Body Size in Early Life and Risk of Diabetes in Adulthood



Dr Md Shajedur Rahman Shawon
Brasenose College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Michaelmas 2019

This thesis is dedicated to
my parents and my wife
for helping me each day to become a better person

Acknowledgements

The work presented in this thesis is only possible because of the 1.3 million women who participated in the Million Women Study, and the many researchers, data management and data entry staff, information technology and support staff of the Cancer Epidemiology Unit at the University of Oxford, past and present, who have worked on this study for more than two decades.

I am most grateful to my supervisors, Professor Dame Valerie Beral and Dr Owen Yang, for giving me the opportunity to study for my DPhil, and for their unfailing academic support, guidance and life advice. I have also benefited immensely from discussions with other colleagues in the Cancer Epidemiology Unit and wider Nuffield Department of Population Health. I would like to thank Jane Green, Gillian Reeves, Benjamin Cairns, Kezia Gaitskell, Dexter Canoy, Sarah Atkinson, Anna Brown, Elinor Smith, Sau Wan Kan, and Julie Schmidt. I am particularly indebted to my fellow DPhil students (Robert Smith, Jae Won Suh, Ai Seon Kuan, Ellie Watts, Seamus Kent, Karima Chiuri, Wemimo Omiyale, Angel Wong, and Yvonne Joko Walburga Fru) who provided continuous mental support and courage during this journey.

I am also very grateful to the Nuffield Department of Population Health not only for the generous funding to undertake my DPhil research but also for the opportunities for training and career support. I would also like to acknowledge the support of the Medical Sciences Division, and of Brasenose College.

I would like to thank my wife, Fariha Binte Hossain, who in the past three years has tolerated me and my somewhat crazy work schedule. My daughter, Ariana Mahnoor made me happy and strong to strive for progress. My parents and my siblings supported my DPhil journey in all possible ways. I would also like to thank my close friends: Gourab Adhikary, Sultan Mahmood Sami, Syed Bahauddin Alam, and Rayhan Ali Mollah for listening to my epidemiology discussions more than they perhaps desired or anticipated.

Finally, I am grateful to Allah for His countless blessings on me. *Alhamdulillah.*

Abstract

Body Size in Early Life and Risk of Diabetes in Adulthood

Md Shajedur Rahamn Shawon, Brasenose College, University of Oxford
A thesis submitted for the degree of Doctor of Philosophy, Michaelmas term 2019

Larger body size in adulthood increases the risk of diabetes in adulthood, whereas, paradoxically, being small at birth is associated with an increased risk of adult-onset diabetes. Direct comparisons of diabetes risks for different trajectories or combinations of body size at birth, childhood, early adulthood and late adulthood will help to better understand the complex relationships between body size throughout life and risk of adult-onset diabetes.

This thesis investigates the effects of early- versus late-life body size on risks of developing diabetes in adulthood among 413,516 women from the Million Women Study, with more than 24,000 diabetes cases identified through electronic linkage to routinely-collected national hospital admissions records. At the beginning of follow-up (mean age 60 [SD 5] years), women reported their birthweight, body size at ages 10 and 20 and current body mass index (BMI), which were validated against values recorded around respective ages. Cox proportional hazards regressions yielded multivariable-adjusted relative risks for adult-onset diabetes by different combinations of these four body size indices.

Within every body size category at younger ages, higher adult BMI sharply increased risks of adult-onset diabetes. The association between low birthweight and higher risks of diabetes was clear, independent and persistent among women with different combinations of body sizes at age 10, 20, and 60 years. Larger body size at ages 10 and 20 years was largely relevant to adult-onset diabetes risk as a predictor of adult adiposity. At every level of birthweight and adult BMI, there were additional risks of diabetes associated with being comparatively thin at age 10 years. However, after allowing for birthweight, size at age 10, and adult BMI, being small at age 20 years had a little additional effect on adult-onset diabetes risk.

In conclusion, as well as the known increased risk with adult adiposity, being small at birth and thin in childhood independently and persistently increase diabetes risk in adulthood.

Contents

List of Figures **xv**

List of Tables **xix**

List of Abbreviations **xxi**

1 Introduction to the thesis **1**
1.1 Background and rationale 2
1.2 Research aims and objectives 4
1.3 Outline of the thesis 6

2 Background **9**
2.1 Diabetes 10
2.1.1 Glucose metabolism and pathophysiology of diabetes 10
2.1.2 Diagnosis of diabetes 11
2.1.3 Classifications of diabetes 13
2.1.4 "Adult-onset diabetes" in this thesis 14
2.2 Burden and trends of diabetes 16
2.3 Established and proposed risk factors of diabetes 18
2.3.1 Genetics 18
2.3.2 Age 18
2.3.3 Ethnicity 19
2.3.4 Family history of diabetes 20
2.3.5 Socioeconomic status 21
2.3.6 Adult adiposity 21
2.3.7 Smoking 23
2.3.8 Alcohol consumption 24
2.3.9 Physical inactivity 24
2.3.10 Co-morbidities 25
2.4 Body size at different ages 26
2.5 Burdens of low birthweight and obesity among children and adults . 26
2.6 Genetics of body size at different ages 27
2.7 Tracking of body size from birth to early life to later life 28
2.8 Conclusion 28

3 Research questions and systematic reviews 29
3.1 Research questions . . . 30
3.1.1 Independent effect, direct effect, and separate effect . . . 31
3.1.2 Fine vs broad stages of early life . . . 32
3.1.3 Multiple mechanisms and apparent associations . . . 33
3.1.4 Choice of body size indicators . . . 35
3.1.5 Effect modification by body size at other ages . . . 36
3.1.6 Measurement errors in body size indices . . . 38
3.1.7 Adult-onset diabetes is not a simple concept . . . 39
3.2 Systematic reviews . . . 39
3.2.1 Birthweight and risk of adult-onset diabetes: systematic review 40
3.2.2 Body size in childhood and risk of adult-onset diabetes: systematic review . . . 48
3.2.3 Body size in early adulthood and risk of adult-onset diabetes: systematic review . . . 53
3.3 Approaches to investigate separate effects of body size at different ages 56
3.4 Conclusion . . . 57

4 The Million Women Study 59
4.1 The Million Women Study . . . 60
4.1.1 Recruitment . . . 61
4.1.2 Re-surveys . . . 63
4.1.3 Electronic record-linkage . . . 63
4.1.4 Ethical approval and funding . . . 65
4.2 Record-linkage databases . . . 66
4.2.1 National Health Services (NHS) Central Registers . . . 66
4.2.2 Hospital Episode Statistics (HES) . . . 66
4.2.3 Scottish Morbidity Records . . . 68
4.2.4 Clinical Practice Research Datalink (CPRD) . . . 68
4.3 Body size at different ages . . . 70
4.3.1 Adult Body mass index (BMI) . . . 72
4.3.2 Birthweight . . . 74
4.3.3 Body size at ages 10 and 20 years . . . 75
4.4 Correlations between body size indices at different ages . . . 78
4.5 Diabetes recording . . . 81
4.5.1 Diabetes in the study questionnaires . . . 81
4.5.2 Diabetes recording in hospital admissions data . . . 83
4.5.3 Diabetes recording in primary care data . . . 84
4.6 Diabetes recording in HES and CPRD . . . 86

4.6.1 Rate of diabetes recording in hospital admissions data and primary care data . . . 86
4.6.2 Probability of first hospital admission with diabetes records among those with a diabetes record in primary care . . . 88
4.6.3 Recording of diabetes in primary care among those with diabetes in hospital admissions . . . 92
4.7 Conclusion . . . 95

5 Analysis framework of prospective investigations 97
5.1 Study baseline . . . 98
5.2 Exclusions and handling missing information . . . 100
5.3 Follow-up . . . 103
5.4 Outcome: adult-onset diabetes . . . 103
5.5 Body size indices . . . 104
5.6 Statistical analysis . . . 104
5.6.1 Checking for proportional hazards assumption . . . 105
5.6.2 Adjustment for covariates and modelling strategy . . . 105
5.6.3 Regression dilution . . . 107
5.6.4 Effect measure modification and subgroups analysis . . . 107
5.6.5 Investigating effect modification by body size at other ages . 108
5.6.6 Additional analyses . . . 110
5.6.7 Presentation of results . . . 112
5.7 Implications in this thesis . . . 112

6 Adult BMI and adult-onset diabetes 113
6.1 Introduction . . . 114
6.2 Aim of this chapter . . . 115
6.3 Methods . . . 116
6.3.1 Study design and participants . . . 116
6.3.2 Adult BMI . . . 116
6.3.3 Outcome . . . 117
6.3.4 Statistical analysis . . . 117
6.4 Results . . . 119
6.4.1 Association between adult BMI and risk of adult-onset diabetes 121
6.4.2 Effect modification by other factors . . . 123
6.4.3 Sensitivity analyses . . . 125
6.5 Discussion . . . 127
6.5.1 Main findings . . . 127
6.5.2 Results in relation to other studies . . . 127

6.5.3 Effects of other factors on the association between adult BMI and diabetes risk . . . 129
6.5.4 Strengths and limitations . . . 129
6.6 Conclusion . . . 131

7 Birthweight and adult-onset diabetes 133
7.1 Introduction . . . 134
7.2 Aim of this chapter . . . 135
7.3 Methods . . . 136
7.3.1 Study design and participants . . . 136
7.3.2 Birthweight . . . 136
7.3.3 Outcome . . . 137
7.3.4 Statistical analysis . . . 137
7.4 Results . . . 140
7.4.1 Association between birthweight and adult-onset diabetes . . . 142
7.4.2 Association of birthweight and risk of adult-onset diabetes among women with different levels of adult adiposity . . . 144
7.4.3 Association between high birthweight and adult-onset diabetes risk . . . 146
7.4.4 Effect modification by other factors . . . 149
7.4.5 Additive interaction between the effects of lower birthweight and higher adult BMI . . . 151
7.4.6 Sensitivity analyses . . . 152
7.5 Discussion . . . 154
7.5.1 Main findings . . . 154
7.5.2 Results in relation to other studies . . . 154
7.5.3 Potential mechanisms . . . 157
7.5.4 Strengths and limitations . . . 157
7.6 Conclusion . . . 159

8 Body size at ages 10 and 20 years and adult-onset diabetes 161
8.1 Introduction . . . 162
8.2 Aim of this chapter . . . 163
8.3 Methods . . . 164
8.3.1 Study design and participants . . . 164
8.3.2 Body size at ages 10 and 20 years . . . 164
8.3.3 Outcome . . . 165
8.3.4 Statistical analysis . . . 165
8.4 Results . . . 168

8.4.1 Risks of adult-onset diabetes associated with body size at ages 10 and 20 years . . . 170
8.4.2 Risks of adult-onset diabetes associated with body size at ages 10 and 20 years among women with different levels of adult adiposity . . . 176
8.4.3 Risks of adult-onset diabetes associated with body size at ages 10 and 20 years among women with different levels of birthweight . . . 178
8.4.4 Effect modification by other factors . . . 180
8.4.5 Sensitivity analysis . . . 183
8.5 Global analysis of body size over the life course . . . 186
8.6 Discussion . . . 189
8.6.1 Main findings . . . 189
8.6.2 Results in relation to other studies . . . 189
8.6.3 Potential mechanisms . . . 191
8.6.4 Strengths and limitations . . . 192
8.7 Conclusion . . . 194

9 Strengths and limitations **195**
9.1 Introduction . . . 196
9.2 Limitations of analyses in this thesis . . . 196
9.2.1 Issues with reported body size indices at different ages . . . 196
9.2.2 Issues with identifying diabetes cases from hospital admissions data . . . 199
9.2.3 Issues related to reverse causation . . . 202
9.2.4 Issues related to residual confounding . . . 203
9.2.5 Issues related to missing data . . . 203
9.2.6 Baseline prevalence of diabetes and "healthy volunteer effect" 204
9.2.7 Issues with analysing correlated body size data . . . 205
9.2.8 Issues with generalising the findings . . . 205
9.3 Strengths of this thesis . . . 207

10 Discussion **209**
10.1 Aim of my DPhil research . . . 210
10.2 Main findings from this thesis . . . 210
10.3 Body size in early life and diabetes in adulthood: the totality of evidence . . . 212
10.3.1 Questions answered in this thesis . . . 212
10.3.2 Questions remaining unanswered in this thesis . . . 212
10.4 Added value of this thesis . . . 213

10.4.1 Validity of evidence 213
10.4.2 Relevance to the field 214
10.4.3 Providing new hypotheses for underlying pathophysiology . . 215
10.4.4 Public health implications 215
10.4.5 Insights on diabetes recording in hospital admissions data . 216
10.5 Recommendations for future research 216
10.6 Conclusions 218

Appendices

A Systematic review **221**
A.1 Association between birthweight and risk of diabetes in adulthood . 221
A.1.1 Search terms 221
A.2 Meta-analysis on the association between birthweight and risk of adult-onset diabetes 229
A.3 Body size in early life and adult-onset diabetes 231
A.3.1 Search terms 231

B Relevant study information **241**
B.1 Million Women Study 3-year re-survey questionnaire 241
B.2 Letter confirming ethical approval for use of MWS-CPRD linked data 246

References **249**

List of Figures

1.1 Approaches taken in this thesis to achieve the research aims and objectives 5

2.1 Feedback loop between pancreatic β-cells and insulin sensitive tissues and pathophysiology of diabetes 11
2.2 Age-adjusted prevalence of diabetes in adults (aged 20-79 years) worldwide in 2017 17
2.3 Doctor-diagnosed diabetes among minority ethnic groups in England and Wales in 2004 20
2.4 Association between adult body mass index and type 2 diabetes . . 22

3.1 Simplistic view of research questions asked in this thesis 30
3.2 Schematic representation of the association between birthweight and risk of developing diabetes in adulthood 32
3.3 Simplified path diagram of the relationships between body size at age 9, 10, and 11 years and risk of developing diabetes in adulthood 33
3.4 Simplified path diagram showing the roles of genetic factors and maternal nutrition on the apparent association between birthweight and risk of developing diabetes in adulthood 34
3.5 Example of path diagrams on the apparent association between birthweight and risk of developing diabetes in adulthood 35
3.6 Example of path diagrams on the apparent associations of BMI and waist:hip ratio with risk of developing diabetes in adulthood 36
3.7 Hypothetical associations between birthweight and risk of diabetes in adulthood when looked separately by adult body size 37
3.8 Errors in reported body size indices and possible effects on their apparent associations with the risk of diabetes in adulthood 38
3.9 Association between birthweight categories and risk of adult-onset diabetes in large-scale prospective studies 44
3.10 Association between birthweight and risk of adult-onset diabetes in large-scale prospective cohort studies: meta-analysis 46

4.1 Location of the 66 NHS breast screening clinics through which Million Women Study participants were recruited 62

4.2 Mean measured BMI versus mean self-reported BMI, in each self-reported BMI category . . . 73
4.3 Mean recorded birthweight versus mean reported birthweight, in each reported birthweight category . . . 74
4.4 Mean measured BMI at age 11 years versus categories of relative body size at age 10 years . . . 76
4.5 Mean measured BMI at age 20 years versus categories of body size at age 20 years . . . 77
4.6 Correlation coefficients between lifetime body size variables in the Million Women Study . . . 78
4.7 Correlations between indices of body size at different ages . . . 79
4.8 Correlations between indices of body size at different ages . . . 80
4.9 Rate of diabetes recording by age in hospital admissions data and primary care data . . . 87
4.10 Schematic diagram of the analysis for comparison of diabetes recorded in the HES with diabetes recorded in CPRD . . . 88
4.11 Cumulative probability of first admission to hospital with diabetes, by time since first record of diabetes in primary care . . . 89
4.12 Diabetes recording in Hospital Episode Statistics (HES) among women who had diabetes recorded in Clinical Practice Research Datalink (CPRD), by body size variables . . . 90
4.13 Diabetes recording in Hospital Episode Statistics (HES) among women who had diabetes recorded in Clinical Practice Research Datalink (CPRD), by socioeconomic and lifestyle variables . . . 91
4.14 Schematic diagram of the analysis for agreement for diabetes recording between hospital admission and primary care . . . 93

5.1 Flowchart of participant exclusions for the analysis . . . 101

6.1 Relative risks (RRs) for diabetes by adult body mass index (BMI) . . . 121
6.2 Relative risks (RRs) for diabetes per 5 kg/m^2 increase in BMI: effect of adjustment by various factors . . . 123
6.3 Relative risks (RRs) for diabetes per 5 kg/m^2 increase in BMI, by various characteristics at baseline . . . 124

7.1 Relative risks (RRs) for diabetes, by birthweight . . . 142
7.2 Relative risks (RRs) for diabetes per 1 kg increase in birthweight: effect of adjustment by various factors . . . 143
7.3 Relative risks (RRs) for diabetes by birthweight and adult body mass index (BMI) . . . 145

7.4 Association between birthweight and risk of diabetes in adulthood, using piece-wise linear spline regressions with cubic knots at 10^{th} 50^{th} and 90^{th} percentiles 147
7.5 Association between birthweight and risk of diabetes in adulthood, using piece-wise linear spline regressions with cubic knots at 2.5, 3.2, and 3.9 kg 148
7.6 Relative risks (RRs) for diabetes per 1 kg increase in birthweight, by various characteristics at baseline 150

8.1 Relative risks (RRs) for diabetes, by body size at ages 10 and 20 years 171
8.2 Relative risks (RRs) for diabetes by body size at age 10 years: effect of adjustment by various factors 173
8.3 Relative risks (RRs) for diabetes by body size at age 20 years: effect of adjustment by various factors 175
8.4 Relative risks (RRs) for diabetes by body size at ages 10 and 20 among women with different levels of adult body mass index (BMI) 177
8.5 Relative risks (RRs) for diabetes by (A) birthweight and body size at age 10 and (B) birthweight and body size at age 20 179
8.6 Relative risks (RRs) for diabetes by body size at age 10 years, by various characteristics at baseline 181
8.7 Relative risks (RRs) for diabetes by body size at age 20 years, by various characteristics at baseline 182
8.8 Relative risks (RRs) for diabetes for per kg decrease in birthweight, thin vs average at age 10 years and per 5 kg/m^2 increase in BMI at age 60 years, by other body size indices 188

A.1 Meta-analysis of the association between birthweight and risk of adult-onset diabetes in all published prospective cohort studies . . . 229

List of Tables

2.1 Subgroups of adult-onset diabetes based on cluster analysis of six related variables* 15

3.1 Summary information of prospective studies with 500 or more cases investigating the association between birthweight and risk of diabetes in adulthood 42
3.2 Key findings from large-scale prospective studies (500 or more incident cases) investigating the association between childhood body size and risk of diabetes in adulthood 51
3.3 Key findings from large-scale prospective studies (500 or more incident cases) investigating the association of body size at early adulthood with risk of diabetes in adulthood 55
3.4 Body size variables and analyses in the large-scale studies examining the roles of body size in early life on adult-onset diabetes risk . . . 58

4.1 Electronic record-linkage to routinely collected NHS healthcare data in the Million Women Study 64
4.2 Details of follow-up in the Million Women Study (n=1,364,308) . . 65
4.3 Lifetime body size variables in the Million Women Study 71
4.4 Information on diabetes in different questionnaires of the Million Women Study 82
4.5 International Classification of Diseases 10$^{\text{th}}$ version (ICD-10) codes for diabetes 83
4.6 Diabetes recording in hospital admissions data in the Million Women Study 84
4.7 Categories of Read Codes for diabetes in the linked primary care data (CPRD) 85
4.8 Description of different dates used in CPRD data 85
4.9 Comparing diabetes recording between hospital admission and primary care 94

5.1 Selected characteristics of women at recruitment and at 3-year resurvey in the Million Women Study 99

5.2 Selected characteristics of women who reported and did not reported birthweight at 3-year re-survey in the Million Women Study 102

6.1 Characteristics of the study population by body mass index (kg/m^2) 120
6.2 Relative risks (RRs) for diabetes for women who were overweight and who were obese 122
6.3 Sensitivity analyses for the association between adult BMI and risk of diabetes in adulthood 126

7.1 Characteristics of the study population by birthweight 141
7.2 Relative risks (RRs) for diabetes, by birth weight including a separate category for ≥4.5 kg 146
7.3 Attributable effects to additive interaction between birthweight and adult BMI on the risk of adult-onset diabetes 151
7.4 Sensitivity analyses for the association between birthweight and risk of diabetes in adulthood 153

8.1 Characteristics of the study population by body size at ages 10 and 20 years 169
8.2 Sensitivity analyses for the association between body size at age 10 and risk of diabetes in adulthood 184
8.3 Sensitivity analyses for the association between body size at age 20 and risk of diabetes in adulthood 185
8.4 Relative risks (RRs) for diabetes by birthweight, body size at age 10, body size at age 20, and BMI at age 60 187

A.1 Search terms for studies investigating the association between birthweight and risk of diabetes in adulthood 222
A.2 Summary information of all included prospective studies investigating the association between birthweight and risk of diabetes in adulthood 223
A.3 Relative risks (RRs) of adult-onset diabetes per 1 kg increase in birthweight, meta-analysis results in subgroups by various factors . 230
A.4 Search terms in EMBASE for studies investigating the association between body size in early life and risk of diabetes in adulthood . . 231
A.5 Search terms in PubMed for studies investigating the association between body size in early life and risk of diabetes in adulthood . . 232
A.6 Summary information of all included prospective studies investigating the association between body size in childhood and risk of diabetes in adulthood 233
A.7 Summary information of all included prospective studies investigating the association of body size in adolescence and in early adulthood with risk of diabetes in adulthood 237

List of Abbreviations

BMI	Body mass index
CDC	Centers for Disease Control
CEU	Cancer Epidemiology Unit (at the University of Oxford)
CIs	Confidence Intervals
CPRD	Clinical Practice Research Datalink
E3N study	Etude Epidémiologique auprès de Femmes de la Mutuelle Générale de l'Education Nationale Study
FBS	Fasting Blood Glucose
GWAS	Genome-wide association studies
g-sCI	Group-specific Confidence Intervals
HbA1c	Glycated haemoglobin
HES	Hospital Episode Statistics
HR	Hazard Ratio
HRT	Hormone Replacement Therapy
ICD-10	International Statistical Classification of Diseases and Related Health Problems 10th revision
IDF	International Diabetes Federation
IDS Scotland	Information Service Division Scotland
IFG	Impaired fasting glucose
IGT	Impaired glucose tolerance
IOTF	International Obesity Task Force
ISAC	Independent Scientific Advisory Committee
IQR	Interquartile range
LADA	Latent autoimmune diabetes of adulthood
MELANY Study	Metabolic, Lifestyle and Nutrition Assessment in Young Adults Study

MR . . . . . Mendelian Randomisation

NDPH . . . . . Nuffield Department of Population Health

NHS . . . . . National Health Service

NSHD . . . . . National Survey of Health and Development

ONS . . . . . Office for National Statistics

PCOS . . . . . Polycystic Ovary Syndrome

SD . . . . . Standard Deviation

SMR . . . . . Scottish Morbidity Records

SNPs . . . . . Single-nucleotide Polymorphism

T2DM . . . . . Type 2 diabetes mellitus

QoF . . . . . Quality and Outcomes Framework

RDR . . . . . Regression Dilution Ratio

RR . . . . . Relative Risk

UK . . . . . United Kingdom

USA . . . . . United States of America

vs . . . . . versus

WHO . . . . . World Health Organisation

Role of the author in this thesis

The work presented in this thesis is my own, conducted under the supervision of Professor Dame Valerie Beral and Dr Owen Yang, except where otherwise stated. The initial concept of investigating the roles of body size in early life on the risk of adult-onset diabetes in the Million Women Study was conceived by Professor Beral, Dr Yang and Dr Miranda Armstrong (she was on the supervision team for first few months of my DPhil study). After the conception of the main aim for the DPhil study, I have been involved in all aspects of the research and have led the systematic review of literature, data analysis on the Million Women Study data, and interpretation of the results, while being supervised by Dr Yang and Professor Beral.

I conducted all searching and summarising of the literature, and wrote all the chapters; and my supervisors contributed to each aspect of research presented in this thesis. The validation of body size variables against recorded values was done previously by Dr Benjamin Cairns *et al.* and Dr Lucy Wright *et al.* from Nuffield Department of Population Health. To compare diabetes recording between primary care and hospital admissions data, I needed permission from an Independent Scientific Advisory Committee (ISAC) assigned by Clinical Practice Research Datalink (CPRD) for further use of an existing CPRD-Million Women Study linked dataset. As the Chief Investigator of this data-linkage, Professor Jane Green submitted the application and did all correspondences with the CPRD team. I helped her in preparing the research protocol for this application.

Other researchers from the Cancer Epidemiology Unit (Professors Jane Green and Professor Gillian Reeves) contributed to the research presented in Chapters 6 to 8, by providing critical comments on the analyses and interpretations of the findings, and commenting on drafts of the corresponding manuscript. Dr Sarah Floud also provided her comments on a poster based on this work. The idea of discussing potential limitations and strengths of this thesis in a separate chapter was given by the examiners of Confirmation of DPhil Status (Professor Jane Armitage and Professor Gillian Reeves).

All codes in Stata are written by me, and Dr Yang coached me how to streamline data management and analysis tasks on Stata. Professor Beral and Dr Yang helped me to design the figures and tables for efficient presentation. For figures drawn in R, I mainly used *Jasper* package developed by Matthew Arnold from Nuffield Department of Population Health.

Publications relating to this thesis

At the time of writing this thesis, one manuscript has been written from work described in this thesis and I am the first author of it. The co-authors of the paper provided valuable suggestions on the methods and contents on various occasions, and therefore, the manuscript may not be considered as to be entirely my own work.

1

Introduction to the thesis

1.1 Background and rationale

Diabetes mellitus (referred to as diabetes hereafter) is a major public health problem with large and increasing impacts on individuals, society and healthcare systems worldwide. The International Diabetes Federation (IDF) estimated that the number of people with diabetes worldwide increased from 108 million in 1980 to 463 million in 2019 [1]. Over the same period, the global prevalence of diabetes in individuals aged 18 years and above increased from 4.7% to 8.5% [2]. Diabetes is a major risk factor for cardiovascular disease and other morbidities, and approximately four million deaths in 2017 were associated with diabetes, and its complications [1]. A significant portion (i.e. 12%) of global health expenditure is spent on treating diabetes in England [3] and worldwide [1].

The relationships between body size over the life course and diabetes are complex. While diabetes may develop over lifetime, body size and its associations with diabetes do not stay the same throughout life. The commonest form of diabetes in adults, type 2 diabetes, has a strong association with being overweight and obese in adulthood [4, 5]. Increasing trends in overweight and obesity globally can largely explain the rapid rise of diabetes burden in recent times [6], although other potential drivers should not be forgotten. There has been accumulating evidence suggesting that the risk of diabetes in adulthood is also associated with body size at birth and in early life [7, 8]. Paradoxical to the adverse effects of high body mass index (BMI) in adulthood on diabetes, being small at birth has been found to be associated with an increased risk of the disease [7, 9]. To help understand these apparently contradictory findings, risks of adult-onset diabetes associated with body size at different ages, for example, at birth, in childhood, and early and late adulthood need to be assessed.

There are issues with analysing body size at different ages over the life course simultaneously [10]. Because body size in early life can predict body size in adulthood [11], it is expected that some parts of the associations between body size in early life and risk of developing diabetes in adulthood are mediated through body size in adulthood. Therefore, results on the associations between body size in early life and the risk of adult-onset diabetes with adjustment for body size in

adulthood from many previous studies [12–16] are difficult to interpret. One way of circumventing the difficulties in separating the effects of early -versus late-life body size on risks of developing diabetes in adulthood is to directly compare the effects for different trajectories or scenarios of body size throughout life [17], for example, small at birth and obese in adulthood. However, large numbers are needed to look into many specific groups reliably, and existing evidence is limited.

Large prospective studies with body size measures for different stages of life can provide substantial insights into how body size in early life may affect the risk of adult-onset diabetes, in addition to the strong risks associated with adult overweight and obesity. The Million Women Study is a population-based prospective study of 1.3 million UK women, recruited in 1996-2001 [18, 19] (details in chapter 4). Leveraging the large size of this cohort and the prospectively collected and validated measures on body size at birth, at ages 10, 20, and around 60 years, this thesis will try to understand the associations of body size in early life and risk of diabetes in adulthood. To identify the details of any specific or consistent associations of body size at different ages with the risk of developing diabetes, I will examine the associations in various combinations of body size at different ages. Such investigation also adds a life course perspective by recognising tracking of body size from early life to later life and their separate effects on adult-onset diabetes risk.

1.2 Research aims and objectives

The aim of the thesis is to understand the associations of body size in early life and the risk of developing diabetes in adulthood. The different stages of early life in this thesis will be: around the time of birth, mid-childhood and early adulthood. To understand how body size at these stages of early life may affect the risk of developing diabetes in adulthood, this thesis has the following objectives:

- To identify the knowledge gaps in current evidence on body size in early life and risk of diabetes in adulthood
- To understand the methodological issues with investigating associations simultaneously for body size at different ages over the life course
- To identify and interpret any separate and joint effects of body size at different ages on the risks of developing diabetes in adulthood

As shown in Figure 1.1, the approaches taken to achieve the research aims and objectives of this thesis are: systematic synthesis of the current evidence on the topic, understanding the data on body size and diabetes in the Million Women Study, and prospective investigations on the associations of body size at different ages and risk of adult-onset diabetes.

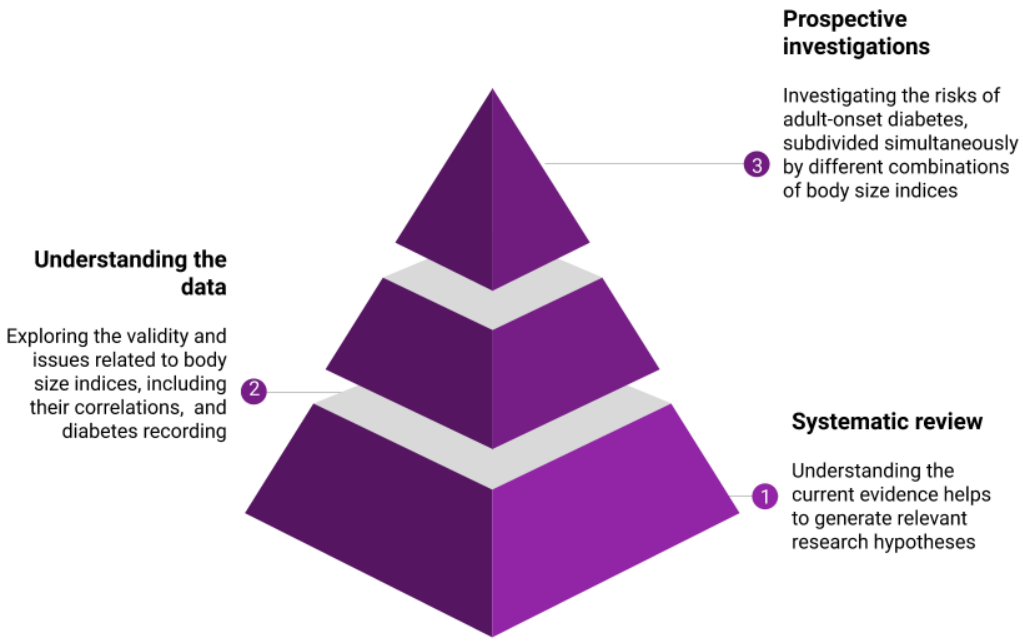


Figure 1.1: Approaches taken in this thesis to achieve the research aims and objectives

1.3 Outline of the thesis

This thesis has ten chapters.

- **Chapter 1: Introduction to the thesis** This chapter provides the rationale of the thesis. It also outlines the research aims, approaches and objectives of this thesis.

- **Chapter 2: Body size in early life and diabetes in adulthood: background** This chapter provides brief backgrounds to the topics of diabetes and body size at different ages across the life course.

- **Chapter 3: Body size in early life and diabetes in adulthood: research questions and systematic review** This chapter outlines the topics that needed careful reflections to understand the research questions involving body size in early life and risk of adult-onset diabetes. Then, it provides systematic reviews of existing evidence on the relationships of birthweight and body size in childhood and in early adulthood with adult-onset diabetes. Altogether, this chapter helps to establish a niche for the current investigations.

- **Chapter 4: The Million Women Study** This chapter provides sufficient details about the primary data source of the thesis, the Million Women Study. It also discusses various issues related to body size at different ages and recording of new diabetes cases which may have implications on the validity and interpretations of the conclusions in this thesis.

- **Chapter 5: Analysis framework of prospective investigations** This chapter introduces the analysis framework for this thesis - describing approaches taken to prospectively analyse data from the Million Women Study for understanding the associations of body size indices at different ages and risk of adult-onset diabetes.

- **Chapter 6: Adult body mass index (BMI) and adult-onset diabetes** Before investigating the effects of body size in early life on the risk of adult-onset diabetes, this chapter examines the association between adult body mass index (BMI) and risk of diabetes in adulthood in the Million Women Study. The findings in this chapter will help to interpret and contextualise the results on body size in early life and risk of diabetes in adulthood.

- **Chapter 7: Birthweight and adult-onset diabetes** This chapter provides a comprehensive understanding on how birthweight can influence the risk of developing diabetes among adults, mainly by investigating the issues pointed out in the systematic review.

- **Chapter 8: Body size at ages 10 and 20 years and adult-onset diabetes** This chapter investigates the associations of body size in childhood (at age 10 years) and in early life (at age 20 years) with the risk of adult-onset diabetes, after accounting for body size at birth and late adulthood. It then also provides a global analysis of the risk of adult-onset diabetes, subdivided by 3 categories of body size at 4 different ages: birthweight, body size at age 10, body size at age 20, and adult BMI.

- **Chapter 9: Strengths and limitations** This chapter discusses issues related to specific conditions in the Million Women Study that could have undermined the validity of the findings in this thesis.

- **Chapter 10: Discussion** This chapter summarises the main findings of this thesis and then discussed the added value of this thesis. It also reflects what and how future research in this topic should move forward.

