sedentary", "sleep"), mark_points = points_to_mark, probs = c(0.25,
    0.5, 0.75)  # This will plot 25, 50 and 75% prediction regions (default is 0.5, 0.9,
    ↪  0.95)
)
```



2.3 Regression modelling

2.3.1 Setting up regression models

The package has a wrapper for linear regression, logistic regression and Cox regression modelling using compositional variables, with adjustment for covariates. Compositional Data Analysis regression modelling works by transforming the variables making up the composition (see "Mechanics of Compositional Data Analysis" below), and using the transformed variables (known as "log-ratio coordinates") in the model. The package automates these steps. For example, to set up a linear regression model for the association between behaviour variables and `BMI`, adjusted for age and sex:

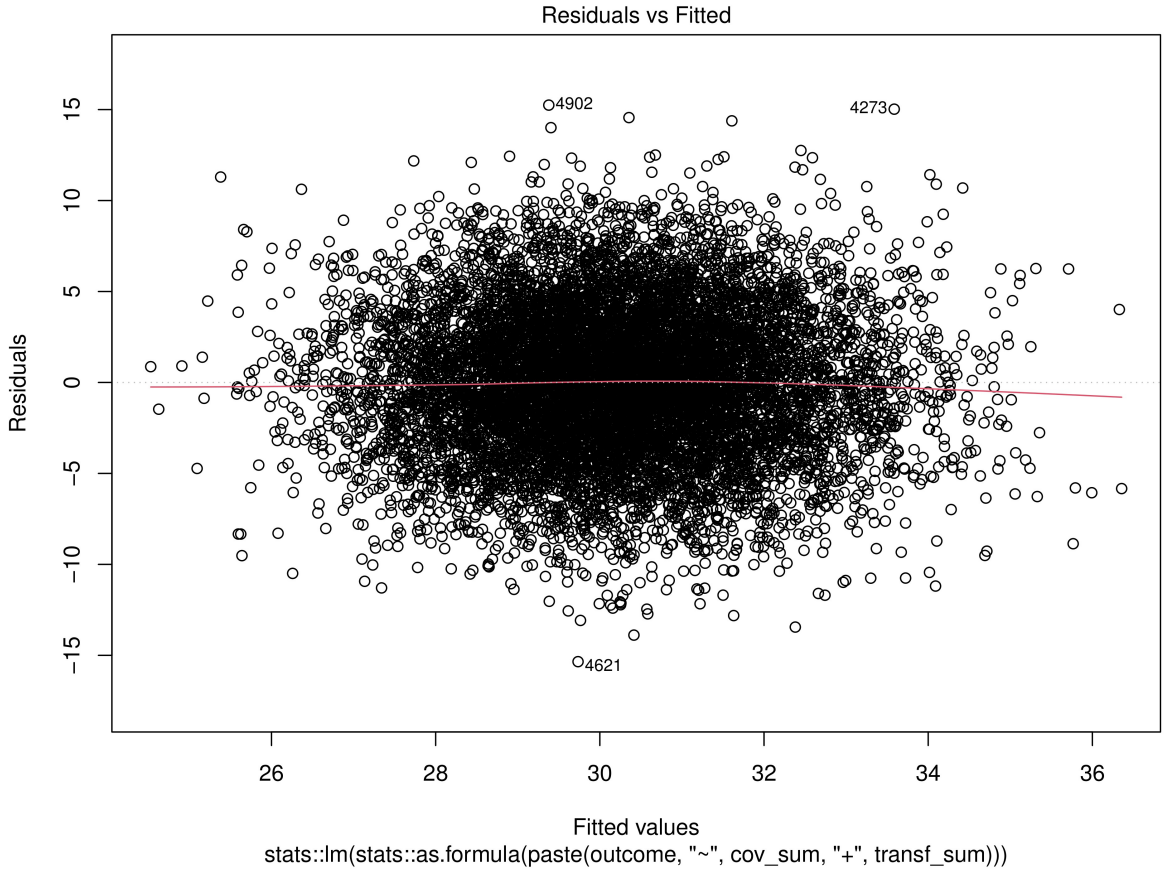
```
lm_BMI <- comp_model(type = "linear", outcome = "BMI", covariates = c("agegroup",
    "sex"), data = simdata, comp_labels = c("vigorous", "moderate",
    "light", "sedentary", "sleep"), det_limit = 0.0012)
#> Zeroes were imputed with detection limit
#>  5e-05  (on the unitless scale) using zCompositions::lrEM
```

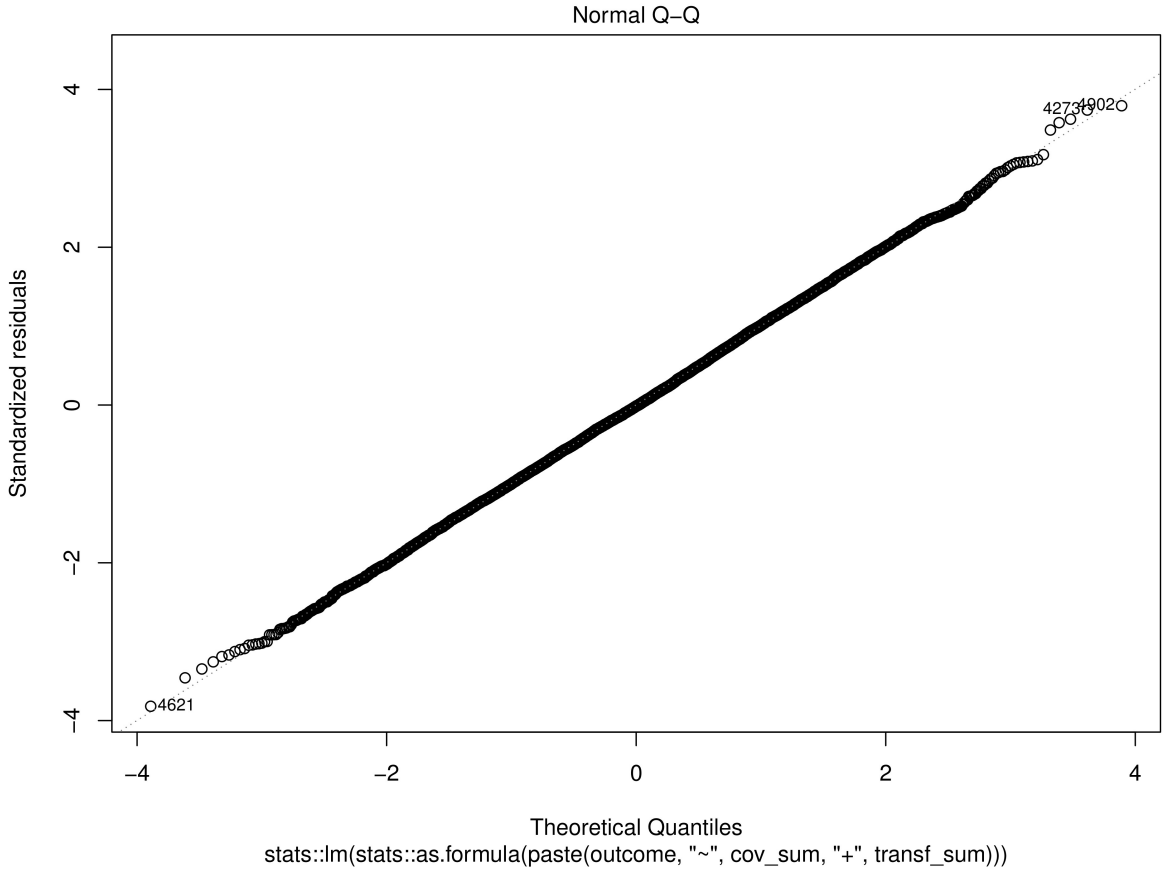
Logistic models can be produced in exactly the same way by setting `type = "logistic"` and using an appropriate binary outcome variable. Cox models can be produced similarly by setting `type = "cox"` and either setting `outcome` as a `Surv` object from the package `survival` or setting the `follow_up_time` and `event` arguments. For examples, see "Regression modelling: further examples."

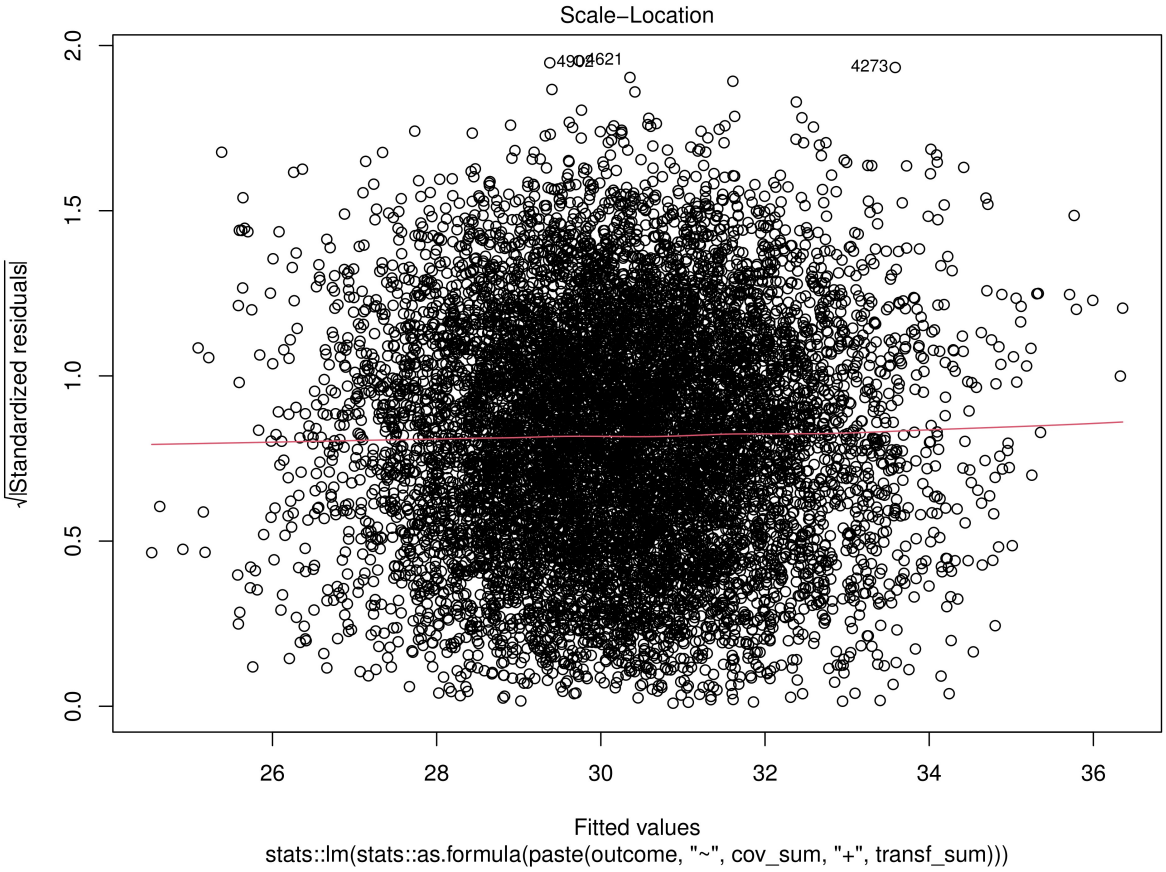
2.3.2 Understanding the model output

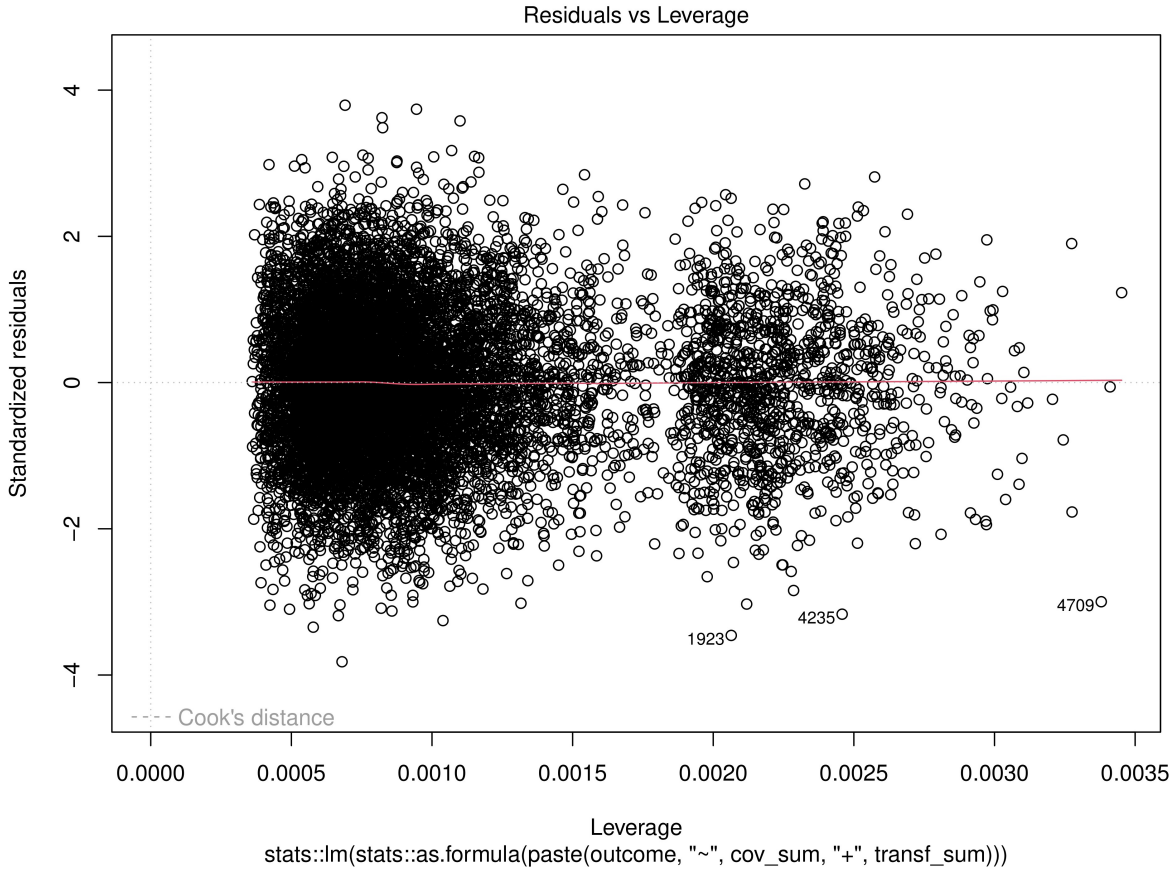
The model object is a standard `lm` object, and we can inspect it in the usual way:

```
plot(lm_BMI)
```




Normal Q–Q
Standardized residuals
Theoretical Quantiles
stats::lm(stats::as.formula(paste(outcome, "~", cov_sum, "+", transf_sum)))
4621
4902
4
2
0
−2
−4


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
4904
4621
4273
26
28
30
32
34
36
Fitted values
stats::lm(stats::as.formula(paste(outcome, "~", cov_sum, "+", transf_sum)))




```
summary(lm_BMI)
#>
#> Call:
#> stats::lm(formula = stats::as.formula(paste(outcome, "~", cov_sum,
#>     "+", transf_sum)), data = data_ready)
#>
#> Residuals:
#>     Min      1Q  Median      3Q     Max
#> -15.347  -2.701  -0.058   2.720  15.246
#>
#> Coefficients:
#>                                Estimate Std. Error t value Pr(>|t|)
#> (Intercept)                     25.6919     0.2207  116.41  < 2e-16 ***
#> agegroupSecond youngest          0.3039     0.1865    1.63  0.10327
#> agegroupMiddle                   0.6979     0.1804    3.87  0.00011 ***
#> agegroupSecond oldest            0.9032     0.1868    4.84  1.3e-06 ***
#> agegroupOldest                   1.0986     0.2384    4.61  4.1e-06 ***
#> sexMale                         -1.0368     0.0804  -12.89  < 2e-16 ***
#> ilr_1_vigorous_vs_parts_2_to_5  -0.3936     0.0271  -14.54  < 2e-16 ***
#> ilr_2_moderate_vs_parts_3_to_5  -1.1253     0.1091  -10.31  < 2e-16 ***
#> ilr_3_light_vs_parts_4_to_5     -0.6550     0.1109   -5.91  3.6e-09 ***
#> ilr_4_sedentary_vs_part_5        2.1267     0.2100   10.13  < 2e-16 ***
```

```
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Residual standard error: 4 on 9990 degrees of freedom
#> Multiple R-squared:  0.132,  Adjusted R-squared:  0.132
#> F-statistic:  169 on 9 and 9990 DF,  p-value: <2e-16
```

The model diagnostics can be interpreted as usual, and in this case all looks fine. However, the summary is hard to interpret because the transformed variables or "coordinates" used in the model do not relate in a straightforward way to the original composition.

2.3.3 Presenting the results: tabulate pivot-coordinate coefficients

A common way to present the results is to present coefficients as usual for covariates, and then produce as many models as there are compositional variables. From each of these models, the coefficient for the first "ilr pivot coordinate" is taken (see "Mechanics of Compositional Data Analysis" below). These ilr pivot coordinate coefficients can be interpreted in terms of reallocation between the particular compositional part and all other parts proportionally. `tab_coefs` is a function which wraps generating one model per compositional variable and outputting a table of model coefficients.