2

Body size in early life and diabetes in adulthood: background

Summary points

- Diabetes is a serious global health problem due to its large burden, and strong associations with morbidities, death and healthcare expenditures.
- Diabetes in adulthood can be heterogeneous with regard to its aetiology, underlying pathophysiology, and clinical presentations.
- The likelihood of developing adult-onset diabetes increases rapidly with age and adult adiposity.
- Size at birth is associated with body size in early life, while early life adiposity is more likely to persist into adulthood.

The aim of this thesis is to understand the associations of body size at different stages of early life on the risk of developing diabetes in adulthood. This chapter provides brief backgrounds to the topics of diabetes and body size over the life course.

2.1 Diabetes

Diabetes is a medical condition characterised by increased levels of glucose in blood (i.e. hyperglycaemia), which, if persistent, can lead to significantly increased risks of developing complications. The long-term complications of diabetes are usually grouped into macrovascular (i.e. those involving large blood vessels) and microvascular (i.e. those involving finer blood vessels). The macrovascular complications of diabetes are coronary artery disease, peripheral arterial disease, and stroke. The microvascular complications are diabetic nephropathy, neuropathy, and retinopathy.

2.1.1 Glucose metabolism and pathophysiology of diabetes

In healthy adults, the glucose level in blood is regulated within a narrow range (between 4.0 to 5.9 mmol/L). Insulin, a hormone secreted from pancreatic β cells, is the primary regulator of glucose metabolism in our body [20–22]. Glucose metabolism is maintained through coordinated actions of multiple organs, particularly liver, intestine, skeletal muscles, and brain. A feedback loop is established between β-cell function and tissue sensitivity to insulin, in which insulin sensitivity affects β-cell response. This feedback loop between β-cell function and tissue sensitivity to insulin are probably mediated through neuronal and humoral mechanisms, but exact mechanisms are still unknown [22].

As shown in Figure 2.1, there are mainly two reasons for which glucose level in blood can rise to an abnormal level. When the feedback loop is disrupted by the destruction of β-cells in the pancreas it leads to absolute insulin deficiency. On the other hand, when increased resistance to actions of insulin in the peripheral tissues develops, initially feedback to β-cells ensures increased insulin secretion to maintain

normal blood glucose. But, when the feedback loop is permanently disrupted due to progressive β-cell failure, the glucose level in blood rises.

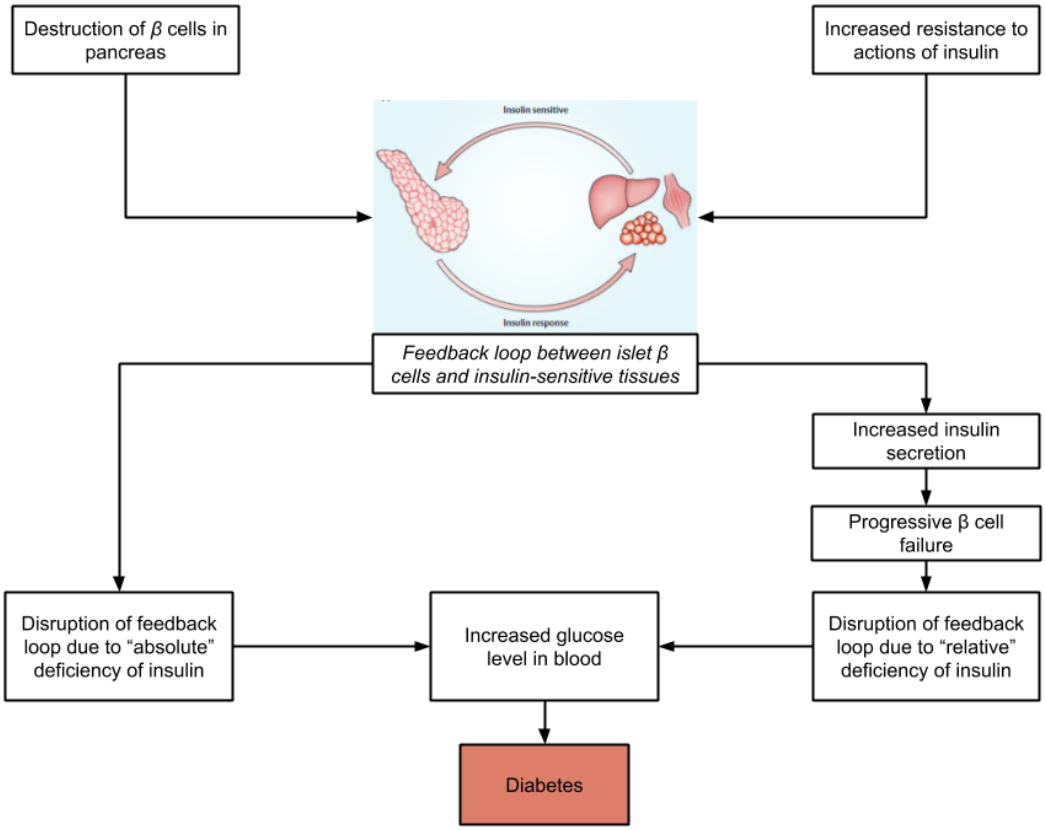


Figure 2.1: Feedback loop between pancreatic β-cells and insulin sensitive tissues and pathophysiology of diabetes

2.1.2 Diagnosis of diabetes

Clinically, diabetes is diagnosed based on arbitrary cut-off points of plasma glucose and Glycated haemoglobin (also known as *HbA1c*), based on probabilities of developing future complications [20, 21]. HbA1c can provide an accurate and objective measure of blood glucose control over a period of weeks to months [21].

According to the American Diabetes Association's 2019 guidelines [23], diabetes is confirmed by either:

- Plasma glucose $\geq$7.0 mmol/L (126 mg/dL) in samples taken in fasting state for at least 8 hours **or**

- Plasma glucose ≥11.1 mmol/L (200 mg/dL) in samples taken 2 hours after a 75 grams glucose load or randomly **or**
- HbA1c≥6.5% (48 mmol/mol)

One abnormal result is necessary to establish a definitive diagnosis of diabetes in individuals with typical symptoms of hyperglycaemia (e.g., increased frequency of urination, increased thirst, dry mouth, tiredness, weight loss etc.), whereas results should be confirmed by repeat testing in individuals with no symptoms. In the UK, the National Institute of Health and Care Excellence (NICE) considers similar diagnostic criteria for diabetes [24].

The term "Prediabetes" is used for those whose plasma glucose levels do not meet the criteria for diabetes but are too high to be considered normal. Studies have shown that individuals with prediabetes have higher risks for diabetes and cardiovascular disease in the future than those with normal plasma glucose levels [23]. These individuals can have impaired fasting glucose (IFG) and/or impaired glucose tolerance (IGT).

According to the American Diabetes Association' 2019 guidelines, diagnostic cut-offs for prediabetes are:

- IFG can be defined as fasting plasma glucose between 100 mg/dL (5.6 mmol/L) and 125 mg/dL (6.9 mmol/L)
- IGT can be defined as plasma glucose 140 mg/dL (7.8 mmol/L) to 199 mg/dL (11.0 mmol/L) in samples taken 2 hours after a 75 grams glucose load
- *HbA1c* 5.7–6.4% (39–47 mmol/mol) is also considered as prediabetes

2.1.3 Classifications of diabetes

Diabetes is a heterogeneous medical condition. The American Diabetes Association [23] broadly classifies diabetes into:

- Type 1 diabetes, 5-10%
- Type 2 diabetes, 90-95%
- Gestational diabetes
- Specific types of diabetes, e.g., monogenic diabetes syndromes (maturity onset diabetes of the young [MODY]), diabetes due to pancreatic diseases (cystic fibrosis, pancreatitis), and drug- or chemical-induced diabetes (e.g. glucocorticoid use)

Type 1 diabetes, previously named as insulin-dependent diabetes" or "juvenile-onset diabetes", accounts for 5-10% of all diabetes cases. The main pathology in type 1 diabetes is cellular-mediated autoimmune destruction of the pancreatic β-cells. In type 1 diabetes patients, one ore more of the following auto-antibodies are present: islet cell autoantibodies, autoantibodies to GAD (GAD65), insulin, the tyrosine phosphatases IA-2 and IA-2β, and ZnT8. About one-third of children with type 1 diabetes present with diabetic ketoacidosis and require emergency hospital admission for management [25].

Type 2 diabetes (previously known as "non-insulin dependent diabetes") accounts for 90–95% of all diabetes cases. It occurs when relative deficiency of insulin occurs as a result of peripheral insulin resistance. Due to its gradual onset, type 2 diabetes frequently goes undiagnosed for many years [20].

Gestational diabetes another form of diabetes that is first diagnosed in the second or third trimester of pregnancy that is not clearly either preexisting type 1 or type 2 diabetes. Generally, this form of diabetes resolves after the end of pregnancy, but they may lead to increased risk of type 2 diabetes in later life [23].

Latent autoimmune diabetes of adulthood (LADA) presents like type 2 diabetes but actually is a form of type 1 diabetes (due to the presence of Glutamic

Acid Decarboxylase autoantibodies (GAD)) [26]. It has an insidious onset but progresses rapidly to treatment with insulin. According to the UK Prospective Diabetes Study, LADA cases are found in between 6% and 10% of all diabetes cases among men and women aged 25–65 years [27].

2.1.4 "Adult-onset diabetes" in this thesis

This thesis concerns the form of diabetes that occurs in adulthood. The traditional paradigms of type 2 diabetes occurring only in adults and type 1 diabetes only in children are no longer valid, but the most common form of diabetes occurred in the middle-aged, and elderly is type 2 diabetes. For example, in the UK Biobank, of all people aged 31 to 60 at time of diabetes diagnosis, 96% have type 2 diabetes and cases of type 2 diabetes increase substantially with increasing age [28]. People are more prone to develop type 2 diabetes with increasing age because of the combined effects of increasing insulin resistance and impaired β-cell function with age. Age-related insulin resistance is associated with adiposity, physical inactivity and sarcopenia [29–32], while it is also evident that both β-cell proliferation and function decline with increasing age [33–35].

Adult-onset diabetes can be heterogeneous with regard to clinical presentations and progression to complications. Researchers in a recent study [36] proposed a more refined classification system for adult-onset diabetes based on cluster analysis of six variables, namely glutamate decarboxylase antibodies (GADA), age at diagnosis, BMI, HbA1c, and estimates of β-cell function and insulin resistance. They identified five clusters of adult-onset diabetes, as shown in table 2.1. While cluster 1 corresponds to the classic type 1 diabetes and LADA, and clusters 4 and 5 correspond to type 2 diabetes, clusters 2 and 3 are severe forms of diabetes previously masked within type 2 diabetes.

Diagnostic confusion can arise for diabetes diagnosed among adults mainly because of possible overlap between common and rarer forms of diabetes (i.e. LADA) [21]. Patients with LADA present like type 2 diabetes at diagnosis but then require treatment with insulin within the first few years after diagnosis and

further tests show that they have GADA antibodies. So, their diagnosis is later changed to type 1 diabetes in routine health data.

Table 2.1: Subgroups of adult-onset diabetes based on cluster analysis of six related variables*

Cluster	Characteristics	Percentage
Cluster 1: Severe autoimmune diabetes (SAID)	Occurred at relatively younger age, in people with a relatively low BMI, poor blood sugar control, insulin deficiency and presence of GADA	6.4%
Cluster 2: Severe insulin-deficient diabetes (SIDD)	GADA-negative but otherwise similar to SAID: low age at onset, relatively low BMI, low insulin secretion, and poor metabolic control	17.5%
Cluster 3: Severe insulin-resistant diabetes (SIRD)	It was characterised by high insulin resistance and high BMI	15.3%
Cluster 4: Mild obesity-related diabetes (MOD)	Characterised by obesity but not insulin resistance	21.6%
Cluster 5: Mild age-related diabetes (MARD)	People were generally older than those in other clusters and had only mild problems with glucose control, similar to MOD	39.1%

*Based on Ahlqvist *et. al* 2018 [36]

The variables used in the analysis were glutamate decarboxylase antibodies (GADA), age at diagnosis, BMI, HbA1c, and estimates of β-cell function and insulin resistance

2.2 Burden and trends of diabetes

Diabetes is the largest and fastest increasing global health problem of this century. According to the International Diabetes Federation (IDF) [1], approximately 463 million adults (aged 20-79 years) were living with diabetes in 2019, which by 2045 this will rise to 700 million. A recent study [37] using the data sources of the latest IDF Diabetes Atlas reported that of the 463 million people living with diabetes in 2019, half were unaware of their condition (i.e. undiagnosed). The prevalence of diabetes is generally higher in men than in women. Global age-standardised diabetes prevalence increased from $4 \cdot 3\%$ in 1980 to $9 \cdot 0\%$ n 2014 in men, and from $5 \cdot 0\%$ to $7 \cdot 9\%$ in women [2]. Most (79%) people with diabetes live in low- and middle-income countries (LMICs) (Figure 2.2), and the proportions of undiagnosed diabetes were much higher in these countries [37]. Globally, approximately 4.0 million deaths are attributable to diabetes, and some 12% of global health expenditure is spent on diabetes [1].

Diabetes prevalence in adults has increased, or at best remained unchanged, in every country since 1980; and the rates of increase are much higher in low and middle-income countries than in high-income countries [2]. A recent systematic review [38] analysed data on 121 separate sex-specific or ethnicity-specific populations and found that between 1990 and 2005, diabetes incidence increased in 66% of populations, whereas an increasing trend was reported in only 33% of populations between 2006 and 2014.

In the UK, 4.7 million people had diabetes (prevalence rate 4-6%) in 2018, and the number of people diagnosed with diabetes has more than doubled in the last 20 years [39]. These numbers are based on the Quality and Outcomes Framework (QoF) data in England, Wales, and Northern Ireland, and the Scottish Diabetes Survey in Scotland. Although the prevalence of diabetes is on the rise, recent reports suggest that the number of new diabetes cases has not increased in the UK in recent years [40, 41].

According to the latest estimates from the National Diabetes Audit Report 2017 [42], every year, diabetes is related to more than 27,000 cases of heart attacks,

almost 100,000 cases of heart failure and approximately 35,600 strokes. Individuals with diabetes are also 20 times more likely to experience an amputation when compared with those without diabetes. According to the current estimates, diabetes is responsible for 5% of all sight loss in the UK [42]. Recent studies [43, 44] have also suggested that people with diabetes have higher risk of suffering from depression and poor quality of life.

The National Health Service (NHS) spends at least £10 billion a year on diabetes, which is 10% of its entire budget [39].

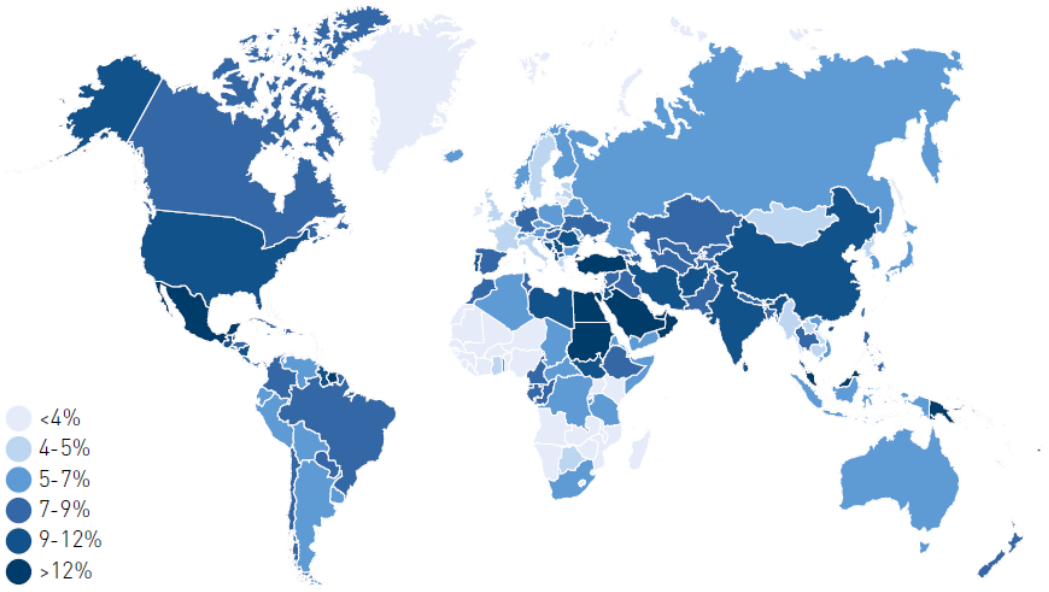


Figure 2.2: Age-adjusted prevalence of diabetes in adults (aged 20-79 years) worldwide in 2017

Figure adapted from the International Diabetes Federation (IDF) Atlas 2017 [1] under the terms of the Creative Commons Attribution License 4.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

2.3 Established and proposed risk factors of diabetes

Adult-onset diabetes is a multi-factorial disease, with many genetic and environmental determinants. This section briefly discusses important established and proposed risk factors of adult-onset diabetes.

2.3.1 Genetics

Adult-onset diabetes, particularly type 2 diabetes has a polygenic architecture. The genetic predisposition of type 2 diabetes is linked to many hundreds of mostly common alleles with small effects [45]. Genome-wide association studies (GWAS) have identified approximately 250 genomic regions (loci) in predisposition to type 2 diabetes [46–49]. These genetic variants range from 50% to 0.02% in minor allele frequency, and their per-allele effect size ranges between odds ratio (OR) 1.04 and OR 8.05 [45]. Genetic variants in the gene *TCF7L2* confer the largest effect on the risk of type 2 diabetes [45]. While many of these loci are related to β-cell function, others have been linked with obesity, insulin resistance, appetite regulation, and concentration of insulin and glucose [22, 45]. However, these genetic variants, together, do not explain much of the genetic basis of diabetes. Genotype risk scores for predicting type 2 diabetes risk performed only slightly better than clinical risk models based on factors like age, BMI, and ethnicity [50–52].

While genetic factors are important determinants of adult-onset diabetes, the recent rise in diabetes burden among adults worldwide cannot be explained by changes in the gene pool. Therefore, understanding the roles of both genetic and environmental factors is crucial to understanding the epidemic.

2.3.2 Age

Diabetes risk increases with increasing age, and therefore a higher burden of diabetes is expected in the ageing populations. The largest number of people living with diabetes worldwide are middle-aged and elderly [1]. A recent study, based on pooled data from 751 population-based studies, estimated that around 40% of the global rise

in the number of adults with diabetes was due to population growth and ageing. The number of new cases of diabetes increases sharply with increasing age until about age 65 years, after which it seems to plateau out [1, 39, 53]. One in every four individuals aged ≥65 years in the US has diabetes [53]. Recent estimates from primary care data in the UK showed that both incidence and prevalence of diabetes peaked at 70-79 years, after adjustment for sex, socioeconomic status, and calendar year [41].

2.3.3 Ethnicity

The risk of adult-onset diabetes varies widely by ethnicity, even within the same country of residence. According to the latest estimates from the Diabetes UK, South Asian and Black people living in the UK are two to four times more likely to develop Type 2 diabetes than white people [39]. Figure 2.3 shows the prevalence of doctor-diagnosed diabetes among minority ethnic groups in England and Wales, and it is clear that the prevalence is approximately three to five times higher in minority ethnic communities than in the while British population. Results from the Nurses' Health Study in the US showed that compared with whites, the age-adjusted relative risks (RRs) were 1.43 (95% CI 1.08–1.90) for Asians, 1.76 (1.32–2.34) for Hispanics, and 2.18 (1.82–2.61) for blacks [54]. Diabetes also occurs much earlier (around 10-12 years younger) in these groups than white population [55].

It is commonly assumed that ethnic differences in the prevalence of diabetes is primarily determined by obesity and lifestyle factors (e.g. diet and exercise) [55]. For example, among South Asian population, type 2 diabetes develops at a lower BMI compared with other ethnic groups [56]. However, a recent prospective investigation [54] showed that the risk of diabetes is significantly higher among ethnic minorities even after differences in BMI was taken into account. The risk of developing microvascular complications of diabetes and lower limb amputations are found to be higher in ethnic minority groups [57].

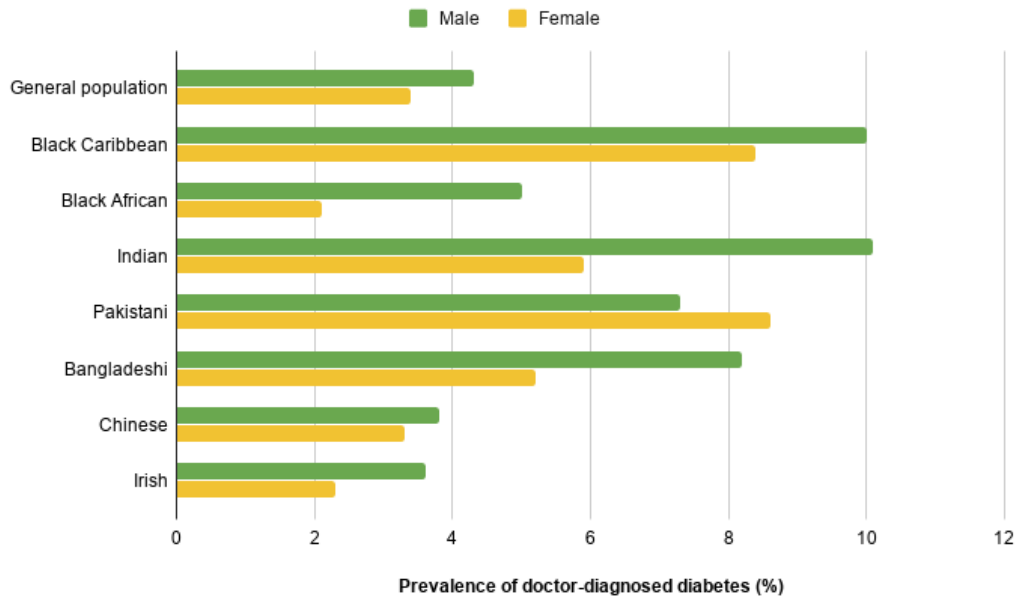


Figure 2.3: Doctor-diagnosed diabetes among minority ethnic groups in England and Wales in 2004

Reproduced from Becker E, Boreham R, Chaudhury M, Craig R, Deverill C, Doyle M et al. Health Survey for England 2004

2.3.4 Family history of diabetes

A family history of diabetes is associated with the development of type 2 diabetes. Individuals with a positive family history of diabetes (in one or more relatives) have a higher risk of diabetes in adulthood. Results from the EPIC-InterAct Study [58] showed that the family history-associated risk of diabetes could not be solely explained by differences in major risk factors, including BMI and physical inactivity. Previous studies have shown that although a positive family history of diabetes is associated with higher average BMI, variation in BMI can explain only 9-21% of the risk associated with family history [59]. It has also been shown that a single question on family history can outweigh the added predictive power of the genetic risk score, indicating its importance in clinical and epidemiological studies [58]. How family history contributes to the risk of adult-onset diabetes is not fully understood, but maternal history of diabetes is associated with both insulin resistance and impaired first-phase insulin secretion [60].

2.3.5 Socioeconomic status

The risk of getting diabetes in adulthood is higher in low socioeconomic groups, whether assessed by educational level, income, or occupation. According to a recent meta-analysis [61], low levels of education (RR 1.41, 95%CI 1.28-1.51), occupation (RR 1.31, 95% CI 1.09-1.57), and income (RR 1.40, 95%CI 1.04-1.88) were associated with increased risk of adult-onset diabetes, when compared with high levels of these determinants. The relationship between low socioeconomic status and the risk of diabetes is complex, and the pathways between them are not yet fully understood. Low socioeconomic position can affect an individual's access to health care services, availability of healthy foods, and various lifestyle choices [61, 62].

Connolly *et al.* [63] reported that in the UK diabetes prevalence in the least deprived quintile was 13.4 per 1000 for men compared with 17.2 per 1000 in the most deprived. For women, the corresponding numbers were 10.8 and 15.5 per 1000. Women in the lower socioeconomic groups tend to have a higher risk of diabetes than men in these groups [64]. The sex differences in relationships between socioeconomic status and diabetes risk could be explained by the fact that women in lower socioeconomic groups are more likely to be obese, physically inactive and experience psycho-social stress than men in these groups [65].

It is important to mention that the socioeconomic trends in diabetes burden seem to be different between high-income countries and LMICs - individuals with high socioeconomic status have higher burden of diabetes [61]. One explanation for such trends can be that unhealthy lifestyles (e.g. physical inactivity and consumption of energy-dense foods) are more common in the high socioeconomic groups. There is also evidence that childhood socioeconomic status is associated with the risk of adult-onset diabetes [66, 67].

2.3.6 Adult adiposity

Adult adiposity is responsible for a large proportion of the risk of developing adult-onset diabetes. Large-scale prospective studies around the world [68–76] have shown that the risk of adult-onset diabetes strongly increases as body mass

index (BMI) in adult life increases. The most recent study [4] based on pooled analysis of individual-level data for 120,813 adults from 16 cohort studies from the USA and Europe showed that when compared to individuals with healthy weight (BMI <25.0 kg/m^2), individuals with overweight (BMI 25.0-29.9 kg/m^2) and individuals with obesity (BMI ≥30 kg/m^2) had 3.3 times (95% CI: 3.0-3.6) and 9.4 times (95% CI: 8.6-10.3) greater risks of developing diabetes during adulthood, respectively. For those with BMI of ≥35 kg/m^2, the risk of diabetes increased to 18.6 times (95% CI: 16.6-20.9) (Figure 2.4).

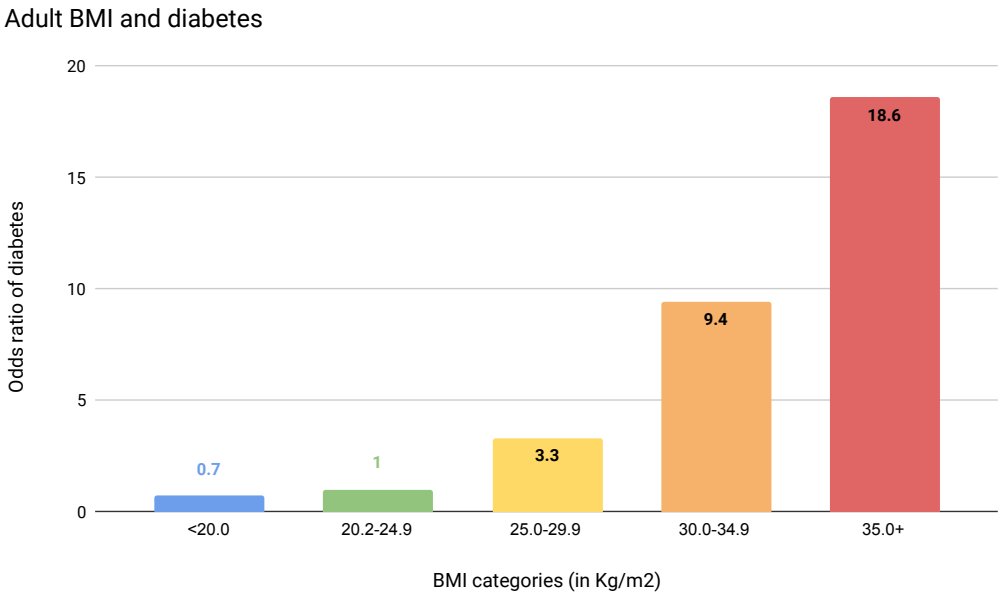


Figure 2.4: Association between adult body mass index and type 2 diabetes

Figure was created based on the results from Kivimaki et al. 2017 [4]

In 2010, a meta-analysis [5] of 18 prospective cohort studies involving 590,251 middle-aged individuals and 15,771 cases of type-2 diabetes reported that the overall relative risk (RR) of diabetes for individuals with obesity compared to those with normal weight was 7.19 (95% confidence intervals [CIs]: 5.74-9.00), and for individuals with overweight was 2.99 (95% CI: 2.42-3.72). When the authors restricted their analysis to studies with at least 400 physician-diagnosed diabetes cases, measured BMI, and at least three main confounding variables (age, family

history of type 2 diabetes, physical activity), the RR was 7.28 (95% CI: 6.47-8.28) for those with obesity and 2.92 (95% CI: 2.57-3.32) for those with overweight.

Although it is well-known that overweight or obese adults have a higher risk of diabetes, the precise mechanism of their association still remains unclear. Recently, an international working group of 32 experts in the pathophysiology, genetics, clinical trials, and clinical care of obesity and/or type 2 diabetes proposed at least three distinct mechanisms linking obesity to insulin resistance: a) increased production of leptin, nonesterified fatty acids (NEFAs), and adipokines/cytokines (e.g. tumor necrosis factor-α, retinol-binding protein-4) by adipose tissue; b) dysmetabolic sequelae of fat depositions in liver and skeletal muscle; and c) mitochondrial dysfunction [77]. Insulin resistance depends not only on the amount of adipose tissue but also on its distribution. Marked dysfunction of pancreatic β-cells have also been found in individuals with obesity who also develop diabetes subsequently [78].

Despite numerous diabetes- and obesity-associated genes having been identified thus far, genetic predictability of these conditions are low (15% for type 2 diabetes and 5% for obesity). This indicates the importance of environmental factors, gene-environment, gene-gene, and epigenetic interactions [77].

2.3.7 Smoking

A recent meta-analysis of 88 prospective studies with nearly 6 million participants concluded that both active and passive smoking were related to increased risks of adult-onset diabetes [79]. Current smokers had 37% (RR 1.37, 95% CI 1.33-1.42) increased the risk for diabetes than never smokers, and there was a dose-response relation for current smoking and diabetes risk. Among never smokers, those with exposure to passive smoking had higher risks (RR 1.22, 95% CI 1.10-1.35) of diabetes than those without exposure to passive smoking. The benefits of smoking cessation to reduce diabetes risk was observed in studies with long term follow-ups, despite an increased risk was observed among those who quit recently [79]. One explanation for such findings could be that recent quitters are more likely to have poor health, and this can lead to reverse causation.

There might be complex relationships between smoking, body size, and diabetes. Smokers tend to have lower BMI than non-smokers [80], while increased BMI is a strong predictor of diabetes. Also, smoking cessation often leads to weight gain, which could explain the increased risk of diabetes among recent quitters. A recent meta-analysis found that after 12 months of quitting smoking, individuals gained 4-5 kg weight on average [81]. However, the increased diabetes risk in recent quitters remained strong even after adjustment for further weight gain [79].

2.3.8 Alcohol consumption

Consuming alcohol in moderate amounts has been shown to be associated with reduced risk of adult-onset diabetes, while heavy drinkers have higher risks of developing the disease [82–84]. Such a reduction in diabetes risk for moderate drinking may be more specific to women than men. A recent meta-analysis found that drinking up to 63 g/day was associated with a lower risk of diabetes, with the highest risk reduction for drinking 10-14 g/day [82]. Most observational studies used nondrinkers, comprised of both never and former drinkers, as the referent category; and former drinkers often have poorer health leading to reverse causation. Therefore, those studies were likely to overestimate the protective effects of moderate alcohol consumption on diabetes. When looked at studies that used a never-drinker reference group, the risk reduction associated with moderate drinking was absent in a recent meta-analysis [82].

2.3.9 Physical inactivity

Physical inactivity has been consistently associated with increased risk of adult-onset diabetes [85], whereas any kind of physical activity, for example, leisure-time activities, vigorous activities, moderate activities, and low-intensity activities appear to be beneficial for reduction of diabetes risk [86, 87]. A recent meta-analysis of 19 prospective cohort studies found that those who were physically inactive were 1.42 times (95% CI 1.25-1.61) more likely to develop diabetes than those who remained active [85]. The relationship between physical inactivity and diabetes risk may partly

be mediated by levels of adiposity. The beneficial effects of physical activity on diabetes seem to be curvilinear, with the greatest reductions observed for low levels of activity, but further risk reductions can be possible at much higher levels [86, 87].

2.3.10 Co-morbidities

Medical conditions like hypertension and Polycystic Ovary Syndrome (PCOS) are thought to be linked with adult-onset diabetes. Individuals with high blood pressure have a higher risk of developing type 2 diabetes. A recent study based on primary care data for 4.1 million individuals reported that those with high blood pressure had a 50% increased risk of incident diabetes than those who did not have high blood pressure [88]. However, the underlying biological pathways connecting hypertension and diabetes are not clear. High blood pressure is associated with chronic inflammation [89] and endothelial dysfunction[90], both of which can play a role in the development of diabetes.

PCOS is a common condition among reproductive-aged women. Women with PCOS generally have ovaries with many follicles and cysts, elevated free testosterone levels, and ovulatory dysfunction. Women who have PCOS are at an increased risk of developing impaired glucose tolerance (IGT), type 2 diabetes, and metabolic syndrome [91]. Metabolic syndrome is a cluster of conditions which includes excess body abdominal fat, higher blood pressure, lipid abnormalities, and elevated glucose levels. These women usually have insulin resistance and therefore higher levels of insulin circulating in the blood. Adolescent obesity is also linked with PCOS, and they both appear to have effects on insulin resistance that are at least partly independent of each other [92].

2.4 Body size at different ages

Body mass index (BMI) is a simple index of weight-for-height that is commonly used to measure body size in adults and children. It is defined as an individual's weight in kilograms divided by the square of their height in meters (kg/m^2). Adults with BMI 25.0-29.9 kg/m^2 are considered as overweight whereas those with BMI $\geq$ 30.0 kg/m^2 are labelled as obese. Defining overweight and obese in children based on their BMI is a complex process, mainly because of the growth patterns of children vary between boys and girls and also by age. While many countries use their own population-specific references for defining overweight and obesity in children and adolescents, several child growth references from the International Obesity Task Force (IOTF) [93], Centers for Disease Control (CDC) [94] and World Health Organization [95] are also commonly used. Recalled information on relative body size in early life instead of BMI was used in many population-based cohorts of middle-aged men and women [96–98].

Birthweight has been commonly used to determine size at birth. It is considered as a surrogate indicator of foetal growth. Low birthweight has been defined by the World Health Organization (WHO) as birthweight of less than 2.5 kg (5.5 pounds) [99]. Birth length and ponderal index (birthweight/birth length^3) can also measure size at birth, but they are less commonly used in epidemiological research.

2.5 Burdens of low birthweight and obesity among children and adults

Low birthweight is a major global health problem, with an estimated $20 \cdot 5$ million babies born with low birthweight in 2015. Most babies with low birthweight are born in low- and middle-income countries. The prevalence of low birthweight was the lowest in European countries (e.g. 6.9% in 2015) and highest in South Asian countries (e.g. 27.0% in 2015).

Worldwide the burden of overweight and obesity is increasing rapidly among both children and adults. According to the recent estimates from the Global Burden

of Disease data, 108 million children had obesity in 2015 [6]. In 2017, among children aged 2 to 15 years in England, 30% were overweight or obese, while 17% were obese [100]. Recent estimates suggest that globally more than 1.9 billion adults (18 years or older) are either overweight or obese [6, 101, 102]. According to the latest report on the Statistics on Obesity, Physical Activity and Diet in England, about two-thirds (i.e. 64%) of adults were overweight or obese, 45% of whom were classified as obese [103].

2.6 Genetics of body size at different ages

Genome-wide association studies (GWAS) for adiposity traits have so far identified more than 300 single-nucleotide polymorphism (SNPs) for BMI, waist-to-hip ratio, body fat and visceral adiposity [104]. A recent meta-analysis based on the Genetic Investigation of ANthropometric Traits (GIANT) consortium identified 97 loci for BMI, which can only explain 2.7% of the variance in BMI [105]. However, the proportion of inter-individual variation due to genetic factors (also known as heritability) for BMI has been estimated to be 40-70% [106]. Some of the identified loci are more strongly associated with obesity (e.g. *FTO*) while others influence thinness more (e.g. *CADM2*) [104]. Large GWAS studies have identified many of the same genetic loci for childhood BMI as in for adult BMI [107, 108].

Studies using genome-wide common variant data from twins [109, 110], mother-child pairs [111, 112], or families [113] have estimated that 24% to 69% of the variance in birthweight can be explained by foetal genetics, whereas maternal genetic factors can explain 3% to 22% of the variations in birthweight. Genome-wide association studies (GWAS) of birthweight so far have identified 65 loci that are robustly associated with birthweight [111, 114]. The majority (n=58) of these independent common variant associations indicated direct effects of foetal genotype [114]. Foetal genetic variants associated with birthweight were found to be enriched for insulin signalling, glucose homeostasis, glycogen biosynthesis, and chromatin remodelling [111]. Although birthweight and childhood BMI share a few loci, birthweight is largely driven by unique loci [104].

2.7 Tracking of body size from birth to early life to later life

Higher birthweight has been associated with higher BMI in children and in adolescents [112, 115]. Birthweight is also positively and linearly associated with BMI in adulthood, but the association becomes weak with increasing age [116, 117]. Evidence from recent Mendelian Randomisation (MR) study suggests that higher birthweight is associated with higher BMI [118]. However, it is not clear whether the associations between birthweight and adult adiposity are genetic in origin, or a result of intrauterine programming. Studies of the Dutch famine [119, 120] and the offspring of diabetic women [121] provide some evidence for an influence of intrauterine programming.

There is evidence that early life adiposity is more likely to persist into adulthood. A recent meta-analysis [11] using BMI data from 16 cohorts suggested that there was a strong association between childhood adiposity and adult adiposity, with children with obesity, were about five times more likely to be obese as adults than those children who were not obese (RR 5.21, 95% CI: 4.50-6.02). The study also suggested that among those who were obese in their adolescence, almost 70% of them were still obese after age 30 years.

2.8 Conclusion

Background information provided in this chapter on diabetes and body size at different stages will help to understand the topic and to formulate the research questions and analysis framework in this thesis.

3

Body size in early life and diabetes in adulthood: research questions and systematic reviews

Summary points

- Careful analyses and interpretations are required to understand the associations of body size at different stages of early life and risk of adult-onset diabetes.
- It is still unclear whether the associations of birthweight, childhood body size and body size in early adulthood with the risk of adult-onset diabetes are independent and persistent across body size at other ages.
- Evidence on direct comparison of diabetes risks associated with different trajectories or combinations of body size across lifetime is limited.

This chapter discusses the research questions asked in this thesis involving body size in early life and risk of developing diabetes in adulthood and related issues that need careful considerations to ask the question right. Then, as a first step to approach the research questions, systematic reviews on the associations between body size in early life and risk of developing diabetes are presented.

3.1 Research questions

The aim of the thesis is to understand the associations of body size at different stages of early life on the risk of developing diabetes in adulthood. The different stages of early life in this thesis will be: around the time of birth, mid-childhood and early adulthood. To understand the associations of body size at these stages of early life on the risk of developing diabetes in adulthood, separate and joint effects of them will be explored. Figure 3.1 shows a simplistic view of the research questions in this thesis. As shown, body size at birth, childhood and early adulthood may or may not have an effect on the risk of developing diabetes in adulthood.

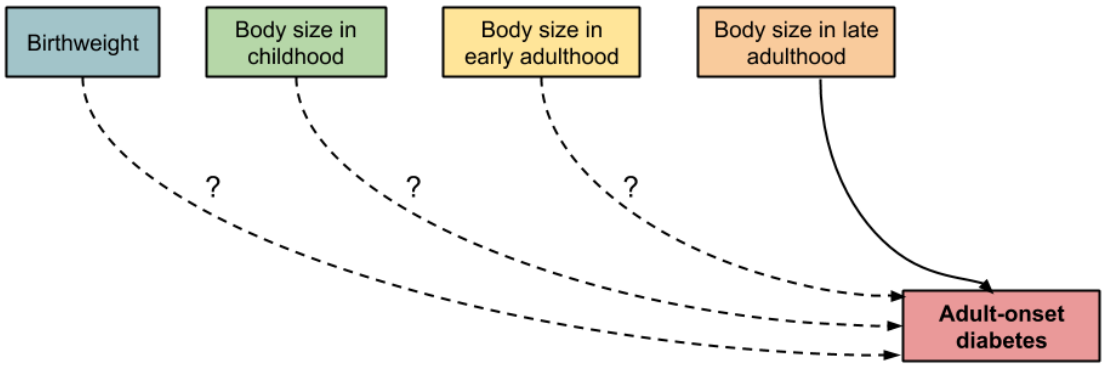


Figure 3.1: Simplistic view of research questions asked in this thesis

The reasons that it is a simplistic view as one might point out after more careful considerations are multi-fold:

3.1.1 Independent effect, direct effect, and separate effect

In epidemiological studies, the independent effect is commonly used to describe the statistical associations between an exposure and an outcome after taken account of a third factor, typically a confounding factor. A confounding factor (e.g., smoking) is associated with both the exposure of interest (e.g., adult adiposity) and the outcome (e.g., diabetes), and cannot be an intermediary step in the causal pathway from the exposure to the outcome. So, the statistical association between adult adiposity and diabetes after taken account of smoking is described as an association that is independent of the effect of smoking. However, the term "independent effect" can cause confusion when the third factor is not a confounding factor, but a mediator in the association. In these situations, the effect through the mediator is a part of the effect of the exposure. After taking account of the third factor (i.e. the mediator), the remaining effect is the effect of exposure to the outcome not mediated by the third factor. This remained effect is recently referred to as “direct effect”, as opposed to the “indirect effect” through the mediator.

For clarification, the separate effects referred to in this thesis concern the examination of the non-confounded and non-mediated direct effects of body size in early life on the adult-onset diabetes risk. As shown in the Figure 3.2, using birthweight as an example, because people with larger birthweight may have larger body size in adulthood, it is possible that there might be some indirect effect of birthweight on diabetes mediated through adult body size. Having recognised this, the main interest of this investigation is to explore whether there are separate effects of birthweight on diabetes that is not mediated through its association with adult body size, and also not confounded by other factors.

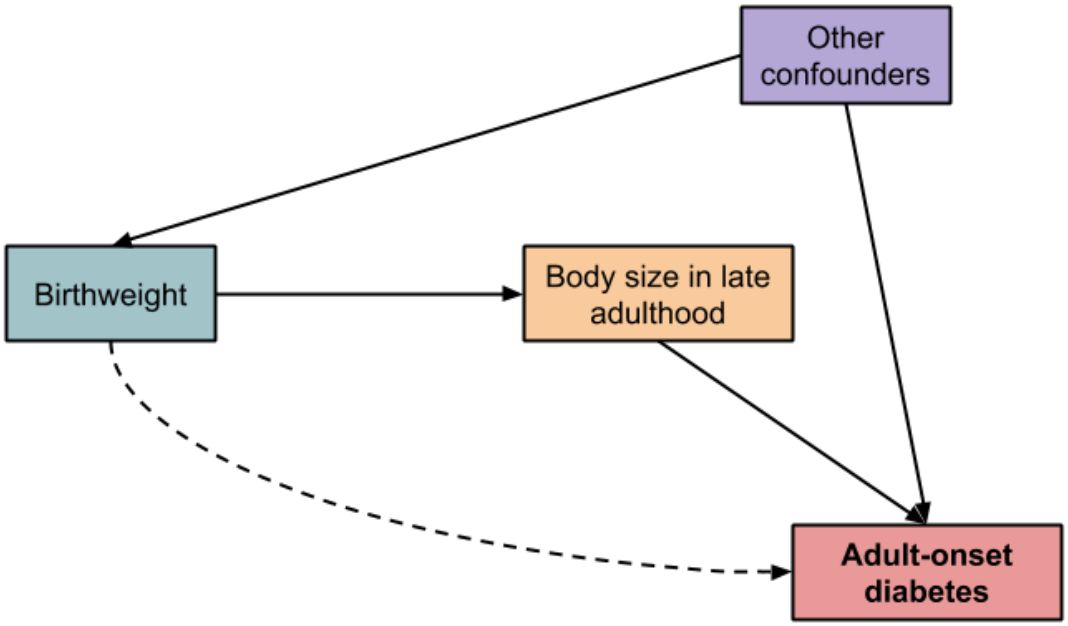


Figure 3.2: Schematic representation of the association between birthweight and risk of developing diabetes in adulthood

3.1.2 Fine vs broad stages of early life

The different stages of early life in this thesis will be: around the time of birth, mid-childhood and early adulthood. When considering the separate effects of body size at different stages of early life on the risk of adult-onset diabetes, it is unclear which stages of early life may have a distinct effect, and the definitions of these stages may mislead the interpretations of the apparent associations. Cutting the early life stages too broadly risks not being able to identify the effect of body size at a critical stage on the risk of diabetes in later life. On the other hand, cutting the early life stages too finely leads to unnecessary investigation with non-meaningful interpretations of the separate effect of body size. For example, as shown in Figure 3.3, investigating the separate effects on body sizes at ages 9, 10, and 11 years on risk of developing diabetes is not meaningful in this context. In this case, body size at age 9 years acts as the background for body size at ages 10 and 11 years. So, estimates obtained after including body size at all these ages in the same

regression model for the risk of adult-onset diabetes will be challenging to interpret, particularly because of their respective conditioning.

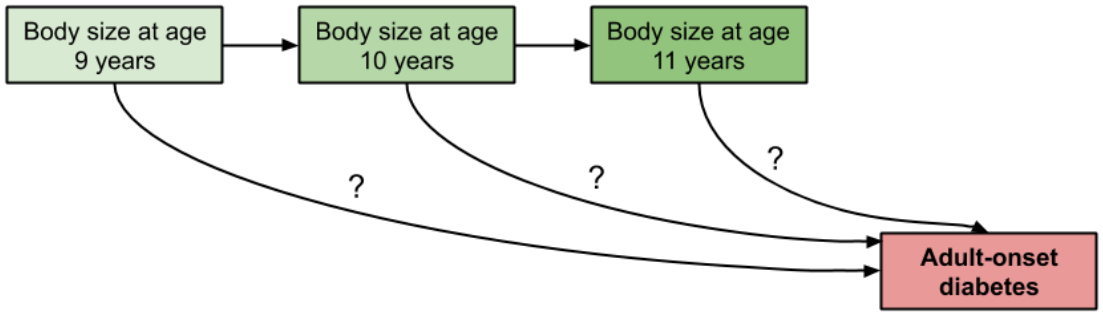


Figure 3.3: Simplified path diagram of the relationships between body size at age 9, 10, and 11 years and risk of developing diabetes in adulthood

3.1.3 Multiple mechanisms and apparent associations

Body size in early life can influence the risk of diabetes in adult life through multiple mechanisms. For example, birthweight is an indicator of pre-natal growth which can be influenced by more than one factor, such as genetics or maternal nutrition [122, 123], which may be the underlying mechanisms that explain the apparent association between birthweight and risk of diabetes in adulthood. As shown in Figure 3.4, genetic factors and maternal nutrition may be the underlying factors that explain the apparent association between birthweight and risk of developing diabetes in adulthood. These factors can be treated as confounders in statistical models, but such treatment might not always be appropriate because these are the potential mechanisms of interest.

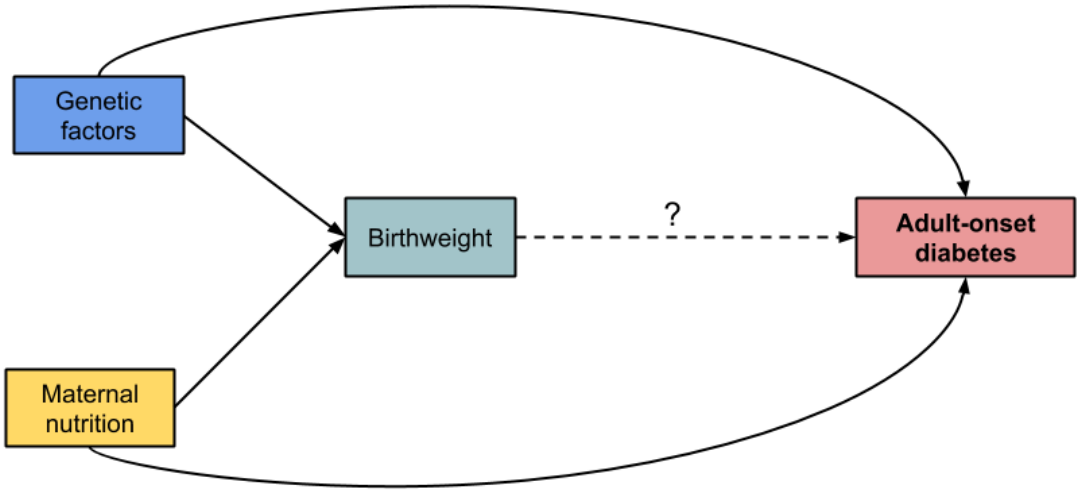


Figure 3.4: Simplified path diagram showing the roles of genetic factors and maternal nutrition on the apparent association between birthweight and risk of developing diabetes in adulthood

Because there might be multiple mechanisms or factors that can determine body size at birth, it is possible that a non-linear association reflects two or more contradictory mechanisms. For example, babies with healthy and better nutrition in foetal life may have higher birthweight, whereas babies born to mothers with gestational diabetes are also known to have abnormally high birthweight. If better nutrition can reduce the risk of developing diabetes in adulthood, one might find that high birthweight is associated with a lower risk of diabetes (as shown in Figure 3.5A). If gestational diabetes in mothers can increase the risk of developing diabetes in adulthood among offspring, then one might find that abnormally high birthweight is associated with a higher risk of diabetes (as shown in Figure 3.5B). Since birthweight can reflect both of these underlying factors, one might observe a combination of the above two associations. As shown in Figure 3.5C, there might be a U-shaped association between birthweight and the risk of developing diabetes in adulthood. In studies where information on these underlying mechanisms or factors influencing a particular body size is not available, the apparent association between that body size and risk of developing diabetes in adulthood should be interpreted carefully.

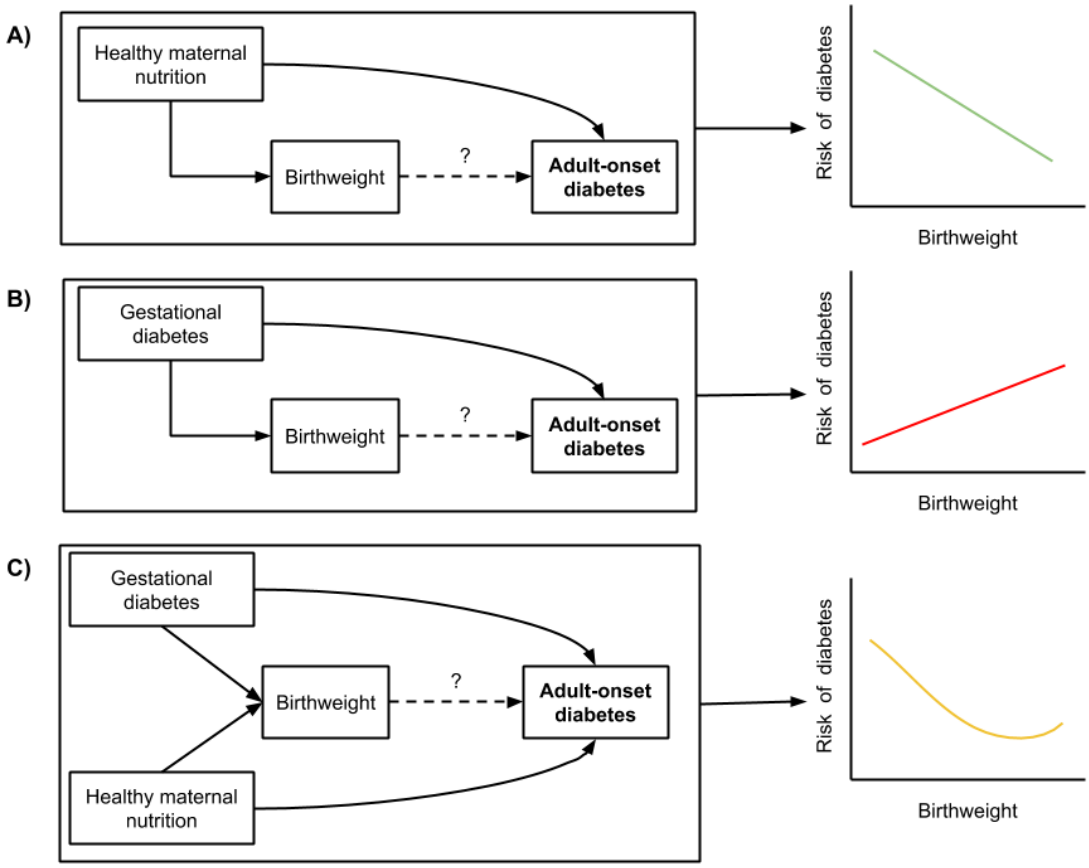


Figure 3.5: Example of path diagrams on the apparent association between birthweight and risk of developing diabetes in adulthood

Diagram A shows healthy maternal nutrition can increase birthweight and decrease the risk of adult-onset diabetes in the child. This can lead to an apparent inverse association between birthweight and the risk of adult-onset diabetes. Diagram B shows gestational diabetes can increase birthweight and also increase the risk of adult-onset diabetes in the child. This can lead to an apparent positive association between birthweight and the risk of adult-onset diabetes. Diagram C shows that both healthy maternal nutrition and gestational diabetes can influence birthweight and hence, there might be an apparent U-shaped association between birthweight and risk of adult-onset diabetes in the child.

3.1.4 Choice of body size indicators

Choice of body size indicators may indicate different underlying mechanisms for an apparent association between body size at a particular age and risk of developing diabetes in adulthood. For example, BMI is commonly used as an indicator of body size, but high BMI may not be an ideal indicator of adiposity, since it may indicate either adiposity (i.e. fat mass) or muscularity (i.e. lean mass) [124]. On the other hand, waist:hip ratio can be a better predictor of visceral adiposity

(i.e. fat distribution) [124, 125]. As shown in Figure 3.6, BMI and waist:hip ratio each reflect different aspects of body size and adiposity. Those with more central distributions of fat might have higher insulin resistance than those with more peripheral distributions [22]. Because it is still unknown which underlying mechanism is the true driver of the association between body size and risk of diabetes in adulthood, apparent associations based on any pre-selected indicator of body size need to be interpreted cautiously.

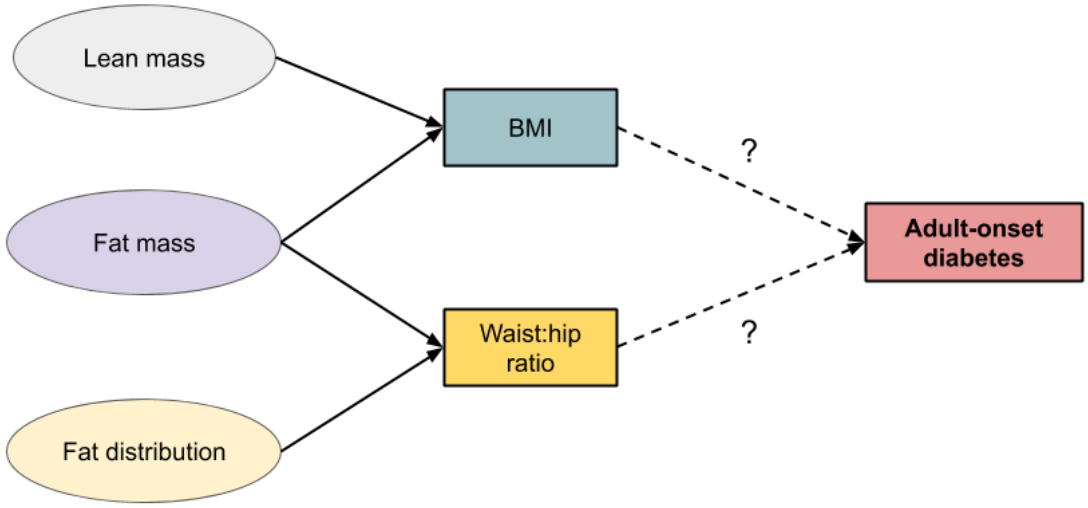


Figure 3.6: Example of path diagrams on the apparent associations of BMI and waist:hip ratio with risk of developing diabetes in adulthood

3.1.5 Effect modification by body size at other ages

Although the aim of the thesis is to examine the separate effect (i.e. the non-confounded and non-mediated effects of body size in early life on the adult-onset diabetes risk), it should be recognised that the effect of body size at a particular age might be modified by weight gain in later life. Because obesity in late adulthood is strongly associated with risk of diabetes [4, 5], if having a low birthweight is found to be associated with a higher risk of developing diabetes in adulthood, then one might have speculated whether this association is largely due to the tendency to gain weight in later life. In this scenario, one may find an apparent association between low birthweight and risk of diabetes in adulthood after adjustment for

adult body size, but when one looks separately by adult body size, the effect of low birthweight can only be seen specifically among those who are overweight or obese as adults, but not among those who remained lean (as shown in Figure 3.7A). On the other hand, if the association between low birthweight and risk of diabetes in adulthood is not due to gaining weight in later life, when one looks at the association separately by adult body size, the effect of low birthweight can be consistently seen among those who are overweight or obese as adults as well as among those who remained lean (as shown in Figure 3.7B).

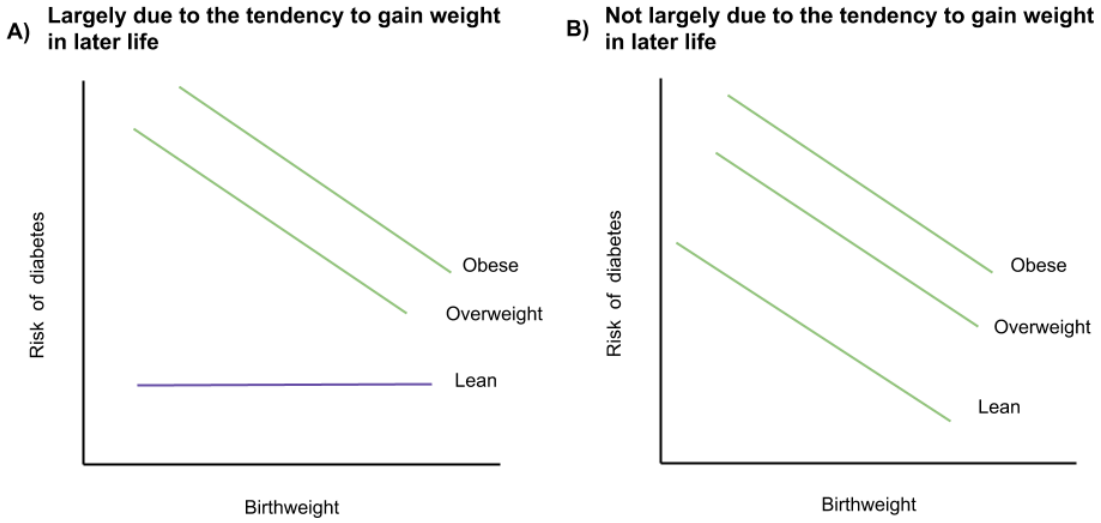


Figure 3.7: Hypothetical associations between birthweight and risk of diabetes in adulthood when looked separately by adult body size

Diagram A shows possible scenario when the effect of birthweight on the risk of developing diabetes in adulthood is largely explained by weight gain in later life. Diagram B shows possible scenario when the effect of birthweight on the risk of developing diabetes in adulthood is not largely explained by weight gain in later life.

Therefore, it would be ideal if one could look separately by individuals with different body size so that the details of any specific or consistent associations can be identified. However, looking into many specific groups results in smaller sample size in each group, and therefore, a large sample size usually required for such details.

3.1.6 Measurement errors in body size indices

As shown in Figure 3.8, errors in the observable body size indices (either measured or reported) can be both random and systematic. Such errors can result in systematic underestimation of any apparent association between particular body size and risk of adult-onset diabetes, also known as "regression dilution". On the other hand, if errors in body size are related to some other factors (or their measurement errors) and these factors are also associated with diabetes risk, then such errors could contribute to the confounding of the apparent association (Figure 3.8). For instance, previous studies [97, 98] reported that when recalled, body sizes in early life are generally under-reported, with more significant under-reporting among overweight or obese adults. Due to this, adult body size can be considered as a confounder of any apparent association between body size in early life and risk of diabetes in adulthood. To minimise such confounding, adult body size should be taken into account in the analysis of the associations between body size in early life and risk of diabetes in adulthood.

Weaker associations between reported and measured values could also result in some loss of study power.

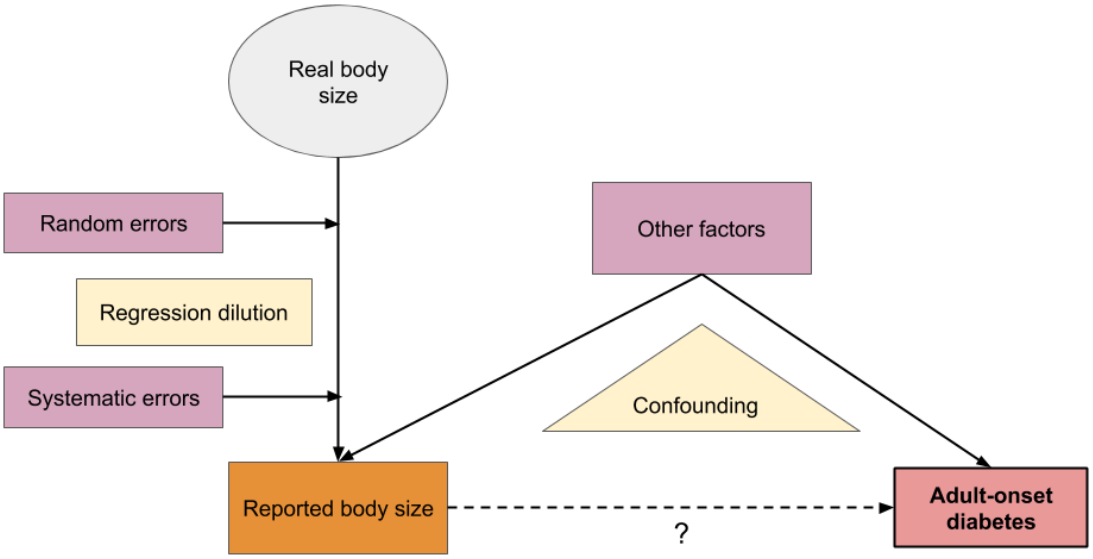


Figure 3.8: Errors in reported body size indices and possible effects on their apparent associations with the risk of diabetes in adulthood

3.1.7 Adult-onset diabetes is not a simple concept

As discussed in the previous chapter, adult-onset diabetes can be heterogeneous with regard to its aetiology, underlying pathophysiology, and clinical presentations. Insulin resistance and β-cell dysfunction are the primary mechanisms of developing diabetes, but the degree of these pathophysiologies can vary depending on the underlying aetiology. Therefore, even if there are associations between body size at different ages and risk of adult-onset diabetes, they might represent different underlying mechanisms. For example, birthweight can affect the risk of diabetes by some mechanisms which are very different from mechanisms linking adult adiposity and diabetes.

The overall understanding of the associations between body size at different stages of early life and the risk of developing diabetes in adulthood should include but not limited to the above-mentioned points.

3.2 Systematic reviews

As the first step, I approached the research question of this thesis by doing systematic reviews for studies on the associations between body size in early life and risk of developing diabetes in adulthood. As mentioned earlier, when considering the separate effects of body size at different stages of early life on the risk of adult-onset diabetes, it is unclear which stages of early life may have distinct effects, and the definitions of these stages may mislead the interpretations of the apparent associations. The three different stages of early life in this thesis are: around the time of birth, mid-childhood and early adulthood. I have conducted systematic reviews sequentially for body size at these three stages.

3.2.1 Birthweight and risk of adult-onset diabetes: systematic review

In this systematic review, I collated evidence from prospective cohort studies on the association between birthweight and the risk of adult-onset diabetes. I looked at how each study was designed and analysed to address various issues of statistical associations (e.g. independent effect, direct effect, and separate effect), measurement errors in birthweight, multiple mechanisms and effect modification by body size in later ages.

Box 3.2.1 briefly describes the methods of systematic search, selection criteria, and analysis of this systematic review.

Search strategy, selection criteria and meta-analysis

I identified published prospective studies of the association between birthweight and risk of adult-onset diabetes through electronic searches in PubMed and EMbase (up to April 2019) with combinations of relevant search terms. I also searched cited references in the included papers for relevant articles. I did not limit the search by publication date, language, source of study population, or study size. Further details about search strategy used and description of the included studies with relevant information are given in Appendix A. For each study, I extracted information on study population, follow-up, assessment of birthweight, and assessment of diabetes.
I extracted the published relative risks (RRs) and 95% confidence intervals (CIs) of incident diabetes per 1 kg increase in birthweight, or sufficient data to allow estimation of such RRs. For each study I extracted the most fully-adjusted RR and adult BMI-adjusted RR, if available. In studies where RRs of diabetes were given only by birthweight categories, I estimated log-linear trend using generalized least square by *glst* package in Stata v15.1 or, using variance weighted least squares if numbers of diabetes cases by birthweight categories were not available [126]. For trend estimation where category mean or median was not available, I estimated the mean birthweight in each birthweight category, where necessary, as category midpoints, or by other methods [127] assuming normal distribution with the grand mean 3.2 kg and standard deviations 0.7 kg, if not given in the published data. Results of sub-groups within studies were combined before they were combined with other studies. Study-specific RRs per 1 kg increase in birthweight for diabetes were combined to provide summary RRs with weights proportional to the inverse of the variance as suggested by Greenland and Longnecker [128]. Chi-square tests were used to assess heterogeneity across studies.

The systematic search yielded nearly 7000 articles, and from them I included 29 prospective cohort studies (in 27 reports) with 32,300 diabetes cases examining the association between birthweight and risk of adult-onset diabetes. Summary information of these studies are shown in Appendix A. The number of diabetes cases ranged from 5 to 12,486, while most studies based on well-established cohorts had 500 or more incident cases of diabetes. Findings from studies with minimum of 500 diabetes cases are presented here and findings from all included studies are given in Appendix A.

Figure 3.1 shows that eight studies (in six reports) had a minimum of 500 incident cases of diabetes. Among them, two studies [129, 130], both from Scandinavia, are record-linkage studies with birthweight information from the birth registry, and with diabetes information from hospital records. Because of the nature of the data-linkage, these studies did not have information beyond age, sex and socioeconomic status. They did not have information on body size beyond birth. By contrast, the other six studies, from US [131, 132], France [133] and China [134], are population-based prospective cohorts of middle-aged individuals with reported birthweight at the time of start of follow-up. Diabetes cases were defined by self-reports (in studies from US [131, 132] and China [134]) or self-reports combined with prescription records (French study [133]) (Figure 3.1). These studies typically have more information on characteristics potentially confounding the association, and provide results with adjustment for them. These include adult BMI-adjusted estimates. In two studies [132, 133], they attempt to investigate the association separately in groups of individuals with different levels of adult BMI, but did not have enough number of cases in each group to have reliable results.