```
tab_coefs(scale_type = "lp", # This argument can be "lp" or "exp" and determines whether
          # coefficients are presented on the scale of the linear predictors ("lp")
          # or are exponentiated ("exp"). Exponentiation gives the Odds Ratio for
          # logistic regression models and the Hazard Ratio for Cox regression models.
          level = 0.95,
          type = "linear",
          outcome = "BMI",
          covariates = c("agegroup", "sex"),
          data = simdata,
          comp_labels = c("vigorous", "moderate", "light", "sedentary", "sleep"),
          det_limit = 0.0012)
#> Zeroes were imputed with detection limit
#>  5e-05  (on the unitless scale) using zCompositions::lrEM
#> Zeroes were imputed with detection limit
#>  5e-05  (on the unitless scale) using zCompositions::lrEM
#> Zeroes were imputed with detection limit
#>  5e-05  (on the unitless scale) using zCompositions::lrEM
#> Zeroes were imputed with detection limit
#>  5e-05  (on the unitless scale) using zCompositions::lrEM
#> Zeroes were imputed with detection limit
#>  5e-05  (on the unitless scale) using zCompositions::lrEM
#> Zeroes were imputed with detection limit
#>  5e-05  (on the unitless scale) using zCompositions::lrEM
#>                                                        fit  2.5 % 97.5 %
#> (Intercept)                                          25.69 25.259 26.125
#> agegroupSecond youngest                               0.30 -0.062  0.669
#> agegroupMiddle                                        0.70  0.344  1.052
#> agegroupSecond oldest                                 0.90  0.537  1.269
#> agegroupOldest                                        1.10  0.631  1.566
#> sexMale                                              -1.04 -1.195 -0.879
#> First pivot coordinate: vigorous vs All other parts  -0.39 -0.447 -0.341
#> First pivot coordinate: moderate vs All other parts  -0.99 -1.207 -0.776
```

```
#> First pivot coordinate: light vs All other parts     -0.14 -0.362  0.089
#> First pivot coordinate: sedentary vs All other parts  2.44  2.132  2.751
#> First pivot coordinate: sleep vs All other parts     -0.92 -1.289 -0.552
```

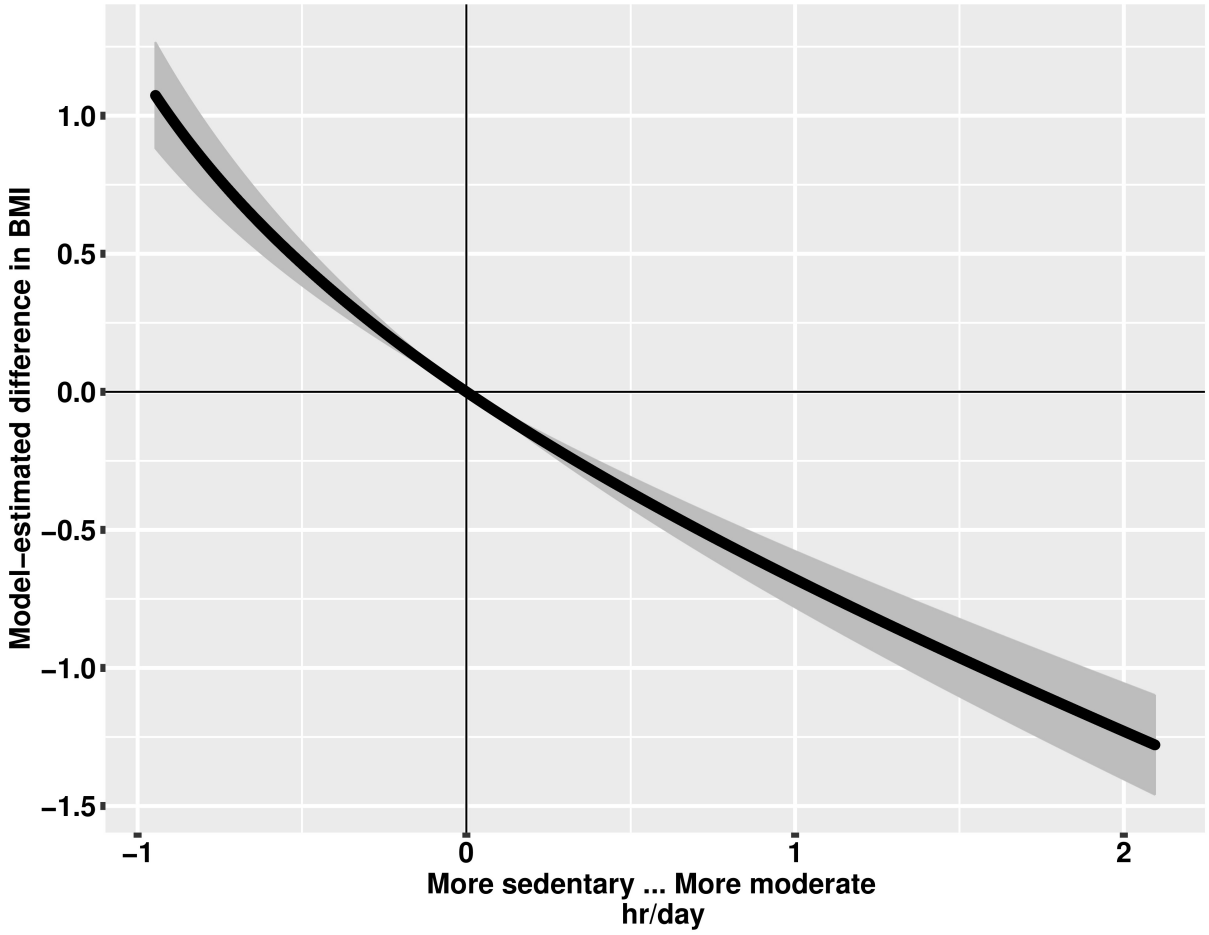
2.3.3.1 Notes

- As these coefficients come from different models they do not parametrise a single model.
- Arguments which would be needed for `comp_model` must be given to this function too, as it regenerates the regression models.
- The function `tab_covariate_coefs` is called within this function, and can be used to extract the coefficients of the covariates only from a model.

2.3.4 Presenting the results: substitution plots

Substitution or reallocation plots show the model-estimated difference in the outcome (relative to at the compositional mean) as the balance between two parts of the composition varies. In the case of time use/movement behaviours, they are sometimes called isotemporal substitution plots. They can be thought of as showing the difference in outcome associated with a hypothetical reallocation of time from one behaviour to another, as estimated by the model. For example:

```
plot_transfers(from_part = "sedentary", to_part = "moderate",
    model = lm_BMI, comp_labels = c("vigorous", "moderate", "light",
        "sedentary", "sleep"), y_label = "Model-estimated difference in BMI",
    units = "hr/day")
#> [1] "Compositional variables not varied in the visualisation were fixed at:"
#> [1] "vigorous : 0.0088 hr/day"
#> [1] "light : 3.5 hr/day"
#> [1] "sleep : 8.8 hr/day"
#> fixed_values will be updated to have compositional parts fixed at the compositional
 ↪  mean. For technical and pragmatic reasons, use of a different reference for the
 ↪  compositional parts is not currently possible.
#> Please note that plotting may take some time.
```



2.3.4.1 Notes

- To customise the appearance of the plot see "Customising plots."
- Results of analyses using a Compositional Data Analysis approach can be presented using reallocations other than 1-1 between two parts. This is not currently implemented in the `plot_transfers` function. However, the `predict_fit_and_ci` function can be used to estimate the outcome for any specified reallocation.
- There is more information on how confidence intervals for this plot are calculated in a vignette (`vignette("derivation_of_CIs_used")`).

2.3.5 Presenting the results: estimates at particular compositions

Another set of estimates that might be of interest are estimates at particular compositions. These can be examined by plotting these in a "forest plot" style plot. This allows assessment of the model-estimated associations with much more complex reallocations between parts.

2.3.5.1 Generating different compositions While there are hugely varied compositions which might be possible, one kind of composition is often of interest. This is compositions where one particular part is increased/decreased at the expense of all other parts proportionally. As this is a common use case, there is a function implemented to do this:

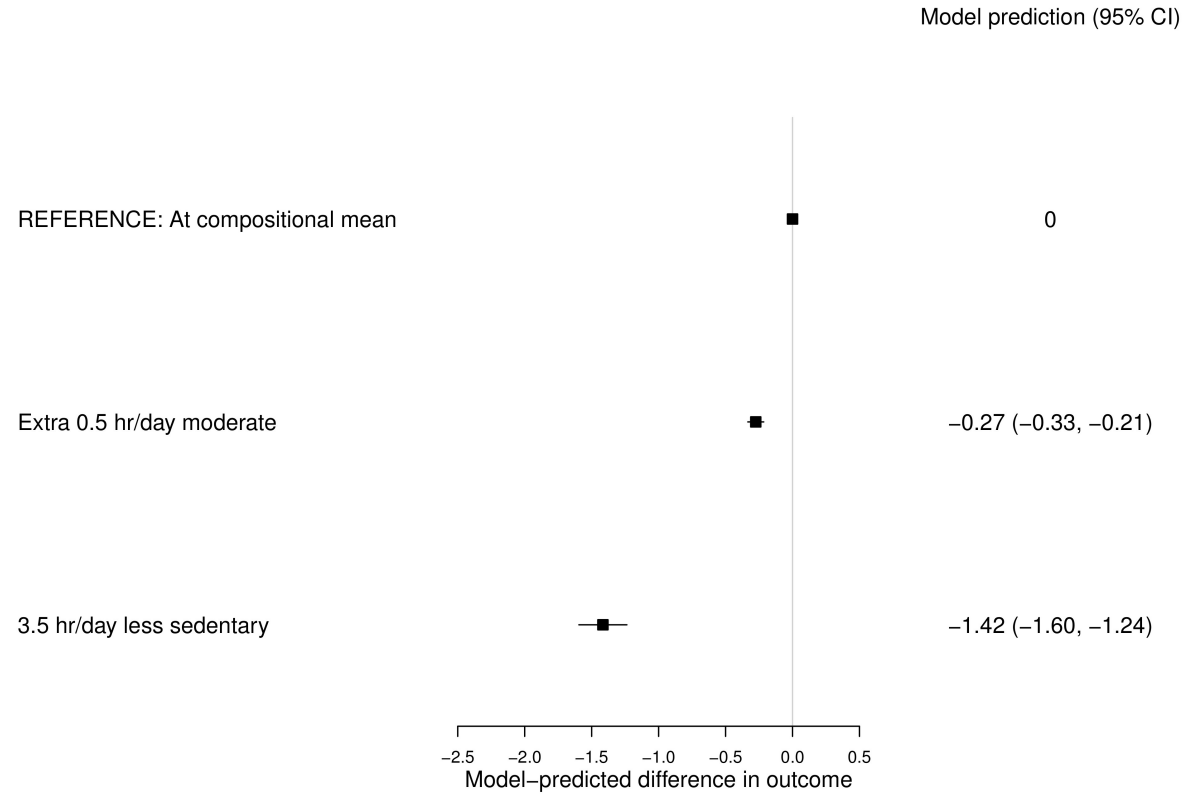
```r
cm_df <- as.data.frame(comp_mean(
  data = simdata,
  comp_labels = c("vigorous", "moderate", "light", "sedentary", "sleep"),
  det_limit = 0.0012,
  units = "hr/day"
  # Note whatever the units are here is the units
  # you should specify the changes in in the next two lines.
))
#> Zeroes were imputed with detection limit
#>  5e-05  (on the unitless scale) using zCompositions::lrEM
new_comp <-
  change_composition(
    composition = cm_df,
    main_part = "moderate",
    main_change = +0.5,
    comp_labels = c("vigorous", "moderate", "light", "sedentary", "sleep")
    )

new_comp2 <-
  change_composition(
    composition = cm_df,
    main_part = "sedentary",
    main_change = -3.5,
    comp_labels = c("vigorous", "moderate", "light", "sedentary", "sleep")
    )

list_for_plot <-
  list("Extra 0.5 hr/day moderate" = new_comp,
       "3.5 hr/day less sedentary" = new_comp2)
print(list_for_plot)
#> $`Extra 0.5 hr/day moderate`
#>   vigorous moderate light sedentary sleep
#> 1   0.0086        2   3.5       9.9   8.6
#>
#> $`3.5 hr/day less sedentary`
#>   vigorous moderate light sedentary sleep
#> 1    0.011      1.9   4.4       6.6    11
```

2.3.5.2 Plotting estimates at particular compositions We now need to plot these, which can be done as follows.

```r
forest_plot_comp(composition_list = list_for_plot, model = lm_BMI,
    comp_labels = c("vigorous", "moderate", "light", "sedentary",
        "sleep"))
```



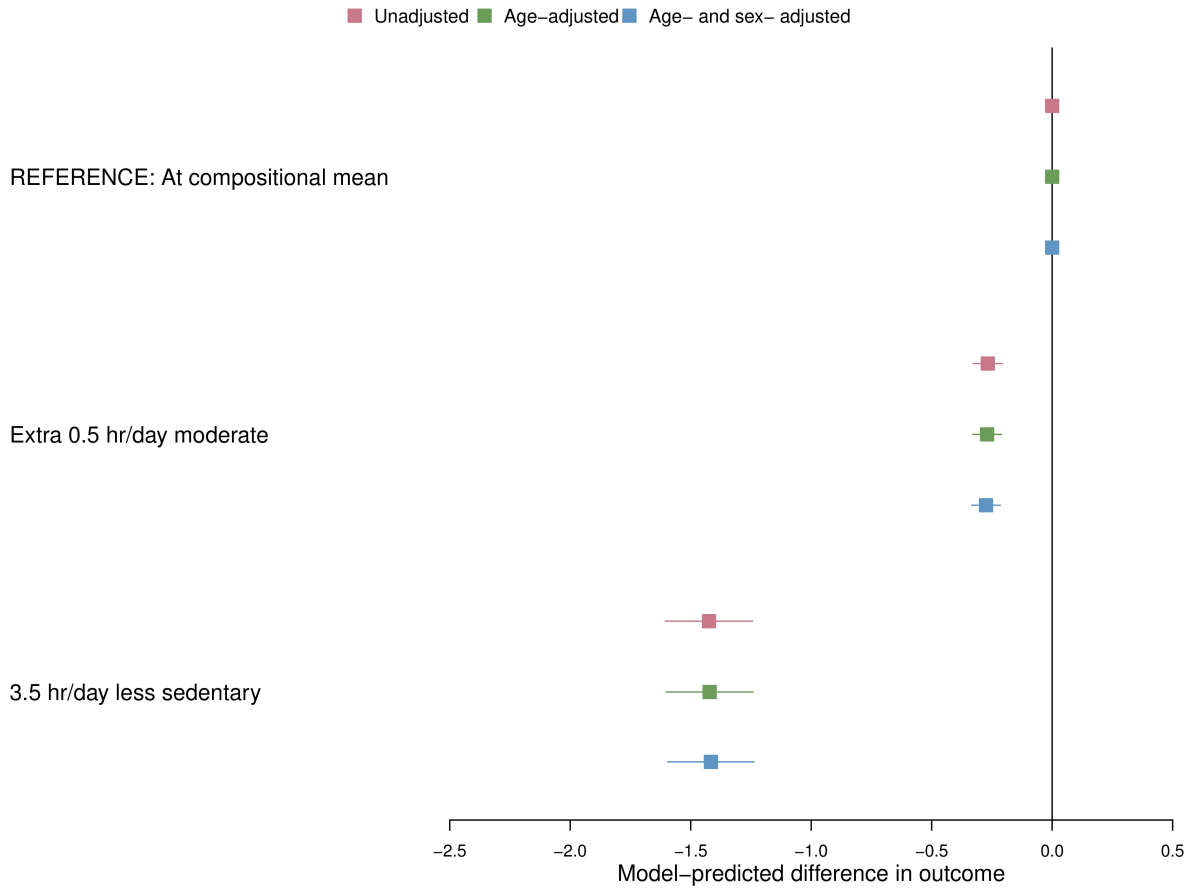
Please note that on the plot with `terms = TRUE` the reference is always the compositional mean *in the dataset used to create the model*. This will agree with the compositional mean calculated above and which we calculated changed compositions relative to if there is no missing data in the dataset used to create the model. To check this, there's a utility function in the package to get the compositional mean directly from the model (although it's a bit ugly):

```
# Set up of the function means it uses this internal from
# the modelling. Hopefully will be streamlined in future...
# :
tl_for_check <- transf_labels(comp_labels = c("vigorous", "moderate",
    "light", "sedentary", "sleep"), transformation_type = "ilr")
# The comp mean calculated from the model:
get_cm_from_model(lm_BMI, comp_labels = c("vigorous", "moderate",
    "light", "sedentary", "sleep"), transf_labels = tl_for_check)$cm
#>   vigorous moderate light sedentary sleep
#> 1  0.00037    0.063  0.15      0.42  0.37
# The comp mean calculated above transformed to unitless
# scale:
cm_df/24
#>   vigorous moderate light sedentary sleep
#> 1  0.00037    0.063  0.15      0.42  0.37
```

The `forest_plot_comp` function can also be used to compare models. To do this, leave `model` at its default

NULL value, and set models_list as a named list of models. For example, to compare lm_BMI with a model adjusted for age_group only and an unadjusted model (very boring in this case as the simulated data doesn't have realistic correlation structure for covariates!):

```
lm_BMI_unadjusted <- comp_model(type = "linear", outcome = "BMI",
    data = simdata, comp_labels = c("vigorous", "moderate", "light",
        "sedentary", "sleep"), det_limit = 0.0012)
#> Zeroes were imputed with detection limit
#>  5e-05  (on the unitless scale) using zCompositions::lrEM
lm_BMI_age_group_only <- comp_model(type = "linear", outcome = "BMI",
    covariates = c("agegroup"), data = simdata, comp_labels = c("vigorous",
        "moderate", "light", "sedentary", "sleep"), det_limit = 0.0012)
#> Zeroes were imputed with detection limit
#>  5e-05  (on the unitless scale) using zCompositions::lrEM
forest_plot_comp(composition_list = list_for_plot, models_list = list(Unadjusted =
 ↪  lm_BMI_unadjusted,
    `Age-adjusted` = lm_BMI_age_group_only, `Age- and sex- adjusted` = lm_BMI),
    comp_labels = c("vigorous", "moderate", "light", "sedentary",
        "sleep"))
```



3 Additional information

This section contains lots of additional information, on:

- Mechanics of Compositional Data Analysis: Log-ratio transformations
- Presenting results via model predictions
- Customising plots
- Regression modelling: further examples
- Getting the numbers out: estimates/ predictions for particular compositions

3.1 Mechanics of Compositional Data Analysis: Log-ratio transformations

The functions above wrap all the "log-ratio mechanics" of using a Compositional Data Analysis approach.

There are several different kinds of transformations which can be used in order to work with compositional data (alr, ilr, clr). These transform the variables into coordinates which are capable of representing all the variation in the data. Different transformations have different pros/cons, and alr and clr can only be used for certain applications. All the functions default to ilr (isometric log ratio) transformation using so-called "pivot coordinates," which is fairly standard in the movement behaviour/time use literature (and is appropriate for the applications which the other transformations are used for). If you want to learn more, see (Filzmoser, Hron, and Templ 2018)(Pawlowsky-Glahn, Egozcue, and Tolosana-Delgado 2011).

The purpose of this package is not to provide this functionality, which is implemented in other R packages, such as `compositions` and `robCompositions`. (The main aim of this package is shown above, in the functions to perform standard epidemiological analyses). This functionality is reimplemented here simply to ensure that future changes to the defaults in those packages do not affect this package.