Table 3.1: Summary information of prospective studies with 500 or more cases investigating the association between birthweight and risk of diabetes in adulthood

Year	Study, country [reference]	Study population at baseline	Follow-up	Assessment of birthweight	Assessment of diabetes
2018	Shanghai Women's Health Study and the Shanghai Men's Health Study, China [134]	11,515 men and 13,569 women aged 40–70 years were recruited in 1996–2000 and in 2002-06, respectively	Follow-up surveys were conducted every 2–3 years since recruitment	Birthweight was self-reported at recruitment	1027 incident cases of T2DM were identified by self-reporting
2015	Copenhagen School Health Records Register, Denmark [129]	113,801 men and 109,298 women born in 1936-1983 were included in 1977 or at age 30 years (median age 30 years)	Followed-up in 1977-2013 (mean age 59 years)	Birthweight was reported by mothers at school entry (age 6-7 years) in school health records	7750 men and 4736 women developed adult T2DM (aged ≥30 years) identified from hospital admission data
2015	Health Professional Follow-up Study, USA [131]	18,305 men were recruited in 1986, mean age 51.3 years	Followed-up in 1986-2010	Birthweight was self-reported on the 1994 follow-up questionnaire	1603 T2DM cases were identified by self-reporting and confirmed by validated supplementary questionnaire
2015	Nurses' Health Study, USA [131]	49,757 women were recruited in 1980, mean age 45.2 years	Followed-up in 1980-2010	Birthweight was self-reported on the 1992 follow-up questionnaire	5381 T2DM cases were identified by self-reporting and confirmed by validated supplementary questionnaire
2015	Nurses' Health Study II, USA [131]	81,732 women were recruited in 1991, mean age 36.0 years	Followed-up in 1991-2011	Birthweight was self-reported on the 1991 follow-up questionnaire	4725 T2DM cases were identified by self-reporting and confirmed by validated supplementary questionnaire
2014	Black Women's Health Study, USA [132]	21,624 black women aged 21-69 years were included in 1995	Followed-up between 1997 and 2013 by biennial questionnaire.	Birthweight was self-reported in 1997 follow-up questionnaire and validated	2388 self-reported diabetes cases were identified
2010	E3N cohort, France [133]	91,453 women born in 1925-1950 and recruited in 1990 (aged 40-65 years)	Followed-up in 1990-2005. Mean years of follow-up 15.1 years	Birthweight was self-reported at study recruitment	2534 diabetes cases were identified by self-reporting or drug reimbursement record linkage
2009	Stockholm Birth Cohort 1925-49, Sweden [130]	6425 men and women born in 1925-1949 in Stockholm and lived in Sweden after 1987 were included	Followed-up from 1987 to 2006 (aged 38 – 81 years)	Birthweight was recorded in birth records	508 diabetes cases were identified in hospital registers

Abbreviations: T2DM: type 2 diabetes mellitus

With or without adjustment for adult body size, all eight studies have reported that lower birthweight is associated with a higher risk of diabetes in adulthood, but not necessarily in a linear fashion. Figure 3.9 shows the results from seven cohorts with separate results for men and women in the Copenhagen School Health Records Register study [129]. As shown in this figure, the category of higher birthweight compared to lower birthweight was associated with a lower risk of diabetes, but in one data-linkage study (Copenhagen School Health Records Register study [129]) women with extreme high birthweight (e.g. 4.5 kg or more) appeared to have an additional risk of diabetes. Two other studies (Nurses' Health Study I and II [131]), reported a significant non-linearity in trend due to some additional risk among women with extremely high birthweight, which became non-significant after adjustment for adult BMI. However, there was no appreciable difference in the category-specific relative risks for the high birthweight category in these two studies (Figure 3.9).

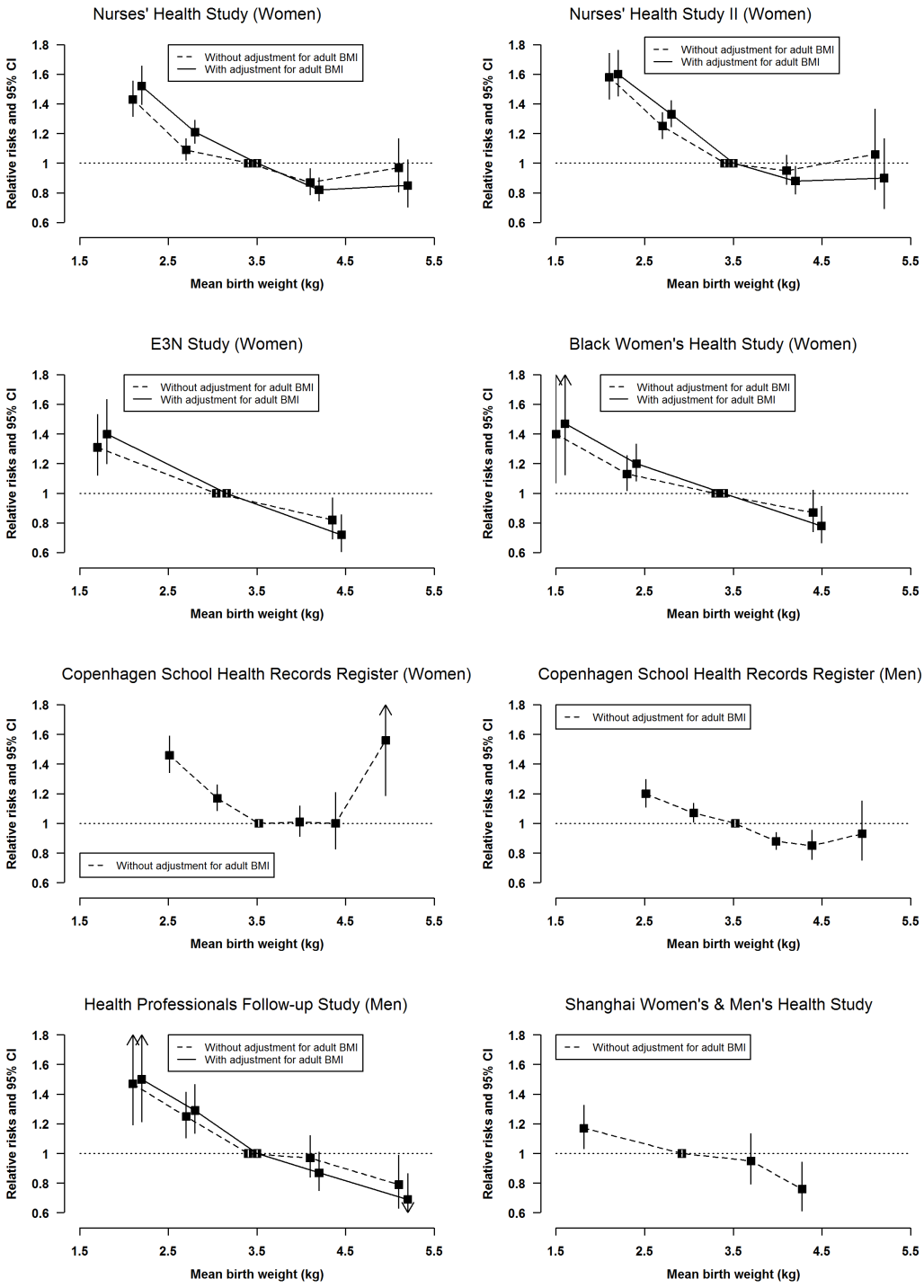


Figure 3.9: Association between birthweight categories and risk of adult-onset diabetes in large-scale prospective studies

Relative risks (RRs) of diabetes are plotted against the mean value of each birthweight category.

Despite the possibility of non-linearity, Figure 3.10 shows that in these studies, there was a general linear trend in the association between birthweight and risk of diabetes in adulthood. When I combined them, for each kg increase in birthweight, the relative risks of developing diabetes in adulthood was 0.82 (95% CI: 0.81-0.84) (Figure 3.10). The inverse relationship between birthweight and risk of adult-onset diabetes remained strong after adjustment for adult BMI (pooled RR for each kg increase in birthweight: 0.77 (95% CI: 0.75-0.79) (Figure 3.10). The associations were also inverse for most of the smaller studies, although only a few reached statistical significance (Supplementary Figure A.1). Only two studies [135, 136], both on native North American populations (i.e. Pima Indians and Saskatchewan native North Americans), showed overall positive associations between birthweight and diabetes in adulthood. It is important to mention that these populations have exceptionally high rates of gestational diabetes, obesity from early ages, and adult-onset diabetes [137, 138].

The Black Women's Health Study [132] looked at the association between birthweight and risk of adult-onset diabetes among obese and nonobese women, whereas the French E3N study [133] looked at this association in women who were of normal weight and overweight as adults. But, the number of cases was small to detect an association within each category of adult adiposity reliably.

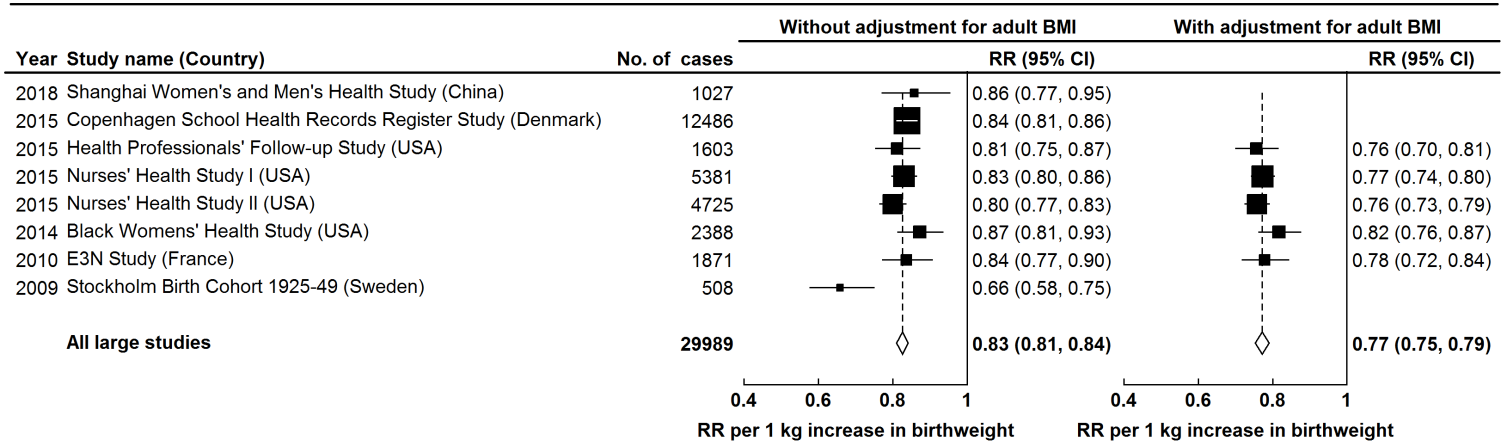


Figure 3.10: Association between birthweight and risk of adult-onset diabetes in large-scale prospective cohort studies: meta-analysis

Study-specific linear trends (per kg increase) for birthweight were extracted from the published reports or estimated using the generalized least square method proposed by Orsini [126]. Two-stage fixed-effects dose-response models were used for combining results. (see subsection 3.2.1 for more details)

3.2.1.1 Questions that have been answered

The inverse association between birthweight and risk of diabetes in adulthood was consistent when looked separately among studies which used measured birthweight and among studies which used self-reported birthweight. The association remained strong after adjustment for adult BMI.

3.2.1.2 Questions to be answered

There are two questions which remained to be answered in the existing evidence. According to the thrifty phenotype hypothesis [139, 140], some people might speculate that the association between low birthweight and risk of diabetes in adulthood can be seen only among those who gained weight in later life, but not among those who remained lean. Ideally, as discussed in section 3.1.5, the association can be examined in subgroups of individuals with different levels of adult BMI. Although two studies [132, 133] attempted to investigate this, they did not have enough number of cases in each group to have reliable results.

Another interesting hypothesis emerged from some studies based on the Helsinki cohort [141–143]. They showed that among babies with low birthweight those who developed diabetes in adulthood had higher rates of postnatal growth, and therefore, some researchers argue that it might be the faster catch up growth that is responsible for the increased risk of diabetes. In this review, one study [129] reported additional risks of diabetes in adulthood for those with extremely high birthweight, which was not clearly replicated in other cohorts. It is possible that some pre-natal mechanisms leading to high birthweight can increase the risk of diabetes in adulthood. However, those with high birthweight are more likely to become obese in adulthood [116, 117] and subsequently have diabetes. Even if high birthweight can be associated with some additional risks of diabetes in adulthood, it may be explained completely by the association between high birthweight and adult obesity. Because the aim of the study is to understand the separate effect of birthweight on diabetes risk in adulthood, it remains unknown whether this apparent additional risk associated with high birthweight, if any, is an effect separate from adult adiposity.

3.2.2 Body size in childhood and risk of adult-onset diabetes: systematic review

In this systematic review, I collated evidence from prospective cohort studies on the association between body size in childhood (age 5-15 years) and the risk of adult-onset diabetes. I looked at how each study was designed and analysed to address various issues of statistical associations (e.g. independent effect, direct effect, and separate effect), measurement errors in body size, multiple mechanisms and effect modification by body size in later ages.

Box 3.2.2 briefly describes the methods of systematic search, selection criteria, and analysis of this systematic review.

Search strategy, selection criteria, and analysis

I identified published prospective studies investigating the associations of body size in early life with adult-onset diabetes through electronic searches in PubMed and EMbase (up to April 2018) with combinations of relevant search terms. I also searched cited references in the included papers for relevant articles. I did not limit the search by publication date, language, source of study population, or study size. Further details about search strategy used and description of all included studies with relevant information are given in section A.3 of Appendix A.

Due to the heterogeneous nature of body size information and associated effect measures, I did not perform any quantitative synthesis and results are presented in a narrative manner. The results from included studies are grouped into body size in childhood (age 5-15 years) and body size in early adulthood (age 16-25 years). Key findings from large-scale studies (with minimum of 500 incident cases of diabetes) are presented in this chapter, while results from smaller studies can be found in the Table A.6 and Table A.7 in Appendix A.

In this systematic review, I included 17 prospective cohort studies which examined the association between body size in childhood ages (5-15 years) and adult-onset diabetes risk. Summary information of these studies are shown in Table A.6 of Appendix A. The number of diabetes cases ranged from 15 to 11,548, while most studies based on well-established cohorts had 500 or more incident cases of diabetes. Findings from studies with minimum of 500 diabetes cases are presented here and findings from all included studies are given in Appendix A.

Five studies had a minimum of 500 incident cases of diabetes. Among them, three studies [141, 144, 145], all from Scandinavia, are record-linkage studies with information on measured BMI in childhood ages, and with diabetes information from hospital admission or prescription registers. Because of the nature of the data-linkage, these studies did not have information beyond age, sex and socioeconomic status (Table 3.2). They did not have information on body size in late adulthood, but one study [144] had body size in early adulthood. By contrast, the other two studies, from US [13] and France [133], are population-based prospective cohorts of middle-aged individuals with reported childhood body size at the time of the start of follow-up. Diabetes cases were defined by either self-reports [13] or self-reports combined with prescription records [133]. These studies typically have more information on characteristics potentially confounding the association, and provide results with adjustment for them. These also include adult BMI-adjusted estimates. The French E3N study [133] attempted to investigate the association of childhood body size with risk of adult-onset diabetes separately in groups of individuals with different levels of birthweight and adult BMI, but did not have enough number of cases in each group to have reliable results.

Although four out of five studies have reported that larger body size in childhood ages is associated with increased risk of diabetes in adulthood, it is not clear whether these associations are independent of later life adiposity (Table 3.2). The Copenhagen School Health Records Register study in Denmark, using measured BMI at ages 7 and 13 years, reported that one could reverse the adverse effects of childhood overweight or obesity for the risk of adult-onset diabetes if they had

normal BMI in their early adulthood [144]. The Nurses' Health Study II used recalled body size information for age 5 and 10 years, and showed that the higher risks of adult diabetes were seen among children who were overweight or obese only if they were also overweight or obese as adults [13]. An analysis of four prospective cohort studies from USA and Finland, also using measured information on BMI, reported that overweight or obese children who became nonobese by adulthood had risks of type 2 diabetes similar to those who were never overweight or obese [146].

Childhood thinness was associated with higher risks of adult-onset diabetes in the French E3N study [133] and only among older women (aged $\geq$66 years at diagnosis) in the Copenhagen School Health Record study [145].

Birthweight had little influence on the association between body size in childhood and risk of adult-onset diabetes [133, 145].

Table 3.2: Key findings from large-scale prospective studies (500 or more incident cases) investigating the association between childhood body size and risk of diabetes in adulthood

Year	Study, country [reference]	Number of participants/cases	Body size variable	Assessment of diabetes	Key findings	Adjustment profile
2018	Copenhagen School Health Record Study, Denmark [144]	62,565/6710	BMI was measured at ages 7 and 13 years	Register-based	Overweight at age 7 years was associated with increased risk of adult T2DM only if it continued to early adulthood. Overweight at age 13 years could increase the risk of T2DM, independently of BMI at early adulthood. Stronger associations were observed for follow-up age 30-60 years than for >60 years	Age and year of birth
2017	Copenhagen School Health Record Study, Denmark [145]	292,827/19,020	BMI was measured at ages 7 and 13 years	Register-based	Higher BMI at ages 7 and 13 was associated with a higher risk of adult T2DM. Among older women (aged 66-83 years), lower BMI also increased the risk of T2DM. Observed associations were stronger in women than in men	Age and year of birth. Birthweight did not modify the associations.
2010	E3N, France [133]	91,453/2534	Recalled body silhouette at 8 years and at menarche	Self-reported or register-based	Lean body sizes at 8 years and at menarche were associated with increased risk of adult diabetes. These associations were persistent in subgroups of women by birthweight and adult BMI	Birth cohort, physical activity, education, prematurity, family history of diabetes, smoking, high cholesterol, hypertension, menopausal status, HRT, oral contraceptive pills, parity, age at first child, and age at menarche.
2010	Nurses' Health Study, USA [13]	109,172/3307	Recalled pictorial diagrams representing body size at ages 5 and 10 years	Self-reported	Larger body sizes at ages 5 and 10 years were associated with increased risk of diabetes. No increased risk of diabetes was observed for larger childhood body sizes among those who did not continue to be overweight in adulthood.	Age, race, smoking, parental history of diabetes, parity, age at first birth, physical activity, birthweight, prematurity, and age at menarche.
2002	Helsinki Birth Cohort 1924-44, Finland [141]	13,517/698	BMI measured at age 11	Register-based	Higher BMI was associated with increased risk of diabetes	Sex, year of birth

Abbreviations: HRT: hormone replacement therapy, T2DM: type 2 diabetes mellitus

3.2.2.1 Questions to be answered

There are several key questions remaining to be answered in the existing evidence. Because larger body size tracks from early life to later life [11], if larger body size in childhood is associated with increased risk of developing diabetes in adulthood, it is important to understand whether this increased risk of adult diabetes is largely driven by adult adiposity, or there is at least one pathway, independent of adult body size, through which body size in childhood can influence the risk of diabetes, i.e. the separate effect. Additionally, it is unclear whether thinner body size in childhood ages can also increase the risk of developing diabetes in adulthood. Because there might be multiple mechanisms or factors that can determine body size in childhood, it is possible that the associations of smaller and larger body size with increased risks of diabetes in adulthood reflect different underlying mechanisms.

Although body size in childhood can have separate effects on the risk of developing adult-onset diabetes, some might speculate that both birthweight and body size in childhood share the same underlying biological mechanisms. In this review, one study [133] reported the risks of adult-onset diabetes for childhood body size separately among women with different levels of birthweight, but it did not have enough number of cases in each group to have reliable results.

3.2.3 Body size in early adulthood and risk of adult-onset diabetes: systematic review

In this systematic review, I collated evidence from prospective cohort studies on the association between body size in early adulthood (age 16-25 years) and the risk of adult-onset diabetes. I looked at how each study was designed and analysed to address various issues of statistical associations (e.g. independent effect, direct effect, and separate effect), measurement errors in body size, multiple mechanisms and effect modification by body size in later ages.

Box 3.2.2 briefly describes the methods of systematic search, selection criteria, and analysis of this systematic review.

In this systematic review, I included 12 prospective cohort studies which examined the association between body size in early adulthood and adult-onset diabetes risk. Summary information of these studies are shown in Table A.7 of Appendix A. The number of diabetes cases ranged from 84 to 13,416, while most studies based on well-established cohorts had 500 or more incident cases of diabetes. Findings from studies with minimum of 500 diabetes cases are presented here and findings from all included studies are given in Appendix A.

Seven studies had a minimum of 500 incident cases of diabetes. Among them, two [147, 148] had recorded information on BMI from military records, one [144] was a record-linkage study with information on measured BMI and four studies [13, 15, 16, 133] had self-reported body size for early adulthood ages. Diabetes cases were defined by either self-reports or from national registers (Table 3.3). While record-linkage studies had minimal information on other factors, population-based prospective cohorts of middle-aged individuals typically had more information on characteristics (including body size in adulthood) potentially confounding the observed associations. One study [148] attempted to investigate the association of body size in early adulthood with risk of adult-onset diabetes separately in groups of individuals with different levels of adult BMI, but did not have enough number of cases in each group to have reliable results.

While larger body size in early adulthood increased the risk of diabetes in adulthood in all studies, adjustment for adult adiposity suggested that much of the increased risk could be explained by the correlations between body size in early and late adulthood (Table 3.3). Moreover, adjustment for body size in later adulthood actually reverse the association between body size early adulthood and risk of adult-onset diabetes in some studies [16, 133, 148]. It was unclear whether smaller body size was associated with an increased risk of developing diabetes in adulthood. The Health Professional Follow-up Study [15] reported that those with BMI <18.5 kg/m^2 at age 21 years had a higher risk of diabetes when compared with those with BMI 18.5-22.9 kg/m^2. This finding was attenuated after weight change between early and late adulthood was taken into account.

The roles of body size in earlier ages, such as at birth and in childhood on the association between body size in early life and risk of adult-onset diabetes is also uncertain (Table 3.3).

3.2.3.1 Questions to be answered

After accounting for body size at other ages, whether body size in early adulthood has a separate effect on the risk of developing diabetes in adulthood is not established in the current review. Larger body size at early adulthood is found to be associated with increased risks of developing diabetes in later adulthood, but it is unclear whether this association is entirely driven by the fact that individuals with larger body size at early adulthood tend to have higher BMI in later adulthood. Statistical adjustment for adult BMI attenuated and in some cases reversed the observed associations between larger body size in early adulthood and risk of adult-onset diabetes. Because early adulthood is an intermediary stage between childhood and late adulthood, one might speculate that body size in early life can reflect the effects of both of these stages. So, body size in both childhood and late adulthood need to taken into account while exploring the separate effect of body size in early adulthood.

Table 3.3: Key findings from large-scale prospective studies (500 or more incident cases) investigating the association of body size at early adulthood with risk of diabetes in adulthood

Year	Study, country [reference]	Number of participants/cases	Body size variable	Assessment of diabetes	Key findings	Adjustment profile
2018	Copenhagen School Health Record Study, Denmark [144]	62,565/6710	BMI was measured between ages 17-26 years	Register-based	Overweight was associated with increased risk of adult T2DM. The risk was greatest among those who were also overweight at ages 7 and 13 years	Age and year of birth
2018	China Kadoorie Biobank, China [16]	482,589/13,416	BMI at age 25 was self-reported	Register-based	BMI was weakly positively associated with diabetes, which was reversed after adjustment for adult BMI. There was no clear association of BMI at age 25 years with incident diabetes within each adult BMI category	Age, study area, education, household income, occupation, smoking, alcohol consumption, physical activity and family history of diabetes
2015	British Regional Heart Study, UK [147]	4846/627	BMI at age 21 was collected from military records	Register-based	Higher BMI was associated with a higher risk of diabetes	Age, smoking, area of residence
2014	Health Professional Follow-up Study, USA [15]	39,909/2735	Self-reported BMI for age 21	Self-reported	Higher BMI at age 21 was associated with increased risk of diabetes. Those with BMI <18.5 kg/m^2 had a higher risk of the disease when compared to those with BMI 18.5-22.9 kg/m^2. Associations were stronger for younger participants	Age, smoking, alcohol intake, and family history of diabetes
2011	MELANY study, Israel [148]	37,674/1173	BMI was measured at age 17 years	Based on fasting plasma glucose level	Higher BMI increased the risk of T2DM, but adjustment for adult BMI completely attenuated the association. Those in the lowest decile for early adulthood BMI tended to have a higher risk of diabetes than those in the higher deciles	Age, family history of diabetes, blood pressure, physical activity, fasting blood glucose, triglyceride level and BMI in adulthood
2010	E3N, France [133]	91,453/2534	Recalled body silhouette at 20-25 year	Self-reported or register-based	Higher body size was associated with increased risk of diabetes but after adjustment for adult BMI the association reversed completely	Birth cohort, physical activity, education, prematurity, family history of diabetes, smoking, high cholesterol, hypertension, menopausal status, HRT, oral contraceptive pills, parity, age at first child, and age at menarche
2010	Nurses' Health Study, USA [13]	109,172/3307	Recalled pictorial diagrams representing body size at age 20 years and self-reported BMI at age 18	Self-reported	Larger body sizes at age 20 and higher BMI at age 18 were associated with increased risk of diabetes	Age, race, smoking, parental history of diabetes, parity, age at first birth, physical activity, birthweight, prematurity, and age at menarche.

Abbreviations: BMI: body mass index, HRT: hormone replacement therapy, T2DM: type 2 diabetes mellitus

3.3 Approaches to investigate separate effects of body size at different ages

As shown in the systematic reviews, the risk of adult-onset diabetes may be influenced by body size in different stages of early life, in addition to the strong influence of adult adiposity. Because body size indices in early life are correlated with each other and with body size in adulthood, it is difficult to separate and interpret the effects of body size at a particular stage on the disease risk [149]. Epidemiological studies have been using different approaches to estimate the separate effects of body size at different ages.

Epidemiological studies investigating the association between body size in early life and risk of developing diabetes in adulthood typically ask the question whether body size at a particular age or stage is independently (of body size at later life) associated with diabetes risk. One of the main potential limitations of this approach is collinearity due to tracking of body size from early life to adulthood [8]. In investigations concerning weight change during a certain period of time, the question asked is whether weight change (i.e. gain or loss) is associated with diabetes risk, independently of baseline weight status. This approach is limited by difficulties in investigating the effects of weight change at multiple periods of time simultaneously [17].

In recent times, a group-based trajectory modelling approach has been increasingly used to understand the dynamic changes of body size over time. Group-based trajectory modelling approach provides a flexible and unique way to identify distinctive clusters of relatively homogeneous trajectory groups based on individual trajectories of body size and also to directly compare the risk of disease across different trajectory groups. This approach of investigation focuses on the relative risk of individuals who experience different developmental courses of body size and therefore circumvents the complexities in disentangling the independent effects of body size in early life versus body size in later life. However, the trajectories derived from statistical modelling may not always accurately reflect each individual's body

size trajectory, and therefore, the results yielded from these studies are difficult to compare with each other [17].

Another approach for life-course analysis of body size is analysis of disease risk by different combinations of body size over the life course, which might be considered as an extension of group-based trajectory modelling approach [8]. This approach allows examining the risks of disease at every level of body size while taking into account the associations between various indices of body size at different ages. However, large numbers are needed to reliably estimate the associations at each level of body size examined.

Table 3.4 shows different approaches used in prospective studies with a minimum 500 cases to investigate separate effects of body size at different ages on the risk of developing diabetes in adulthood. So far, most studies examined the effects of body size at a particular age with adjustment for body size at later ages. Only a few studies had results presented in subgroups of individuals defined by body size at one or more stages [13, 132, 133, 144, 148]. The most comprehensive analyses to date examined risk separately in subgroups defined by their birthweight, size at age 8 years, and adult BMI in about 2000 women with adult-onset diabetes; statistical power was limited, and few associations were statistically significant [133]. A recent study [150] based on the Nurses' Health Study used group-based trajectory modeling of body size and reported that compared to those who remained lean throughout life, individuals who are either persistently heavy over the life course or lean in early life and then become heavy in middle life have substantially higher risks of diabetes as adults.

3.4 Conclusion

This chapter outlines the topics that needed careful reflections to understand the research questions involving body size in early life and risk of adult-onset diabetes. Then, systematic reviews highlight the questions to be answered. This chapter will serve as a mainstay of this thesis and will help to devise the analysis framework, interpreting the findings and placing the findings in context.

Table 3.4: Body size variables and analyses in the large-scale studies examining the roles of body size in early life on adult-onset diabetes risk

Year	Study, country [reference]	Birthweight	Body size in childhood	Body size in early adulthood	Body size in late adulthood	How results were presented
2018	Copenhagen School Health Records Register, Denmark [144]		+	+		Results were presented in subgroups
2017	Copenhagen School Health Records Register, Denmark [145]	+				
2017	Nurses' Health Study and Health Professionals Follow-up Study, USA [150]		+	+	+	Group-Based Trajectory
2015	Copenhagen School Health Records Register, Denmark [129]	+	+			Results were adjusted in regression models
2015	Health Professional Follow-up Study, USA [131]	+			+	Results were adjusted in regression models
2015	Nurses' Health Study, USA [131]	+			+	Results were adjusted in regression models
2015	Nurses' Health Study II, USA [131]	+			+	Results were adjusted in regression models
2015	British Regional Heart study, UK [147]			+		
2014	Black Women's Health Study, USA [132]	+			+	Results were presented in subgroups
2014	Health Professional Follow-up Study, USA [15]			+	+	Results were adjusted in regression models
2011	MELANY study, Israel [148]			+	+	Results were presented in subgroups
2010	E3N cohort, France [133]	+	+	+	+	Results were presented in few subgroups
2010	Nurses' Health Study II, USA [13]		+	+	+	Results were presented in few subgroups
2009	Swedish Birth Cohort 1925-49, Sweden [130]	+				
2002	Helsinki Birth Cohort Study 1924-44, Finland [141]	+	+			No adjustments for each other
1999	Nurses' Health Study, USA [12]	+			+	Results were adjusted in regression models

4

The Million Women Study

Summary points

- The Million Women Study is a prospective cohort of 1.3 million women aged 50 years and over in the UK, recruited between 1996 and 2001 through the NHS breast cancer screening programme.
- Participants are followed for health events (e.g. hospital admissions and cancers), emigration, and death through electronic linkage to national registers up until 31 December 2017.
- Women reported their birthweight, body sizes ages 10 and 20 and body mass index at age 60 years, which were previously validated against values recorded around respective ages.
- Among the women with newly diagnosed diabetes in primary care, around 80% had at least one hospital admission with mention of diabetes during the next 10 years.

This chapter provides details about the primary data source of the thesis, the Million Women Study. It includes details about study design, settings, participants and follow-up of this cohort. It also introduces the databases which are electronically-linked to the Million Women Study. As the aim of this thesis is to investigate the associations between body size in early life and risk of adult-onset diabetes, I describe how information about body size at different ages over the lifetime was collected, and how new cases of diabetes were identified in this study. I also discuss the issues related to these measures, which may have implications on the validity and interpretations of the conclusions in this thesis.

4.1 The Million Women Study

The Million Women Study is a population-based prospective study of 1.3 million women aged 50 years and above in England and Scotland [18, 19]. It was established with a primary aim of obtaining robust observational evidence on the effects of the use of different types of menopausal hormone replacement therapy on the risk of breast cancer. At recruitment, and through the subsequent four re-survey questionnaires (occurring at every 3-5 years), the investigators collected information on a wide range of sociodemographic, reproductive, and lifestyle factors that could affect women's health. All participating women have been followed for health events (including hospital admissions, cancer and deaths) through electronic record-linkage to routinely collected National Health Service (NHS) data. For around 8% of all participants, additional record-linkage to the primary care data is available. There are also several validation studies in subsets of the cohort to understand various methodological issues associated with different measures (including reported information on questionnaires and clinical outcomes recorded in linked data). Availability of information from all these sources has made this cohort suitable for large-scale epidemiological investigations on many research questions beyond its primary aim [18].

The coordinating centre for the study is based in the Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford.

4.1.1 Recruitment

The Million Women Study recruited about one in every four UK women aged 50-64 years between 1 May 1996 and 31 December 2001. The participants were recruited through the UK's NHS Breast Screening Programme, which routinely invited UK women of this age group to attend free mammographic breast screening every three years. Sixty-six breast cancer screening centres participated in the recruitment process, and the recruitment questionnaire was sent to women with their postal invitation to breast screening. Every screening centre in England, Scotland and Wales was invited to participate in recruiting participants for the Million Women Study. Although nearly all centres showed enthusiasm for the study, practical problems, for example, difficulties in packaging the study questionnaire together with the letters typically posted to women precluded the involvement of some centres including those in London ([19]. Almost half of the women invited by the participating screening centres joined the study either by bringing their completed questionnaire with them when they were screened or by posting their questionnaire to the study coordinating centre [19].

To understand whether the inclusion of a questionnaire with the screening invitation would affect the uptake of breast screening, a pilot study [151] was conducted among 6000 women living in Oxford and West London. The uptake of breast screening was similar among those who were and were not sent a study questionnaire along with their postal invitation letter. Analysis of questionnaires sent in the initial phase of the recruitment showed that about 75% of the women attended the screening at the participating centres [19]. The national uptake rate of breast screening was approximately 75% during the same period [152]. A previous report [153] also showed that women who did not attend breast screening were more likely to come from deprived areas. Among those who attended the screening, around 66% returned the questionnaire with full identification details and signed permission to follow-up [19]. Most of the characteristics of these women were broadly similar to those of all UK women of the same age at that time [18].

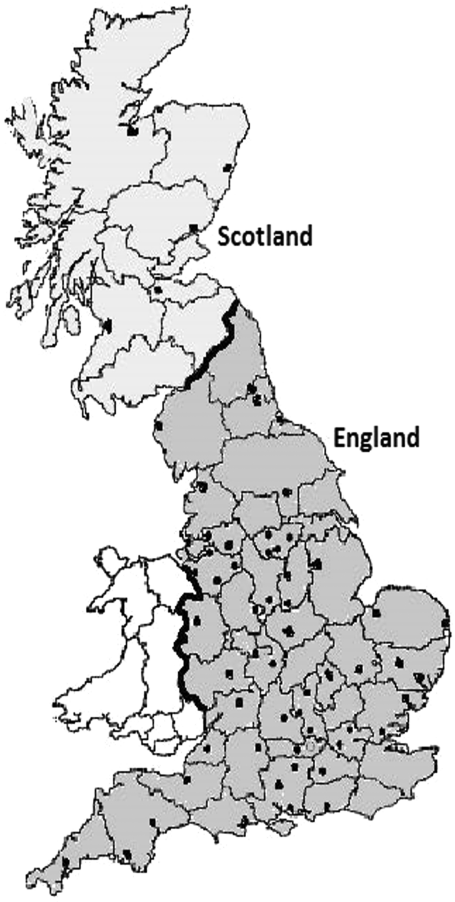


Figure 4.1: Location of the 66 NHS breast screening clinics through which Million Women Study participants were recruited

Figure adapted from Banks et al. 1999 [19] under the terms of the Creative Commons Attribution License 2.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

4.1.2 Re-surveys

Postal re-survey questionnaires were sent to all consenting and surviving participants approximately 3, 8, 12, and 15 years, on average, after recruitment. These questionnaires were used to collect information on various factors that may change over time (e.g. body weight, smoking, physical exercise etc.) and to collect new information on other factors including body size in early life and medical conditions including diabetes. Some women also provided information about diet and cognitive function through online questionnaires. All these questionnaires are available to view on the study website (`www.millionwomenstudy.org/questionnaires/`).

4.1.3 Electronic record-linkage

Women in the Million Women Study are followed by electronic record-linkage to routinely collected NHS registers on death, emigration, cancer, and hospital admissions. For a subset of 102,076 (8% of total) women in England, primary care data was available through linkage to the Clinical Practice Research Datalink (CPRD). Table 4.1 lists general electronic record-linkages to routinely collected NHS data for the Million Women Study participants along with their data provider and follow-up period.

Table 4.1: Electronic record-linkage to routinely collected NHS healthcare data in the Million Women Study

	Data provider	Population with data	Follow-up period (at the time of analysis)	Data
Deaths, emigrations, and other loss to follow-up	England: ONS/NHS Digital Scotland: ISDScotland	Entire cohort	Up until December 2017	ICD-10 cause of death
Cancer registrations	England: ONS/NHS Digital Scotland: ISDScotland	Entire cohort	Up until 31 December 2017	ICD-9 or ICD-10 codes for cancer site
Hospital inpatient/daypatient admissions	England: ONS/NHS Digital (database: HES) Scotland: ISDScotland (database: SMR)	Entire cohort	Up until December 2017	ICD-10 diagnoses OPCS4 procedures
Primary care	England: Clinical Practice Research Datalink	102,000 women in England	Up until 31 December 2012	Clinical and diagnostic Read codes

ONS = Office for National Statistics [https://www.ons.gov.uk/]

NHS Digital (formerly the Health and Social Care Information Centre) = [https://digital.nhs.uk/].

HES = Hospital Episode Statistics [154]

ISDScotland = Information Services Division Scotland (NHS National Services Scotland) [155]

ICD = International Classification of Diseases tenth version [156]

This table was adapted from Green et al. 2018 [18] under the terms of the Creative Commons Attribution License 4.0,
which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

By the time of analysis of this thesis, 16% of women had died, and 89% had had at least one hospital admission with any cause (up until 31 December 2017). Only 1.4% of the cohort had been lost to follow-up by emigration, withdrawal from the NHS, or for some other reason. Up until 31 December 2017, 21% (n=286,786) of women had an incident cancer (excluding non-melanoma skin cancer) (Table 4.2).

Table 4.2: Details of follow-up in the Million Women Study (n=1,364,308)

	Numbers
Up until 31 December 2017	
Number of deaths (%)	216,969 (15.9)
Emigrated (%)	7,317 (0.5)
Other loss to follow-up (%)	12,482 (0.9)
Total number of hospital admissions	9,114,124
Women with at least one hospital admission (%)	1,208,007 (88.5)
Women with incident cancer (excluding non-melanoma skin cancer) (%)	286,786 (21.0)

4.1.4 Ethical approval and funding

The ethical approval for the Million Women Study was taken from the Oxford and Anglia Multi-Centre Research Ethics Committee (MREC 97/01) (now known as the East of England Cambridge South Research Ethics Committee). All participants gave signed consent for re-contact and follow-up through their health records in NHS registers. I received a separate approval from the Independent Scientific Advisory Committee (ISAC) to use the primary care data (i.e. CPRD) for this thesis (ISAC protocol: 16_156). Letter confirming ethical approval from ISAC can be found in Appendix B.

The Million Women Study is funded by Cancer Research UK (grant C570/A16491) and the Medical Research Council (grant MR/K02700X/1). I was funded by a DPhil studentship from the Nuffield Department of Population Health to undertake research presented in this thesis.

4.2 Further details on record-linkage databases

To provide sufficient background on follow-up and diabetes recording in the Million Women Study, I briefly described the relevant databases in which participants were followed-up by electronic record-linkage for deaths, emigration, cancers, hospital admissions, and primary care in this section.

4.2.1 National Health Services (NHS) Central Registers

The Million Women Study participants are linked to the NHS Central Registers for information on deaths, emigrations, and cancers. In England, these data are provided by the Office for National Statistics (ONS) [154] and in Scotland, they are provided by the Information Services Division Scotland (ISDScotland) [155]. Follow-up for deaths is based on mortality data provided by ONS and ISDScotland. Mortality data include the date and underlying cause (using International Classification of Diseases tenth version [ICD-10] codes) [156] for each death. Information on changes in registration status due to emigration or other loss to follow-up is based on the migration data. Follow-up for cancer is based on cancer registration statistics data (provided by ONS and ISDScotland) which include information on primary cancers, primary cancer sites and secondary cancers (i.e. metastasis). Cancer site is recorded using the ICD-9 or ICD-10 codes.

At the time of analysis for this thesis, the end date of follow-up for deaths and emigrations was 31 December 2017 in both England and Scotland; and for cancers, it was 31 December 2015.

4.2.2 Hospital Episode Statistics (HES)

For the Million Women Participants in England, information on hospital admissions was obtained from Hospital Episode Statistics (HES). HES is a data warehouse managed by the NHS Digital (formerly the Health and Social Care Information Centre) [154]. It contains details of all hospital admissions, outpatient appointments and accident and emergency attendances at NHS hospitals in England. HES data also include information on private patients treated in NHS hospitals, patients

resident outside of England, and care delivered by treatment centres funded by the NHS [154].

There are three primary datasets for HES: admitted inpatient data (1997 onwards), outpatient data (2003 onwards), and accident and emergency data (2008 onwards) [157]. All Million Women Study participants are linked to the admitted inpatient dataset, which provides information about each finished consultant episode (commonly known as "episode") for each woman between 1 April 1997 and 31 December 2017. An episode is a continuous period of care provided by a particular consultant at a particular hospital provider. During a "spell" of treatment (i.e. the time a patient is in hospital, from admission to discharge), if a patient is transferred to another consultant, another provider or both, then a new episode will be generated. Therefore, many times a single hospital stay can be represented by more than one episode and thus will have multiple HES records for that stay [158].

For each episode, information is provided on clinical diagnoses (including up to 11 primary diagnoses and up to 41 other diagnoses), coded using the ICD-10, and up to 24 operations and procedures using OPCS-4 (Office of Population, Censuses and Surveys: Classification of Interventions and Procedures, 4^{th} revision) [159]. Clinicians involved in clinical care for admitted patients generate clinical notes, and trained-clinical coders convert clinical notes to ICD-10 and OPCS-4 codes.

HES data have been valuable in addressing diverse research questions. Although previous studies [160, 161] found that many clinicians are unaware of the use and value of the data generated from clinical notes, HES data can be used for generating sufficiently accurate evidence [160, 162, 163]. In 2001, Campbell *et al.* [164] reported that the accuracy of the UK routinely collected data was overall high (84% for diagnostic codes and 97% for procedures). A more recent systematic review [165] found that since the introduction of Payment by Results (PbR) in 2002, the accuracy rates of routinely collected datasets in the UK have improved significantly. PbR is a system of paying NHS healthcare providers a standard national price or tariff for each patient seen or treated [166]. It is important to remember that despite general accuracy, there are specific issues with HES depending on how it is used.

4.2.3 Scottish Morbidity Records

For the women recruited in Scotland (n=120,000), information on hospital admissions was available from the Scottish Morbidity Records (SMR), managed by the ISDScotland [155]. SMR has four records: SMR00 for outpatient attendance, SMR01 for general acute inpatient and day cases, SMR02 for maternity inpatient and day cases, and SMR04 for mental health inpatient and day cases [167]. The Million Women Study participants are linked to the SMR01 record, which is an episode-based record relating to all inpatients and day cases discharged from non-psychiatric, non-obstetric wards in Scottish hospitals. Like HES in England, clinical notes are converted to standard clinical codes (using the ICD10 and OPCS4 classifications) by trained-clinical coders. According to the latest report from ISDScotland [168], the average accuracy rate for diagnostic codes is 88% and for main operation/procedure is 94%.

4.2.4 Clinical Practice Research Datalink (CPRD)

For about 8% of the Million Women Study participants in England (n=102,076), primary care data was available through linkage to the Clinical Practice Research Datalink (CPRD) for the period from 1 January 1990 to 31 December 2012. CPRD is an electronic health record containing anonymised patient data on diagnoses, symptoms, test results, prescriptions, and patient history for patients registered with an NHS General Practitioner (GP) practice in England [169]. The database known as CPRD GOLD contains data from GP practices using Vision software. CPRD is managed by a division in the Medicines and Healthcare products Regulatory Agency (MHRA). CPRD contains medical records with varying observation period, depending on when each individual joins or leaves a particular GP practice that contributes data to CPRD. Nearly 11 million people (approximately 7% of the UK population) have records in CPRD, and they are broadly representative of the general population in terms of age, sex, and ethnicity [170, 171]. CPRD contains different datasets on practices, staffs, and patients. Practice dataset contains details about region, collection date and "up to standard" date (i.e. date at which the

practice data is deemed to be of research quality). The staff dataset contains details about healthcare providers. For patients, information is collected on demographics, consultations, clinical symptoms, diagnostic test results, prescriptions and referrals to the secondary care [172].

Clinical data are recorded using a hierarchical system of Read Codes (version 2), consisting of nearly 96,000 codes [173]. Read Codes are a coded thesaurus of clinical terms, which have been used in the NHS since 1985. These codes provide a semi-standard vocabulary for clinicians to record patient findings and procedures, in health and social care IT systems across primary and secondary care [174].

CPRD has recently looked at data quality at both practice and patient levels. Data from a practice qualifies for research use if they fulfil the criteria based on death recording and continuity in the data. Patients' data is flagged as acceptable if it meets certain quality standards [175]. Problems with miscoding and misclassification in CPRD have been documented in previous studies [176, 177], especially before 2004, when the Quality and Outcomes Framework (QoF) was introduced [178]. QoF is a scheme by which GPs are partly funded through the achievements of set target, and it has been found to incentivise GPs' clinical coding details and accuracy. In April 2016, QoF was decommissioned in Scotland. Recent reviews of validation studies using CPRD data reported high positive predictive values for a wide range of outcomes [179, 180].

4.3 Body size at different ages

This section discusses body size indices at different ages over the life course in the Million Women Study, and their validation. In the recruitment questionnaire, the Million Women Study participants reported their current height in feet and inches and current weight in stones and pounds. In the 3-year re-survey questionnaire, they were asked about their body size for their current age and when they were young. Table 4.3 lists all the body size variables reported in these two questionnaires with percentages of valid information for each of these variables.

In this thesis, these body size indices from the Million Women Study will be used: birthweight, body size at age 10, body size at age 20, and body mass index (BMI) when participants completed the 3-year re-survey questionnaire. These body size indices were previously compared against values recorded in subsamples of the cohort at around birth, at ages 11, 20, and 66 years [96, 181]. I will use the categories of different body size indices from those comparison studies so that the recorded values for each category can be used.

As discussed in the previous chapter, the likely effects of errors in the reported body size indices should be considered while investigating the associations between body size in early life and risk of developing diabetes in adulthood.

Table 4.3: Lifetime body size variables in the Million Women Study

Variables	Unit of reporting	Unit of analysis	% with non-missing information [a]
Recruitment questionnaire			
Height	Feet and inches	Centimetres	98.4
Weight	Stones and pounds	Kilograms	95.8
Body mass index	Derived	Kg/m^2	94.8
3-year re-survey questionnaire			
Weight	Stones and pounds	Kilograms	93.5
Body mass index	Derived	Kg/m^2	91.9
Waist circumference	Inches	Centimetres	64.5
Hip circumference	Inches	Centimetres	65.3
Waist-to-hip ratio	Derived	None	62.9
Waist-to-height ratio	Derived	None	63.9
Birthweight	Pounds and ounces	Kilograms	56.2
Relative body size at age 10 years	Thin, average, plump	Thin, average, plump	99.4
Clothes size at age 20 years	UK size <8, 10, 12, 14, 16, 18+	Small ($\leq$ size 12), medium (size 14), and large ($\geq$ size 16)	98.8
Clothes size at resurvey	UK size <8, 10, 12, 14, 16, 18+	Small ($\leq$ size 12), medium (size 14), and large ($\geq$ size 16)	98.6

[a] Of those who completed the 3-year re-survey (n=866,532)

4.3.1 Adult Body mass index (BMI)

The reported height from recruitment questionnaire and weight from the relevant questionnaire were converted to centimetres (cm) and kilograms (kg) to derive body mass index (BMI) at recruitment and at each re-survey for participating women. BMI is usually referred to as self-reported or measured depending on how information on weight and height was collected, but BMI itself is usually not directly reported or measured. Around 92% of all women who completed the 3-year re-survey had BMI information (mean 26.1 [SD 4.6] kg/m^2).

To understand the validity of estimated BMI based on reported weight and height, Wright *et al.* [181] compared BMI based on reported height and weight from different questionnaires with BMI estimated from measured height and weight in a subsample of the Million Women Study. About nine years after the recruitment, in 2008, approximately 14,000 women were invited to attend a GP appointment to have height and weight measured. A total of approximately 4000 women visited their GPs and had their height and weight measured at around age 66 years on average. Reported BMI at recruitment and at 3-year re-survey were highly correlated with measured BMI in 2008 (Pearson correlation 0.85 and 0.88, respectively). However, on average, reported BMIs at recruitment and at 3-year re-survey were 1.4 kg/m^2 and 1.1 kg/m^2 lower than the mean BMI based on measured height and weight at age 66 years. While these numbers indicate some form of systematic under-reporting in the estimated BMI based on reported weight and height, it is important to note that the weight and height measurements were taken approximately six years after they filled the 3-year resurvey questionnaire. Figure 4.2 shows the mean measured BMIs for each of the categories of reported BMI at 3-year re-survey, and the mean reported BMIs were lower than mean measured BMIs in each BMI category.

Previous studies [182, 183] reported that self-reported BMI tended to be underestimated because of underestimation of weight and overestimation of height. The magnitude of underestimation is usually greater among individuals with higher BMI [184, 185].

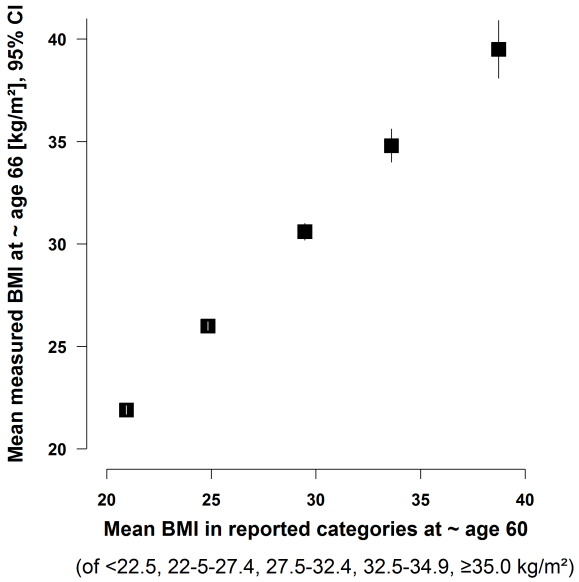


Figure 4.2: Mean measured BMI versus mean self-reported BMI, in each self-reported BMI category

Figure was recreated using results from Wright et al. 2015 [181]

Regression dilution ratios (RDRs) can be estimated to quantify the apparent attenuation of estimates due to reporting errors [186, 187]. RDR is a non-parametric estimate of the ratio of such a biased estimate to the coefficient that would be found if analyses were conducted using true values of the variable of interest. This ratio depends on both random error and systematic errors but is largely independent of the true coefficient [96]. RDR for BMI reported at 3-year re-survey was 0.96 (95% CI: 0.95-0.98) [181], suggesting that the likely attenuation of estimates of linear disease-exposure associations would not be large.

In the Million Women Study, reported adult BMI has been used to investigate associations with a wide range of outcomes, including all-cause mortality [188], coronary heart disease [189], stroke [190], venous thromboembolism [191], musculoskeletal problems [192–195], liver cirrhosis [196], various cancers [197–201], hospital admissions [202], and hospital costs [203].

4.3.2 Birthweight

About three years after recruitment in the Million Women Study, in 2001 (IQR:2000-2003), 56% women reported their birthweight in pounds which were then converted to kg for analysis. The average birthweight was 3.2 (SD 0.7) kg.

Birthweight reported by the participants was compared against weight recorded around the time of birth in a subset of women (n=541) who also participated in the MRC National Survey of Health and Development (NSHD) study [204]. Reported birthweight was found to be highly correlated (Pearson correlation: 0.78) with birthweight recorded around the time of birth. On average, reported birthweight was 0.1 (SD 0.4) kg lower than recorded birthweight (i.e. systematic under-reporting). However, there was substantial over-reporting among women with high birthweight [96]. Figure 4.3 shows the mean recorded birthweight in each category of reported birthweight in the comparison study. For birthweights of <2.5 kg and $\geq$ 4.0 kg, recorded birthweight showed significant differences.

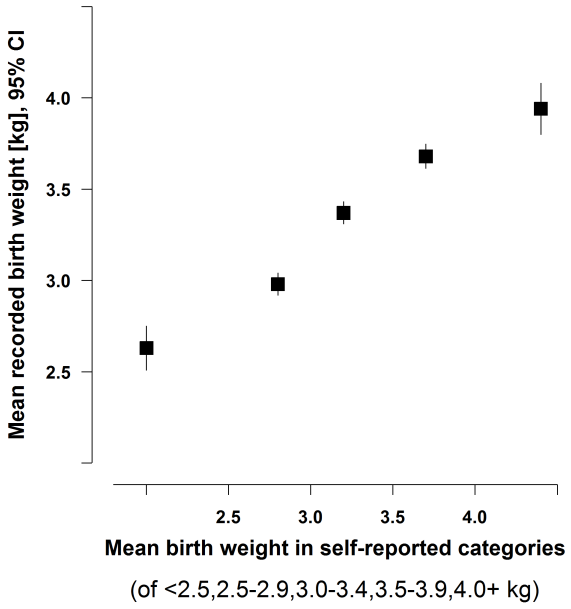


Figure 4.3: Mean recorded birthweight versus mean reported birthweight, in each reported birthweight category

Figure was recreated using results from Cairns et al. 2011 [96]

The likely effects of reporting errors in birthweight on the estimates of disease-exposure associations were found to be substantial (RDR 0.59, 95% CI: 0.50-0.67) [96]. In the absence of birthweight records, epidemiological studies in different populations have been using birthweight reported in middle ages to estimate its association with various disease outcomes [205–207]. In the Million Women Study, reported birthweight had been used previously to estimate associations with adult cancer incidence [208], and hip and knee replacement [209].

4.3.3 Body size at ages 10 and 20 years

In the 3-year re-survey questionnaire, the Million Women Study participants were asked about their body size at age 10 years relative to their peers (thin, average, plump) and their UK clothes size (8 or less, 10, 12, 14, 16, 18+) when they were 20 years old. For body size at age 20 years, women were grouped into three categories: small ($\leq$ size 12), medium (size 14), and large ($\geq$ size 16).

Recalled body size information at ages 10 and 20 were compared with BMI derived from recorded height and weight at age 11 and 20 years among 541 Million Women Study participants who also participated in the NSHD study [96]. Reported ordinal body sizes at these two ages were moderately correlated with recorded BMI at relevant ages (Spearman's correlation: 0.51 for body size at age 10 and 0.63 for body size at age 20, P<0.001 for each). As shown in Figure 4.4, mean measured BMI for thin, average, and plump categories at age 10 was 16.2, 17.5, and 20.4 kg/m^2, respectively. The average recorded BMI for being small, average, and large at age 20 was 19.9, 21.3, and 24.3 kg/m^2, respectively (Figure 4.5). Although it was not possible to assess the likely effects of reporting errors for body size at ages 10 and 20, clear associations between categories of body size and mean measured BMIs were observed. Previous studies have reported that when recalled, body sizes in early life are generally under-reported, with more significant under-reporting among overweight or obese adults [210, 211].

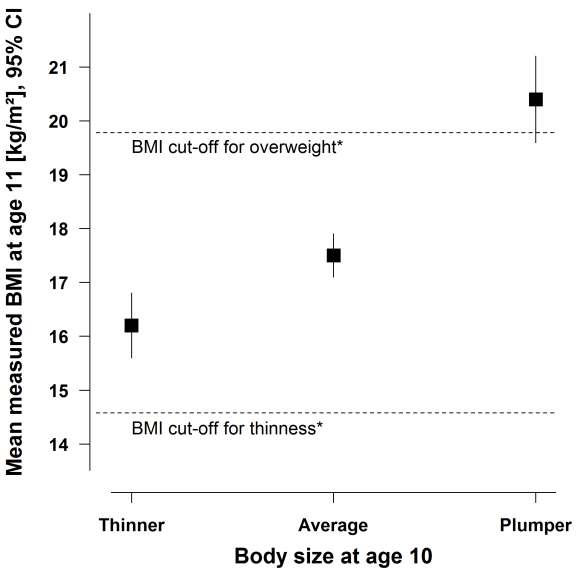


Figure 4.4: Mean measured BMI at age 11 years versus categories of relative body size at age 10 years

Figure was recreated using results from Cairns et al. 2011 [96]

Distinct BMI means associated with each level of body size will help in clinical and public health interpretations of such measures. According to the International Obesity Task Force (IOTF) classification [93], BMI cut-offs for thinness and overweight were 14.7 kg/m^2 and 19.8 kg/m^2, respectively for girls at age 11 years. For age 20, the BMI cut-offs for thinness and overweight were 18.5 kg/m^2 and 25.0 kg/m^2, respectively. Therefore, it is important to acknowledge that the results for reported body size at ages 10 and 20 years in the Million Women Study cannot be interpreted for extreme levels of thinness and overweight.

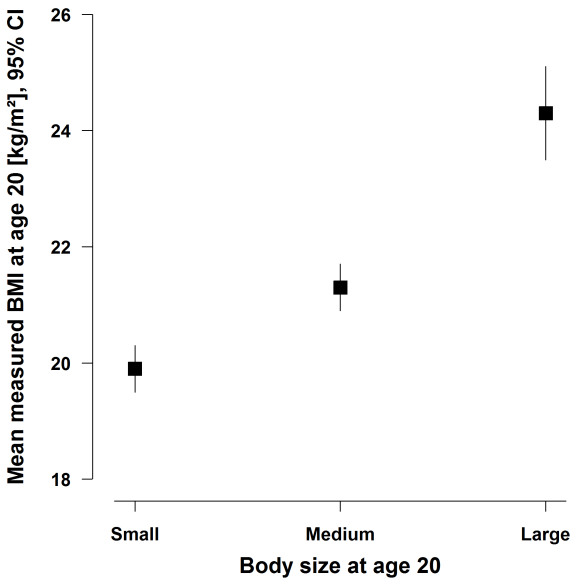


Figure 4.5: Mean measured BMI at age 20 years versus categories of body size at age 20 years

Figure was recreated using results from Cairns et al. 2011 [96]

Previously in the Million Women Study, investigators used the reported body size at ages 10 and 20 years for examining their associations with postmenopausal endometrial cancer [212], lymphoid malignancies [213], and hip-knee replacement [209].

4.4 Correlations between body size indices at different ages

As mentioned in earlier chapters, correlations between body size indices at different ages are essential to understand the effects of body size at a particular stage on the risk of developing diabetes.

Figure 4.6 shows the correlation between indices of body size at different ages in the Million Women Study. There were weak correlations between body size indices at any two ages (Spearman's correlation ranging from 0.12 to 0.37) except between birthweight and adult BMI (0.02).

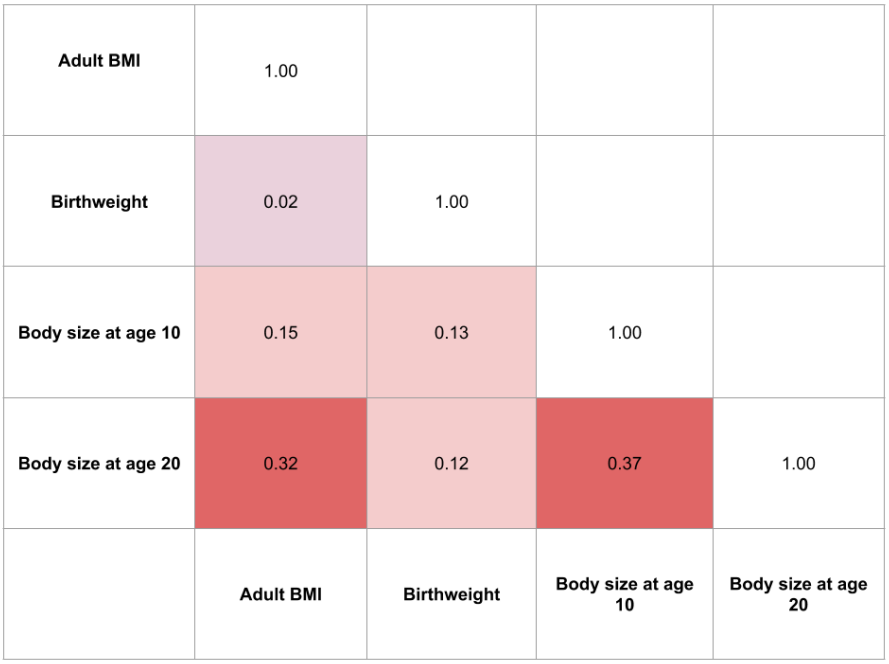


Figure 4.6: Correlation coefficients between lifetime body size variables in the Million Women Study

Pearson correlations were used for evaluating the linear relationship between two continuous variables (e.g. birthweight and adult BMI). Spearman's correlations were used for assessing relationships involving an ordinal variable (e.g. body size at age 10, body size at age 20). All the coefficients were significant, $p<0.0001$

Further examinations in Figure 4.7 and Figure 4.8 suggest that indices of body size at different ages were associated with each other in varying degrees, except birthweight and adult BMI. Those with a birthweight of <2.5 kg and

≥4.0 kg had higher adult BMI on average than those with birthweight of 2.5-4.0 kg, making the overall relationship between birthweight and adult BMI non-linear (Figure 4.7). Therefore, the correlation coefficient of 0.02 is not a good reflection of their association.

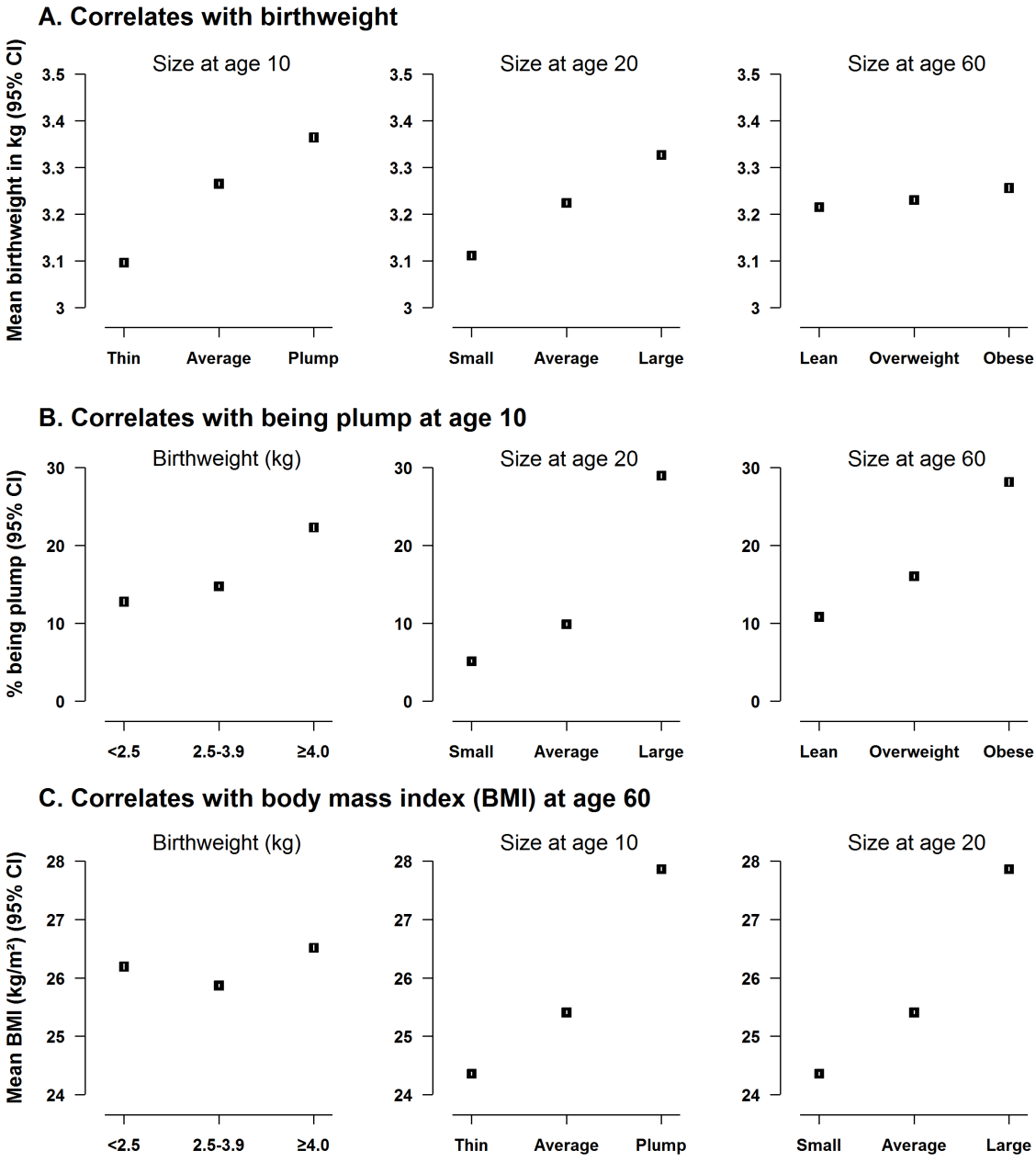


Figure 4.7: Correlations between indices of body size at different ages

Figure 4.8 has values for being thin at age 10, and for being small or large at age 20 years.

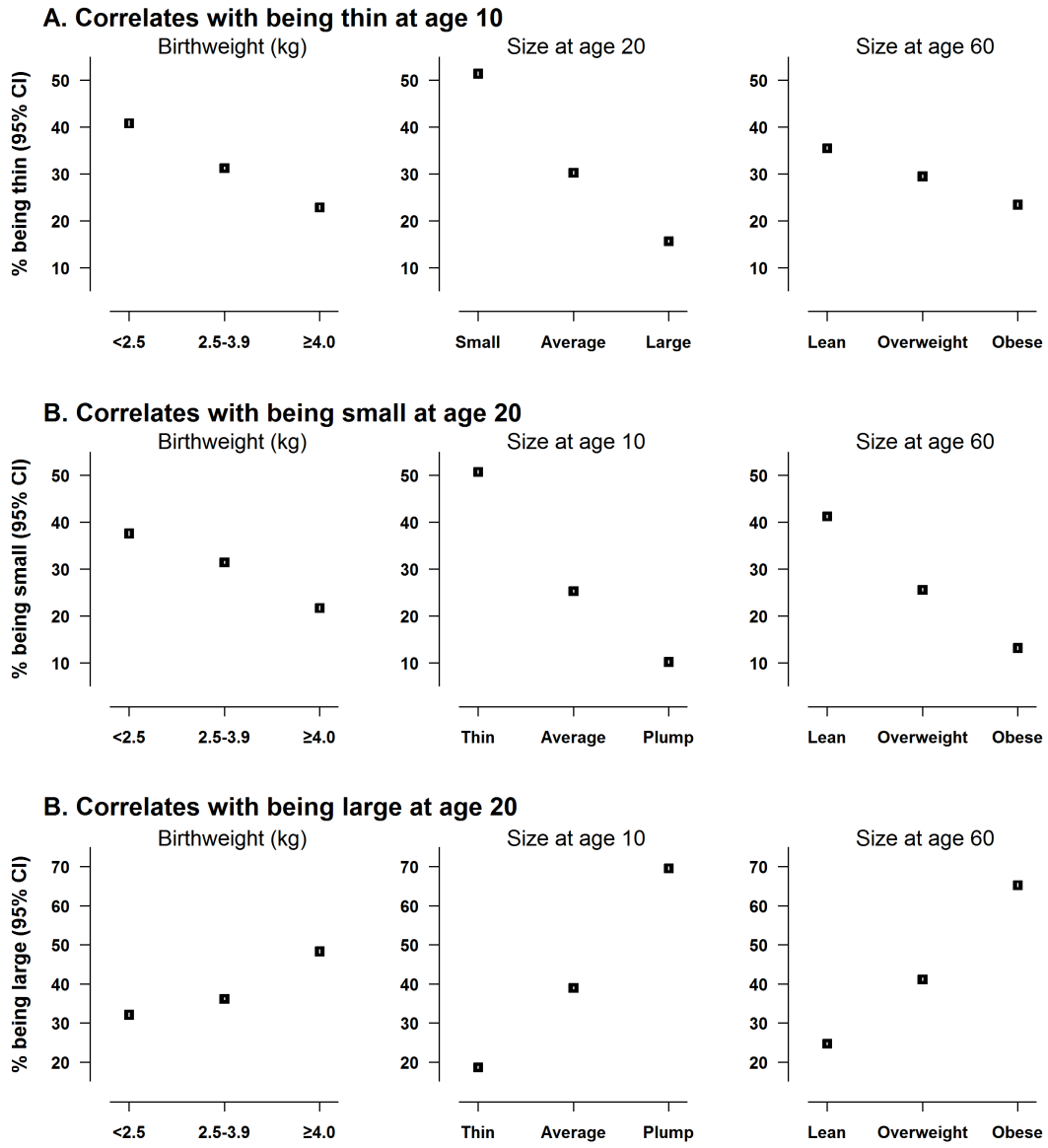


Figure 4.8: Correlations between indices of body size at different ages

Overall, the availability of body size indices at around birth, mid-childhood, early adulthood and late adulthood for a large number of participants in the Million Women Study allows examination of the risk of adult-onset diabetes separately among women with different combinations of body size so that the details of any specific or consistent associations can be identified. Although there were some correlations between reported and recorded body size measures, possible reporting errors warrant cautious interpretation of any apparent associations with the risk of adult-onset diabetes.