```
data_ilr_impute_zeroes <- transform_comp(data = simdata, comp_labels = c("vigorous",
    "moderate", "light", "sedentary", "sleep"), transformation_type = "ilr",
    det_limit = 0.0012)
#> Zeroes were imputed with detection limit
#>  5e-05  (on the unitless scale) using zCompositions::lrEM
head(data_ilr_impute_zeroes)
#>   X    sex        agegroup follow_up_time event BMI disease vigorous moderate
#> 1 1 Female   Second oldest            24  TRUE  29       0  6.6e-05    0.015
#> 2 2   Male Second youngest            23 FALSE  27       0  1.8e-04    0.047
#> 3 3 Female          Middle            21 FALSE  31       0  2.0e-03    0.062
#> 4 4   Male          Middle            17 FALSE  33       0  8.4e-04    0.076
#> 5 5 Female   Second oldest            14 FALSE  35       0  9.4e-05    0.040
#> 6 6 Female          Middle            14 FALSE  23       0  8.9e-04    0.085
#>   light sedentary sleep all_activity ilr_1_vigorous_vs_parts_2_to_5
#> 1  0.10      0.48  0.41          2.8                          -6.8
#> 2  0.23      0.37  0.35          6.6                          -6.2
#> 3  0.12      0.42  0.39          4.4                          -4.1
#> 4  0.17      0.41  0.34          6.0                          -4.9
#> 5  0.15      0.46  0.34          4.7                          -6.7
#> 6  0.18      0.25  0.48          6.4                          -4.9
#>   ilr_2_moderate_vs_parts_3_to_5 ilr_3_light_vs_parts_4_to_5
#> 1                           -2.5                       -1.19
#> 2                           -1.6                       -0.37
#> 3                           -1.3                       -0.99
#> 4                           -1.2                       -0.63
#> 5                           -1.7                       -0.77
```

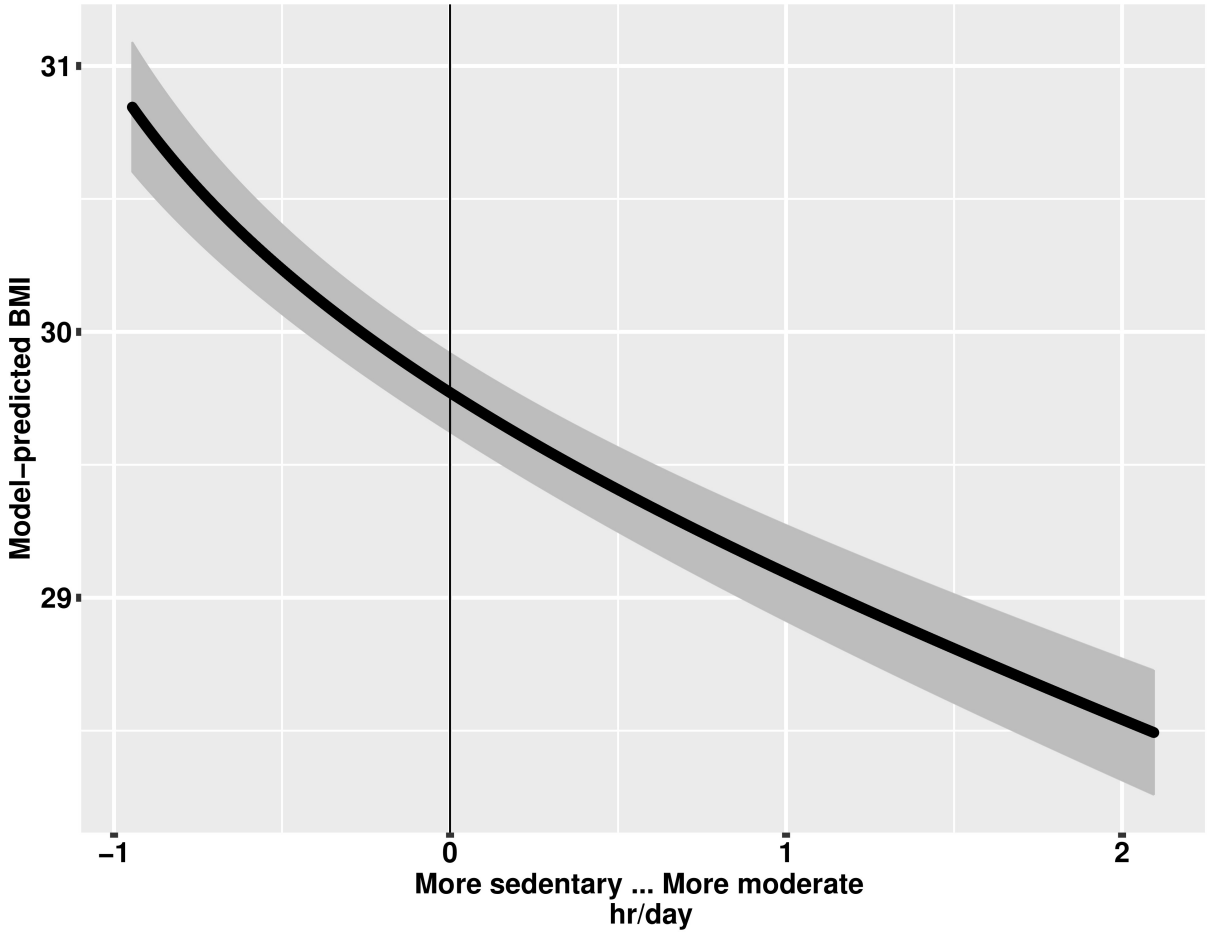
```
#> 6                            -1.0                        -0.53
#>   ilr_4_sedentary_vs_part_5
#> 1                     0.111
#> 2                     0.037
#> 3                     0.048
#> 4                     0.136
#> 5                     0.204
#> 6                    -0.470
```

There are also a few other things we can do with this function, like specifying a given "first part" (which is useful when it comes to reporting, and is used therefore in `tab_coefs`). To learn more, run `?transform_comp`.

3.2 Presenting results via full model predictions

An alternative which may sometimes be useful when interpreting models is to consider model-predicted (mean) outcomes for different input variable values, including values of covariates, rather than differences associated with a particular difference in compositional variables. Functions related to plotting (e.g. the `plot_transfers` function) can do this by setting `terms = FALSE`:

```
plot_transfers(from_part = "sedentary", to_part = "moderate",
    model = lm_BMI, comp_labels = c("vigorous", "moderate", "light",
        "sedentary", "sleep"), y_label = "Model-predicted BMI",
    units = "hr/day", terms = FALSE)
#> [1] "Compositional variables not varied in the visualisation were fixed at:"
#> [1] "vigorous : 0.0088 hr/day"
#> [1] "light : 3.5 hr/day"
#> [1] "sleep : 8.8 hr/day"
#> fixed_values will be updated to have compositional parts fixed at the compositional
 ↪  mean. For technical and pragmatic reasons, use of a different reference for the
 ↪  compositional parts is not currently possible.
#> Note that the confidence intervals on these predictions include uncertainty driven by
 ↪  other, non-compositional variables. To look at compositional variables only, use
 ↪  terms = TRUE
#> [1] "Covariate values were fixed at: "
#> [1] "agegroup : Middle"
#> [1] "sex : Male"
#> Please note that plotting may take some time.
#> Warning: Removed 1 rows containing missing values (geom_hline).
```



3.2.0.1 Notes

- The predictions and confidence intervals on this plot capture the predictions and uncertainty on them due to all variables in the model, including the covariates at the set fixed values.
- However, confidence intervals are still based on confidence intervals for the mean outcome from the model (not prediction intervals).
- The `fixed_values` argument can be used to set the values at which covariates are fixed (which if not set default to mean or modal values for continuous and categorical variables respectively)
- Other functions, such as `forest_plot_comp`, can also output predicted outcomes by setting `terms = FALSE`.

3.3 Customising plots

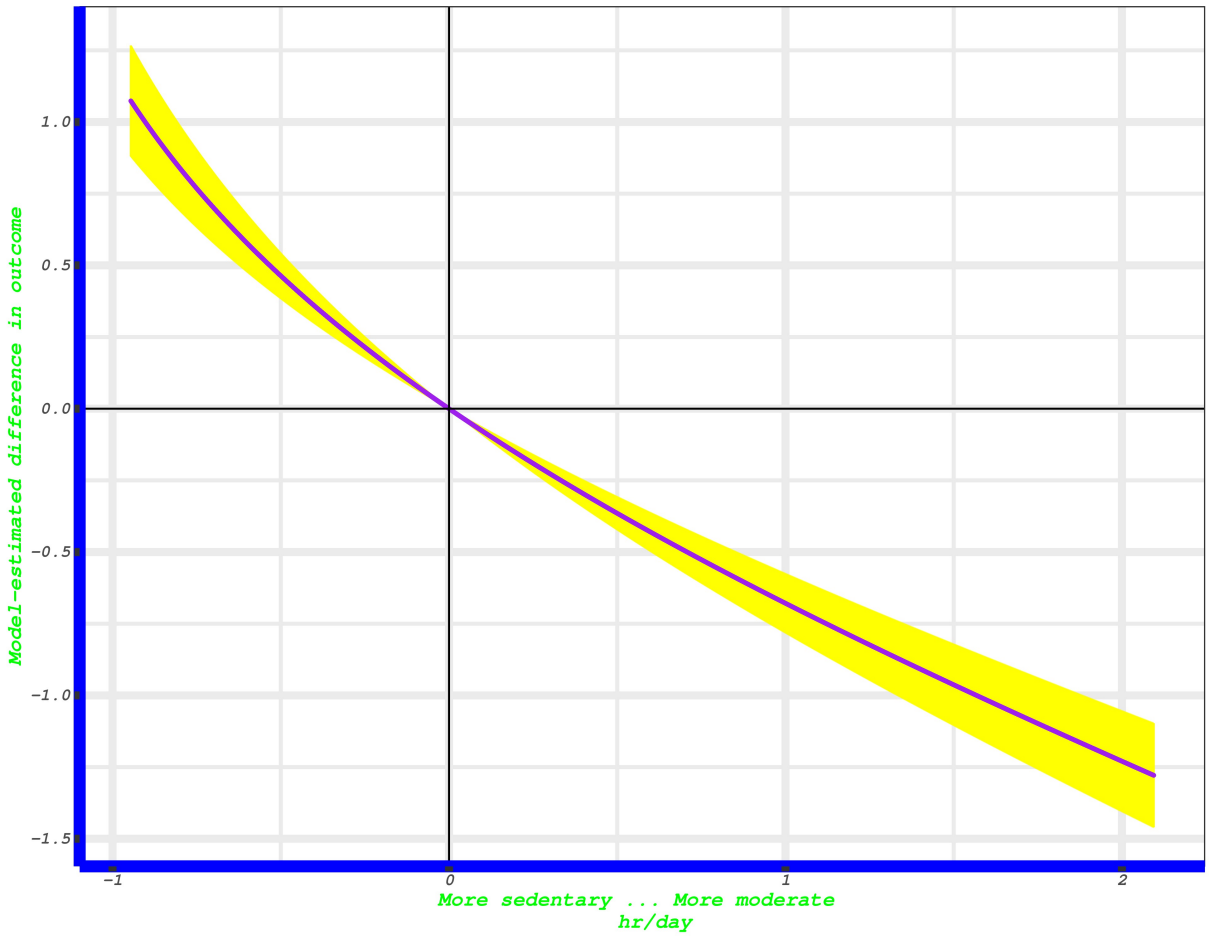
All the plotting functions can have settings changed to make more visually attractive plots (e.g. changing text size, line size, text colour). `plot_transfers` and `simplex_plotting` are based on `ggplot2` and its extension `ggtern`.

- To specify non-data aspects of the plot (e.g. line, axis and text size and colours), set the optional `theme` argument. Note that you will need the `ggplot2` package loaded in order to do this.
- For `plot_transfers`, to specify characteristics of the plotted line, set the `point_specification` argument as a `geom_point` object.

- For `plot_transfers`, to specify colours of the error bars, set the `error_bar_colour` argument as the name of a colour.
- For `plot_transfers`, if the line looks "gappy" (likely at the ends), increase `granularity` (and if it takes too long to plot, decrease `granularity`).

For an extreme example of implementing some of this:

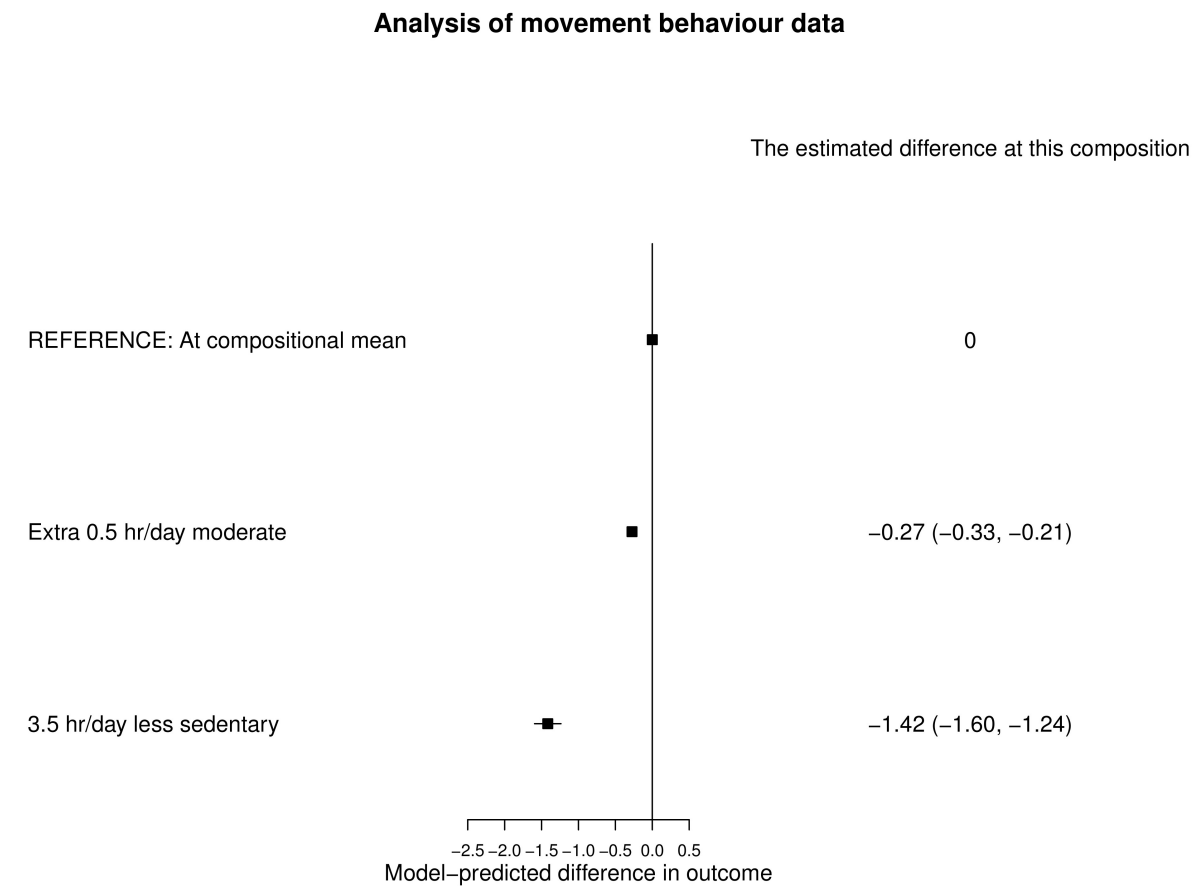
```
library(ggplot2)  # it's easier to specify e.g. themes if ggplot2 is loaded.
plot_transfers(from_part = "sedentary", to_part = "moderate",
    model = lm_BMI, comp_labels = c("vigorous", "moderate", "light",
        "sedentary", "sleep"), y_label = "Model-estimated difference in outcome",
    units = "hr/day", point_specification = geom_point(size = 0.5,
        colour = "purple"), error_bar_colour = "yellow", theme = theme_bw() +
        theme(line = element_line(colour = "red", size = 2),
            text = element_text(family = "mono", face = "bold.italic",
                colour = "green"), axis.line = element_line(colour = "blue",
                size = 3)))
#> [1] "Compositional variables not varied in the visualisation were fixed at:"
#> [1] "vigorous : 0.0088 hr/day"
#> [1] "light : 3.5 hr/day"
#> [1] "sleep : 8.8 hr/day"
#> fixed_values will be updated to have compositional parts fixed at the compositional
 ↪  mean. For technical and pragmatic reasons, use of a different reference for the
 ↪  compositional parts is not currently possible.
#> Please note that plotting may take some time.
```



Slightly differently, `forest_plot_comp` is based on the `forestplot` package, so takes arguments specifying the plot in the same way as that function. There are several arguments which set particular aspects of the forest plot. In addition, any argument that the basic `forestplot` function can take, which isn't already set, can be set as an argument to `forest_plot_comp`.

```
library(forestplot) # it's easier to specify options such as fpColors if the forestplot
 ↪  package is loaded.
#> Loading required package: grid
#> Loading required package: magrittr
#>
#> Attaching package: 'magrittr'
#> The following objects are masked from 'package:testthat':
#>
#>     equals, is_less_than, not
#> Loading required package: checkmate
forest_plot_comp(
  composition_list = list_for_plot,
  model = lm_BMI,
  comp_labels = c("vigorous", "moderate", "light", "sedentary", "sleep"),
  pred_name = "The estimated difference at this composition",
  # This sets the name with which to label the column. ,
  title = "Analysis of movement behaviour data",
  # This sets the overall plot title.
```

```
  col = forestplot::fpColors(zero = "black") # This sets the colour of the line at 0 (or
   ↪  1, if it's the plot of a logistic or Cox model)
)
```



For additional information on how to customise the functions, see the manual by running e.g. `?plot_transfers`.

3.4 Regression modelling: further examples

As well as linear models, we can produce logistic and Cox models:

```
log_outcome <- comp_model(type = "logistic", outcome = "disease",
    covariates = c("agegroup", "sex"), data = simdata, comp_labels = c("vigorous",
        "moderate", "light", "sedentary", "sleep"), det_limit = 0.0012)
#> Zeroes were imputed with detection limit
#>  5e-05  (on the unitless scale) using zCompositions::lrEM
cox_outcome <- comp_model(type = "cox", event = "event", follow_up_time =
 ↪  "follow_up_time",
    covariates = c("agegroup", "sex"), data = simdata, comp_labels = c("vigorous",
        "moderate", "light", "sedentary", "sleep"), det_limit = 0.0012)
#> Zeroes were imputed with detection limit
#>  5e-05  (on the unitless scale) using zCompositions::lrEM
```

3.5 Getting the numbers out: estimates for particular compositions

To get the numbers out at particular compositions, we can use the `predict_fit_and_ci` function (which is wrapped in the plotting functions). This takes a data frame argument for `new_data`. It returns a data frame with the fit and confidence interval appended (in addition to some intermediate values used in calculating the estimate).