4.5 Diabetes recording

In the Million Women Study, some information about diabetes is available from reported data in the study questionnaires and from record-linkage data on hospital admissions (for the whole cohort) and on primary care (for only around 8% of all participants). In this section, I will discuss how prevalent and new cases of diabetes can be identified in this cohort, and possible issues related to them.

4.5.1 Diabetes in the study questionnaires

Investigators of the Million Women Study collected information on a wide range of medical conditions, including diabetes through the recruitment and subsequent re-survey questionnaires. Table 4.4 lists the variables related to diabetes in different questionnaires. There are variables on diabetes diagnosis, treatment for diabetes, insulin use, and family history of diabetes. While these variables can be used to identify women who already have diabetes, they are not well-suited for identifying new cases of diabetes during the follow-up. Information on family history of diabetes can be used as an essential covariate for the analysis of body size and risk of developing diabetes in adulthood.

Table 4.4: Information on diabetes in different questionnaires of the Million Women Study

Variables related to diabetes	Recruitment	3-year re-survey	8-year re-survey	12-year re-survey	15-year re-survey
Related to diabetes diagnosis					
Ever diagnosed with diabetes	+			+	+
Diagnosed with diabetes in the last 5 years		+	+		
Age at first diagnosis of diabetes		+	+	+	
Related to diabetes treatment					
Ever treated for diabetes					
Currently treated for diabetes	+	+			
Age at first treatment for diabetes		+			
Regular intake of diabetes medications				+	
Recent use of insulin		+	+		
Related to family history of diabetes					
Maternal history of diabetes		+	+	+	+
Paternal history of diabetes		+	+	+	+
Sibling history of diabetes			+	+	+

4.5.2 Diabetes recording in hospital admissions data

As mentioned previously, all participating women in the Million Women Study are followed up through record-linkage to routinely collected NHS databases on hospital admissions, i.e. HES in England [154] and SMR in Scotland [167]. Information about all inpatient and day-care admissions with corresponding dates and ICD-10 codes [156] for primary and other clinical diagnoses for each admission episode are available in these records. Women who have had diabetes can be identified in hospital admissions data by using the ICD-10 codes for diabetes listed in Table 4.5.

Table 4.5: International Classification of Diseases 10th version (ICD-10) codes for diabetes

ICD-10 code	Type of diabetes
E10	Type 1 or Insulin-dependent diabetes mellitus
E11	Type 2 or Non-insulin-dependent diabetes mellitus
E12	Malnutrition-related diabetes mellitus
E13	Other specified diabetes mellitus
E14	Unspecified diabetes mellitus

Table 4.6 shows the details about diabetes recording among the Million Women Study participants in the hospital admissions data. There were 9.1 million episodes of hospital admissions until 31 December 2017 and about 89% women had at least one hospital admission episode. Among all episodes of hospital admissions, 10% had a code for diabetes. Diabetes was recorded as an "other" diagnosis in almost all of these hospital admissions (97%). The majority (88%) of hospital admissions with diabetes codes had an ICD-10 code (E11) for type 2 diabetes. For a small proportion of women (<2%), different ICD-10 codes for diabetes was recorded in subsequent hospital admissions.

Table 4.6: Diabetes recording in hospital admissions data in the Million Women Study

	Numbers (%)
Up until 31 December 2017	
Number of women	1,364,308
Total number of hospital admissions	9,114,124
Women with at least one hospital admission	1,208,007 (88.5)
Among total hospital admissions	
Hospital admissions with diabetes codes	865,841 (9.5)
Among hospital admissions with diabetes codes	
Diabetes recorded as an "other" diagnosis	839,866 (97.0)
ICD-10 code (E11) for type 2 diabetes	761,940 (88.0)
ICD-10 code (E10) for type 1 diabetes	69,267 (8.0)
ICD-10 code (E12-14) for other types of diabetes	34,634 (4.0)
Among women with diabetes codes in hospital admissions	
Percentage of different ICD-10 codes in subsequent admissions	1.9%

4.5.3 Diabetes recording in primary care data

For about 8% of the Million Women Study participants, data linkage with CPRD is established up until 31 December 2012. CPRD has information from General Practitioner (GP) practices on clinical diagnoses, diagnostic test results, prescription, referral to secondary care, and other factors, recorded using Read codes [173]. Diabetes cases can be identified in CPRD if any of the diabetes-specific Read Codes are available. According to the Read code system, diabetes diagnoses codes start with the capital letter C. Before 2006, codes beginning with C10 were used to identify patients with diabetes. In 2006, Read codes for diabetes were refined to include the type: C10E0 to C10EP for type 1 and C10F0 to C10EQ for type 2) [214]. In addition to the diagnostic codes for diabetes (e.g. C codes), there are codes available in CPRD which also suggest a diabetes diagnosis (e.g. 'Diabetes mellitus with unspecified complication' or 'Seen in diabetic clinic') or strongly imply a diabetes diagnosis (e.g. 'Cellulitis in diabetic foot', 'Diabetic nephropathy') [174].

Here, I used the Read Code list for identifying diabetes from Tate *et al.* [40], which combined both diagnostic codes and codes suggestive of a diabetes diagnosis. I also added Read Codes from the ClinicalCodes repository [215], which holds code lists for all published electronic medical record studies. All these Read Codes were grouped into four categories, as listed in Table 4.7.

Table 4.7: Categories of Read Codes for diabetes in the linked primary care data (CPRD)

Category	Description
Category 1	Type 1 diabetes with a diagnostic (C) code
Category 2	Type 2 diabetes with a diagnostic (C) code
Category 3	General C code – type not specified
Category 4	Read codes suggestive of a diagnosis of diabetes or a diabetes diagnosis is strongly implied

Adapted from Tate et al. [40]

CPRD also has several definitions of dates which are important to identify any clinical outcome, including diabetes [175]. Details about these dates in CPRD data is given in Table 4.8.

Table 4.8: Description of different dates used in CPRD data

Date	Definition
Current Registration Date	Date the patient's current period of registration with the practice began (date of the first 'permanent' record after the latest transferred out period)
Transfer Out Date	Date the patient transferred out of the practice, if relevant. Empty for patients who have not transferred out
Death Date	Date of death of patient – derived using a CPRD algorithm
Last Collection Date	Date of the last collection for the practice
Up To Standard Date	Date at which the practice data is deemed to be of research quality. Derived using a CPRD algorithm that primarily looks at practice death recording and gaps in the data
Event Date	Date associated with the event, as entered by the GP

Information collected from the CPRD GOLD Data Specification [175]

Using the codes for diabetes and corresponding dates in HES and CPRD, I will compare diabetes recording in hospital data and in primary care data in the next section.

4.6 Comparison of diabetes recording in hospital admissions data and in primary care data

Diabetes is largely diagnosed and managed in primary care in the UK [39], but only a small number of women in the Million Women Study had linkage to CPRD data. On the other hand, diabetes recording in hospital admissions data provides a unique opportunity for epidemiological investigations on diabetes with virtually complete long-term follow-up, but such method can have several limitations, including but not limited to under-reporting and delayed recording [216]. According to the UK National Diabetes Inpatient Audit, over a million people with diabetes were admitted to hospital in England in 2017, and 92% of them were not admitted because of their diabetes [42]. Of all inpatient spells in the English Hospital Episode Statistics, 6.2% have diabetes coded at discharge [217]. Comparison between Scottish National Diabetes Register and hospital admissions data from 2000 to 2007 showed that only 59% of those with a previous diagnosis of diabetes had diabetes recorded in their hospital admissions [218].

In this section, I compare diabetes recording in hospital admissions data and in primary care data to better understand potential issues with using hospital admissions record for identifying new cases of diabetes in the Million Women Study.

4.6.1 Rate of diabetes recording in hospital admissions data and primary care data

Rates of diabetes recording by age at diagnosis in hospital admissions data and primary care data were estimated among 85,435 women from the Million Women Study who had no record of diabetes at or before recruitment and also had linked primary care data. As shown in Figure 4.9, the age-specific rates of diabetes recording were almost half in hospital admissions data when compared to primary care data for all age groups. In hospital admissions data, age-specific rates ranged between 1.7 per 1000 women per year for age group 55-59 years and 10.2 per 1000 women per year for age group $\geq$75 years. In primary care data, the age-specific

rates for diabetes for the age group 55-59 years and ≥75 years were 2.7 and 23.1 respectively per 1000 women per year.

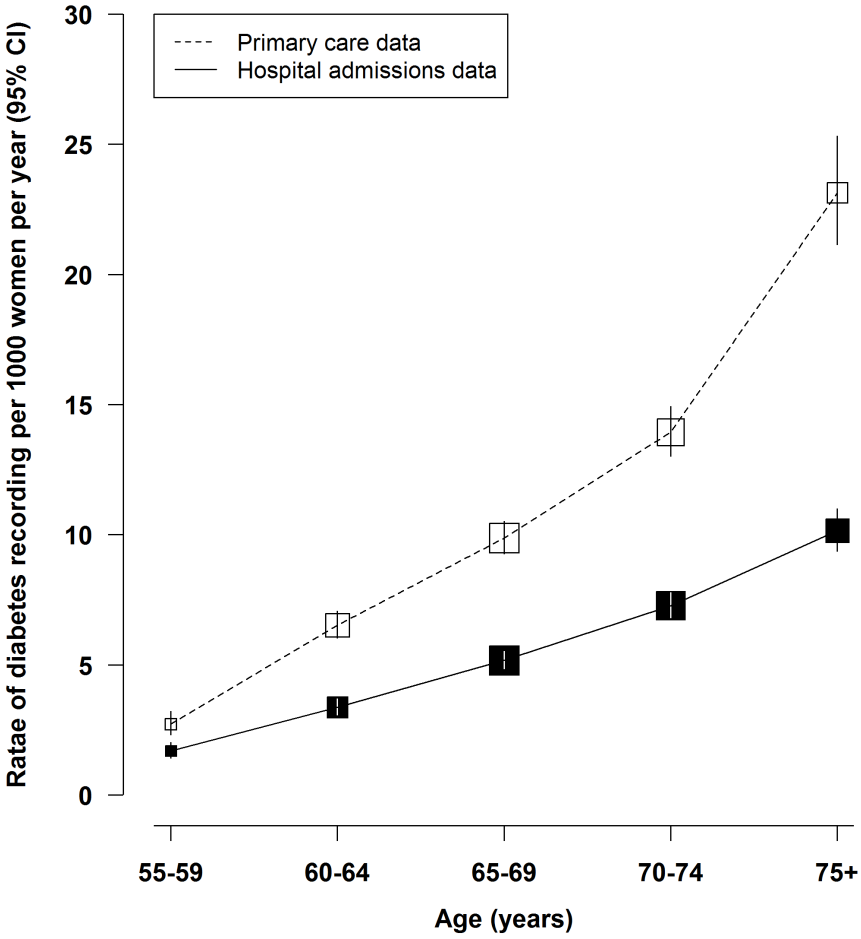


Figure 4.9: Rate of diabetes recording by age in hospital admissions data and primary care data

Analysis was based on 85,435 women who had no prior record of diabetes at at or before recruitment and also had linked primary care data.

4.6.2 Probability of first hospital admission with diabetes records among those with a diabetes record in primary care

One of the main concerns, when one relies on diabetes record in hospital admissions data to identify new diabetes cases, is under-detection. This can happen mainly for two reasons: firstly, not everyone with diabetes requires hospital admission, and secondly, even if they had hospital admission, the records might not include diabetes. A recent study [219] on the UK Biobank found that less than half of the individuals identified with type 2 diabetes in primary care had evidence of type 2 diabetes in hospital admissions data.

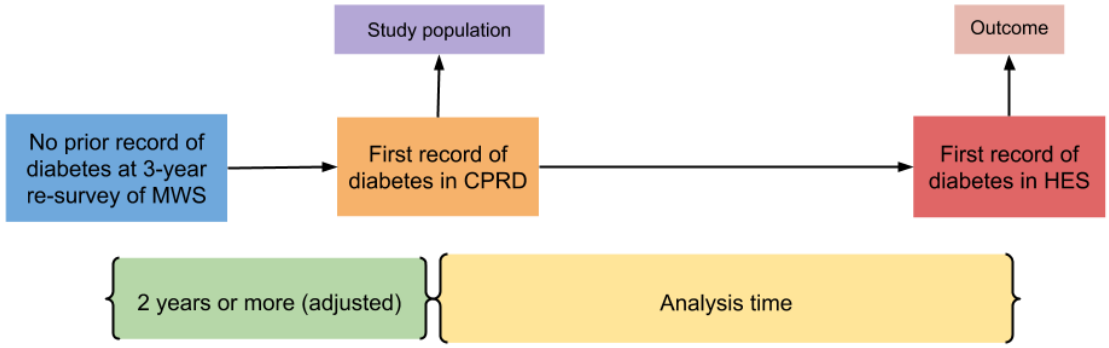


Figure 4.10: Schematic diagram of the analysis for comparison of diabetes recorded in the HES with diabetes recorded in CPRD

Abbreviations: CPRD: Clinical Practice Research Datalink; HES: Hospital Episode Statistics; MWS: Million Women Study

Here, I examined the probability of first hospital admission with diabetes by time since the first record of diabetes in primary care in England, and estimate the detection rate by time of follow-up. As shown in the Figure 4.10, among those who completed the 3-year re-survey and had new diabetes recorded in CPRD were followed up for their first hospital admission with diabetes record. New diabetes cases after 3-year re-survey were assumed when: 1) women reported that they did not have or treated for diabetes in the 3-year re-survey and recruitment questionnaires, 2) had no diabetes record previously in CPRD, and 3) at least 2 years of follow-up in CPRD prior to the diabetes recording. I looked at the

cumulative probability of the first admission to hospital with diabetes, by time since the first record of diabetes in primary care.

Among those who did not have a record of diabetes at or before the 3-year re-survey, 1765 women had a first record of diabetes after at least 2 years of follow-up in CPRD. The median year of diabetes recording in CPRD was 2011 (IQR: 2007-2012). By the end of 10 years into follow-up, approximately 80% of new diabetes recorded in CPRD was also detected in hospital admissions data (Figure 4.11). The median time for a new diabetes to be detected in hospital admissions record was 3.7 (95% CI 3.4-4.1) years.

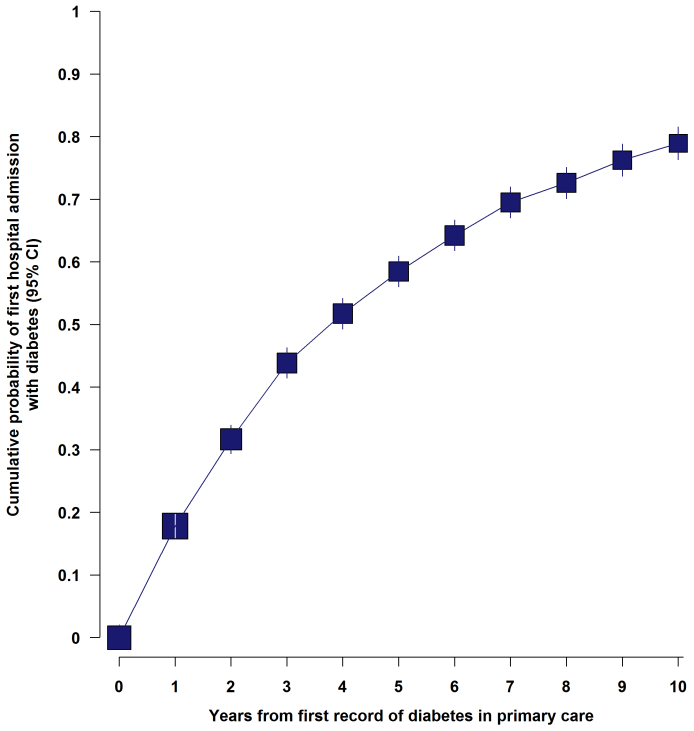


Figure 4.11: Cumulative probability of first admission to hospital with diabetes, by time since first record of diabetes in primary care

The likelihoods of hospital admission and of diabetes recording in hospital admission can be influenced by age, socioeconomic status, smoking, physical activity, and BMI [220, 221]. For this thesis, it is important to understand how much the probabilities of diabetes recording in hospital admission can vary by different body size indices and other socioeconomic and lifestyle variables.

As shown in Figure 4.12, the median time lags for diabetes diagnosis at hospital admissions for women with BMI <25.0, 25.0-29.9, and ≥30.0 kg/m^2 were 5.0 (95% CI: 4.1-6.9), 4.1 (95% CI: 3.5-4.6), and 3.3 (95% CI: 2.4-3.9) years, respectively. The median times taken for diabetes to be recorded in hospital admissions were 4.2 (95% CI: 3.3-5.5), 4.1 (95% CI: 3.4-4.6), and 3.2 (95% CI: 2.4-3.9) years for birthweight of <2.5, 2.5-3.4, and ≥3.5 kg, respectively. The corresponding numbers for body size at ages 10 were: thin (3.9 [3.4-4.4]), average (3.8 [3.4-4.3]), and plump (3.6 [3.0-4.6]); and for body size at age 20 years were: small (4.0 [3.4-4.9]), medium (4.2 [3.5-4.9)], and large (3.5 [3.2-4.0]) years (Figure 4.12).

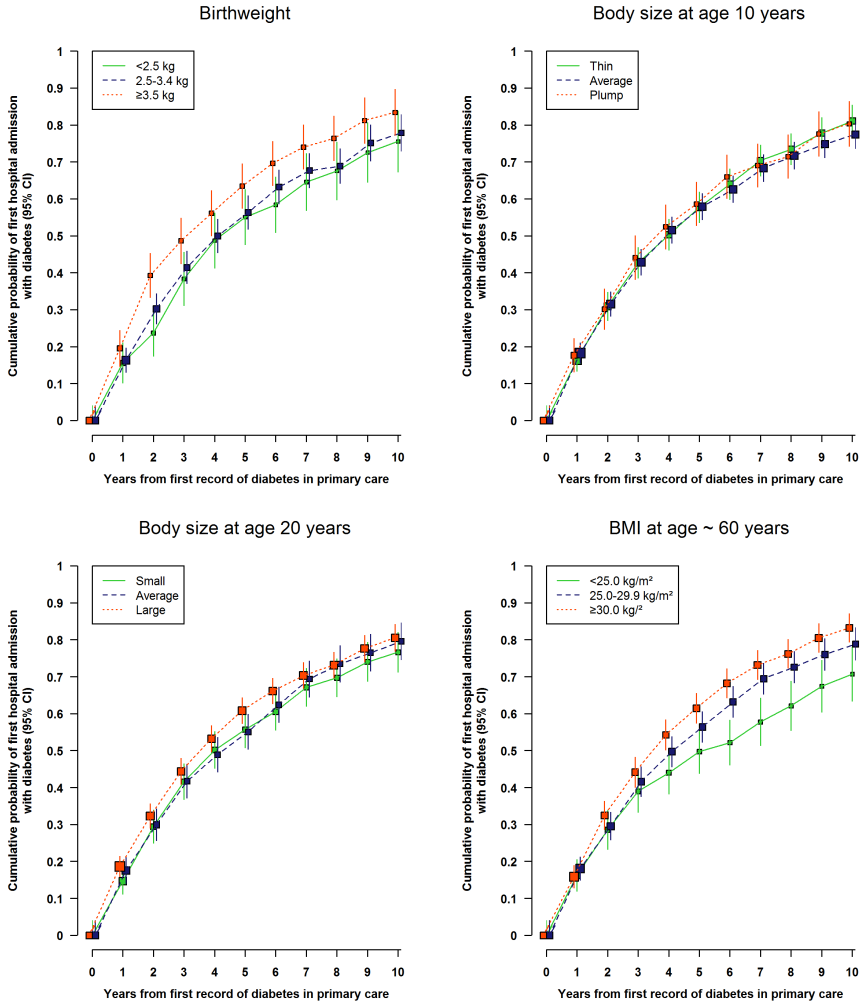


Figure 4.12: Diabetes recording in Hospital Episode Statistics (HES) among women who had diabetes recorded in Clinical Practice Research Datalink (CPRD), by body size variables

As shown in Figure 4.13, the cumulative probabilities of first hospital admission with diabetes among those who had diabetes in CPRD differed by levels of social deprivation, smoking, alcohol drinking, and strenuous exercise, which can confound the associations of body size at different ages and risk of adult-onset diabetes.

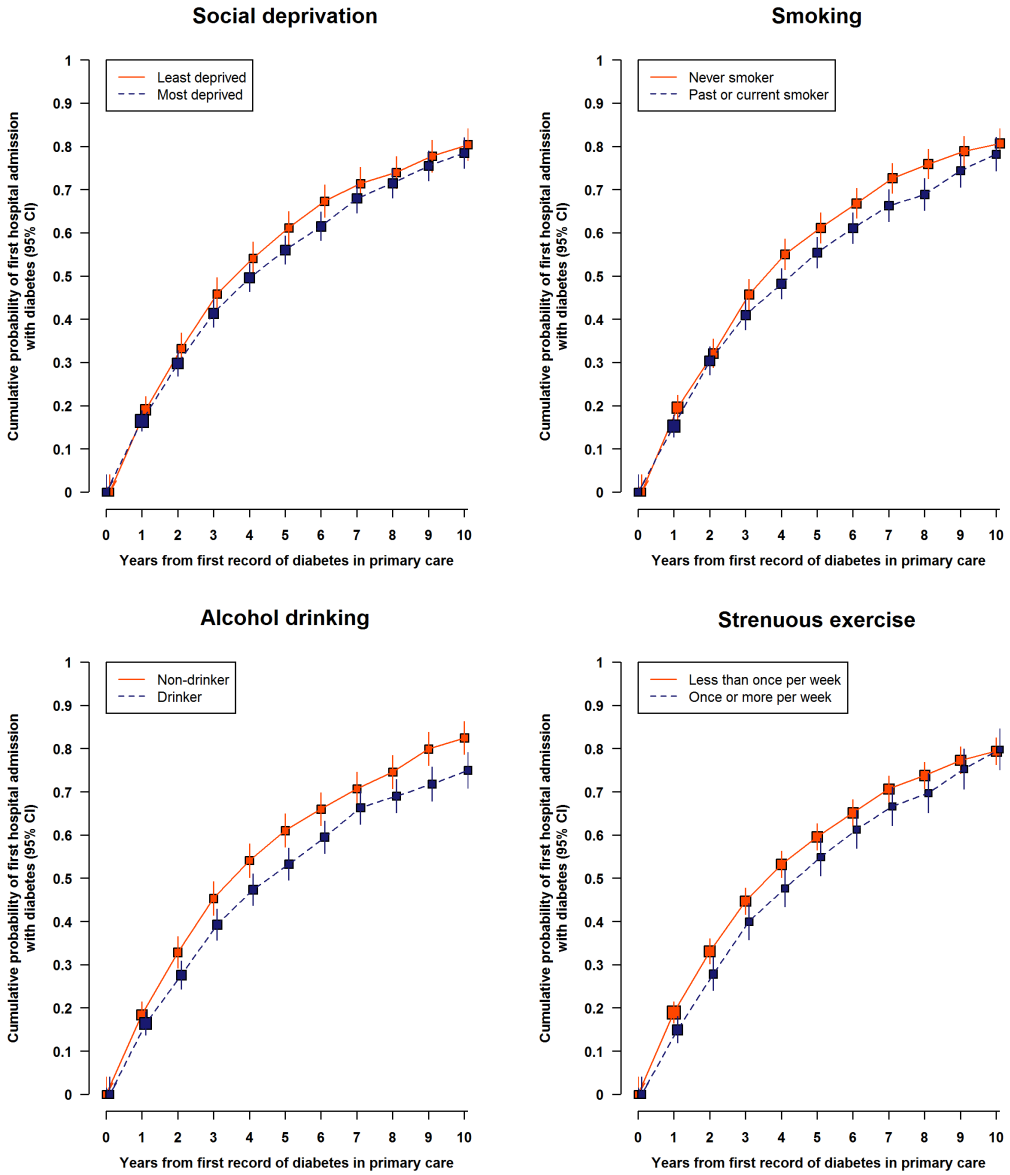


Figure 4.13: Diabetes recording in Hospital Episode Statistics (HES) among women who had diabetes recorded in Clinical Practice Research Datalink (CPRD), by socioeconomic and lifestyle variables

Overall, these results show small but non-trivial differences in detecting diabetes in hospital admissions data in the Million Women Study by body size at different ages and other characteristics. These differences should be taken into consideration when investigating the associations between body size in early life and risk of developing diabetes in adulthood. The likely effects of these differences are especially important when the effect size is small but perhaps less problematic in associations with moderate to large effect sizes.

4.6.3 Recording of diabetes in primary care among those with diabetes in hospital admissions

Another issue with diabetes recording in hospital admissions is whether anyone with a diabetes code in hospital admissions actually has diabetes. Ideally, this can be explored by reviewing hospital records, which was not possible for the Million Women Study. However, if a woman with diabetes recorded in hospital admission also has a diabetes record in her primary care data, then it is assumable that diabetes codes in her hospital admissions actually reflect her diabetes.

Here, I compared diabetes recorded in hospital admission data with diabetes recorded in primary care data. Observation periods in CPRD and HES differ; therefore, for fair comparisons in diabetes recording between hospital admission and primary care, I restricted the analysis to women with overlapping observation period. To be considered as overlapping observation period, the observation period in CPRD had to cover at least 12 months before and 12 months after the first HES record of diabetes. This is explained with the schematic diagram for two hypothetical women in Figure 4.14.

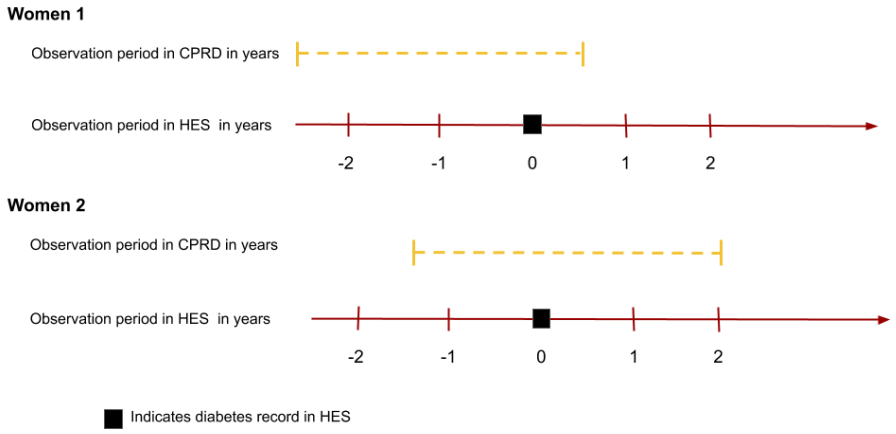


Figure 4.14: Schematic diagram of the analysis for agreement for diabetes recording between hospital admission and primary care

For women 1, the observation period in CPRD covers one year before the record of diabetes in HES but not one year after; so she was excluded from the analysis. For women 2, the observation period in CPRD covers one year both before and after the record of diabetes in HES, so she was included in the analysis.

Among those who had linked data for both CPRD and HES, 9,386 women had a first record of diabetes in HES. Among them, 3,655 women had an overlapping period of observation in CPRD. Among women with any HES diagnoses of diabetes, 95% had a record of diabetes available in CPRD. The proportions for the availability of CPRD record of diabetes differed between women with HES record of diabetes with mention of type1, type 2, and other types. The proportions for the availability of CPRD record of diabetes for women with HES record for type 2 diabetes, type 1 diabetes, and other diabetes types were 98%, 96%, and 85%, respectively (Table 4.9). The proportions of women with available CPRD record of diabetes were much higher ($\geq$98%) among women who had two or more hospital admissions mentioning diabetes.

Table 4.9: Comparing diabetes recording between hospital admission and primary care

	Women with HES record for diabetes		
	Number of admissions recording diabetes		
	All	**1**	**2+**
Women with any HES record for diabetes			
CPRD record of diabetes available	3481 (95.2%)	366 (76.3%)	3115 (98.1%)
CPRD record of diabetes not available	174 (4.8%)	114 (23.8%)	60 (1.9%)
Total (=100%)	3655	480	3175
Women with HES record for Type 2 diabetes			
CPRD record of diabetes available	3057 (95.7%)	330 (79.5%)	2727 (98.1%)
CPRD record of diabetes not available	138 (4.3%)	85 (20.5%)	53 (1.9%)
Total (=100%)	3195	415	2780
Women with HES record for type 1 diabetes			
CPRD record of diabetes available	253 (97.7%)	15 (88.2%)	238 (98.3%)
CPRD record of diabetes not available	6 (2.3%)	2 (11.8%)	4 (1.7%)
Total (=100%)	259	17	242
Women with HES records only for diabetes, not otherwise specified			
CPRD record of diabetes available	171 (85.1%)	21 (43.8%)	150 (98.0%)
CPRD record of diabetes not available	30 (14.9%)	27 (56.3%)	3 (2.0%)
Total (=100%)	201	48	153

4.7 Conclusion

In summary, this chapter shows that the Million Women Study can be used to prospectively investigate the associations of body size at birth, childhood, early adulthood, and later adulthood with the risk of developing adult-onset diabetes. The large size of this study can offer substantial insights into the separate effects of body sizes at different ages by subdividing women simultaneously according to their body size at each age. Careful considerations are required in the analysis framework of this thesis to deal with the potential issues in the Million Women Study, including but not limited to correlations among body size indices, their measurement errors, and under-detection of diabetes in hospital admissions data.

5

Analysis framework of prospective investigations

Summary

- To ensure prospective follow-up of the participants, the 3-year re-survey was considered as the baseline for all analyses in this thesis.
- 413,516 women without known diabetes or cancer at baseline reported information on their body size at birth and at ages 10, 20, and 60 years; and were followed-up to the earliest of the first hospital admission with diabetes, emigration, death, or end of follow-up (31 December 2017).
- Cox proportional hazards regression models with attained age as the underlying time variable were used to yield multivariable-adjusted relative risks (RRs) for adult-onset diabetes.
- RRs of adult-onset diabetes will be estimated for different combinations of body size throughout life

The aim of the thesis is to understand the associations of body size at different stages of early life on the risk of developing diabetes in adulthood. The research questions in this thesis and relevant conceptual frameworks for understanding those questions were outlined in chapter 3. Then, in chapter 4, the Million Women Study was discussed with special attention paid to the validity of body size indices at different ages and completeness of diabetes recording in the hospital admissions data. Now, this chapter introduces the analysis framework for this thesis - describing methodological approaches taken to analyse data from the Million Women Study for investigating the associations of body size indices at different ages and risk of adult-onset diabetes.

5.1 Study baseline

As described in the previous chapter, the Million Women Study is a prospective cohort of 1.3 million women aged 50 years and over in the UK, recruited between 1996 and 2001. Participants of this cohort are followed for health events (e.g. hospital admissions and cancers), emigration, and death through electronic linkage to national registers. About three years after recruitment, in median year 2001 (IQR 2000-03), participants reported their birthweight, body size at age 10 (in relative size), body size at age 20 (in UK clothes size), and body size in adulthood (in body mass index [BMI]) in the 3-year re-survey questionnaire. To ensure prospective follow-up of the participants, the 3-year re-survey was considered as the baseline for all analyses in this thesis. A copy of the 3-year re-survey questionnaire is in Appendix B.

A total of 866,530 women (64% of those at recruitment) returned the 3-year re-survey questionnaire. Possible differences between participants and non-participants are important for the interpretation of the findings and generalisation to the background population. Table 5.1 shows some key characteristics of women who responded to the recruitment and 3-year re-survey. Participation from the least deprived group (based on quintiles of Townsend index [222]) remained similar in

Table 5.1: Selected characteristics of women at recruitment and at 3-year re-survey in the Million Women Study

	Recruitment	**3-year re-survey**
Number of women	1,364,308	866,530
Year of birth, median (IQR)	1942 (1938-1946)	1942 (1938-1946)
Year of completing survey, median (IQR)	1998 (1998-1999)	2001 (2000-2003)
Current age, mean (SD)	56.7 (4.9)	60.0 (5.0)
Social deprivation (%)		
Least deprived	20.0	19.8
Most deprived	21.7	16.8
Highest educational attainment (%)		
None	42.7	37.3
Secondary or technical	41.8	45.3
Tertiary	12.9	15.4
Smoking status (%)		
Never	50.3	52.8
Former	26.4	33.4
Current	18.1	12.0
Drinkers (%)	78.7	67.2
Strenuous exercise, twice or more per week (%)	20.1	21.8
Current BMI (kg/m^2), mean (SD)	26.2 (4.7)	26.1 (4.6)

Percentages include participants with missing data on characteristics; percentage of missing data is less than 3% for all characteristics except for smoking status (5%).

both surveys whereas those who were in the most deprived group were less likely to participate in the 3-year re-survey. The participation rate in the 3-year re-survey also dropped for those who had no education (from 43% to 37%)and were current smoker (from 20% to 12%). The proportion of drinkers decreased from recruitment to 3-year re-survey, while BMI on average remained almost the same in both surveys.

5.2 Exclusions and handling missing information

Figure 5.1 shows the flowchart of participant exclusions for the analyses in this thesis. The Million Women Study undergoes several standard exclusions: those with unsigned study questionnaires, those with missing information on age or who were aged <48 years at recruitment, those with unmatched record at the Office for National Statistics, and those with missing information on the follow-up date. Participants were also excluded if they had any invasive cancer other than non-melanoma skin cancer prior to study baseline (i.e. 3-year re-survey). For prospective investigation on adult-onset diabetes, women were also excluded if they reported diabetes or taking insulin at or before study baseline. Women were excluded if they had any record of diabetes in the linked hospital admissions data prior to the baseline.

So that the same women were compared in all analyses, participants with missing information on birthweight, body size at age 10, body size at age 20, and adult BMI were also excluded from the main analysis. Women with missing data on covariates (usually less than 5%) were included as a separate category. The final study population for analysis was 413,516.