```
predict_fit_and_ci(model = lm_BMI, new_data = new_comp, comp_labels = c("vigorous",
    "moderate", "light", "sedentary", "sleep"))
#>   vigorous moderate light sedentary sleep BMI agegroup  sex
#> 1  0.00036    0.083  0.14      0.41  0.36  30   Middle Male
#>   ilr_1_vigorous_vs_parts_2_to_5 ilr_2_moderate_vs_parts_3_to_5
#> 1                          -5.7                             -1
#>   ilr_3_light_vs_parts_4_to_5 ilr_4_sedentary_vs_part_5
#> 1                        -0.8                       0.1
#>   fit.ilr_1_vigorous_vs_parts_2_to_5 fit.ilr_2_moderate_vs_parts_3_to_5
#> 1                              0.027                               -0.3
#>   fit.ilr_3_light_vs_parts_4_to_5 fit.ilr_4_sedentary_vs_part_5
#> 1                               0                      2.1e-16
#>   se.fit.ilr_1_vigorous_vs_parts_2_to_5 se.fit.ilr_2_moderate_vs_parts_3_to_5
#> 1                                0.0019                                0.029
#>   se.fit.ilr_3_light_vs_parts_4_to_5 se.fit.ilr_4_sedentary_vs_part_5   df
#> 1                                  0                          2e-17 9990
#>   residual.scale   fit lower_CI upper_CI
#> 1              4 -0.27    -0.33    -0.21
```

As noted above, this can be used to look at any reallocations between parts which are of interest.

References

Filzmoser, P., K. Hron, and M. Templ. 2018. *Applied Compositional Data Analysis.* https://doi.org/10.1007/978-3-319-96422-5.

Lund Rasmussen, C., J. Palarea-Albaladejo, M. S. Johansson, P. Crowley, M. L. Stevens, N. Gupta, K. Karstad, and A. Holtermann. 2020. "Zero problems with compositional data of physical behaviors: a comparison of three zero replacement methods." *International Journal of Behavioral Nutrition and Physical Activity.* https://doi.org/10.1186/s12966-020-01029-z.

Palarea-Albaladejo, J., and J. A. Martín-Fernández. 2015. "ZCompositions - R package for multivariate imputation of left-censored data under a compositional approach." https://doi.org/10.1016/j.chemolab.2015.02.019.

Pawlowsky-Glahn, V., J. J. Egozcue, and R. Tolosana-Delgado. 2011. "Lecture notes on Compositional Data Analysis," no. May: 1–100. https://doi.org/10.13140/RG.2.1.1296.2728.

Appendix D: Within-person variation in compositional data

D.1 Within-person variation in compositional data

The challenges presented by within-person variation in compositional data are perhaps most clearly illustrated in the case where data is closed, that is, all summing to the same number or expressed as proportions, as in device-measured 24-hour movement behaviour data closed to 24 hours. Nonetheless, the approach developed still applies in the context of compositional data not in this form (which, by definition of compositional data, should be expressible in this form without loss of relevant information).

Suppose time in different behaviours with true values $(X_1, \ldots, X_D)$ is measured with error or within-person variation as

$$(X_1^*, \ldots, X_D^*) = (X_1, \ldots, X_D) + (U_1, \ldots, U_D) \tag{D.1}$$

where $(U_1, \ldots, U_D)$ is an error term.

Both $(X_1^*, \ldots, X_D^*)$ and $(X_1, \ldots, X_D)$ are vectors of time in different 24-hour movement behaviours, so are subject to the constant sum constraint (for convenience, they

will be assumed throughout to be closed to 1, but the development is general). This means $(U_1, \ldots, U_D)$ must sum to 0:

$$\begin{aligned} U_1 + \cdots + U_D &= (X_1^* - X_1) + \cdots + (X_D^* - X_D) \\ &= (X_1^* + \cdots + X_D^*) - (X_1 + \cdots + X_D) \\ &= 1 - 1 \\ &= 0. \end{aligned}$$

A good measurement error model would seem obliged to account for this constraint on the possible distribution of $(U_1, \ldots, U_D)$.

D.2 Equivalence with perturbation on the simplex

Here, it will be shown that additive measurement error respecting the constant sum constraint (i.e. $(U_1, \ldots, U_D)$) is equivalent to perturbation using the Aitchison geometry on the simplex (as introduced in Section 5.2). This will be shown in two parts. First, it will be shown that arbitrary additive measurement error respecting the constant sum constraint can be written as perturbation by a positive-valued vector ($\Rightarrow$). Second, it will be shown that an arbitrary perturbation by a positive-valued vector can be written as additive measurement error respecting the constant sum constraint ($\Leftarrow$).

($\Rightarrow$) Equation (D.1) may be written using the perturbation operation as:

$$\begin{aligned}
(X_1^*, \ldots, X_D^*) &= (X_1, \ldots, X_D) + (U_1, \ldots, U_D) \\
&= (X_1 + U_1, \ldots, X_D + U_D) \\
&= \left(X_1\Big(\frac{X_1 + U_1}{X_1}\Big), \ldots, X_D\Big(\frac{X_D + U_D}{X_D}\Big) \right) \\
&= \mathcal{C}\left(X_1\Big(\frac{X_1 + U_1}{X_1}\Big), \ldots, X_D\Big(\frac{X_D + U_D}{X_D}\Big) \right) \\
&= (X_1, \ldots, X_D) \oplus \left(\frac{X_1 + U_1}{X_1}, \ldots, \frac{X_D + U_D}{X_D} \right).
\end{aligned}$$

As each X_i is positive and $X_i + U_i = X_i^*$ is positive, $\left(\frac{X_1+U_1}{X_1}, \ldots, \frac{X_D+U_D}{X_D} \right)$ is a positive-valued vector.

($\Leftarrow$) Suppose an arbitrary positive-valued vector, $(\alpha_1, \ldots, \alpha_D)$. It will be shown that perturbation by this vector can be written as additive measurement error respecting the constant sum constraint. Writing $S = X_1\alpha_1 + \cdots + X_D\alpha_D$, we have:

$$\begin{aligned}
(X_1^*, \ldots, X_D^*) &= (X_1, \ldots, X_D) \oplus (\alpha_1, \ldots, \alpha_D) \\
&= \mathcal{C}\left(X_1\alpha_1, \ldots, X_D\alpha_D \right) \\
&= \left(\frac{X_1\alpha_1}{S}, \ldots, \frac{X_D\alpha_D}{S} \right) \\
&= \left(X_1\left(\frac{\alpha_1}{S}\right), \ldots, X_D\left(\frac{\alpha_D}{S}\right) \right) \\
&= \left(X_1\left(1 + \frac{(\alpha_1 - S)}{S}\right), \ldots, X_D\left(1 + \frac{(\alpha_D - S)}{S}\right) \right) \\
&= \left(X_1 + \frac{X_1(\alpha_1 - S)}{S}, \ldots, X_D + \frac{X_D(\alpha_D - S)}{S} \right) \\
&= \left(X_1, \ldots, X_D \right) + \left(\frac{X_1(\alpha_1 - S)}{S}, \ldots, \frac{X_D(\alpha_D - S)}{S} \right).
\end{aligned}$$

Then, considering the error term:

$$\begin{aligned}\frac{X_1(\alpha_1 - S)}{S} + \cdots + \frac{X_D(\alpha_D - S)}{S} &= \frac{X_1\alpha_1 - X_1S + \cdots + X_D\alpha_D - X_DS}{S} \\ &= \frac{X_1\alpha_1 + \cdots + X_D\alpha_D}{S} - \frac{X_1S + \cdots + X_DS}{S} \\ &= 1 - (X_1 + \cdots + X_D) \\ &= 1 - 1 \\ &= 0.\end{aligned}$$

That is, perturbation by this arbitrary positive-valued vector can be rewritten as additive measurement error respecting the constant sum constraint.

D.3 Consequences when working with isometric log-ratio coordinates

The last section showed an equivalence between measurement error respecting the constant sum constraint and perturbation on the simplex. In other words, a natural way to describe measurement error or within-person variation for compositional data is via perturbation on the simplex: the measured value is the true value perturbed by the measurement error/ within-person variation vector.

The isometric isomorphism between the simplex with the Aitchison geometry and Euclidean space (introduced in Section 5.2) means that if measurement error is represented by perturbation on the simplex, after isometric log-ratio transformation it will be represented straightforwardly by the addition of an unconstrained measurement error vector to the true value. This suggests that, within the CoDA regression framework, it will be natural to work with a measurement error model for within-person

variation based around unconstrained additive measurement error in the isometric log-ratio coordinates. Working with isometric log-ratio coordinates will therefore allow standard epidemiological approaches for additive measurement error in regression modelling to be applied. This is the basis of the approach in Chapter 6.

References

1. Walmsley, R. *et al.* Reallocation of Time between Device-Measured Movement Behaviours and Risk of Incident Cardiovascular Disease. *British Journal of Sports Medicine.* ISSN: 0306-3674, 1473-0480. pmid: 34489241. https://bjsm.bmj.com/content/early/2021/09/06/bjsports-2021-104050 (2021).
2. *Epicoda* 2020. https://github.com/OxWearables/epicoda.
3. *Ukb_download_and_prep_template* 2020. https://github.com/OxWearables/ukb_download_and_prep_template.
4. *biobankAccelerometerAnalysis* https://github.com/OxWearables/biobankAccelerometerAnalysis.
5. Labarthe, D. R. Cardiovascular Health and Disease. *Handbook of Epidemiology,* 2121–2157. https://ezproxy-prd.bodleian.ox.ac.uk:2120/referencework entry/10.1007/978-0-387-09834-0_35 (2014).
6. Danaei, G. & Rahimi, K. *Coronary Heart Disease: Epidemiology and Prevention* 3603–3616. https://ezproxy-prd.bodleian.ox.ac.uk:2989/view/10.1093/med/9780198746690.001.0001/med-9780198746690-chapter-365 (2022) (Oxford University Press).
7. *Cerebrovascular Disease – Classifications, Symptoms, Diagnosis and Treatments* https://www.aans.org/ (2021).
8. NHS. *Stroke* nhs.uk. https://www.nhs.uk/conditions/stroke/ (2021).
9. Feigin, V. L. *et al.* Global, Regional, and National Burden of Stroke and Its Risk Factors, 1990–2019: A Systematic Analysis for the Global Burden of Disease Study 2019. *The Lancet Neurology* **20,** 795–820. ISSN: 1474-4422, 1474-4465. pmid: 34487721. https://www.thelancet.com/journals/laneur/article/PIIS1474-4422(21)00252-0/fulltext (2021).
10. Roth, G. A. *et al.* Global Burden of Cardiovascular Diseases and Risk Factors, 1990–2019: Update From the GBD 2019 Study. *Journal of the American College of Cardiology* **76,** 2982–3021. ISSN: 0735-1097. https://www.sciencedirect.com/science/article/pii/S0735109720377755 (2020).
11. GBD 2019 Diseases and Injuries Collaborators. Ischaemic Heart Disease — Level 3 Cause. *The Lancet* **396** (2020).

12. GBD 2019 Diseases and Injuries Collaborators. Stroke — Level 3 Cause. *The Lancet* **396** (2020).

13. Bhatnagar, P., Wickramasinghe, K., Wilkins, E. & Townsend, N. Trends in the Epidemiology of Cardiovascular Disease in the UK. *Heart* **102,** 1945–1952. ISSN: 1355-6037, 1468-201X. pmid: 27550425. https://heart.bmj.com/content/102/24/1945 (2016).

14. Smolina, K., Wright, F. L., Rayner, M. & Goldacre, M. J. Determinants of the Decline in Mortality from Acute Myocardial Infarction in England between 2002 and 2010: Linked National Database Study. *BMJ* **344,** d8059. ISSN: 0959-8138, 1468-5833. pmid: 22279113. https://www.bmj.com/content/344/bmj.d8059 (2012).

15. Office for National Statistics. *Leading Causes of Death, UK* (2020). https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/causesofdeath/articles/leadingcausesofdeathuk/latest (2022).

16. Public Health England. A Review of Recent Trends in Mortality in England, 86 (2018).

17. Mensah, G. A. *et al.* Decline in Cardiovascular Mortality: Possible Causes and Implications. *Circulation Research* **120,** 366–380. ISSN: 1524-4571. pmid: 28104770 (2017).

18. Raleigh, V. S. *Trends in Life Expectancy in EU and Other OECD Countries: Why Are Improvements Slowing?* (OECD, Paris, 2019). https://www.oecd-ilibrary.org/social-issues-migration-health/trends-in-life-expectancy-in-eu-and-other-oecd-countries_223159ab-en (2021).

19. NICE-The National Institute for Health and Care Excellence. *Cardiovascular Disease Prevention* NICE impact (2021), 21.

20. NICE-The National Institute for Health and Care Excellence. *BNF: British National Formulary - NICE* https://bnf.nice.org.uk/treatment-summary/cardiovascular-disease-risk-assessment-and-prevention.html (2021).

21. World Health Organization. *WHO Report on the Global Tobacco Epidemic 2021: Addressing New and Emerging Products* (World Health Organization, Geneva, 2021).

22. World Health Organization. *WHO Guidelines on Physical Activity and Sedentary Behaviour* (World Health Organization, Geneva, 2020).

23. Caspersen, C. J. & Christenson, G. M. Physical Activity , Exercise , and Physical Fitness : Definitions and Distinctions for Health-Related Research (April 1985).

24. Bouchard, C., Blair, S. N. & Haskell, W. L. *Physical Activity and Health* 796 pp. Google Books: tO96DwAAQBAJ (Human Kinetics, 2012).

25. Schmid, D. & Leitzmann, M. F. Physical Activity Epidemiology. *Handbook of Epidemiology,* 1927–2002. https://ezproxy-prd.bodleian.ox.ac.uk:2120/referenceworkentry/10.1007/978-0-387-09834-0_67 (2014).

26. Paluch, A. E. *et al.* in *Physical Activity and Public Health Practice* (eds Ainsworth, B. E. & Macera, C. A.) (Taylor & Francis Group, 2012).

27. Holtermann, A., Straker, L., Lee, I.-M., van der Beek, A. J. & Stamatakis, E. Long Overdue Remarriage for Better Physical Activity Advice for All: Bringing Together the Public Health and Occupational Health Agendas. *British Journal of Sports Medicine* **54,** 1377–1378. ISSN: 0306-3674, 1473-0480. pmid: 32376672. https://bjsm.bmj.com/content/54/23/1377 (2020).

28. Ainsworth, B. E. *et al.* 2011 Compendium of Physical Activities: A Second Update of Codes and MET Values. *Medicine and Science in Sports and Exercise* **43,** 1575–1581. ISSN: 01959131 (2011).

29. Strain, T. *et al.* Levels of Domain-Specific Physical Activity at Work, in the Household, for Travel and for Leisure among 327 789 Adults from 104 Countries. *British Journal of Sports Medicine* **54,** 1488–1497. ISSN: 14730480. pmid: 33239355. http://bjsm.bmj.com/ (2020).

30. Dumuid, D., Olds, T. & Sawyer, S. M. Moving beyond More: Towards a Healthy Balance of Daily Behaviours. *The Lancet* **0.** ISSN: 0140-6736, 1474-547X. https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(21)01600-7/abstract (2021).

31. Tremblay, M. S., Esliger, D. W., Tremblay, A. & Colley, R. Incidental Movement, Lifestyle-Embedded Activity and Sleep: New Frontiers in Physical Activity Assessment, 11 (2021).

32. Blair, S. N. *et al.* A Tribute to Professor Jeremiah Morris: The Man Who Invented the Field of Physical Activity Epidemiology. *Annals of Epidemiology* **20,** 651–660. ISSN: 1047-2797. https://www.sciencedirect.com/science/article/pii/S1047279710001250 (2010).

33. Morris, J., Heady, J., Raffle, P., Roberts, C. & Parks, J. Coronary Heart-Disease and Physical Activity of Work. ISSN: 0140-6736. pmid: 13110075 (1953).

34. Morris, J. N. & Heady, J. A. Mortality in Relation to the Physical Activity of Work: A Preliminary Note on Experience in Middle Age. *British journal of industrial medicine* **10,** 245–254. ISSN: 00071072 (1953).

35. Heady, M., Morris, J. & Raffle, P. Physique of London Busmen: Epidemiology of Uniforms. *The Lancet* **268,** 569–570. ISSN: 0140-6736. https://ezproxy-prd.bodleian.ox.ac.uk:2056/science/article/pii/S0140673656920499 (1956).

36. Paffenbarger, R. S., Blair, S. N. & Lee, I. M. A History of Physical Activity, Cardiovascular Health and Longevity: The Scientific Contributions of Jeremy N Morris, DSc, DPH, FRCP. *International Journal of Epidemiology* **30,** 1184–1192. ISSN: 03005771. pmid: 11689543. https://academic.oup.com/ije/article/30/5/1184/724212 (2001).

37. Morris, J. N., Pollard, R., Everitt, M. G., Chave, S. P. W. & Semmence, A. M. Vigorous Exercise in Leisure Time: Protection against Coronary Heart Disease. *The Lancet* **316,** 1207–1210. ISSN: 0140-6736, 1474-547X. https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(80)92476-9/abstract (1980).

38. Armstrong, M. E. G. *et al.* Reported Frequency of Physical Activity in a Large Epidemiological Study: Relationship to Specific Activities and Repeatability over Time. *BMC medical research methodology* **11,** 97. ISSN: 1471-2288. pmid: 21831330 (2011).

39. Bennett, D. A. *et al.* Association of Physical Activity With Risk of Major Cardiovascular Diseases in Chinese Men and Women. *JAMA Cardiology,* 1–10. ISSN: 2380-6583. pmid: 29117341. http://cardiology.jamanetwork.com/article.aspx?doi=10.1001/jamacardio.2017.4069 (2017).

40. Manson, J. E. *et al.* A Prospective Study of Walking as Compared with Vigorous Exercise in the Prevention of Coronary Heart Disease in Women. *The New England Journal of Medicine* **341,** 650–658. ISSN: 0028-4793. pmid: 10460816 (1999).

41. Kraus, W. E. *et al.* Physical Activity, All-Cause and Cardiovascular Mortality, and Cardiovascular Disease. *Medicine and Science in Sports and Exercise* **51,** 1270–1281. ISSN: 1530-0315. pmid: 31095084 (2019).

42. Department of Health & Human Services & 2018 Physical Activity Guidelines Advisory Committee. *2018 Physical Activity Guidelines Advisory Committee Scientific Report* (U.S. Department of Health and Human Services, 2018, Washington, DC, 2018), 779. pmid: 19178654.

43. Bahls, M., Baurecht, H., Hanssen, H. & Van Craenenbroeck, E. M. How to Establish Causality between Physical Inactivity and Mortality? *European Journal of Preventive Cardiology* **29,** e266–e267. ISSN: 2047-4873. https://doi.org/10.1093/eurjpc/zwac021 (2022).

44. Mora, S., Cook, N., Buring, J. E., Ridker, P. M. & Lee, I.-M. Physical Activity and Reduced Risk of Cardiovascular Events: Potential Mediating Mechanisms. *Circulation* **116,** 2110–2118. ISSN: 1524-4539. pmid: 17967770 (2007).

45. Hamer, M. *et al.* Physical Activity and Inflammatory Markers over 10 Years: Follow-up in Men and Women from the Whitehall II Cohort Study. *Circulation.* ISSN: 00097322. pmid: 22891048 (2012).

46. Fiuza-Luces, C. *et al.* Exercise Benefits in Cardiovascular Disease: Beyond Attenuation of Traditional Risk Factors. *Nature Reviews Cardiology* **15,** 731–743. ISSN: 17595010. pmid: 30115967. http://dx.doi.org/10.1038/s41569-018-0065-1 (2018).