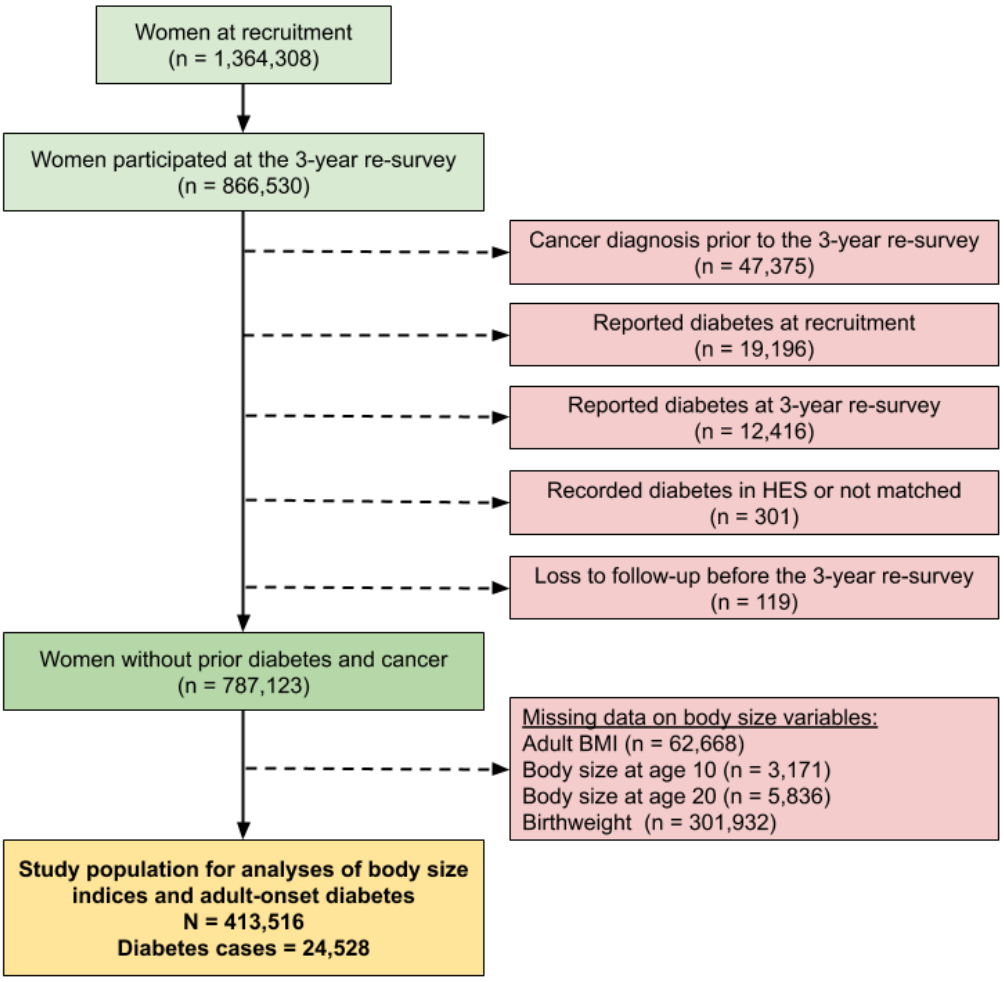


Figure 5.1: Flowchart of participant exclusions for the analysis

It should also be noted that 44% of those who completed the 3-year re-survey questionnaire did not report their birthweight, which might influence the representativeness of the study population to the underlying cohort. To help understand the possible issues with representation, I looked at some key characteristics among those who reported and who did not report birthweight (Table 5.2). Women who had valid information for birthweight were less likely to be socially deprived, uneducated, non-drinker, current smoker, and physically inactive in compared to women who did not report birthweight. Both groups had similar adult BMI on average (Table 5.2).

Table 5.2: Selected characteristics of women who reported and did not reported birthweight at 3-year re-survey in the Million Women Study

	Did not report birthweight	**Reported birthweight**
Number of women	379516	487016
Year of birth, median (IQR)	1942 (1937-1945)	1943 (1938-1946)
Year of completing survey, median (IQR)	2001 (2000-2003)	2001 (2001-2003)
Current age, mean (SD)	60.5 (5.0)	59.7 (4.9)
Social deprivation (%)		
Least deprived	20.7	18.2
Most deprived	22.4	15.7
Highest educational attainment (%)		
None	42.7	33.0
Secondary or technical	41.7	48.1
Tertiary	13.2	17.2
Current smoker (%)	12.5	11.5
Current alcohol drinkers (%)	65.5	68.5
Strenuous exercise, twice or more per week (%)	20.6	22.8
Current BMI (kg/m^2), mean (SD)	26.1 (4.5)	26.2 (4.7)

In this thesis, all analyses were conducted at complete-case-basis. Complete-case analysis is commonly used to deal with missing data in cohort studies, but it can give rise to biased estimates if participants with missing data and complete data differ in relation to exposure and outcome [223, 224]. Other methods based on advanced statistical procedures (e.g. multiple imputation) can produce biased estimates because of their failure in dealing with the uncertainty in the data [225–227]. Additional analysis was performed among participants with information on other body size variables by considering a separate category for missing values in birthweight.

5.3 Follow-up

As discussed in the previous chapter, record-linkage to the national registers on hospital admissions, cancers, emigration, and death ensures virtually complete and long-term follow-up for participants of the Million Women Study. After relevant exclusions at baseline, the remaining women were followed-up for diabetes in the linked hospital admissions data up until 31 December 2017. Person-years were calculated from the date they completed the 3-year re-survey questionnaire to the earliest of the following: first hospital admission with diabetes (the main outcome, described next), emigration, death, or end of follow-up.

5.4 Outcome: adult-onset diabetes

Among women with no prior record of diabetes at baseline, new cases of adult-onset diabetes were identified during follow-up using hospital admissions data. Adult-onset diabetes was defined as the first hospital admission during follow-up with International Classification of Diseases 10^{th} revision [ICD-10] codes for diabetes (E10-E14) [156]. Women who had hospital admissions with diabetes code E11 was specified as type 2. Similarly, those with ICD-10 code E10 were considered as type 1 diabetes, and those with codes E12-E14 were considered as "other" type.

As mentioned in chapter 4, for a small number of women (<2%), diabetes code in subsequent hospital admissions was different from the code recorded in the first admission. For these women, diabetes type was assigned according to the lowest code. For example, if a woman had a code for type 2 diabetes (E11) in her first hospital admission and had a code for type 1 diabetes (E10) in the subsequent hospitalisations, then she was considered to have type 1 diabetes. As mentioned earlier, diagnostic confusion can arise for adult-onset diabetes mainly because of possible overlap between common and rarer forms of diabetes (i.e. LADA) [21]. Patients with LADA present like type 2 diabetes at diagnosis but then require treatment with insulin within the first few years after diagnosis and further tests show that they have GADA antibodies. So, their diagnosis is later changed to type 1

diabetes in routine health data. However, the date of the earliest hospital admission (irrespective of diabetes code) was used as the event date for adult-onset diabetes.

5.5 Body size indices

Women included for analysis in this thesis reported their birthweight, body sizes ages 10 and 20 and adult BMI, which were compared against values recorded around respective ages (details in chapter 4). For the main analyses, women were grouped into five categories of reported birthweight (<2.5, 2.5-2.9, 3.0-3.4 [reference group], 3.5-3.9, and ≥4.0 kg), three categories of body size at age 10 years (thin, average, plump), three categories of size at age 20 years (UK clothes size ≤12 [small], 14 [medium], ≥16 [large]), and five categories of adult BMI (<22.5, 22.5-27.4, 27.5-32.4, 32.5-34.9, and ≥35.0 kg/m^2). As shown in chapter 4, corresponding recorded values for every category of each of these body size indices were assigned from previous validation studies [96, 181].

To examine the risk of adult-onset diabetes by two or more body size variables together, women were classified according to the joint categories of those body size variables. With this approach, I estimated the risk of adult-onset diabetes for every combination of body size indices with reference to a common baseline risk group.

5.6 Statistical analysis

As mentioned earlier, the main interest of this thesis is to explore whether there are separate effects of body size in early life on the risk of developing diabetes in adulthood, which is not mediated through their associations with adult body size, and also not confounded by other factors. Issues like measurement errors in body size indices, multiple mechanisms and effect modification by body size in later ages also needed to be considered in the statistical analysis.

I used Cox proportional hazards regression models with attained age as the underlying time variable to estimate hazard ratios (referred to as relative risks

[(RRs] hereafter) for adult-onset diabetes. For each body size index, category-specific RRs of adult-onset diabetes were estimated with reference to a baseline group. Where necessary, log-linear trends for quantitative body size indices (e.g. birthweight and adult BMI) were estimated by assigning mean recorded values from the validation studies to each self-reported category.

5.6.1 Checking for proportional hazards assumption

Tests and graphs based on the Schoenfeld Residuals were used for verifying whether the Cox models satisfy the assumption of proportional hazards. Additionally, log cumulative hazard function was plotted over follow-up time to check for any potential violation of proportionality of hazards, and when it occurs during follow-up, in case of any such violation [228].

5.6.2 Adjustment for covariates and modelling strategy

To ensure that comparisons were made among women who were as similar as possible, all analyses were stratified by year of birth (before 1939, 1939-1944, in or after 1945), single year of completing the baseline questionnaire (2000 or earlier, 2001, 2002, 2003, 2004, 2005), ten regions of residence in the UK, and social deprivation (according to the Townsend index [222] in quintiles), and were adjusted for smoking (never, past, current), alcohol intake (non-drinkers, $<$7, 7-14, and $\geq$15 drinks per week), strenuous exercise (none, once or more per week), adult height (less than 160, 160-170, 170 cm or more), age at menarche (11 or less, 12, 13, 14, 15 or more years), and maternal history of diabetes (yes, no). Where appropriate, analyses were also adjusted for body size at the other ages.

These factors were adjusted in the models based on the following: prior knowledge about their possible confounding effects on the associations of body size indices and adult-onset diabetes risk, or understandings on how they can be related to measurement errors and diabetes identification. Adjustment for all these variables was done separately, and then simultaneously to see whether their inclusion made any appreciable changes to the main estimates for exposures. Log likelihood ratio

test yielded changes in chi-square statistics between regression models, which was used to understand whether additional adjustment for each factor contributed significantly to the model fit.

As mentioned in chapter 2, age is one of the strongest risk factors for diabetes [229], therefore for all Cox regression models, attained age was used as the underlying time variable. Attained age can be expressed as functions of age at baseline and follow-up time. All analyses were also stratified by year of birth and year at completing the baseline survey, which effectively controlled for the age when women had entered into the study.

Socioeconomic variables like region of residence and social deprivation can influence lifetime body sizes as well as risk of diabetes (either by affecting the incidence of the disease [230, 231] or by affecting hospital admission rate [232–234]). All analyses in this thesis were stratified by ten regions of residence in the UK (Oxford, East Anglia, South West, Thames, West Midlands, Northern and Yorkshire, Trent, North West (Mersey), North West (Manchester/ Lancashire, and Scotland) and by quintiles of social deprivation based on the Townsend deprivation index (calculated by combining information from 1991 census data and postcode) [222].

All analyses were also adjusted for adult lifestyle variables: smoking, alcohol intake, and strenuous exercise. As discussed in chapter 2, adult lifestyle variables are related to the risk of diabetes in adulthood, but due to temporality issues, it is unlikely that they are associated with body size in early life. Therefore, they should not be regarded as typical confounders while investigating the associations of birthweight and body size in early life with the risk of adult-onset diabetes. However, there might be complex interactions between lifestyle factors, adult BMI and diabetes, which might affect the associations of body size in early life with the risk of adult-onset diabetes. Adjustment for these factors has another methodological advantage - any potential reporting errors in lifetime body size variables by these factors can be taken care of.

I also adjusted for age at menarche (11 or less, 12, 13, 14, 15 or more years) in adjusted models because previous studies showed that earlier age at menarche

was associated with both being small at birth [235] and being large at pre-pubertal ages [236]. Additionally, some studies suggested that those who undergo menarche at earlier ages tended to be obese both in early adulthood and in later adulthood, and be at higher risk of diabetes in adulthood [237]. Maternal history of diabetes is also an important factor to be considered as a confounder because it can be related to individual's size at birth and at early childhood [238] as well as the risk of type 2 diabetes in later life [239].

Where appropriate, analyses in this thesis were also adjusted for body size indices at other ages. However, adjusting for body size at other ages in regression models might have several methodological limitations. As explained in chapter 3, adjustment for body size in later life that is likely to be on the causal path from body size in early life and adult-onset diabetes (i.e. mediator) can effectively take account of the effect of the body size which is mediated through the body size in later life. This often pulls the effect estimates towards the null [240]. Another issue with standard adjustments for closely spaced measurements is the difficulty in the interpretation of results [10].

5.6.3 Regression dilution

As discussed in previous chapters, reported body size indices at different ages over lifetime are generally subjected to substantial errors. This can result in systematic underestimation of the associations between the body sizes reported at baseline and subsequent disease risk (also known as "regression dilution" bias) [241, 242]. In subsamples of the Million Women Study, body size indices were compared against measured or recorded values at around similar ages [96, 181]. In this thesis, to correct for regression dilution for a given body size variable, I used mean recorded value for each category to estimate log-linear trends or to plot category-specific RRs in figures.

5.6.4 Effect measure modification and subgroups analysis

Effect modification is applied to situations in which the magnitude of the effect of an exposure of interest differs depending on the level of a third variable. The

presence of effect modification by any factor can be explored by reporting the stratum-specific estimates separately.

To explore any possible effect modification, I estimated the RRs for diabetes by each of the body size variables in subgroups of women defined by year of birth (before 1939, 1939-1945, in or after 1946), tertiles of social deprivation index, adult height (less than 160 cm, 160-170 cm, 170 cm or more), age at menarche (less than 13 years, 13 years or more), smoking status (never, past, current), alcohol intake (non-drinkers, drinkers), strenuous exercise (none, once or more per week), and maternal history of diabetes (yes, no). These analyses were limited to women with valid information on these nine factors of interest, and therefore, the case numbers were slightly different from previous analyses. Statistical significance tests for effect modification were carried out via likelihood ratio test, by comparing the model with the product terms for the interaction between the variable of interest and body size index with the model without the interaction term. P-values for heterogeneity were calculated after adjustment for multiple testing [243]. Due to the large size of the Million Women Study, all interaction tests yielded statistically significant results but whether they reflected any meaningful effect modifications were not clear. Therefore, results of any interaction tests were not reported in the thesis.

5.6.5 Investigating effect modification by body size at other ages

Chapter 3 explains why it is ideal for examining the effects of body size at a particular age on the risk of adult-onset diabetes separately among individuals with different body size so that the details of any specific or consistent associations can be identified. The large size of the Million Women Study allowed me to analyse the risk of diabetes in different combinations of body size over the life course. Using this approach, I first looked at risks of developing diabetes in adulthood by birthweight and adult BMI, then by birthweight and body size at age 10 or 20 years with adjustment for adult BMI. Finally, I examined the risk of adult-onset diabetes, subdivided

simultaneously by 3 categories of body size at 4 different ages: birthweight, body size at age 10, body size at age 20, and adult BMI (i.e. into 3x3x3x3=81 subgroups).

5.6.6 Additional analyses

Like any other observational epidemiological studies, there are conditions in the Million Women Study (mainly discussed in chapter 4), which could have undermined the validity of the findings to a significant extent. I conducted a number of additional analyses to assess the robustness of the findings and to interpret the current findings in context. The rationale, results, and interpretation of these analyses are thoroughly discussed in chapter 9.

Box 5.6.6 shows the list of additional analyses which were conducted for each of the results chapters in this thesis. The aims of these additional analyses were: 1) to understand the effects of under-reporting and delayed recording of diabetes when using hospital data by using diabetes recorded in Clinical Practice Research Datalink (CPRD) in a sub-sample, 2) to investigate the associations of body size in early life with type 2 diabetes, 3) to minimise the effects of possible reverse causation, and 4) to exclude any potential confounding related to general ill-health or changes in behaviour.

List of additional analyses conducted for each result chapter

To understand the effects of under-reporting and delayed recording of diabetes when using hospital data

- Using CPRD-recorded diabetes in a sub-sample

To investigate the associations of body size in early life with type 2 diabetes

- Restricting to diabetes coded specifically as type-2 (ICD-10 code: E11)

To minimise the effects of possible reverse causation

- Excluding first five years of follow-up
- Excluding first ten years of follow-up

To reduce potential confounding by general ill-health or changes in behaviour

- Excluding women who rated their health as fair or poor
- Excluding women who reported a change in their diet due to illness

For additional analyses using linked data from CPRD, the start date of follow-up was considered as the latest date of completing the 3-year re-survey questionnaire, current registration date, and up to standard date. Person-years were calculated from the start date of follow-up to the earliest of the following: first CPRD record with diabetes, last collection date, transfer out date, or date of death. Details about various dates specific to the analysis of CPRD data are given in Table 4.8 in the previous chapter. New cases of diabetes were identified if women without prior known diabetes had at least one consultation with one or more diabetes-specific Read Codes during follow-up in the CPRD data. The corresponding date of consultation was considered as the event date. Because changes in methods of data collection and outcome definition can lead to a period effect in prospective studies, and CPRD data has gone through significant changes due to introduction of Quality of Framework (QoF) in 2004 [178], I additionally adjusted for calendar year for analyses using CPRD data.

5.6.7 Presentation of results

When more than two groups were compared, group-specific CIs (g-sCIs) were estimated for each group to allow direct comparison between any two groups, even if neither is the reference group. To calculate g-sCIs, the floating absolute risk method was is used, which does not alter the value of RRs but provides a 95% CI for each RR based on the amount of data in that category [244]. Since the reference groups are large, allowing for statistical variations, it makes little difference to the CIs for any of the RRs. Hence the terms "CIs" and "g-sCIs" can be used interchangeably, and although g-sCIs were calculated, they are referred to simply as CIs. When two categories are compared or where trends are estimated, conventional 95% CIs are calculated.

Where results are presented in figures, the RRs and CIs are shown as squares and lines with the area of each square inversely proportional to the variance of the logarithm of the corresponding RR, which shows the amount of statistical information involved with each estimate. In all figures, the RRs are shown in normal scale.

I used Stata version 15.1 [245] for all analyses while most figures were drawn in R [246].

5.7 Implications in this thesis

Using the methodological approaches described in this chapter, the Million Women Study data will be analysed to investigate the associations of risk of developing diabetes in adulthood with adult BMI (chapter 6), birthweight (chapter 7), and body size at ages 10 and 20 years (chapter 8).

6

Adult body mass index and risk of adult-onset diabetes

Summary

- The risk of adult-onset diabetes increased strongly with increasing adult BMI. When compared with women with healthy weight (BMI <25.0 kg/m^2), being overweight (BMI 25.0-29.9 kg/m^2) and being obese (BMI ≥30 kg/m^2) were associated with 2.6 times and 6.6 times higher risk of developing diabetes in adult life, respectively.
- The associations between adult adiposity and diabetes risk remained strong even after accounting for various socioeconomic and lifestyle factors.
- Considering adult BMI while examining the effects of body size in early life on diabetes risk in adulthood will help to interpret the findings and will add a life course perspective to the thesis.

6.1 Introduction

This thesis aims to investigate whether there are separate effects of body size in different stages of early life on the risk of developing diabetes in adulthood that is not mediated through their associations with adult body size, and also not confounded by other factors. Therefore, before investigating the effects of body size in early life on the risk of adult-onset diabetes, it is pertinent to understand the association between adult body mass index (BMI) and risk of diabetes in adulthood in the Million Women Study.

Epidemiological, and clinical studies over the past 30 years have demonstrated a strong link between adult body mass index (BMI) and increased risk of diabetes [68–76]. As mentioned previously in chapter 2, meta-analyses based on prospective cohort studies showed that when compared with individuals with healthy weight (BMI <25.0 kg/m^2), those with overweight (BMI 25.0-29.9 kg/m^2) or obesity (BMI ≥30 kg/m^2) had much greater risks of developing diabetes during adulthood [4, 5]. Because adult BMI is a mediator in the associations between body size in early life and risk of adult-onset diabetes, examining how adult BMI is related to the risk of adult-onset diabetes is important to interpret the associations between body size in early life and risk of adult-onset diabetes. Furthermore, understanding the influences of body size in early life on adult-onset diabetes risk in relation to a well-established adult exposure like BMI will add a life course perspective to this thesis.

Another important aspect of examining the association between adult BMI and risk of diabetes is to understand the possible factors that may modify the associations examined in this thesis. In chapter 2, I discussed different factors that are related to the risk of developing diabetes in adulthood and can also confound, possibly through some complex interactions with adult BMI, the associations between body size in early life and risk of adult-onset diabetes. Therefore, examining the effects of adjustment for various socioeconomic and lifestyle factors on the association between adult BMI and risk of diabetes in adulthood is needed.

As shown in chapter 4, adult BMI influences diabetes recording in hospital admissions data more than any other body size variables. This chapter also

examines the effects of adult BMI on the risk of diabetes recorded from hospital admissions data and primary care data, which will help to contextualise the findings in the following chapters.

6.2 Aim of this chapter

The aim of this chapter is to understand the association between adult BMI and risk of adult-onset diabetes in the Million Women Study. To achieve this aim, I will systematically perform the following tasks:

1. Estimate the category-specific risks of adult-onset diabetes for adult BMI
2. Estimate the log-linear trend (per 5 kg/m^2 increase) for the association between adult BMI and adult-onset diabetes
3. Examine the effects of adjustment for various socioeconomic and lifestyle factors on the association between adult BMI and adult-onset diabetes
4. Assess any potential effect modification by other factors through examining the association separately among subgroups of women defined by those factors

6.3 Methods

The details about the Million Women Study and the framework of analysis in this thesis have been described respectively in chapter 4 and chapter 5, and are summarised below.

6.3.1 Study design and participants

The Million Women Study is a population-based prospective cohort of 1.3 million women aged 50 years and over in the UK [18, 19]. In brief, between 1996 and 2001 the Million Women Study recruitment questionnaire was included together with a postal invitation letter for routine breast cancer screening, sent to women aged 50-64 years in 66 National Health Services (NHS) screening centres. Half of the invited women completed the recruitment questionnaire and gave written consent for re-contact and for follow-up through other health records. All participants have a unique National Health Service (NHS) number, and using NHS number and date of birth, they are linked electronically to routinely collected national databases on hospital admissions, cancers, and deaths provided by NHS Digital [154] in England and the Information Services Division [155] in Scotland. The study was approved by the Oxford and Anglia Multi-Centre Research Ethics Committee (MREC 97/5/001). The study questionnaires can be found on the study website (www.millionwomenstudy.org).

6.3.2 Adult BMI

BMI is usually referred to as reported or measured depending on how information on weight and height was collected, but BMI itself is usually not directly reported or measured. About three years after recruitment, when participants were aged about 60 years on average, they reported their current weight. BMI at around age 60 years was then derived from weight reported in the 3-year re-survey and height reported at recruitment. Estimated BMI based on reported weight and height at age 60 years was previously validated against measured BMI at around age 66 years in a subsample of women [181]. For analysis, women were grouped into five categories of reported adult BMI: <22.5, 22.5-27.4, 27.5-32.4, 32.5-34.9,

and ≥35.0 kg/m^2. Mean measured BMIs within each category of reported BMI were 21.9, 26.0, 30.6, 34.8, and 39.5 kg/m^2, respectively.

6.3.3 Outcome

New cases of adult-onset diabetes were identified if women with no prior record of diabetes had a hospital admission with International Classification of Diseases 10$^{\text{th}}$ revision (ICD-10) codes for diabetes (E10-E14). The date of the earliest hospital admission with diabetes record was taken as the event date for incident diabetes. To validate hospital diagnoses of diabetes, I compared information on diabetes recorded in primary care for 7% of the cohort linked to the Clinical Practice Research Datalink (CPRD) [169] and examined the time between the first record of diabetes in primary care and in hospital data (more details in chapter 4).

6.3.4 Statistical analysis

Analyses included women with information on all four indices of body size, who had no prior record of diabetes (either self-reported or hospital recorded) or of cancer (other than non-melanoma skin cancer) at study baseline. Person-years were calculated from the date the baseline questionnaire was completed to the earliest of the following: first hospital admission with diabetes, emigration, death, loss to follow-up, or end of follow-up (31 December 2017).

I used Cox proportional hazards regression models with attained age as the underlying time variable to estimate relative risks (RRs) of diabetes recorded in hospital admissions. Group-specific confidence intervals (g-sCIs) were estimated because more than two categories of adult BMI were compared [244], but they are referred to simply as CIs (details on why the terms "CIs" and "g-sCIs" can be used interchangeably is given in subsection 5.6.7 in chapter 5). RRs for adult-onset diabetes per 5 kg/m^2 increase in adult BMI were also estimated. To allow for measurement errors and regression dilution bias [241] for reported adult BMI, mean measured values within each self-reported category were used to plot the RRs in figures or to estimate log-liner trends.

To ensure that comparisons were made within women who were as similar as possible, all analyses were stratified by year of birth (before 1939, 1939-1945, after 1495), single year of completing the baseline questionnaire (2000 or earlier, 2001, 2002, 2003, 2004, 2005), ten regions of residence in the UK, and social deprivation (according to the Townsend index [222] in quintiles), and were adjusted for smoking (never, past, current), alcohol intake (non-drinkers, <7, 7-14, and ≥15 drinks per week), strenuous exercise (none, once or more per week), and maternal history of diabetes (yes, no). So that the same women were compared in all analyses, those with missing or non-applicable values (usually less than 5%) for each adjustment variable were treated as a separate category. All adjustment variables were reported at baseline, except social deprivation and strenuous exercise reported at the recruitment.

I explored the effects of adjustment for potential confounding factors on the RR per 5 kg/m^2 increase in BMI in an analysis restricted to 343,251 women with full information on all adjustment variables.

To assess any further potential for effect modification by other factors, I also assessed the association of adult BMI with adult-onset diabetes separately in subgroups defined by various individual characteristics, including year of birth, age at follow-up, socioeconomic status, height, age at menarche, parity, smoking, alcohol intake, strenuous exercise, and maternal history of diabetes. I also looked at the association by dividing women according to the age at diagnosis (<65, 65-74, and ≥75 years).

Sensitivity analyses were done to deal with the issues related to using hospital data for identifying adult-onset diabetes and minimising any reverse causation bias and potential confounding by general ill-health or changes in behaviour.

I used Stata version 15.1 for all analyses. All statistical tests were two-sided.

6.4 Results

The 413,516 women included in this analysis had a mean age at baseline of 59.5 (SD 4.8) years and a median year of birth of 1943 (IQR: 1938-1946). Table 6.1 shows characteristics of the study population, including measured BMI, by five categories of BMI, reported at study baseline. Those with higher BMI tended to have mothers who had diabetes, to be of lower socioeconomic status, to smoke less currently, to drink less alcohol and to be less physically active than those with lower BMI. Based on BMI measured in the validation sample, the mean BMI in the study population was 26.8 (standard error [SE] 0.1) kg/m^2.

During 6.1 million person-years of follow-up, a mean of 15 (SD 3) years per woman, 88% (n=364,526) had at least one hospital admission, and among them, 6.7% (n=24,528) had a hospital record of diabetes, first recorded at mean age 70 (SD 6) years. The large majority (93%, 22,784/24,528) had a hospital diagnosis of type 2 diabetes (ICD-10 code E11).

Table 6.1: Characteristics of the study population by body mass index (kg/m^2)

	Body mass index category (kg/m^2)					
	<22.5	**22.5-27.4**	**26.5-32.4**	**32.5-34.9**	**≥35.0**	**All women**
Number of women (%)	85087 (20.6)	204313 (49.4)	88284 (21.3)	17802 (4.3)	18030 (4.4)	413516 (100.0)
Mean measured BMI (SE)[a]	21.9 (0.1)	26.0 (0.1)	30.6 (0.2)	34.8 (0.4)	39.5 (0.7)	26.8 (0.1)
Median year of birth (IQR)	1943 (1939-1946)	1943 (1938-1946)	1943 (1938-1946)	1943 (1939-1946)	1944 (1939-1947)	1943 (1938-1946)
Mean age at baseline (SD)	59.3 (4.9)	59.6 (4.8)	59.8 (4.8)	59.4 (4.7)	59.1 (4.6)	59.5 (4.8)
Maternal history of diabetes (%)	5.5	6.9	8.5	9.6	10.7	7.3
Characteristics at baseline						
Mean adult height in cm (SD)	163.7 (6.5)	162.7 (6.3)	161.7 (6.6)	161.4 (6.6)	160.2 (7.7)	162.5 (6.6)
Lowest socioeconomic quintile (%)	15.5	16.9	21.3	24.2	29.0	18.4
Current smoker (%)	14.9	11.5	9.9	9.2	8.3	11.6
Nondrinker (%)	27.3	27.2	34.1	41.2	47.9	30.2
Mean alcohol drinks per week, in drinkers (SD)	7.7 (6.1)	7.3 (6.0)	6.9 (6.1)	6.4 (6.1)	6.3 (6.4)	7.2 (6.1)
Strenuous exercise at least once per week (%)	50.1	45.8	37.4	31.9	27.2	43.5
Follow-up						
Person-years (in thousands)	1280.3	3071.0	1287.0	251.2	243.9	6133.5
Mean years of follow-up per woman (SD)	15.0 (3.0)	15.0 (2.9)	14.6 (3.3)	14.1 (3.6)	13.5 (4.0)	14.8 (3.1)
Number of women with newly recorded diabetes (%)	1408 (1.7)	7887 (3.9)	8460 (9.6)	2796 (15.7)	3977 (22.1)	24528 (5.9)

Abbreviations: IQR=interquartile range, SD=standard deviation, SE=standard error)

All values except social deprivation and strenuous exercise were assessed when information about body mass index was recorded

[a] BMI was measured in a validation sample (more details in Chapter 4)

6.4.1 Association between adult BMI and risk of adult-onset diabetes

As expected, the risk of adult-onset diabetes increased sharply with adult BMI (Figure 6.1). The adjusted RR of diabetes for BMI ≥35 kg/m^2 vs. BMI <22.5 kg/m^2 was 13.5 (95% CI 13.1-13.9; p<0.0001). After adjustment for sociodemographic and lifestyle factors, compared with women with a normal weight (BMI <25.0 kg/m^2), the risk of developing diabetes in women who were overweight (BMI 25.0-29.9 kg/m^2) was almost 2.6 times higher (95% CI 2.54-2.65; p<0.0001), and almost 6.6 times higher for women who were obese (BMI ≥30.0 kg/m^2) (Table 6.2). The fully-adjusted RR for adult-onset diabetes per 5 kg/m^2 increase in BMI was 2.03 (95% CI 2.01-2.06).

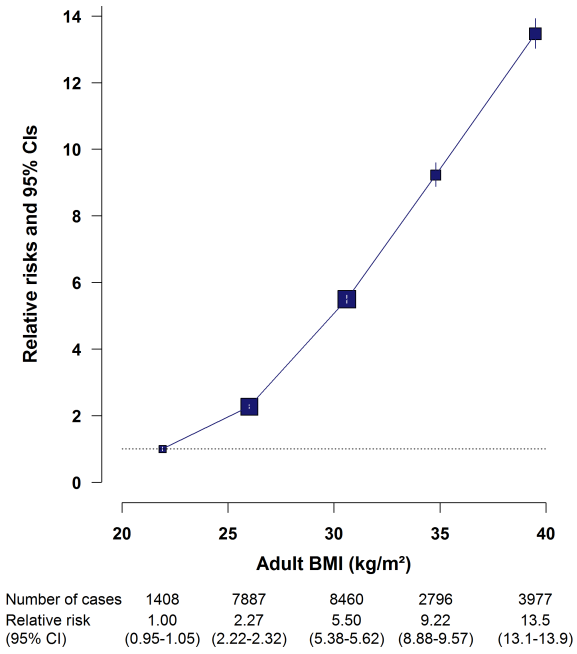


Figure 6.1: Relative risks (RRs) for diabetes by adult body mass index (BMI)

RRs with 95% confidence intervals (CIs) are adjusted for age, year of birth, year at baseline, region of residence, social deprivation, smoking, alcohol intake, strenuous exercise, and maternal history of diabetes. RRs are plotted against the mean measured BMI in each category at about six years after the baseline.

Table 6.2: Relative risks (RRs) for diabetes for women who were overweight and who were obese

BMI categories	No. of women/ cases	Adjusted RR (95% CI)	
		Minimally-adjusted model[a]	Fully-adjusted model[b]
Normal weight (<25.0 kg/m^2)	193690/4477	1.00 (0.97-1.03)	1.00 (0.97-1.03)
Overweight (25.0-29.9 kg/m^2)	153367/9651	2.68 (2.62-2.73)	2.59 (2.54-2.65)
Obese (≥30.0 kg/m^2)	66459/10400	7.11 (6.97-7.25)	6.55 (6.42-6.68)
Trend (per 5 kg/m^2)[c]	413516/24528	2.10 (2.08-2.13)	2.03 (2.01-2.06)

[a] Stratified by year of birth, year of baseline, region, and social deprivation

[b] Additionally adjusted for smoking, alcohol intake, strenuous exercise, and maternal history of diabetes

[c] Linear trend across BMI categories was estimated by assigning the mean measured BMI to each category and modelling this variable as continuous variable.

I explored the individual effects of adjustment for each of the socioeconomic and lifestyle variables on the RR for diabetes per 5 kg/m^2 increase in BMI in an analysis restricted to 343251 women with full information on all adjustment variables (Figure 6.2. Additional adjustments for each of these factors had little effect individually on the RR for diabetes. Compared with the RR (2.17, 95% CI 2.14-2.19) that was adjustment for age, year of birth, year of baseline, and region, adjustment for all factors simultaneously reduced the RR slightly to 2.04 (95% CI 2.02-2.07).