47. Moreira, J. B., Wohlwend, M. & Wisløff, U. Exercise and Cardiac Health: Physiological and Molecular Insights. *Nature Metabolism.* ISSN: 25225812. pmid: 32807982 (2020).

48. Febbraio, M. A. Health Benefits of Exercise - More than Meets the Eye! *Nature Reviews Endocrinology.* ISSN: 17595037. pmid: 28051119 (2017).

49. Cornelissen, V. A. & Smart, N. A. Exercise Training for Blood Pressure: A Systematic Review and Meta-analysis. *Journal of the American Heart Association* **2,** e004473. https://www.ahajournals.org/doi/10.1161/jaha.112.004473.

50. Public Health England. *Health Matters: Physical Activity - Prevention and Management of Long-Term Conditions* GOV.UK. https://www.gov.uk/government/publications/health-matters-physical-activity/health-matters-physical-activity-prevention-and-management-of-long-term-conditions (2021).

51. Guthold, R., Stevens, G. A., Riley, L. M. & Bull, F. C. Worldwide Trends in Insufficient Physical Activity from 2001 to 2016: A Pooled Analysis of 358 Population-Based Surveys with 1·9 Million Participants. *The Lancet Global Health* **6,** e1077–e1086. ISSN: 2214109X. pmid: 30193830 (2018).

52. Lee, I. M. *et al.* Effect of Physical Inactivity on Major Non-Communicable Diseases Worldwide: An Analysis of Burden of Disease and Life Expectancy. *The Lancet* **380,** 219–229. ISSN: 1474547X. pmid: 22818936 (2012).

53. U.S. Department of Health and Human Services. *Physical Activity and Health: A Report of the Surgeon General* (1996). https://www.cdc.gov/nccdphp/sgr/.

54. *Better Health* nhs.uk. https://www.nhs.uk/better-health/ (2021).

55. *Couch to 5K App Hits 5 Million Downloads* GOV.UK. https://www.gov.uk/government/news/couch-to-5k-app-hits-5-million-downloads (2021).

56. Dunstan, D. W., Howard, B., Healy, G. N. & Owen, N. Too Much Sitting - A Health Hazard. *Diabetes Research and Clinical Practice* **97,** 368–376. ISSN: 01688227. pmid: 22682948. http://dx.doi.org/10.1016/j.diabres.2012.05.020 (2012).

57. Lund Rasmussen, C. *et al.* Light-Intensity Physical Activity Derived from Count or Activity Types Is Differently Associated with Adiposity Markers. *Scandinavian Journal of Medicine and Science in Sports* **30,** 1966–1975. ISSN: 16000838. pmid: 32516858. /pmc/articles/PMC7540429/ (2020).

58. Tremblay, M. S. *et al.* Sedentary Behavior Research Network (SBRN) - Terminology Consensus Project Process and Outcome. *International Journal of Behavioral Nutrition and Physical Activity* **14,** 1–17. ISSN: 14795868 (2017).

59. Carskadon, M. A. & Dement, W. C. in *Principles and Practice of Sleep Medicine* 15–24.e3 (Elsevier, 2017). https://linkinghub.elsevier.com/retrieve/pii/B9780323242882000027 (2021).

60. Falck, R. S., Davis, J. C., Li, L., Stamatakis, E. & Liu-Ambrose, T. Preventing the '24-Hour Babel': The Need for a Consensus on a Consistent Terminology Scheme for Physical Activity, Sedentary Behaviour and Sleep. *British Journal of Sports Medicine* **56,** 367–368. ISSN: 0306-3674, 1473-0480. pmid: 34556466. https://bjsm.bmj.com/content/56/7/367 (2022).

61. Bauman, A. E., Petersen, C. B., Blond, K., Rangul, V. & Hardy, L. L. in *Sedentary Behaviour Epidemiology* (eds Leitzmann, M. F., Jochem, C. & Schmid, D.) 73–106 (Springer International Publishing, Cham, 2018). http://link.springer.com/10.1007/978-3-319-61552-3_4 (2021).

62. Healy, G. N. in *Physical Activity and Public Health Practice* (eds Ainsworth, B. E. & Macera, C. A.) (Taylor & Francis Group, 2012).

63. López-Valenciano, A. *et al.* Changes in Sedentary Behaviour in European Union Adults between 2002 and 2017. *BMC Public Health* **20,** 1206. ISSN: 1471-2458. https://doi.org/10.1186/s12889-020-09293-1 (2020).

64. Stamatakis, E., de Rezende, L. F. & Rey-Lopez, J. P. in *Sedentary Behaviour Epidemiology* 215–244 (2018).

65. Chastin, S. F. *et al.* How Does Light-Intensity Physical Activity Associate with Adult Cardiometabolic Health and Mortality? Systematic Review with Meta-Analysis of Experimental and Observational Studies. *British Journal of Sports Medicine* **53,** 370–376. ISSN: 14730480. pmid: 29695511 (2019).

66. Tremblay, M. S. & Ross, R. How Should We Move for Health? The Case for the 24-Hour Movement Paradigm. *CMAJ* **192,** E1728–E1729. ISSN: 0820-3946, 1488-2329. pmid: 33288508. https://www.cmaj.ca/content/192/49/E1728 (2020).

67. Martínez-González, M. A., Martínez, J. A., Hu, F. B., Gibney, M. J. & Kearney, J. Physical Inactivity, Sedentary Lifestyle and Obesity in the European Union. *International Journal of Obesity and Related Metabolic Disorders: Journal of the International Association for the Study of Obesity* **23,** 1192–1201. pmid: 10578210 (1999).

68. Dempsey, P. C. *et al.* New Global Guidelines on Sedentary Behaviour and Health for Adults: Broadening the Behavioural Targets. *International Journal of Behavioral Nutrition and Physical Activity* **17,** 1–12. ISSN: 14795868. pmid: 33239026. https://doi.org/10.1186/s12966-020-01044-0 (2020).

69. LaCroix, A. Z. *et al.* Association of Light Physical Activity Measured by Accelerometry and Incidence of Coronary Heart Disease and Cardiovascular Disease in Older Women. *JAMA Network Open* **2,** e190419. ISSN: 2574-3805. http://jamanetworkopen.jamanetwork.com/article.aspx?doi=10.1001/jamanetworkopen.2019.0419 (2019).

70. Dempsey, P. C. *et al.* Prospective Associations of Accelerometer-Measured Physical Activity and Sedentary Time With Incident Cardiovascular Disease, Cancer, and All-Cause Mortality. *Circulation* **141,** 1113–1115. https://www.ahajournals.org/doi/full/10.1161/CIRCULATIONAHA.119.043030 (2020).

71. Jefferis, B. J. *et al.* Does Total Volume of Physical Activity Matter More than Pattern for Onset of CVD? A Prospective Cohort Study of Older British Men. *International journal of cardiology* **278,** 267–272. ISSN: 1874-1754 (gqw, 8200291 2019).

72. Qiu, S. *et al.* Does Objectively-Measured Light-Intensity Physical Activity Reduce the Risk of Cardiovascular Mortality? A Meta-Analysis. *European heart journal. Quality of care & clinical outcomes.* ISSN: 2058-1742 (2020).

73. Itani, O., Jike, M., Watanabe, N. & Kaneita, Y. Short Sleep Duration and Health Outcomes: A Systematic Review, Meta-Analysis, and Meta-Regression. *Sleep Medicine Reviews* **32,** 246–256. ISSN: 15322955. http://dx.doi.org/10.1016/j.sleep.2016.08.006 (2016).

74. Jike, M., Itani, O., Watanabe, N., Buysse, D. J. & Kaneita, Y. Long Sleep Duration and Health Outcomes: A Systematic Review, Meta-Analysis and Meta-Regression. *Sleep Medicine Reviews* **39,** 25–36. ISSN: 15322955 (2018).

75. Stamatakis, E. *et al.* Is the Time Right for Quantitative Public Health Guidelines on Sitting? A Narrative Review of Sedentary Behaviour Research Paradigms and Findings. *British Journal of Sports Medicine* **91,** 399–404 (2018).

76. Yerramalla, M. S. *et al.* Association of Daily Composition of Physical Activity and Sedentary Behaviour with Incidence of Cardiovascular Disease in Older Adults. *International Journal of Behavioral Nutrition and Physical Activity* **18,** 83. ISSN: 1479-5868. https://doi.org/10.1186/s12966-021-01157-0 (2021).

77. Amagasa, S. *et al.* Is Objectively Measured Light-Intensity Physical Activity Associated with Health Outcomes after Adjustment for Moderate-to-Vigorous Physical Activity in Adults? A Systematic Review. *International Journal of Behavioral Nutrition and Physical Activity* **15,** 65. ISSN: 1479-5868. https://doi.org/10.1186/s12966-018-0695-z (2018).

78. Lee, I.-M. & Shiroma, E. J. Using Accelerometers to Measure Physical Activity in Large-Scale Epidemiological Studies: Issues and Challenges. *British Journal of Sports Medicine* **48,** 197–201. ISSN: 0306-3674. pmid: 24297837. http://bjsm.bmj.com/lookup/doi/10.1136/bjsports-2013-093154 (2014).

79. Prince, S. *et al.* A Comparison of Direct versus Self-Report Measures for Assessing Physical Activity in Adults: A Systematic Review. *International Journal of Behavioral Nutrition and Physical Activity* **5.** ISSN: 02533820. pmid: 18990237 (2008).

80. Sabia, S. *et al.* Association between Questionnaire-and Accelerometer-Assessed Physical Activity: The Role of Sociodemographic Factors. *American Journal of Epidemiology* **179,** 781–790. ISSN: 14766256. pmid: 24500862 (2014).

81. Ferrari, P., Friedenreich, C. & Matthews, C. E. The Role of Measurement Error in Estimating Levels of Physical Activity. *American Journal of Epidemiology* **166,** 832–840. ISSN: 00029262. pmid: 17670910 (2007).

82. Ekelund, U., Dalene, K. E., Tarp, J. & Lee, I.-M. Physical Activity and Mortality: What Is the Dose Response and How Big Is the Effect? *British Journal of Sports Medicine* **54,** 1125–1126. ISSN: 0306-3674, 1473-0480. pmid: 31964630. https://bjsm.bmj.com/content/54/19/1125 (2020).

83. Grøntved, A. & Hu, F. B. Television Viewing and Risk of Type 2 Diabetes, Cardiovascular Disease, and All-Cause Mortality: A Meta-analysis. *JAMA* **305,** 2448–2455. ISSN: 0098-7484. https://doi.org/10.1001/jama.2011.812 (2011).

84. Gore, S. A., Foster, J. A., DiLillo, V. G., Kirk, K. & Smith West, D. Television Viewing and Snacking. *Eating Behaviors* **4,** 399–405. ISSN: 1471-0153. https://www.sciencedirect.com/science/article/pii/S1471015303000539 (2003).

85. Ekelund, U. *et al.* Dose-Response Associations between Accelerometry Measured Physical Activity and Sedentary Time and All Cause Mortality: Systematic Review and Harmonised Meta-Analysis. *Bmj* **366,** l4570. ISSN: 0959-8138. http://www.bmj.com/lookup/doi/10.1136/bmj.l4570 (2019).

86. Craig, C. L. *et al.* International Physical Activity Questionnaire: 12-Country Reliability and Validity. *Medicine and Science in Sports and Exercise.* ISSN: 01959131. pmid: 12900694 (2003).

87. Armstrong, T. & Bull, F. Development of the World Health Organization Global Physical Activity Questionnaire (GPAQ). *Journal of Public Health* **14,** 66–70. ISSN: 09431853. https://link.springer.com/article/10.1007/s10389-006-0024-x (2006).

88. Chen, K. Y. & Bassett, D. R. The Technology of Accelerometry-Based Activity Monitors: Current and Future. *Medicine and Science in Sports and Exercise* **37,** S490–500. ISSN: 0195-9131. pmid: 16294112 (11 Suppl 2005).

89. Troiano, R. P., McClain, J. J., Brychta, R. J. & Chen, K. Y. Evolution of Accelerometer Methods for Physical Activity Research. *British Journal of Sports Medicine* **48,** 1019–1023. ISSN: 14730480. pmid: 24782483. /pmc/articles/PMC4141534/ (2014).

90. Arem, H., Keadle, S. K. & Matthews, C. E. Invited Commentary: Meta-Physical Activity and the Search for the Truth. *American Journal of Epidemiology* **181,** 656–658. ISSN: 0002-9262. https://doi.org/10.1093/aje/kwu472 (2015).

91. Colbert, L. H., Matthews, C. E., Havighurst, T. C., Kim, K. & Schoeller, D. A. Comparative Validity of Physical Activity Measures in Older Adults. *Medicine and Science in Sports and Exercise* **43,** 867–876. ISSN: 1530-0315. pmid: 20881882 (2011).

92. Møller, N. C., Kristensen, P. L., Wedderkopp, N., Andersen, L. B. & Froberg, K. Objectively Measured Habitual Physical Activity in 1997/1998 vs 2003/2004 in Danish Children: The European Youth Heart Study. *Scandinavian Journal of Medicine & Science in Sports* **19,** 19–29. ISSN: 1600-0838. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0838.2008.00774.x (2009).

93. Doherty, A. *et al.* Large Scale Population Assessment of Physical Activity Using Wrist Worn Accelerometers: The UK Biobank Study. *PLoS ONE* **12,** 1–14. ISSN: 19326203. pmid: 28146576 (2017).

94. Burchartz, A. *et al.* Assessing Physical Behavior through Accelerometry – State of the Science, Best Practices and Future Directions. *Psychology of Sport and Exercise* **49,** 101703. ISSN: 14690292 (April 2020).

95. Troiano, R. P. *et al.* Physical Activity in the United States Measured by Accelerometer. *Medicine and Science in Sports and Exercise* **40,** 181–188. ISSN: 01959131. pmid: 18091006. https://pubmed.ncbi.nlm.nih.gov/18091006/ (2008).

96. *The Objective Monitoring of Physical Activity: Contributions of Accelerometry to Epidemiology, Exercise Science and Rehabilitation* (eds Shephard, R. J. & Tudor-Locke, C.) http://link.springer.com/10.1007/978-3-319-29577-0 (2021) (Springer International Publishing, Cham, 2016).

97. Evenson, K. R., Wen, F. & Herring, A. H. Associations of Accelerometry-Assessed and Self-Reported Physical Activity and Sedentary Behavior with All-Cause and Cardiovascular Mortality among US Adults. *American Journal of Epidemiology.* ISSN: 14766256. pmid: 27760774 (2016).

98. *DAPA Measurement Toolkit* https://dapa-toolkit.mrc.ac.uk/physical-activity/objective-methods/accelerometers (2021).

99. *AX3* http://www.axivity.com/product/1 (2021).

100. Rowlands, A. V. *et al.* Accelerometer-Assessed Physical Activity in Epidemiology: Are Monitors Equivalent? *Medicine & Science in Sports & Exercise.* ISSN: 0195-9131. pmid: 28976493 (2017).

101. Wijndaele, K. *et al.* Utilization and Harmonization of Adult Accelerometry Data: Review and Expert Consensus. *Medicine and Science in Sports and Exercise* **47,** 2129–2139. ISSN: 0195-9131. pmid: 25785929. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4731236/ (2015).

102. Hart, T. L., Swartz, A. M., Cashin, S. E. & Strath, S. J. How Many Days of Monitoring Predict Physical Activity and Sedentary Behaviour in Older Adults? *International Journal of Behavioral Nutrition and Physical Activity* **8,** 62. ISSN: 1479-5868. https://doi.org/10.1186/1479-5868-8-62 (2011).

103. Matthews, C. E., Ainsworth, B. E., Thompson, R. W. & Bassett, D. R. J. Sources of Variance in Daily Physical Activity Levels as Measured by an Accelerometer. *Medicine & Science in Sports & Exercise* **34,** 1376–1381. ISSN: 0195-9131. https://journals.lww.com/acsm-msse/Fulltext/2002/08000/Sources_of_variance_in_daily_physical_activity.21.aspx (2002).

104. Kerr, J. *et al.* Objective Assessment of Physical Activity: Classifiers for Public Health. *Med Sci Sports Exerc.* **48,** 951–957. ISSN: 1530-0315. pmid: 27089222 (2017).

105. Thompson, D., Batterham, A. M., Peacock, O. J., Western, M. J. & Booso, R. Feedback from Physical Activity Monitors Is Not Compatible with Current Recommendations: A Recalibration Study. *Preventive Medicine* **91,** 389–394. ISSN: 10960260 (2016).

106. Pedišić, Ž. & Bauman, A. Accelerometer-Based Measures in Physical Activity Surveillance: Current Practices and Issues. *British Journal of Sports Medicine.* ISSN: 14730480. pmid: 25370153 (2015).

107. Migueles, J. H. *et al.* Comparability of Published Cut-Points for the Assessment of Physical Activity: Implications for Data Harmonization. *Scandinavian Journal of Medicine and Science in Sports* **29,** 566–574. ISSN: 16000838. pmid: 30548545 (2019).

108. Farrahi, V., Niemelä, M., Kangas, M., Korpelainen, R. & Jämsä, T. Calibration and Validation of Accelerometer-Based Activity Monitors: A Systematic Review of Machine-Learning Approaches. *Gait and Posture* **68,** 285–299. ISSN: 18792219. pmid: 30579037. https://pubmed.ncbi.nlm.nih.gov/30579037/ (2019).

109. Montoye, A. H. K., Mudd, L. M., Biswas, S. & Pfeiffer, K. A. Energy Expenditure Prediction Using Raw Accelerometer Data in Simulated Free Living. *Medicine & Science in Sports & Exercise* **47,** 1735–1746. ISSN: 0195-9131. https://journals.lww.com/acsm-msse/Fulltext/2015/08000/Energy_Expenditure_Prediction_Using_Raw.23.aspx (2015).

110. Montoye, A. H., Begum, M., Henning, Z. & Pfeiffer, K. A. Comparison of Linear and Non-Linear Models for Predicting Energy Expenditure from Raw Accelerometer Data. *Physiological Measurement* **38,** 343–357. ISSN: 13616579. pmid: 28107205. https://iopscience.iop.org/article/10.1088/1361-6579/38/2/343 (2017).

111. Montoye, A. H., Westgate, B. S., Fonley, M. R. & Pfeiffer, K. A. Cross-Validation and out-of-Sample Testing of Physical Activity Intensity Predictions with a Wrist-Worn Accelerometer. *Journal of Applied Physiology* **124,** 1284–1293. ISSN: 15221601. pmid: 29369742. http://www.jappl.org (2018).

112. Bastian, T. *et al.* Automatic Identification of Physical Activity Types and Sedentary Behaviors from Triaxial Accelerometer: Laboratory-Based Calibrations Are Not Enough. *Journal of Applied Physiology* **118,** 716–722. ISSN: 8750-7587. pmid: 25593289. https://www.physiology.org/doi/10.1152/japplphysiol.01189.2013 (2015).

113. Narayanan, A., Desai, F., Stewart, T., Duncan, S. & Mackay, L. Application of Raw Accelerometer Data and Machine-Learning Techniques to Characterize Human Movement Behavior: A Systematic Scoping Review. *Journal of physical activity & health* **17,** 360–383. ISSN: 1543-5474 (Electronic). pmid: 32035416 (2020).

114. Willetts, M., Hollowell, S., Aslett, L., Holmes, C. & Doherty, A. Statistical Machine Learning of Sleep and Physical Activity Phenotypes from Sensor Data in 96,220 UK Biobank Participants. *Scientific Reports* **8.** ISSN: 20452322. pmid: 29784928 (2018).

115. Clevenger, K. A. *et al. A Systematic Review and Repository of Novel Methods for Estimating Physical Activity and Energy Expenditure from Accelerometer Data* in. International Conference on Ambulatory Monitoring of Physical Activity and Movement (Online, 2021), 1.

116. Dohrn, I.-M., Sjostrom, M., Kwak, L., Oja, P. & Hagstromer, M. Accelerometer-Measured Sedentary Time and Physical Activity-A 15 Year Follow-up of Mortality in a Swedish Population-Based Cohort. *Journal of science and medicine in sport. Erratum in: J Sci Med Sport. 2019 Mar;22(3):246; PMID: 30509864 [https://www.ncbi.nlm.nih.gov/pubmed/30509864]* **21,** 702–707. ISSN: 1878-1861 (2018).

117. Bellettiere, J. *et al.* Sedentary Behavior and Cardiovascular Disease in Older Women: The OPACH Study. *Circulation* **139,** 1036–1046. ISSN: 15244539. pmid: 31031411 (2019).

118. Ramakrishnan, R. *et al.* Accelerometer Measured Physical Activity and the Incidence of Cardiovascular Disease : Evidence from the UK Biobank Cohort Study. *PLoS Medicine* **2,** 1–16. http://dx.doi.org/10.1371/journal.pmed.1003487 (2021).

119. LaMonte, M. J. *et al.* Accelerometer-Measured Physical Activity and Mortality in Women Aged 63 to 99. *Journal of the American Geriatrics Society* **66,** 886–894. ISSN: 1532-5415 (2018).

120. Ensrud, K. E. *et al.* Objective Measures of Activity Level and Mortality in Older Men. *Journal of the American Geriatrics Society* **62,** 2079–87. ISSN: 1532-5415 (2014).

121. Ballin, M., Nordstrom, P., Niklasson, J. & Nordstrom, A. Associations of Objectively Measured Physical Activity and Sedentary Time with the Risk of Stroke, Myocardial Infarction or All-Cause Mortality in 70-Year-Old Men and Women: A Prospective Cohort Study. *Sports medicine* **51,** 339–349. ISSN: 1179-2035 (2021).

122. Chastin, S. F. M., Palarea-Albaladejo, J., Dontje, M. L. & Skelton, D. A. Combined Effects of Time Spent in Physical Activity, Sedentary Behaviors and Sleep on Obesity and Cardio-Metabolic Health Markers: A Novel Compositional Data Analysis Approach. *Plos One* **10,** e0139984. ISSN: 1932-6203. pmid: 26461112. http://dx.plos.org/10.1371/journal.pone.0139984 (2015).

123. Pedišić, Ž., Dumuid, D. & Olds, T. S. Integrating Sleep, Sedentary Behaviour, and Physical Activity Research in the Emerging Field of Time-Use Epidemiology: Definition, Concepts, Statistical Methods, Theoretical Framework, and Future Directions, 18 (2017).

124. Mekary, R. A., Willett, W. C., Hu, F. B. & Ding, E. L. Isotemporal Substitution Paradigm for Physical Activity Epidemiology and Weight Change. *American Journal of Epidemiology* **170,** 519–527. ISSN: 00029262. pmid: 19584129 (2009).

125. Dumuid, D. *et al.* The Compositional Isotemporal Substitution Model: A Method for Estimating Changes in a Health Outcome for Reallocation of Time between Sleep, Physical Activity and Sedentary Behaviour. *Statistical Methods in Medical Research* **28,** 846–857. ISSN: 14770334. pmid: 29157152 (2017).

126. Millard, L. A. C., Tilling, K., Gaunt, T. R., Carslake, D. & Lawlor, D. A. Association of Physical Activity Intensity and Bout Length with Mortality: An Observational Study of 79,503 UK Biobank Participants. *PLOS Medicine* **18,** e1003757. ISSN: 1549-1676. https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1003757 (2021).

127. Dumuid, D. *et al.* Compositional Data Analysis in Time-Use Epidemiology : What , Why , How. *International Journal of Environmental Research and Public Health* **17,** 2220 (2020).

128. Migueles, J. H. *et al.* GRANADA Consensus on Analytical Approaches to Assess Associations with Accelerometer-Determined Physical Behaviours (Physical Activity, Sedentary Behaviour and Sleep) in Epidemiological Studies. *British Journal of Sports Medicine* **0,** bjsports-2020–103604. ISSN: 0306-3674. http://bjsm.bmj.com/ (2021).

129. Dumuid, D. *et al.* Adiposity and the Isotemporal Substitution of Physical Activity, Sedentary Time and Sleep among School-Aged Children: A Compositional Data Analysis Approach. *BMC Public Health.* ISSN: 14712458. pmid: 29499689 (2018).

130. Pedišić, Ž. Measurement Issues and Poor Adjustments for Physical Activity and Sleep Undermine Sedentary Behaviour Research - the Focus Should Shift to the Balance between Sleep, Sedentary Behaviour, Standing and Activity. *Kinesiology* **46,** 135–146. ISSN: 1331-1441 (2014).

131. Rosenberger, M. E. *et al.* The 24-Hour Activity Cycle: A New Paradigm for Physical Activity. *Medicine and Science in Sports and Exercise* **51,** 454–464. ISSN: 1530-0315. pmid: 30339658 (2019).

132. Arnold, K. F., Berrie, L., Tennant, P. W. G. & Gilthorpe, M. S. A Causal Inference Perspective on the Analysis of Compositional Data. *International Journal of Epidemiology.* ISSN: 0300-5771 (2020).

133. Aitchison, J. The Statistical Analysis of Compositional Data. *The Statistical Analysis of Compositional Data* **44,** 139–177 (1982).

134. Scealy, J. L. & Welsh, A. H. Colours and Cocktails: Compositional Data Analysis 2013 Lancaster Lecture. *Australian and New Zealand Journal of Statistics.* ISSN: 1467842X (2014).

135. Dumuid, D. *et al.* Compositional Data Analysis for Physical Activity, Sedentary Time and Sleep Research. *Statistical Methods in Medical Research* **27,** 096228021771083. ISSN: 0962-2802. pmid: 28555522. http://journals.sagepub.com/doi/10.1177/0962280217710835 (2017).

136. *Compositional Data Analysis in the Geosciences; from Theory to Practice* in (eds Buccianti, A., Mateu-Figueras, G. & Pawlowsky-Glahn, V.) (2004).

137. Tsilimigras, M. C. & Fodor, A. A. Compositional Data Analysis of the Microbiome: Fundamentals, Tools, and Challenges. *Annals of Epidemiology* **26,** 330–335. ISSN: 18732585. pmid: 27255738 (2016).

138. Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V. & Egozcue, J. J. Microbiome Datasets Are Compositional: And This Is Not Optional. *Frontiers in Microbiology* **8,** 1–6. ISSN: 1664302X (2017).

139. Reimann, C. *et al.* The Concept of Compositional Data Analysis in Practice - Total Major Element Concentrations in Agricultural and Grazing Land Soils of Europe. *Science of the Total Environment* **426,** 196–210. ISSN: 00489697. pmid: 22503552 (2012).

140. Janssen, I. *et al.* A Systematic Review of Compositional Data Analysis Studies Examining Associations between Sleep, Sedentary Behaviour, and Physical Activity with Health Outcomes in Adults. *Applied physiology, nutrition, and metabolism = Physiologie appliquee, nutrition et metabolisme* **45,** S248–S257. ISSN: 17155320. pmid: 33054342. https://pubmed.ncbi.nlm.nih.gov/33054342/ (2020).

141. Ross, R. *et al.* Canadian 24-Hour Movement Guidelines for Adults Aged 18-64 Years and Adults Aged 65 Years or Older: An Integration of Physical Activity, Sedentary Behaviour, and Sleep. *Applied Physiology, Nutrition, and Metabolism = Physiologie Appliquee, Nutrition Et Metabolisme* **45,** S57–S102. ISSN: 1715-5320. pmid: 33054332 (10 (Suppl. 2) 2020).

142. Fox, M. P., MacLehose, R. F. & Lash, T. L. in *Applying Quantitative Bias Analysis to Epidemiologic Data* 75–103 (Springer, Cham, 2021). https://ezproxy-prd.bodleian.ox.ac.uk:2120/chapter/10.1007/978-3-030-82673-4_4 (2022).

143. Smith, L. H. Selection Mechanisms and Their Consequences: Understanding and Addressing Selection Bias. *Current Epidemiology Reports* **7,** 179–189. ISSN: 2196-2995. https://link.springer.com/10.1007/s40471-020-00241-6 (2020).

144. Sattar, N. & Preiss, D. Reverse Causality in Cardiovascular Epidemiological Research: More Common Than Imagined? *Circulation* **135,** 2369–2372. ISSN: 1524-4539. pmid: 28606949 (2017).

145. Floud, S. *et al.* Body Mass Index, Diet, Physical Inactivity, and the Incidence of Dementia in 1 Million UK Women. *Neurology* **94,** e123–e132. ISSN: 0028-3878, 1526-632X. pmid: 31852815. https://n.neurology.org/content/94/2/e123 (2020).

146. Barker, J. *et al.* Physical Activity of UK Adults with Chronic Disease: Cross-Sectional Analysis of Accelerometer-Measured Physical Activity in 96 706 UK Biobank Participants. *International Journal of Epidemiology,* 1–8. ISSN: 0300-5771 (2019).

147. Fox, M. P., MacLehose, R. F. & Lash, T. L. in *Applying Quantitative Bias Analysis to Epidemiologic Data* 1–24 (Springer, Cham, 2021). https://ezproxy-prd.bodleian.ox.ac.uk:2120/chapter/10.1007/978-3-030-82673-4_1 (2022).

148. Lee, I. M. *et al.* Association of Step Volume and Intensity with All-Cause Mortality in Older Women. *JAMA Internal Medicine* **02215.** ISSN: 21686106 (2019).

149. Schmid, D., Ricci, C., Baumeister, S. E. & Leitzmann, M. F. Replacing Sedentary Time with Physical Activity in Relation to Mortality. *Medicine and Science in Sports and Exercise.* ISSN: 15300315. pmid: 26918559 (2016).

150. Kuster, R. P., von Rosen, P., Grooten, W. J. A., Dohrn, I.-M. & Hagstromer, M. Self-Reported and Device-Measured Physical Activity in Leisure Time and at Work and Associations with Cardiovascular Events-A Prospective Study of the Physical Activity Paradox. *Int J Environ Res Public Health* **18.** ISSN: 1660-4601 and 1660-4601 (2021).

151. Dohrn, I.-M., Welmer, A.-K. & Hagstromer, M. Accelerometry-Assessed Physical Activity and Sedentary Time and Associations with Chronic Disease and Hospital Visits - a Prospective Cohort Study with 15 Years Follow-Up. *The international journal of behavioral nutrition and physical activity* **16,** 125. ISSN: 1479-5868 (2019).

152. Nguyen, S. *et al.* Accelerometer-Derived Daily Life Movement Classified by Machine Learning and Incidence of Cardiovascular Disease in Older Women: The OPACH Study. *Journal of the American Heart Association* **11,** e023433. https://ezproxy-prd.bodleian.ox.ac.uk:3349/doi/10.1161/JAHA.121.023433 (2022).

153. Yerramalla, M. S. *et al.* Objectively Measured Total Sedentary Time and Pattern of Sedentary Accumulation in Older Adults: Associations with Incident Cardiovascular Disease and All-Cause Mortality. *The Journals of Gerontology: Series A,* glac023. ISSN: 1079-5006, 1758-535X. https://academic.oup.com/biomedgerontology/advance-article/doi/10.1093/gerona/glac023/6517536 (2022).

154. Dohrn, I. M., Kwak, L., Oja, P., Sjöström, M. & Hagströmer, M. Replacing Sedentary Time with Physical Activity: A 15-Year Follow-up of Mortality in a National Cohort. *Clinical Epidemiology* **10,** 179–186. ISSN: 11791349. pmid: 29416378 (2018).

155. Freedson, P. S., Melanson, E. & Sirard, J. Calibration of the Computer Science and Applications, Inc. Accelerometer. *Medicine and science in sports and exercise* **30,** 777–81. ISSN: 0195-9131. pmid: 9588623 (1998).

156. Evenson, K. R. *et al.* Calibrating Physical Activity Intensity for Hip-Worn Accelerometry in Women Age 60 to 91 Years: The Women's Health Initiative OPACH Calibration Study. *Preventive Medicine Reports* **2,** 750–756. ISSN: 2211-3355. pmid: 26527313 (2015).

157. Button, K. S. *et al.* Power Failure: Why Small Sample Size Undermines the Reliability of Neuroscience. *Nature Reviews Neuroscience* **14,** 365–376. ISSN: 1471-0048. https://www.nature.com/articles/nrn3475 (5 2013).

158. Kline, C. E. *et al.* Physical Activity and Sleep: An Updated Umbrella Review of the 2018 Physical Activity Guidelines Advisory Committee Report. *Sleep Medicine Reviews* **58,** 101489. ISSN: 1087-0792. https://www.sciencedirect.com/science/article/pii/S1087079221000745 (2021).

159. Chastin, S. *et al.* Joint Association between Accelerometry-Measured Daily Combination of Time Spent in Physical Activity, Sedentary Behaviour and Sleep and All-Cause Mortality: A Pooled Analysis of Six Prospective Cohorts Using Compositional Analysis. *British Journal of Sports Medicine,* bjsports-2020–102345. ISSN: 0306-3674. https://bjsm.bmj.com/lookup/doi/10.1136/bjsports-2020-102345 (2021).

160. Tomova, G. D., Arnold, K. F., Gilthorpe, M. S. & Tennant, P. W. G. Adjustment for Energy Intake in Nutritional Research: A Causal Inference Perspective. *The American Journal of Clinical Nutrition* **115,** 189–198. ISSN: 0002-9165. https://doi.org/10.1093/ajcn/nqab266 (2022).

161. Allen, N. *et al.* UK Biobank: Current Status and What It Means for Epidemiology. *Health Policy and Technology* **1,** 123–126. ISSN: 22118837 (2012).

162. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Medicine* **12.** ISSN: 15491676. pmid: 25826379 (2015).

163. Collins, R. What Makes UK Biobank Special? *The Lancet* **379,** 1173–1174. ISSN: 01406736 (2012).

164. Manolio, T. A. *et al.* New Models for Large Prospective Studies: Is There a Better Way? *American Journal of Epidemiology* **175,** 859–866. ISSN: 0002-9262. https://doi.org/10.1093/aje/kwr453 (2012).

165. Littlejohns, T. J., Sudlow, C., Allen, N. E. & Collins, R. UK Biobank: Opportunities for Cardiovascular Research. *European Heart Journal* **40,** 1158–1166. ISSN: 0195-668X. https://doi.org/10.1093/eurheartj/ehx254 (2019).

166. Bycroft, C. *et al.* The UK Biobank Resource with Deep Phenotyping and Genomic Data. *Nature* **562,** 203–209. ISSN: 1476-4687. https://www.nature.com/articles/s41586-018-0579-z. (7726 2018).

167. Backman, J. D. *et al.* Exome Sequencing and Analysis of 454,787 UK Biobank Participants. *Nature* **599,** 628–634. ISSN: 1476-4687. https://www.nature.com/articles/s41586-021-04103-z (7886 2021).

168. *Whole Genome Sequencing Data on 200,000 UK Biobank Participants Available Now* https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/news/whole-genome-sequencing-data-on-200-000-uk-biobank-participants-available-now (2022).

169. *Ethics* https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us/ethics (2022).

170. *Basis of Your Participation - Privacy Notice for UK Biobank Participants* https://www.ukbiobank.ac.uk/explore-your-participation/basis-of-your-participation (2022).

171. *Register* https://www.ukbiobank.ac.uk/enable-your-research/register (2022).

172. *Approved Research: Statistical Machine Learning of Wearable Sensor Data to Predict Disease Outcomes* https://www.ukbiobank.ac.uk/enable-your-research/approved-research/statistical-machine-learning-of-wearable-sensor-data-to-predict-disease-outcomes (2022).

173. *Biomedical Research Computing — University of Oxford, Medical Sciences Division* https://www.medsci.ox.ac.uk/divisional-services/support-services-1/bmrc (2022).

174. UK Biobank. *Physical Activity Monitor (Accelerometer)* 2016.

175. Strain, T. *et al.* Wearable-Device-Measured Physical Activity and Future Health Risk. *Nature Medicine* **26,** 1385–1391. ISSN: 1546170X. pmid: 32807930. https://doi.org/10.1038/s41591-020-1012-3 (2020).

176. UK Biobank. *: External Info : Data_providers_and_dates* https://biobank.ctsu.ox.ac.uk/crystal/exinfo.cgi?src=Data_providers_and_dates (2022).

177. Woodfield, R., Grant, I., UK Biobank Stroke Outcomes Group, UK Biobank Follow-Up and Outcomes Working Group & Sudlow, C. L. M. Accuracy of Electronic Health Record Data for Identifying Stroke Cases in Large-Scale Epidemiological Studies: A Systematic Review from the UK Biobank Stroke Outcomes Group. *PLOS ONE* **10,** e0140533. ISSN: 1932-6203. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0140533 (2015).

178. Rannikmäe, K. *et al.* Accuracy of Identifying Incident Stroke Cases from Linked Health Care Data in UK Biobank. *Neurology* **95,** e697–e707. ISSN: 0028-3878, 1526-632X. pmid: 32616677. https://n.neurology.org/content/95/6/e697 (2020).

179. Wood, A. *et al.* Linked Electronic Health Records for Research on a Nationwide Cohort of More than 54 Million People in England: Data Resource. *BMJ* **373,** n826. ISSN: 1756-1833. pmid: 33827854. https://www.bmj.com/content/373/bmj.n826 (2021).

180. Fry, A. *et al.* Study Design Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. **186,** 1026–1034 (2017).

181. Doherty, A. *et al.* GWAS Identifies 14 Loci for Device-Measured Physical Activity and Sleep Duration. *Nature Communications* **9.** ISSN: 1769-4493. pmid: 30531941 (2018).

182. Staudenmayer, J., He, S., Hickey, A., Sasaki, J. & Freedson, P. Methods to Estimate Aspects of Physical Activity and Sedentary Behavior from High-Frequency Wrist Accelerometer Measurements. *Journal of Applied Physiology* **119,** 396–403. ISSN: 8750-7587. pmid: 26112238. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4538283/ (2015).

183. Johnston, W. *et al.* Recommendations for Determining the Validity of Consumer Wearable and Smartphone Step Count: Expert Statement and Checklist of the INTERLIVE Network. *Br J Sports Med* **0,** 1–14. http://dx.doi.org/10.1136/bjsports-2020-103147 (2020).

184. Ellis, K., Kerr, J., Godbole, S., Staudenmayer, J. & Lanckriet, G. Hip and Wrist Accelerometer Algorithms for Free-Living Behavior Classification. *Medicine and Science in Sports and Exercise.* ISSN: 15300315 (2016).

185. Winkler, E. A. H. *et al.* Identifying Adults' Valid Waking Wear Time by Automated Estimation in activPAL Data Collected with a 24 h Wear Protocol. **37,** 1653–1668. ISSN: 0967-3334. https://doi.org/10.1088/0967-3334/37/10/1653 (2016).

186. Kelly, P., Fitzsimons, C. & Baker, G. Should We Reframe How We Think about Physical Activity and Sedentary Behavior Measurement? Validity and Reliability Reconsidered. *International Journal of Behavioral Nutrition and Physical Activity* **13.** ISSN: 14795868. pmid: 26931142 (2016).

187. Gershuny, J. *et al.* Testing Self-Report Time-Use Diaries against Objective Instruments in Real Time. *Sociological Methodology* **50,** 318–349. ISSN: 0081-1750, 1467-9531. http://journals.sagepub.com/doi/10.1177/0081175019884591 (2020).

188. Kelly, P. *et al.* An Ethical Framework for Automated, Wearable Cameras in Health Behavior Research. *American Journal of Preventive Medicine* **44,** 314–319. ISSN: 07493797. pmid: 23415131. http://dx.doi.org/10.1016/j.amepre.2012.11.006 (2013).

189. *Wearable-Camera-Desktop-Browser* Oxford Wearables Group, 2020. https://github.com/OxWearables/wearable-camera-desktop-browser (2022).

190. Van Hees, V. T. *et al.* Autocalibration of Accelerometer Data for Free-Living Physical Activity Assessment Using Local Gravity and Temperature: An Evaluation on Four Continents. *Journal of Applied Physiology.* ISSN: 8750-7587. pmid: 25103964 (2014).

191. Mannini, A., Intille, S. S., Rosenberger, M., Sabatini, A. M. & Haskell, W. Activity Recognition Using a Single Accelerometer Placed at the Wrist or Ankle. *Medicine and Science in Sports and Exercise* **45,** 2193–2203. ISSN: 1530-0315. pmid: 23604069 (2013).

192. Zhang, S., Rowlands, A. V., Murray, P. & Hurst, T. L. Physical Activity Classification Using the GENEA Wrist-Worn Accelerometer. *Medicine and Science in Sports and Exercise* **44,** 742–748. ISSN: 01959131 (2012).

193. Vähä-Ypyä, H., Vasankari, T., Husu, P., Suni, J. & Sievänen, H. A Universal, Accurate Intensity-Based Classification of Different Physical Activities Using Raw Data of Accelerometer. *Clinical Physiology and Functional Imaging* **35,** 64–70. ISSN: 1475097X (2015).

194. Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J. *Classification and Regression Trees* (1984).

195. Breiman, L. Bagging Predictors. *Machine Learning* **24,** 123–140. ISSN: 08856125 (1996).

196. Forney, G. D. The Viterbi Algorithm. *Proceedings of the IEEE* **61,** 268–278. ISSN: 15582256 (1973).

197. Yuan, H. *et al. Self-Supervised Learning for Human Activity Recognition Using 700,000 Person-days of Wearable Data* 2022. arXiv: 2206.02909 [cs, eess]. http://arxiv.org/abs/2206.02909 (2022).

198. Chen, K. *et al.* Deep Learning for Sensor-based Human Activity Recognition: Overview, Challenges, and Opportunities. *ACM Computing Surveys* **54,** 1–40. ISSN: 0360-0300, 1557-7341. https://dl.acm.org/doi/10.1145/3447744 (2022).

199. *The Python Language Reference — Python 3.6.15 Documentation* https://docs.python.org/3.6/reference/ (2022).

200. R Core Team. R: A Language and Environment for Statistical Computing. *Vienna, Austria.* https://www.r-project.org/ (2019).

201. Pawlowsky-Glahn, V., Egozcue, J. J. & Tolosana-Delgado, R. Lecture Notes on Compositional Data Analysis, 1–100. http://ima.udg.edu/Recerca/EIO/inici (2011).

202. Clarke, A. E. & Janssen, I. A Compositional Analysis of Time Spent in Sleep, Sedentary Behaviour and Physical Activity with All-Cause Mortality Risk. *International Journal of Behavioral Nutrition and Physical Activity* **18,** 1–12. ISSN: 14795868. pmid: 33549100. https://doi.org/10.1186/s12966-021-01092-0 (2021).

203. Gupta, N. *et al.* The Physical Activity Paradox Revisited: A Prospective Study on Compositional Accelerometer Data and Long-Term Sickness Absence. *International Journal of Behavioral Nutrition and Physical Activity* **17,** 1–9. ISSN: 14795868. pmid: 32690043 (2020).

204. Stamatakis, E. *et al.* Any Public Health Guidelines Should Always Be Developed from a Consistent, Clear Evidence Base. *British Journal of Sports Medicine* **53,** 1555–1556. ISSN: 14730480. pmid: 30760456. http://bjsm.bmj.com/ (2019).

205. Filzmoser, P., Hron, K. & Templ, M. *Applied Compositional Data Analysis* (2018).

206. Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G. & Barceló-Vidal, C. Isometric Logratio Transformations for Compositional Data Analysis. *Mathematical Geology* **35,** 279–300. ISSN: 08828121 (2003).

207. Aitchison, J. & Egozcue, J. J. Compositional Data Analysis: Where Are We and Where Should We Be Heading? *Mathematical Geology.* ISSN: 08828121. pmid: 25246403 (2005).

208. Palarea-Albaladejo, J. & Martín-Fernández, J. A. ZCompositions - R Package for Multivariate Imputation of Left-Censored Data under a Compositional Approach. *Chemometrics and Intelligent Laboratory Systems* **143,** 85–96. ISSN: 18733239 (2015).

209. Sagelv, E. H. *et al.* The Bidirectional Associations between Leisure Time Physical Activity Change and Body Mass Index Gain. The Tromsø Study 1974–2016. *International Journal of Obesity,* 1–14. ISSN: 0307-0565. http://www.nature.com/articles/s41366-021-00853-y (2021).

210. Canchola, A. J. *et al. Cox Regression Using Different Time Scales* in. Western Users of SAS Software (San Francisco, California, 2003), 6.

211. Cologne, J. *et al.* Proportional Hazards Regression in Epidemiologic Follow-up Studies: An Intuitive Consideration of Primary Time Scale. *Epidemiology* **23,** 565–573. ISSN: 1044-3983. JSTOR: 23214300 (2012).

212. Lund Rasmussen, C. *et al.* Zero Problems with Compositional Data of Physical Behaviors: A Comparison of Three Zero Replacement Methods. *International Journal of Behavioral Nutrition and Physical Activity.* ISSN: 1479-5868 (2020).

213. Woodward, M. *Epidemiology: Study Design and Data Analysis* (CRC Press, Boca Raton, Florida ; London, England ; New York, New York, 2014).

214. McGregor, D. E., Palarea-Albaladejo, J., Dall, P. M. P., Hron, K. & Chastin, S. F. Cox Regression Survival Analysis with Compositional Covariates: Application to Modelling Mortality Risk from 24-h Physical Activity Patterns. *Statistical Methods in Medical Research* **29,** 1447–1465. ISSN: 0962-2802. pmid: 31342855. http://journals.sagepub.com/doi/10.1177/0962280219864125 (2019).

215. Mekary, R. A. & Ding, E. L. Isotemporal Substitution as the Gold Standard Model for Physical Activity Epidemiology: Why It Is the Most Appropriate for Activity Time Research. *International Journal of Environmental Research and Public Health* **16,** 11–13. ISSN: 16604601 (2019).

216. Therneau, T. M. A Package for Survival Analysis in S. Version 2.38. *CRAN website - http://cran.r-project.org/package=survival* (2015).

217. Therneau, T. M. & Grambsch, P. M. *Modeling Survival Data: Extending the Cox Model* (Springer, New York, 2000).

218. Wickham, H. The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software* **40,** 1–29. ISSN: 1548-7660. https://doi.org/10.18637/jss.v040.i01 (2011).

219. Dowle, M. & Srinivasan, A. *Data.Table: Extension of 'data.Frame'.* 2019. https://cran.r-project.org/package=data.table.

220. Ren, K. *Rlist: A Toolbox for Non-Tabular Data Manipulation* version 0.4.6.2. 2021. https://CRAN.R-project.org/package=rlist (2022).

221. Warnes, G. R., Bolker, B. & Lumley, T. Gtools: Various R Programming Tools. *R package version 3.8.1.* (2018).

222. The Emerging Risk Factors Collaboration. Separate and Combined Associations of Body-Mass Index and Abdominal Adiposity with Cardiovascular Disease: Collaborative Analysis of 58 Prospective Studies. *The Lancet* **377,** 1085–1095. ISSN: 0140-6736. https://www.sciencedirect.com/science/article/pii/S0140673611601050 (2011).

223. Fox, M. P., MacLehose, R. F. & Lash, T. L. in *Applying Quantitative Bias Analysis to Epidemiologic Data* 25–55 (Springer, Cham, 2021). https://ezproxy-prd.bodleian.ox.ac.uk:2120/chapter/10.1007/978-3-030-82673-4_2 (2022).

224. Lash, T. L. *et al.* Good Practices for Quantitative Bias Analysis. *International Journal of Epidemiology* **43,** 1969–1985. ISSN: 0300-5771. https://doi.org/10.1093/ije/dyu149 (2014).

225. Ding, D., Lee, I.-M., Bauman, A. E., Ekelund, U. & Stamatakis, E. Dietary Risk versus Physical Inactivity: A Forced Comparison with Policy Implications? *The Lancet* **397,** 1709–1710. ISSN: 0140-6736, 1474-547X. pmid: 33965086. https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(21)00801-1/fulltext (2021).

226. Bennett, D. A., 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 Medical Research Methodology.* ISSN: 14712288. pmid: 28927376 (2017).

227. Keogh, R. H. *et al.* STRATOS Guidance Document on Measurement Error and Misclassification of Variables in Observational Epidemiology: Part 1—Basic Theory and Simple Methods of Adjustment. *Statistics in Medicine,* 1–35. ISSN: 10970258. pmid: 32246539 (December 2018 2020).

228. Carroll, R. J., Ruppert, D., Stefanski, L. A. & Crainiceanu, C. M. *Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition* http://ebookcentral.proquest.com/lib/oxford/detail.action?docID=274076 (2021) (CRC Press LLC, London, United Kingdom, 2006).

229. Keogh, R. H. & White, I. R. A Toolkit for Measurement Error Correction, with a Focus on Nutritional Epidemiology. *Statistics in Medicine* **33,** 2137–2155. ISSN: 10970258 (2014).

230. Prospective Studies Collaboration. Age-Specific Relevance of Usual Blood Pressure to Vascular Mortality: A Meta-Analysis of Individual Data for One Million Adults in 61 Prospective Studies. *The Lancet* **360,** 1903–1913. ISSN: 01406736. https://linkinghub.elsevier.com/retrieve/pii/S0140673602119118 (2002).

231. Frost, C. & Thompson, S. G. Correcting for Regression Dilution Bias: Comparison of Methods for a Single Predictor Variable. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **163,** 173–189. ISSN: 1467-985X. https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-985X.00164 (2000).

232. Shaw, P. A. *et al.* STRATOS Guidance Document on Measurement Error and Misclassification of Variables in Observational Epidemiology: Part 2—More Complex Methods of Adjustment and Advanced Topics. *Statistics in Medicine,* 1–32. ISSN: 10970258. pmid: 32246531 (February 2020).

233. Spiegelman, D., McDermott, A. & Rosner, B. Regression Calibration Method for Correcting Measurement-Error Bias in Nutritional Epidemiology. *The American Journal of Clinical Nutrition* **65,** 1179S–1186S. ISSN: 0002-9165, 1938-3207. https://academic.oup.com/ajcn/article/65/4/1179S/4655713 (1997).

234. Keadle, S. K. *et al.* Reproducibility of Accelerometer-Assessed Physical Activity and Sedentary Time. *American Journal of Preventive Medicine* **52,** 541–548. ISSN: 1873-2607. pmid: 28062274 (2017).

235. Saint-Maurice, P. F. *et al.* Reproducibility of Accelerometer and Posture-derived Measures of Physical Activity. *Medicine and science in sports and exercise* **52,** 876. pmid: 31688655. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7078017/ (2020).

236. Aadland, E. & Ylvisåker, E. Reliability of Objectively Measured Sedentary Time and Physical Activity in Adults. *PLoS ONE* **10,** 1–13. ISSN: 19326203 (2015).

237. Jaeschke, L., Steinbrecher, A., Jeran, S., Konigorski, S. & Pischon, T. Variability and Reliability Study of Overall Physical Activity and Activity Intensity Levels Using 24 H-Accelerometry-Assessed Data. *BMC Public Health* (2018).

238. Aitchison, J. & Shen, S. M. Measurement Error in Compositional Data. *Journal of the International Association for Mathematical Geology* **16,** 637–650. ISSN: 1573-8868. https://doi.org/10.1007/BF01029322 (1984).

239. Hallman, D. M., Mathiassen, S. E., van der Beek, A. J., Jackson, J. A. & Coenen, P. Calibration of Self-Reported Time Spent Sitting, Standing and Walking among Office Workers: A Compositional Data Analysis. *International journal of environmental research and public health* **16.** ISSN: 1660-4601. pmid: 31461868 (2019).

240. Rosner, B., Willett, W. C. & Spiegelman, D. Correction of Logistic Regression Relative Risk Estimates and Confidence Intervals for Systematic Within-Person Measurement Error. *Statistics in Medicine* **8,** 1051–1069. ISSN: 0277-6715. https://ezproxy-prd.bodleian.ox.ac.uk:2152/doi/abs/10.1002/sim.4780080905 (1989).

241. Rosner, B., Spiegelman, D. & Willett, W. C. Correction of Logistic Regression Relative Risk Estimates and Confidence Intervals for Measurement Error: The Case of Multiple Covariates Measured with Error. *American Journal of Epidemiology* **132,** 734–745. ISSN: 1476-6256, 0002-9262. https://academic.oup.com/aje/article/102293/CORRECTION (1990).

242. Orloff, J. & Bloom, J. *Bootstrap Confidence Intervals* 2014. https://ocw.mit.edu/courses/18-05-introduction-to-probability-and-statistics-spring-2014/resources/mit18_05s14_reading24/ (2022).

243. *The Python Language Reference — Python 3.7.13 Documentation* https://docs.python.org/3.7/reference/ (2022).

244. Tyrrell, J. *et al.* Genetic Predictors of Participation in Optional Components of UK Biobank. *Nature Communications* **12,** 1–13. ISSN: 20411723. pmid: 33563987. https://doi.org/10.1038/s41467-021-21073-y (2021).

245. Infante-Rivard, C. & Cusson, A. Reflection on Modern Methods: Selection Bias—a Review of Recent Developments. *International Journal of Epidemiology* **47,** 1714–1722. ISSN: 0300-5771. https://doi.org/10.1093/ije/dyy138 (2018).

246. Munafò, M. R., Tilling, K., Taylor, A. E., Evans, D. M. & Davey Smith, G. Collider Scope: When Selection Bias Can Substantially Influence Observed Associations. *International Journal of Epidemiology* **47,** 226–235. ISSN: 0300-5771. pmid: 29040562. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5837306/ (2018).

247. Rothman, K. J., Gallacher, J. E. & Hatch, E. E. Why Representativeness Should Be Avoided. *International Journal of Epidemiology* **42,** 1012–1014. ISSN: 0300-5771. https://doi.org/10.1093/ije/dys223 (2013).

248. Batty, G. D., Gale, C. R., Kivimäki, M., Deary, I. J. & Bell, S. Comparison of Risk Factor Associations in UK Biobank against Representative, General Population Based Studies with Conventional Response Rates: Prospective Cohort Study and Individual Participant Meta-Analysis. *The BMJ* **368,** 1–8. ISSN: 17561833. pmid: 32051121 (2020).

249. Bradley, V. & Nichols, T. E. Addressing Selection Bias in the UK Biobank Neurological Imaging Cohort, 2022.01.13.22269266. ISSN: 2226-9266. https://www.medrxiv.org/content/10.1101/2022.01.13.22269266v2 (2022).

250. Stamatakis, E. *et al.* Is Cohort Representativeness Passé? Poststratified Associations of Lifestyle Risk Factors with Mortality in the UK Biobank. *Epidemiology* **32,** 179–188. ISSN: 1044-3983. https://journals.lww.com/epidem/fulltext/2021/03000/is_cohort_representativeness_pass___poststratified.4.aspx (2021).

251. Van Buuren, S. & Groothuis-Oudshoorn, K. Mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software* **45,** 1–67. ISSN: 1548-7660. https://doi.org/10.18637/jss.v045.i03 (2011).

252. Sterne, J. A. C. *et al.* Multiple Imputation for Missing Data in Epidemiological and Clinical Research: Potential and Pitfalls. *BMJ* **338,** b2393. ISSN: 0959-8138, 1468-5833. pmid: 19564179. https://www.bmj.com/content/338/bmj.b2393 (2009).

253. Van Buuren, S. *et al. Mice: Multivariate Imputation by Chained Equations* version 3.14.0. 2021. https://CRAN.R-project.org/package=mice (2022).

254. Gkatzionis, A. & Burgess, S. Contextualizing Selection Bias in Mendelian Randomization: How Bad Is It Likely to Be? *International Journal of Epidemiology* **48,** 691–701. ISSN: 0300-5771. https://academic.oup.com/ije/article/48/3/691/5132989 (2019).

255. Kujala, U. M. Is Physical Activity a Cause of Longevity? It Is Not as Straightforward as Some Would Believe. A Critical Analysis. *British journal of sports medicine.* ISSN: 14730480. pmid: 29545237 (2018).

256. Matthews, C. E. *et al.* Exploration of Confounding Due to Poor Health in an Accelerometer-Mortality Study. *Medicine and Science in Sports and Exercise* **52,** 2546–2553. ISSN: 1530-0315. pmid: 32472927 (2020).

257. Tarp, J. *et al.* Accelerometer-Measured Physical Activity and Sedentary Time in a Cohort of US Adults Followed for up to 13 Years: The Influence of Removing Early Follow-up on Associations with Mortality. *International Journal of Behavioral Nutrition and Physical Activity* **17,** 1–8. ISSN: 14795868. pmid: 32169059 (2020).

258. Strain, T. *et al.* Impact of Follow-up Time and Analytical Approaches to Account for Reverse Causality on the Association between Physical Activity and Health Outcomes in UK Biobank. *International Journal of Epidemiology* **49,** 162–172 (2019).

259. Carstensen, B., Plummer, M., Laara, E. & Hills, M. *Epi: Statistical Analysis in Epidemiology* version 2.47. 2022. https://CRAN.R-project.org/package=Epi (2022).

260. Plummer, M. & Carstensen, B. Lexis: An R Class for Epidemiological Studies with Long-Term Follow-Up. *Journal of Statistical Software* **38,** 1–12. ISSN: 1548-7660. https://doi.org/10.18637/jss.v038.i05 (2011).

261. Therneau, T., Crowson, C. & Atkinson, E. *Using Time Dependent Covariates and Time Dependent Coefficients in the Cox Model* Vignette for 'survival' package (2022), 31. https://cran.r-project.org/web/packages/survival/vignettes/timedep.pdf.

262. Hong, L. S. & Lewington, S. *Lexis Expansion- Age-at-Risk Adjustment for Survival Analysis* in. PhUSE (2013), 5.

263. Hamer, M., Bauman, A., Bell, J. A. & Stamatakis, E. Examining Associations between Physical Activity and Cardiovascular Mortality Using Negative Control Outcomes. *International Journal of Epidemiology* **48,** 1161–1166. ISSN: 14643685. pmid: 30541040 (2019).

264. Van Der Weele, T. J. & Ding, P. Sensitivity Analysis in Observational Research: Introducing the E-Value. *Annals of Internal Medicine* **167,** 268–274. ISSN: 15393704. pmid: 28693043 (2017).

265. Ding, P. & Van Der Weele, T. J. Sensitivity Analysis Without Assumptions. *Epidemiology (Cambridge, Mass.)* **27,** 368–377. ISSN: 1044-3983. pmid: 26841057. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4820664/ (2016).

266. DiPietro, L. *et al.* Physical Activity, Injurious Falls, and Physical Function in Aging: An Umbrella Review. *Medicine and science in sports and exercise* **51,** 1303–1313. ISSN: 0195-9131. pmid: 31095087. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6527126/ (2019).

267. Mathur, M. B., Ding, P., Riddell, C. A. & Van Der Weele, T. J. Web Site and R Package for Computing E-values. *Epidemiology* **29,** e45. ISSN: 15315487 (2018).

268. Mahmood, S. *et al.* A Quantitative Bias Analysis to Estimate Measurement Error-Related Attenuation of the Association between Self-Reported Physical Activity and Colorectal Cancer Risk. *International Journal of Epidemiology* **49,** 153–161. ISSN: 14643685. pmid: 31687751. https://academic.oup.com/ije/article/49/1/153/5612999 (2020).

269. Clarke, R. *et al.* Underestimation of Risk Associations Due to Regression Dilution in Long-Term Follow-up of Prospective Studies. *American Journal of Epidemiology* **150,** 341–353. ISSN: 0002-9262. pmid: 10453810 (1999).

270. Lim, S., Wyker, B., Bartley, K. & Eisenhower, D. Measurement Error of Self-Reported Physical Activity Levels in New York City: Assessment and Correction. *American Journal of Epidemiology* **181,** 648–655. ISSN: 0002-9262. https://doi.org/10.1093/aje/kwu470 (2015).

271. Kivimäki, M. *et al.* Physical Inactivity, Cardiometabolic Disease, and Risk of Dementia: An Individual-Participant Meta-Analysis. *BMJ* **365,** l1495. ISSN: 0959-8138, 1756-1833. pmid: 30995986. https://www.bmj.com/content/365/bmj.l1495 (2019).

272. DiPietro, L. *et al.* Advancing the Global Physical Activity Agenda: Recommendations for Future Research by the 2020 WHO Physical Activity and Sedentary Behavior Guidelines Development Group. *International Journal of Behavioral Nutrition and Physical Activity* **17,** 1–11. ISSN: 14795868. pmid: 33239105. https://doi.org/10.1186/s12966-020-01042-2 (2020).

273. Kerr, J. *et al.* Using the SenseCam to Improve Classifications of Sedentary Behavior in Free-Living Settings. *American Journal of Preventive Medicine* **44,** 290–296. ISSN: 07493797. pmid: 23415127. http://dx.doi.org/10.1016/j.amepre.2012.11.004 (2013).

274. Rose, S. Machine Learning for Prediction in Electronic Health Data. *JAMA Network Open* **1,** e181404. ISSN: 2574-3805. https://doi.org/10.1001/jamanetworkopen.2018.1404 (2018).

275. Chan Chang, S. *et al.* Capture-24: Activity Tracker Dataset for Human Activity Recognition. https://ora.ox.ac.uk/objects/uuid:99d7c092-d865-4a19-b096-cc16440cd001 (2021).

276. LeCun, Y., Bengio, Y. & Hinton, G. Deep Learning. *Nature* **521,** 436–444. ISSN: 1476-4687. https://www.nature.com/articles/nature14539 (7553 2015).

277. Zhuang, F. *et al.* A Comprehensive Survey on Transfer Learning. *Proceedings of the IEEE* **109,** 43–76. ISSN: 1558-2256 (2021).

278. Finlayson, S. G. *et al.* The Clinician and Dataset Shift in Artificial Intelligence. *New England Journal of Medicine* **385,** 283–286. ISSN: 0028-4793. https://doi.org/10.1056/NEJMc2104626 (2021).

279. Van Hees, V. T. *et al.* Estimating Sleep Parameters Using an Accelerometer without Sleep Diary. *Scientific Reports* **8,** 1–11. ISSN: 20452322 (2018).

280. Straker, L. M. *et al.* Rationale, Design and Methods for the 22 Year Follow-up of the Western Australian Pregnancy Cohort (Raine) Study. *BMC Public Health* **15,** 663. ISSN: 1471-2458. https://doi.org/10.1186/s12889-015-1944-6 (2015).

281. Esliger, D. W. *et al.* Validation of the GENEA Accelerometer. *Medicine and Science in Sports and Exercise* **43,** 1085–1093. ISSN: 1530-0315. pmid: 21088628 (2011).

282. Fraysse, F. *et al.* Physical Activity Intensity Cut-Points for Wrist-Worn GENEActiv in Older Adults. *Frontiers in Sports and Active Living* **2.** ISSN: 2624-9367. https://www.frontiersin.org/article/10.3389/fspor.2020.579278 (2021).

283. Lyden, K., Kozey, S. L., Staudenmeyer, J. W. & Freedson, P. S. A Comprehensive Evaluation of Commonly Used Accelerometer Energy Expenditure and MET Prediction Equations. *European Journal of Applied Physiology* **111,** 187–201. ISSN: 14396319. pmid: 20842375 (2011).

284. Bach, K. *et al.* A Machine Learning Classifier for Detection of Physical Activity Types and Postures During Free-Living. *Journal for the Measurement of Physical Behaviour* **-1,** 1–8. ISSN: 2575-6605, 2575-6613. https://journals.humankinetics.com/view/journals/jmpb/aop/article-10.1123-jmpb.2021-0015/article-10.1123-jmpb.2021-0015.xml (aop 2022).

285. Dumuid, D. *et al.* Balancing Time Use for Children's Fitness and Adiposity: Evidence to Inform 24-Hour Guidelines for Sleep, Sedentary Time and Physical Activity. *PLOS ONE* **16,** e0245501. ISSN: 1932-6203. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0245501 (2021).

286. Cappuccio, F. P., Cooper, D., D'Elia, L., Strazzullo, P. & Miller, M. A. Sleep Duration Predicts Cardiovascular Outcomes: A Systematic Review and Meta-Analysis of Prospective Studies. *European Heart Journal* **32,** 1484–1492. ISSN: 0195-668X. https://doi.org/10.1093/eurheartj/ehr007 (2011).

287. Loh, R., Stamatakis, E., Folkerts, D., Allgrove, J. E. & Moir, H. J. Effects of Interrupting Prolonged Sitting with Physical Activity Breaks on Blood Glucose, Insulin and Triacylglycerol Measures: A Systematic Review and Meta-analysis. *Sports Medicine* **50,** 295–330. ISSN: 1179-2035. pmid: 31552570 (2020).

288. Augustin, N. H., Mattocks, C., Cooper, A. R., Ness, A. R. & Faraway, J. J. Modelling Fat Mass as a Function of Weekly Physical Activity Profiles Measured by Actigraph Accelerometers. *Physiological Measurement* **33,** 1831–1839. ISSN: 09673334 (2012).

289. Augustin, N. H., Mattocks, C., Faraway, J. J., Greven, S. & Ness, A. R. Modelling a Response as a Function of High-Frequency Count Data: The Association between Physical Activity and Fat Mass. *Statistical Methods in Medical Research* **26,** 2210–2226. ISSN: 14770334. pmid: 26187735 (2017).

290. Lindsay, T. *et al.* Joint Associations between Objectively Measured Physical Activity Volume and Intensity with Body Fatness: The Fenland Study. *International Journal of Obesity* **46,** 169–177. ISSN: 1476-5497. https://www.nature.com/articles/s41366-021-00970-8 (1 2022).

291. Price, A. J. *et al.* Differences in Risk Factors for 3 Types of Stroke. *Neurology* **90,** e298–e306. ISSN: 0028-3878. pmid: 29321237. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5798656/ (2018).

292. Wade, K. H., Richmond, R. C. & Davey Smith, G. Physical Activity and Longevity: How to Move Closer to Causal Inference. *British Journal of Sports Medicine.* ISSN: 14730480. pmid: 29545236 (2018).

293. Lynch, B. M. *et al.* Approaches to Improve Causal Inference in Physical Activity Epidemiology. *Journal of Physical Activity and Health.* ISSN: 15435476 (2020).

294. Van Smeden, M., Lash, T. L. & Groenwold, R. H. H. Reflection on Modern Methods: Five Myths about Measurement Error in Epidemiological Research. *International Journal of Epidemiology,* 1–10. ISSN: 0300-5771 (2019).

295. Hernán, M. A. & Robins, J. M. *Causal Inference: What If* (Chapman & Hall/CRC, Boca Raton, 2020).

296. White, T. *et al.* Estimating Energy Expenditure from Wrist and Thigh Accelerometry in Free-Living Adults: A Doubly Labelled Water Study. *International Journal of Obesity* **43,** 2333–2342. ISSN: 14765497. http://dx.doi.org/10.1101/370247 (2019).

297. Hernán, M. A., Hernández-Díaz, S. & Robins, J. M. A Structural Approach to Selection Bias. *Epidemiology* **15,** 615–625. ISSN: 1044-3983. JSTOR: 20485961 (2004).

298. Van Alten, S., Domingue, B. W., Galama, T. J. & Marees, A. T. Reweighting the UK Biobank to Reflect Its Underlying Sampling Population Substantially Reduces Pervasive Selection Bias Due to Volunteering, 2022.05.16.22275048. https://www.medrxiv.org/content/10.1101/2022.05.16.22275048v1 (2022).

299. Sabia, S. *et al.* Physical Activity , Cognitive Decline , and Risk of Dementia : 28 Year Follow-up of Whitehall II Cohort Study. *BMJ (Online)* **2709,** 1–12. ISSN: 17561833. pmid: 28642251. https://pubmed.ncbi.nlm.nih.gov/28642251/ (June 2017).

300. Kyu, H. H. *et al.* Physical Activity and Risk of Breast Cancer, Colon Cancer, Diabetes, Ischemic Heart Disease, and Ischemic Stroke Events: Systematic Review and Dose-Response Meta-Analysis for the Global Burden of Disease Study 2013. *BMJ (Online).* ISSN: 17561833. pmid: 27510511 (2016).

301. Greenland, S. Commentary: An Argument against E-values for Assessing the Plausibility That an Association Could Be Explained Away by Residual Confounding. *International Journal of Epidemiology* **49,** 1501–1503. ISSN: 0300-5771. https://doi.org/10.1093/ije/dyaa095 (2020).

302. MacLehose, R. F., Ahern, T. P., Lash, T. L., Poole, C. & Greenland, S. The Importance of Making Assumptions in Bias Analysis. *Epidemiology* **32,** 617–624. ISSN: 1044-3983. https://journals.lww.com/epidem/Fulltext/2021/09000/The_Importance_of_Making_Assumptions_in_Bias.1.aspx (2021).

303. Blum, M. R., Tan, Y. J. & Ioannidis, J. P. A. Use of E-values for Addressing Confounding in Observational Studies—an Empirical Assessment of the Literature. *International Journal of Epidemiology* **49,** 1482–1494. ISSN: 0300-5771. https://doi.org/10.1093/ije/dyz261 (2020).

304. Ioannidis, J. P., Tan, Y. J. & Blum, M. R. Limitations and Misinterpretations of E-Values for Sensitivity Analyses of Observational Studies. *Annals of Internal Medicine* **170,** 108–111. ISSN: 0003-4819. https://www.acpjournals.org/doi/10.7326/M18-2159 (2019).

305. Bauman, A. E. *et al.* Physical Activity 2: Correlates of Physical Activity: Why Are Some People Physically Active and Others Not? *The Lancet* **380,** 258–271. ISSN: 01406736. pmid: 22818938 (2012).

306. Fox, M. P., MacLehose, R. F. & Lash, T. L. in *Applying Quantitative Bias Analysis to Epidemiologic Data* 197–231 (Springer, Cham, 2021). https://ezproxy-prd.bodleian.ox.ac.uk:2120/chapter/10.1007/978-3-030-82673-4_7 (2022).

307. Fox, M. P., MacLehose, R. F. & Lash, T. L. in *Applying Quantitative Bias Analysis to Epidemiologic Data* 233–290 (Springer, Cham, 2021). https://ezproxy-prd.bodleian.ox.ac.uk:2120/chapter/10.1007/978-3-030-82673-4_8 (2022).

308. Fox, M. P., MacLehose, R. F. & Lash, T. L. in *Applying Quantitative Bias Analysis to Epidemiologic Data* 291–327 (Springer, Cham, 2021). https://ezproxy-prd.bodleian.ox.ac.uk:2120/chapter/10.1007/978-3-030-82673-4_9 (2022).

309. Fox, M. P., MacLehose, R. F. & Lash, T. L. in *Applying Quantitative Bias Analysis to Epidemiologic Data* 415–440 (Springer, Cham, 2021). https://ezproxy-prd.bodleian.ox.ac.uk:2120/chapter/10.1007/978-3-030-82673-4_12 (2022).

310. Lawlor, D. A., Tilling, K. & Davey Smith, G. Triangulation in Aetiological Epidemiology. *International Journal of Epidemiology* **45,** 1866–1886. ISSN: 0300-5771. https://doi.org/10.1093/ije/dyw314 (2016).

311. *Reproducible Analytical Pipelines - Data in Government* https://dataingovernment.blog.gov.uk/2017/03/27/reproducible-analytical-pipeline/ (2022).

312. Goldacre, B., Morley, J. & Hamilton, N. *Better, Broader, Safer: Using Health Data for Research and Analysis* (2022), 112.

313. Chaput, J.-P., Olds, T. & Tremblay, M. S. Public Health Guidelines on Sedentary Behaviour Are Important and Needed: A Provisional Benchmark Is Better than No Benchmark at All. *British Journal of Sports Medicine* (2018).

314. Latimer-Cheung, A. E. *et al.* The Canadian 24-Hour Movement Guidelines for Children and Youth: Implications for Practitioners, Professional and Organizations and Strategies for Dissemination and Implementation. ISSN: 1715-5312. https://tspace.library.utoronto.ca/handle/1807/72619 (2016).

315. Thompson, D., Peacock, O., Western, M. & Batterham, A. M. Multidimensional Physical Activity: An Opportunity, Not a Problem. *Exercise and Sport Sciences Reviews* **43,** 67–74. ISSN: 15383008. pmid: 25607280. /pmc/articles/PMC4374165/ (2015).

316. Ekelund, U. *et al.* Joint Associations of Accelerometer Measured Physical Activity and Sedentary Time with All-Cause Mortality: A Harmonised Meta-Analysis in More than 44 000 Middle-Aged and Older Individuals. *British Journal of Sports Medicine* **54,** 1499–1506. ISSN: 14730480. pmid: 33239356. http://dx.doi.org/10.1136/bjsports-2020-103270 (2020).

317. Sallis, J. F., Owen, N. & Fotheringham, M. J. Behavioral Epidemiology: A Systematic Framework to Classify Phases of Research on Health Promotion and Disease Prevention. *Annals of Behavioral Medicine.* ISSN: 08836612. pmid: 11253440 (2000).

318. Axivity. *AX9 Data Sheet* 2015. https://axivity.com/files/resources/AX9_Data_Sheet.pdf.

319. Stevens, M. L. *et al.* Thigh-Worn Accelerometry for Measuring Movement and Posture across the 24-Hour Cycle: A Scoping Review and Expert Statement. *BMJ Open Sport & Exercise Medicine* **6,** e000874. ISSN: 2055-7647. https://bmjopensem.bmj.com/content/6/1/e000874 (2020).

320. Stamatakis, E. *et al.* Emerging Collaborative Research Platforms for the next Generation of Physical Activity, Sleep and Exercise Medicine Guidelines: The Prospective Physical Activity, Sitting, and Sleep Consortium (ProPASS). *British Journal of Sports Medicine* **54,** 435–437. ISSN: 14730480. pmid: 31076396. http://bjsm.bmj.com/ (2020).

321. Crowley, P. *et al.* Comparison of Physical Behavior Estimates from Three Different Thigh-Worn Accelerometers Brands: A Proof-of-Concept for the Prospective Physical Activity, Sitting, and Sleep Consortium (ProPASS). *International Journal of Behavioral Nutrition and Physical Activity* **16,** 65. ISSN: 1479-5868. https://doi.org/10.1186/s12966-019-0835-0 (2019).

322. Stewart, T. *et al.* A Dual-Accelerometer System for Classifying Physical Activity in Children and Adults. *Medicine & Science in Sports & Exercise* **50,** 2595–2602. ISSN: 0195-9131. https://journals.lww.com/acsm-msse/Fulltext/2018/12000/A_Dual_Accelerometer_System_for_Classifying.25.aspx (2018).

323. British Heart Foundation Data Science Centre. *Workshop Report – How Can Consumer Wearables Transform Our Understanding of Cardiovascular Disease?* (Zenodo, 2022). https://zenodo.org/record/5827260 (2022).

324. *The BED Study* The BED Study. https://www.bedstudy.co.nz (2022).

325. Karpe, F. *et al.* Cohort Profile: The Oxford Biobank. *International Journal of Epidemiology* **47,** 21–21g. ISSN: 0300-5771. https://doi.org/10.1093/ije/dyx132 (2018).

326. Arnold, K. F. *et al.* Reflection on Modern Methods: Generalized Linear Models for Prognosis and Intervention—Theory, Practice and Implications for Machine Learning. *International Journal of Epidemiology* **49,** 2074–2082. ISSN: 0300-5771. https://doi.org/10.1093/ije/dyaa049 (2020).

327. Hippisley-Cox, J., Coupland, C. & Brindle, P. Development and Validation of QRISK3 Risk Prediction Algorithms to Estimate Future Risk of Cardiovascular Disease: Prospective Cohort Study. *BMJ* **357,** j2099. ISSN: 1756-1833. pmid: 28536104. https://www.bmj.com/content/357/bmj.j2099 (2017).

328. Stefanick, M. L. *et al.* Women's Health Initiative Strong and Healthy Pragmatic Physical Activity Intervention Trial for Cardiovascular Disease Prevention: Design and Baseline Characteristics. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences* **76,** 725–734. ISSN: 1758-535X. pmid: 33433559 (2021).

329. Bradford Hill, A. The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine* **58,** 295–300. ISSN: 0035-9157. pmid: 14283879. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1898525/ (1965).

330. Dempsey, P. C. *et al.* Sedentary Behavior and Chronic Disease: Mechanisms and Future Directions. *Journal of Physical Activity & Health* **17,** 52–61. ISSN: 1543-5474. pmid: 31794961 (2020).

331. Dunstan, D. W., Dogra, S., Carter, S. E. & Owen, N. Sit Less and Move More for Cardiovascular Health: Emerging Insights and Opportunities Nature Reviews — Cardiology. *Nature Reviews Cardiology,* 1–12. www.nature.com/nrcardio (2021).

332. Carley, D. W. & Farabi, S. S. Physiology of Sleep. *Diabetes Spectrum : A Publication of the American Diabetes Association* **29,** 5–9. ISSN: 1040-9165. pmid: 26912958. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4755451/ (2016).

333. The Turing Way Community *et al. The Turing Way: A Handbook for Reproducible Data Science* https://zenodo.org/record/3233986 (2021) (Zenodo, 2019).

334. Migueles, J. H., Rowlands, A. V., Huber, F., Sabia, S. & van Hees, V. T. GGIR: A Research Community–Driven Open Source R Package for Generating Physical Activity and Sleep Outcomes From Multi-Day Raw Accelerometer Data. *Journal for the Measurement of Physical Behaviour* **2,** 188–196. ISSN: 2575-6605, 2575-6613. https://journals.humankinetics.com/view/journals/jmpb/2/3/article-p188.xml (2019).

335. Wilson, G. *et al.* Good Enough Practices in Scientific Computing. *PLOS Computational Biology* **13,** e1005510. ISSN: 1553-7358. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005510 (2017).

336. Van den Boogaart, K. G., Tolosana-Delgado, R. & Bren, M. *Compositions: Compositional Data Analysis* version 2.0-4. 2022. https://CRAN.R-project.org/package=compositions (2022).

337. Templ, M. *et al. robCompositions: Compositional Data Analysis* version 2.3.1. 2021. https://CRAN.R-project.org/package=robCompositions (2022).

338. Stanford, T. *The Deltacomp Package* 2021. https://github.com/tystan/deltacomp (2022).

339. McGregor, D. & Chastin, S. *OpenCoDa* 2021. opencoda.net.

340. *Snapshot Tests* https://testthat.r-lib.org/articles/snapshotting.html (2022).

341. *Code Complete, 2nd Edition* https://learning.oreilly.com/library/view/code-complete-2nd/0735619670/ (2022) ().