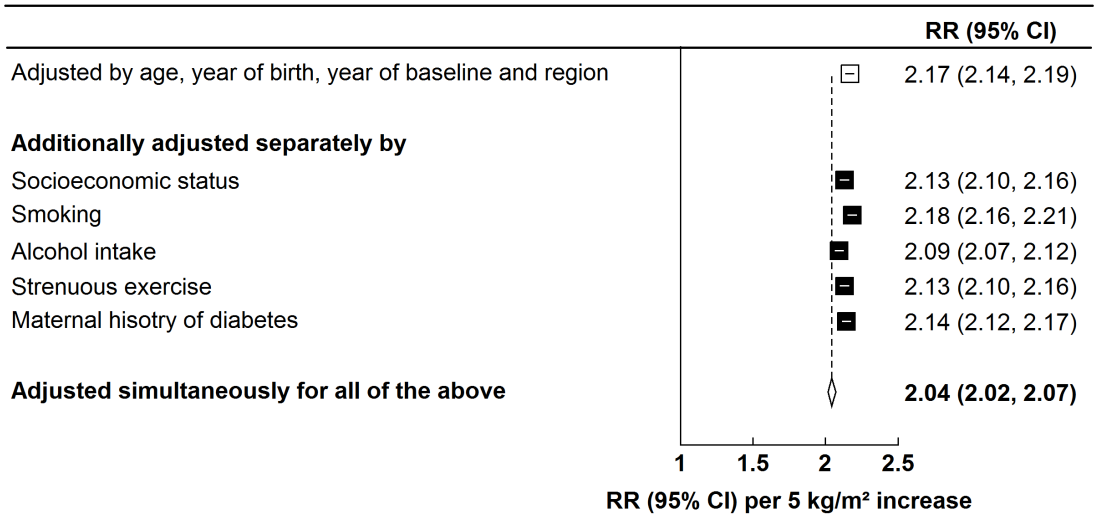


Figure 6.2: Relative risks (RRs) for diabetes per 5 kg/m^2 increase in BMI: effect of adjustment by various factors

Analysis restricted to 343,251 women (including 19528 cases of adult-onset diabetes) with information on all adjustment variables. Likelihood ratio tests showed that additional adjustments for each factor contributed significantly to the model fit (P<0.001).

6.4.2 Effect modification by other factors

Figure 6.3 shows the RR for diabetes per 5 kg/m^2 increase in BMI in subgroups of women defined by their year of birth, age at diagnosis, social deprivation, adult height, age at menarche, parity, smoking, alcohol intake, strenuous exercise, and maternal history of diabetes. RR for diabetes per 5 kg/m^2 increase in BMI was 2.30 (95% CI 2.23-2.37) in women aged less than 65 years at follow-up while it was 1.98 (95% CI 1.94-2.02) in women aged 75 years or more, indicating stronger associations between adult BMI and diabetes among those who developed diabetes in younger ages. The associations between BMI and adult-onset diabetes were slightly stronger in never smokers than in smokers, in drinkers than in non-drinkers, and in those who were physically active than those who were not. Evidence of heterogeneity in the effect of BMI on diabetes was also observed by maternal history of diabetes.



Figure 6.3: Relative risks (RRs) for diabetes per 5 kg/m^2 increase in BMI, by various characteristics at baseline

The dotted line represents the RR per 5 kg/m^2 increase in BMI for all women. RRs are adjusted for age, year of birth, year at baseline, region of residence, social deprivation, smoking, alcohol intake, strenuous exercise, and maternal history of diabetes.

6.4.3 Sensitivity analyses

Table 6.3 shows the adult BMI-associated RRs for diabetes in various sensitivity analyses: diabetes recorded in primary care (i.e. CPRD) data, restricting to diabetes coded especially as type-2, excluding initial years of follow-up, excluding women with poor or fair health, and excluding women who changed their diet due to illness.

In analysis with diabetes recorded in primary care, RRs for diabetes by adult BMI was much lower than RRs using hospital admissions data. For instance, the RR for women with BMI of $\geq$35.0 kg/m^2 vs BMI<22.5 kg/m^2 was 8.69 (95% CI: 7.84-9.63) in primary care data, while the RR for the same BMI group was 13.5 (95% CI: 13.1-13.9) in hospital admissions data. Such differences in the RRs for diabetes by adult BMI in primary care data and hospital data could be due to the fact that adult adiposity increases the rate of hospital admission and subsequently, the probability of diabetes recording.

When the analysis was restricted to type 2 diabetes only, the BMI category-specific RRs of diabetes were only slightly higher than those observed in the main analysis. The RRs did not change when the first 5 years of follow-up were excluded, but after excluding the first 10 years of follow-up, the RRs were slightly attenuated. Exclusions of those with poor or fair health and those who changed their diet due to illness did not make appreciable differences in the adult BMI-associated RRs of diabetes.

Table 6.3: Sensitivity analyses for the association between adult BMI and risk of diabetes in adulthood

BMI categories	No. of women/ cases	Adjusted RR (95% CI)
Restricting to diabetes coded specifically as type-2 (ICD-10 code: E11)		
<22.5 kg/m^2	84952/1273	1.00 (0.95-1.06)
22.5-27.4 kg/m^2	203681/7255	2.31 (2.26-2.37)
27.5-32.4 kg/m^2	87699/7875	5.67 (5.54-5.79)
32.5-34.9 kg/m^2	17616/2610	9.56 (9.19-9.93)
≥35.0 kg/m^2	17816/3763	14.16 (13.7-14.6)
Diabetes recorded in primary care (i.e. CPRD) data		
<22.5 kg/m^2	9364/270	1.00 (0.88-1.13)
22.5-27.4 kg/m^2	21945/1170	1.81 (1.71-1.92)
27.5-32.4 kg/m^2	9316/1106	3.96 (3.73-4.21)
32.5-34.9 kg/m^2	1875/311	5.68 (5.05-6.38)
≥35.0 kg/m^2	1791/425	8.69 (7.84-9.63)
Excluding the first 5 years of follow-up		
<22.5 kg/m^2	83055/1250	1.00 (0.95-1.06)
22.5-27.4 kg/m^2	200085/7090	2.31 (2.25-2.36)
27.5-32.4 kg/m^2	85621/7425	5.51 (5.38-5.63)
32.5-34.9 kg/m^2	17064/2421	9.16 (8.80-9.53)
≥35.0 kg/m^2	16961/3327	13.05 (12.6-13.5)
Excluding the first 10 years of follow-up		
<22.5 kg/m^2	79565/809	1.00 (0.93-1.07)
22.5-27.4 kg/m^2	191424/4546	2.31 (2.24-2.38)
27.5-32.4 kg/m^2	79810/4394	5.22 (5.07-5.37)
32.5-34.9 kg/m^2	15431/1373	8.47 (8.03-8.94)
≥35.0 kg/m^2	14756/1752	11.38 (10.8-11.9)
Excluding women who rated their health as poor or fair at baseline		
<22.5 kg/m^2	68382/927	1.00 (0.94-1.07)
22.5-27.4 kg/m^2	162393/5270	2.34 (2.28-2.40)
27.5-32.4 kg/m^2	62227/4984	5.62 (5.46-5.77)
32.5-34.9 kg/m^2	11037/1461	9.59 (9.11-10.10)
≥35.0 kg/m^2	9527/1745	13.67 (13.0-14.3)
Excluding women women who changed their diet due to illness		
<22.5 kg/m^2	75803/1167	1.00 (0.94-1.06)
22.5-27.4 kg/m^2	179850/6311	2.23 (2.18-2.29)
27.5-32.4 kg/m^2	73996/6521	5.47 (5.34-5.60)
32.5-34.9 kg/m^2	14402/2104	9.33 (8.94-9.74)
≥35.0 kg/m^2	14072/2898	13.69 (13.2-14.2)

Analyses were adjusted for age, year of birth, year of baseline, region, deprivation index, smoking,alcohol intake, strenuous exercise, and maternal diabetes

6.5 Discussion

6.5.1 Main findings

Analyses in this chapter showed that the risk of adult-onset diabetes increased strongly with increasing adult BMI. When compared with women with healthy weight, being overweight and being obese was associated with 2.6 times and 6.6 times higher risk of developing diabetes in adult life, respectively. The associations between adult adiposity and diabetes risk remained strong even after accounting for various socioeconomic and lifestyle factors. Results from a wide range of sensitivity analyses highlighted that the association between adult BMI and risk of adult-onset diabetes remained strong, although careful considerations would be required to interpret the results in the following chapters using diabetes cases from hospital admissions data.

6.5.2 Results in relation to other studies

Large-scale prospective studies around the world [68–76] have shown that the risk of adult-onset diabetes strongly increases as BMI in adult life increases. As expected, the risk of adult-onset diabetes increased sharply with adult BMI in the Million Women Study. The analyses in this chapter showed that the RR of adult-onset diabetes for women with obesity compared to those with normal weight was 6.55 (95% CI: 6.42-6.68), whereas the RR was 2.59 (2.54-2.65) for women with overweight. The adjusted RR of diabetes for BMI $\geq$35 kg/m^2 vs BMI $<$22.5 kg/m^2 was 13.5 (95% CI 13.1-13.9).

A recent pooled analysis of individual-level data for 120,813 adults from 16 cohort studies from the USA and Europe showed that when compared to individuals with healthy weight, for those with BMI of $\geq$ 35 kg/m^2, the risk of diabetes increased to 18.6 times (95% CI: 16.6-20.9) [4]. In that study, those with overweight and obesity had 3.3 times (95% CI: 3.0-3.6) and 9.4 times (95% CI: 8.6-10.3) higher risks of developing diabetes during adulthood, respectively. Results from that study are not directly comparable to the results reported in this chapter, particularly because of somewhat different reference groups. Nevertheless, in both

studies, it can be seen that higher BMI was associated with substantial increases in the risk of adult-onset diabetes.

It is understandable that differences in sample size, method of assessment of BMI and method of identifying diabetes cases can be related to variations in RRs between studies. In a meta-analysis by Abdullah *et al.*, when results from studies with at least 400 physician-diagnosed diabetes cases, measured BMI, and at least three main confounding variables (age, family history of type 2 diabetes, physical activity) were combined, the RR of diabetes was 7.28 (95% CI: 6.47-8.28) among those with obesity and 2.92 (95% CI: 2.57-3.32) among those with overweight. Although the magnitudes of RRs were not the same between studies, strong associations between high BMI in adulthood and risk of diabetes were consistently observed.

The strong positive link between BMI and diabetes is also well supported in Mendelian randomisation (MR) analyses - for each unit increase in BMI based on 96 genetic variants in a recent two-sample MR study, odds ratios (ORs) for type 2 diabetes were found to be between 1.19 and 1.38 [247]. Other MR studies [248–250] with different sets of instrument variables for BMI also observed strong associations between BMI and type 2 diabetes.

Data from some studies [70, 71] have suggested that central adiposity, often measured using waist circumference and waist-to-hip ratio, is strongly and independently (of BMI) associated with type 2 diabetes among adults. Although researchers from the Million Women Study collected self-reported data for these variables, I did not look into the roles of central adiposity on the risk of adult-onset diabetes because this is beyond the scope of this thesis and additional validations of those measures are required.

6.5.3 Effects of other factors on the association between adult BMI and diabetes risk

The findings of weaker associations of BMI for the risk of diabetes diagnosed in older ages than for the risk of diabetes diagnosed in younger ages could suggest that the aging process may outweigh the effects of BMI on the risk of diabetes [144]. However, it is important to recognise that the absolute risk (i.e. incidence rate) of diabetes was actually much higher among older women than younger women.

Results from this chapter also suggested that socioeconomic and lifestyle factors had little effect on the observed association between adult BMI and risk of diabetes. Unhealthy lifestyles such as smoking, alcohol intake, and physical inactivity are important confounders of the association between BMI and adult-onset diabetes because they are closely liked to BMI and can also influence the risk of diabetes. These factors can increase the underlying risk of diabetes among those with a healthy weight, which can partially explain why the RRs for diabetes per 5 kg/m^2 increase in BMI were slightly lower in smokers than in never smokers, and in those who were physically inactive than who were active. Also, non-drinkers in this cohort might have a different disease risk profile than drinkers, which could explain the observed heterogeneity in the diabetes risk associated with BMI. However, the biological mechanisms explaining these variations in BMI-related diabetes risk are unclear.

6.5.4 Strengths and limitations

Strengths and limitations concerning the prospective investigations in the Million Women Study are described in details in chapter 9. The strengths and limitations that are specific to investigating the effects of adult BMI are discussed below.

As discussed in chapter 3, major limitations of using reported adult BMI might include issues with measurement errors, choice of indicator, and multiple mechanisms. A previous report [181] from the Million Women Study compared reported and measured BMI at relevant ages in a subsample of participants. Although reported BMI was highly correlated to measured BMI, there was systematic under-reporting of BMI, which was substantial in the highest quintile of measured BMI. However,

the use of mean measured BMI for each reported category in the analysis and presentation of results is expected to minimise the effects of measurement errors and associated regression dilution bias [241]. There are also concerns about BMI as an indicator of adiposity, mainly because it cannot distinguish between lean and fat mass [124]. Choice of adiposity indicator may also indicate a different underlying mechanism for the association between adult adiposity and the risk of developing diabetes.

As shown in chapter 4, the probabilities of diabetes recording in hospital admission varied largely by levels of adult BMI. The median time lags for diabetes diagnosis at hospital admissions for women with BMI <25.0, 25.0-29.9, and ≥30.0 kg/m^2 were 5.0 (95% CI: 4.1-6.9), 4.1 (95% CI: 3.5-4.6), and 3.3 (95% CI: 2.4-3.9) years, respectively (more details in Figure 4.12). Therefore, using hospital admissions data for identifying new diabetes cases warrants cautious interpretations of the association between adult BMI and risk of adult-onset diabetes. Additional analysis using diabetes recorded in CPRD yielded lower RRs for diabetes by adult BMI than RRs using hospital admissions data.

Given the observational nature of this study, there are also possibilities of residual confounding (due to unmeasured or imperfect assessment of measured factors) and reverse causation. A wide range of sensitivity analyses was conducted to deal with the issues related to using hospital data and minimising any reverse causation bias and potential confounding by general ill-health or changes in behaviour, and similar results in those analyses indicate the robustness of the findings. The prospective follow-up for diabetes ensured that any errors in reported adult BMI were unlikely to be substantially different between those who developed and did not develop diabetes during follow-up. Record-linkage allowed a completed and unbiased follow-up for the whole cohort, with only 1.5% loss to follow-up and at least one hospital admission for 89% of women.

6.6 Conclusion

A strong association between adult BMI and risk of developing diabetes has been established in this chapter. This suggests that adult BMI should be taken into account to help to understand the separate effects of body size in early life on diabetes risk in adulthood. The findings of this chapter will also help with the interpretation of the results for body size in early life in the context of adiposity in later life, and thereby, will add a life course perspective to the thesis.

7

Birthweight and risk of adult-onset diabetes

Summary points

- Low birthweight was associated with an increased risk of adult-onset diabetes.
- The inverse associations between birthweight and risk of adult-onset diabetes were strong when looked separately among women who were of normal weight, and who were overweight and obese.
- The slight additional risk associated with extremely high birthweight was largely attributable to its association with adult adiposity.

7.1 Introduction

The systematic review in chapter 3 showed that with or without adjustment for adult body mass index (BMI), being born with a lower birthweight has been generally found to be associated with a higher risk of diabetes in adulthood. There are two major hypotheses concerning the underlying mechanisms for the association between low birthweight and increased risk of diabetes in adulthood: 1) Thrifty phenotype hypothesis, 2) Foetal insulin hypothesis. According to the thrifty phenotype hypothesis [139, 140], poor nutrition in foetal and early infant life reprogrammes individual's metabolism to become nutritionally thrifty by inducing permanent changes in the development and function of pancreatic β-cell and overall glucose-insulin metabolism. Such changes, combined with effects of ageing, adiposity and physical inactivity, predispose to the later development of diabetes in adulthood. On the other hand, the foetal insulin hypothesis [251] suggests that genetic variants affecting fetal pancreas development and β-cell dysfunction can lead to both reduced foetal growth and a higher risk of diabetes in adulthood. Recent studies [252–254] provided evidence for such shared genetic basis for low birthweight and type 2 diabetes.

According to the thrifty phenotype hypothesis, it has been hypothesised that increased risk of adult-onset diabetes associated with low birthweight may be due to increased life-long susceptibility to obesity-related diabetogenesis. As discussed in chapter 3, this can be explored by examining the association between low birthweight and the risk of developing diabetes in adulthood in subgroups of individuals with different levels of adult BMI. However, evidence on whether having a lower birthweight is associated with a higher risk of diabetes in adults who are not overweight or obese is limited. Two studies [132, 133] attempted to investigate this, but they did not have enough number of cases in each group to have reliable results.

Some recent studies [9, 129, 131] have suggested that there might be some additional risks of adult-onset diabetes for those with extremely high birthweight. It is possible that some pre-natal mechanisms leading to high birthweight can increase the risk of diabetes in adulthood. On the other hand, those with high birthweight

are more likely to have high BMI in adulthood [116, 117] and subsequently have diabetes. Even if high birthweight can be associated with some additional risks of diabetes in adulthood, it is unclear whether the additional risks of high birthweight is an effect separate from adult adiposity.

7.2 Aim of this chapter

The overall aim of this chapter is to provide a comprehensive understanding of how birthweight can influence the risk of developing diabetes among adults, mainly by investigating the issues pointed out in the systematic review. To achieve this aim, I will systematically perform the following tasks:

1. Investigate the association between birthweight and adult-onset diabetes by estimating category-specific risks and log-linear trend (per 1 kg increase)
2. Examine the association between extremely high birthweight and risk of adult-onset diabetes
3. Explore whether there is any confounding or effect modification by various socioeconomic and lifestyle factors on the association between birthweight and adult-onset diabetes
4. Investigate the association between birthweight and risk of adult-onset diabetes separately among women with different levels of adult adiposity
5. Explore whether there is an additive interaction between the effects of lower birthweight and higher adult BMI on the risk of adult-onset diabetes

7.3 Methods

The details about the Million Women Study and the framework of analysis in this thesis have been described in chapter 4 and chapter 5, and are summarised below.

7.3.1 Study design and participants

The Million Women Study is a population-based prospective cohort of 1.3 million women aged 50 years and over in the UK [18, 19]. In brief, between 1996 and 2001 the Million Women Study recruitment questionnaire was included together with a postal invitation letter for routine breast cancer screening, sent to women aged 50-64 years in 66 National Health Services (NHS) screening centres. Half of the invited women completed the recruitment questionnaire and gave written consent for re-contact and for follow-up through other health records. All participants have a unique National Health Service (NHS) number, and using NHS number and date of birth, they are linked electronically to routinely collected national databases on hospital admissions, cancers, and deaths provided by NHS Digital [154] in England and the Information Services Division [155] in Scotland. The study was approved by the Oxford and Anglia Multi-Centre Research Ethics Committee (MREC 97/5/001). The study questionnaires can be found on the study website (www.millionwomenstudy.org).

7.3.2 Birthweight

About three years after recruitment, in median year 2001 (IQR 2000-03), participants reported their birthweight. Previously, recalled birthweight was compared against birthweight recorded around the time of birth from the MRC National Survey of Health and Development (NSHD) in a subsample of the Million Women Study [96] (details can be found in subsection 4.3.2 in chapter 4). For most analyses, women were grouped into five categories of birthweight: <2.5, 2.5-2.9, 3.0-3.4 (reference group), 3.5-3.9, and ≥4.0 kg. The mean recorded birthweights in each of these categories were 2.63, 2.98, 3.37, 3.68, and 3.94 kg, respectively Figure 4.3. For analyses on high birthweight, I divided the category of ≥4.0 kg into two categories: 4.0-4.4 and ≥4.5 kg.

7.3.3 Outcome

New cases of adult-onset diabetes were identified if women with no prior record of diabetes had a hospital admission with International Classification of Diseases 10th revision (ICD-10) codes for diabetes (E10-E14). The date of the earliest hospital admission with diabetes record was taken as the event date for incident diabetes. To validate hospital diagnoses of diabetes, I compared information on diabetes recorded in primary care for 7% of the cohort linked to the Clinical Practice Research Datalink (CPRD) [169] and examined the time between first record of diabetes in primary care and in hospital data (more details in chapter 4).

7.3.4 Statistical analysis

Analyses included women with information on all four indices of body size, who had no prior record of diabetes (either self-reported or hospital recorded) or of cancer (other than non-melanoma skin cancer) at study baseline. Person-years were calculated from the date the baseline questionnaire was completed to the earliest of the following: first hospital admission with diabetes, emigration, death, loss to follow-up, or end of follow-up (31 December 2017).

I used Cox proportional hazards regression models with attained age as the underlying time variable to estimate relative risks (RRs) of diabetes recorded in hospital admissions. Group-specific confidence intervals (g-sCIs) were estimated because more than two categories of birthweight were compared [244], but they are referred to simply as CIs (details on why the terms "CIs" and "g-sCIs" can be used interchangeably is given in subsection 5.6.7 in chapter 5). RRs for adult-onset diabetes per 1 kg increase in birthweight were also estimated. To allow for measurement errors and regression dilution bias [241] for reported birthweight, mean measured values within each self-reported category were used to plot the RRs in figures or to estimate log-liner trends.

To ensure that comparisons were made within women who were as similar as possible, all analyses were stratified by year of birth (before 1939, 1939-1945, after 1495), single year of completing the baseline questionnaire (2000 or earlier,

2001, 2002, 2003, 2004, 2005), ten regions of residence in the UK, and social deprivation (according to the Townsend index [222] in quintiles), and were adjusted for adult height (<155, 155-159, 160-164, 165-169, 170-174, and ≥175 cm), age at menarche (11 or less, 12, 13, 14, 15 or more years), smoking (never, past, current), alcohol intake (non-drinkers, <7, 7-14, and ≥15 drinks per week), strenuous exercise (none, once or more per week), and maternal history of diabetes (yes, no). So that the same women were compared in all analyses, those with missing or non-applicable values (usually less than 5%) for each adjustment variable were treated as a separate category. All adjustment variables were reported at baseline, except social deprivation and strenuous exercise which were reported at the recruitment.

I explored the effects of adjustment for potential confounding factors on the RR per 1 kg increase in birthweight in an analysis restricted to 343,251 women with full information on all adjustment variables.

To further investigate the shape and possible non-linearity of the association between birthweight and adult-onset diabetes, I performed piece-wise linear spline regressions, with 3 knot points at 10^{th} 50^{th} and 90^{th} percentiles of birthweight (at 2.3, 3.2, and 4.1 kg). I also used 3 knot points at 2.5, 3.2, and 3.9 kg, because the mean (SD) birthweight in our study population is 3.2 (0.7) kg. A reference point of 3.2 kg was chosen for the graphical representation. I investigated the spline regression models with and without adjustment for adult BMI. Potential non-linear association of birthweight with risk of diabetes in adulthood was tested using the likelihood ratio test by comparing the model with the linear term only with the model with both linear and cubic spline terms.

I also cross-classified all women into 5 X 3 categories based on level of birth weight (<2.5, 2.5-2.9, 3.0-3.4, 3.5-3.9, and ≥4.0 kg) and adult BMI (<25.0, 25.0-29.9, and ≥30.0 kg/m^2). Women with birthweight of 3.0-3.4 kg and adult BMI <25 kg/m^2 were considered as the reference category. This approach of analysis estimated the RRs for adult diabetes for different combinations of birthweight and adult BMI with reference to common baseline risk.

The additive interaction between lower birthweight and higher adult BMI was assessed by estimating an index called the "relative excess risk due to interaction (RERI)" [255]. I then examined the decomposition of the joint effect - proportions attributable to lower birthweight alone, to higher adult BMI alone, and to their interactions.

To assess any further potential for effect modification by other factors, I also assessed the association of birthweight with adult-onset diabetes separately in subgroups defined by various individual characteristics, including year of birth, age at follow-up, socioeconomic status, height, age at menarche, parity, smoking, alcohol intake, strenuous exercise, and maternal history of diabetes. I also looked at the association by dividing women according to the age at diagnosis ($<$65, 65-74, and $\geq$75 years).

Sensitivity analyses were done to deal with the issues related to using hospital data for identifying adult-onset diabetes and minimising any reverse causation bias and potential confounding by general ill-health or changes in behaviour.

I used Stata version 15.1 for all analyses. All statistical tests were two-sided.

7.4 Results

A total of 413,516 women without known diabetes or cancer at baseline (mean age 60 [SD 5] years) reported their birthweight. Table 7.1 shows the characteristics of the study population by birthweight. Women with birthweight of <2.5 kg and ≥4.0 kg had slightly greater average BMIs at age 60 (26.2 and 26.5 kg/m^2, respectively) than women with birthweights of 2.5-4.0 kg. Women of lower birthweights were more likely to be relatively thin at age 10 or small at age 20, while women of higher birthweights were more likely to be plump or large at ages 10 and 20 years. Characteristics at study baseline were not strongly related to birthweight, although those of birthweights <2.5 kg and ≥4.0 kg were more likely to be of lowest socioeconomic status, current smokers, and somewhat less physically active than women with intermediate birthweights. Women with high birthweight were also more likely to be taller and to have mothers with diabetes (Table 7.1).

During 6.1 million person-years of follow-up, a mean of 15 (SD 3) years per woman, 88% (n=364,526) had at least one hospital admission, and among them, 6.7% (n=24,528) had a hospital record of diabetes, first recorded at mean age 70 (SD 6) years. The large majority (93%, 22,784/24,528) had a hospital diagnosis of type 2 diabetes (ICD-10 code E11).

Table 7.1: Characteristics of the study population by birthweight

	Birthweight categories (in kg)					
	<2.5	**2.5-2.9**	**3.0-3.4**	**3.5-3.9**	**≥4.0**	**All women**
Number of women (%)	54187 (13.1)	78583 (19.0)	145386 (35.2)	78359 (18.9)	57001 (13.8)	413516 (100.0)
Mean weight recorded at birth[a] (SD)	2.6 (0.4)	3.0 (0.2)	3.4 (0.2)	3.7 (0.2)	3.9 (0.5)	3.2 (0.7)
Median year of birth (IQR)	1942 (1938-1946)	1943 (1938-1946)	1943 (1939-1946)	1943 (1939-1946)	1942 (1938-1946)	1943 (1938-1946)
Mean age at baseline (SD)	59.7 (4.8)	59.5 (4.8)	59.4 (4.8)	59.3 (4.8)	60.0 (5.0)	59.5 (4.8)
Maternal history of diabetes (%)	6.9	6.5	6.8	7.6	9.3	7.3
Mean age at menarche (SD)	13.3 (1.6)	12.8 (1.5)	12.4 (1.6)	13.1 (1.6)	12.9 (1.5)	12.7 (1.6)
Characteristics at baseline						
Mean adult height in cm (SD)	160.3 (6.6)	161.2 (6.4)	162.7 (6.3)	163.6 (6.4)	164.5 (6.5)	162.5 (6.6)
Lowest socioeconomic quintile (%)	20.8	18.1	17.5	17.7	19.8	18.4
Current smoker (%)	12.3	11.1	11.2	11.5	12.8	11.6
Nondrinker (%)	33.5	30.6	28.8	29.1	31.5	30.2
Mean alcohol drinks per week, in drinkers (SD)	6.8 (5.8)	7.1 (6.0)	7.3 (6.1)	7.4 (6.2)	7.4 (6.3)	7.2 (6.1)
Strenuous exercise at least once per week (%)	41.1	43.3	44.8	44.2	41.8	43.5
Mean BMI in kg/m^2 (SD)	26.2 (4.6)	25.8 (4.4)	25.8 (4.4)	26.0 (4.5)	26.5 (4.7)	26.0 (4.5)
Follow-up						
Person-years (in thousands)	793.4	1165.8	2162.6	1167.9	843.9	6133.5
Mean years of follow-up per woman (SD)	14.6 (3.3)	14.8 (3.1)	14.9 (3.0)	14.9 (3.0)	14.8 (3.1)	14.8 (3.1)
Number of women with newly recorded diabetes (%)	4453 (8.2)	5229 (6.7)	8091 (5.6)	3834 (4.9)	2921 (5.1)	24528 (5.9)

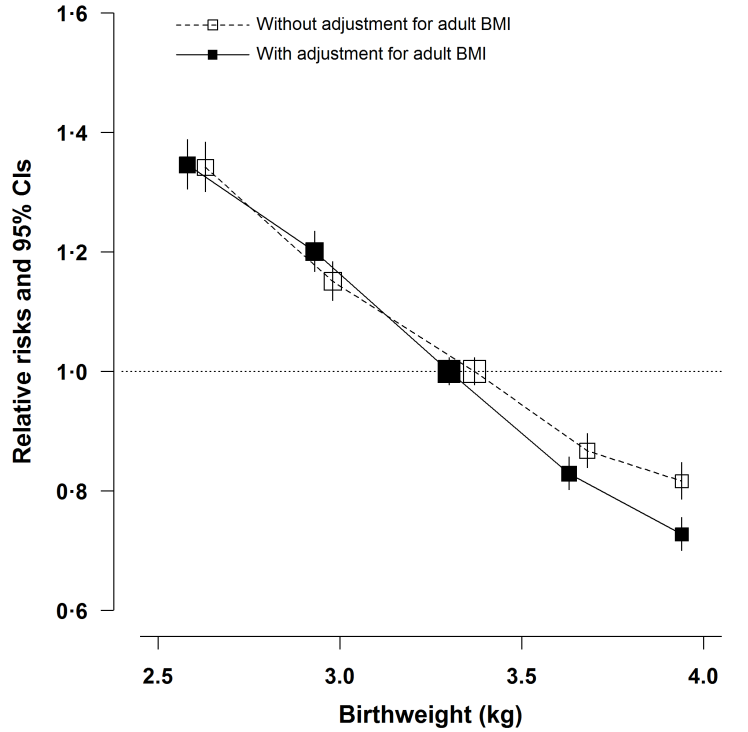
All values except social deprivation and strenuous exercise were assessed when information about birthweight was reported.

[a] Birthweight was recorded around the time of birth in a subsample (see methods)

Abbreviations: BMI = body mass index; SD = standard deviation

7.4.1 Association between birthweight and adult-onset diabetes

Women with lower birthweight were more likely to have diabetes in their adult life. When compared to those with birthweight 3.0-3.4 kg, RRs for diabetes for those with birthweight <2.5 kg and ≥4.0 kg were 1.35 (95% CI 1.30-1.38; p<0.0001) and 0.73 (95% CI 0.70-0.75; p<0.0001), respectively (Figure 7.1). Further adjustment for adult BMI did not change the association between birthweight and risk of adult-onset diabetes substantially. RR of diabetes for each kilogram increase in birthweight was 0.68 (95% CI: 0.66-0.70) without adjustment for adult BMI and 0.62 (95% CI: 0.60-0.64) with adjustment for adult BMI.



Number of cases	4453	5229	8091	3834	2921
RR (95% CI) without BMI adjustment	1.34(1.30-1.38)	1.15(1.12-1.18)	1.00(0.98-1.02)	0.87(0.84-0.89)	0.82(0.79-0.85)
RR (95% CI) with BMI adjustment	1.35(1.31-1.39)	1.20(1.17-1.23)	1.00(0.98-1.02)	0.83(0.80-0.86)	0.73(0.70-0.75)

Figure 7.1: Relative risks (RRs) for diabetes, by birthweight

RRs with 95% confidence intervals (95% CIs) are adjusted for age, year of birth, year at baseline, region of residence, social deprivation, smoking, alcohol intake, strenuous exercise, adult height, age at menarche and maternal history of diabetes. RRs are plotted against the mean measured birthweight around the time of birth in each category.

Figure 7.2 shows the effect of adjustment by potential confounding variables on the RR for adult-onset diabetes per 1 kg increase in birthweight in an analysis restricted to those with full information on all adjustment variables (n=338,512). Compared with the risk with adjustment for age, year of birth, year of baseline, and region (RR 0.66, 95% CI 0.63-0.68), additional adjustment for all factors simultaneously changed the RR slightly to 0.61 (0.59-0.64).

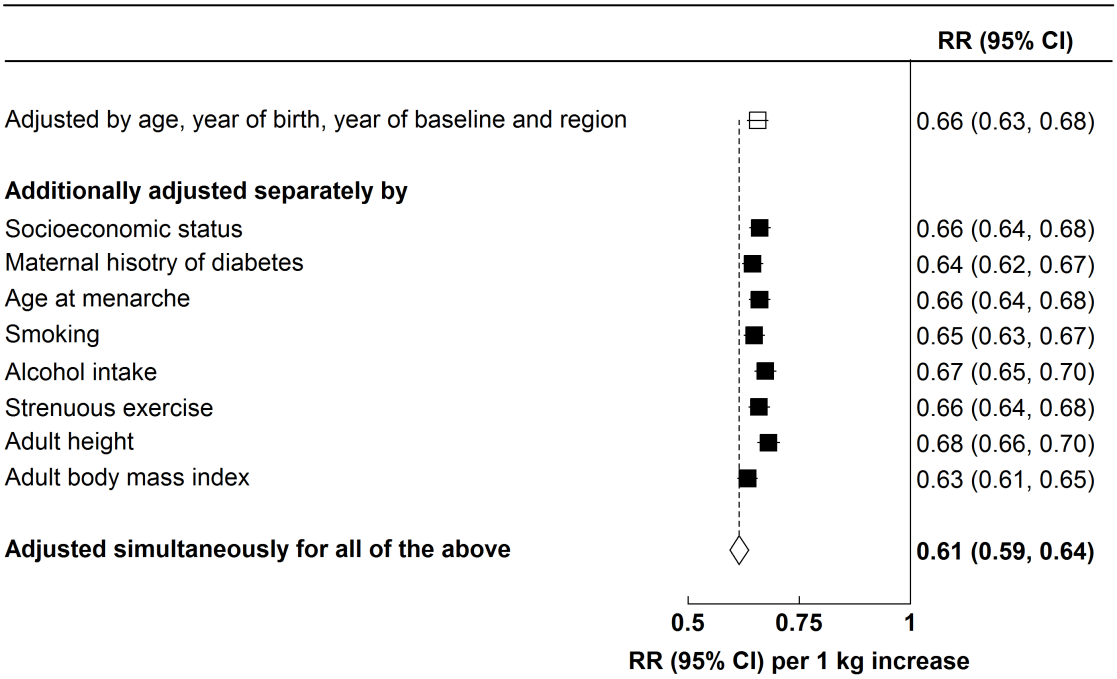
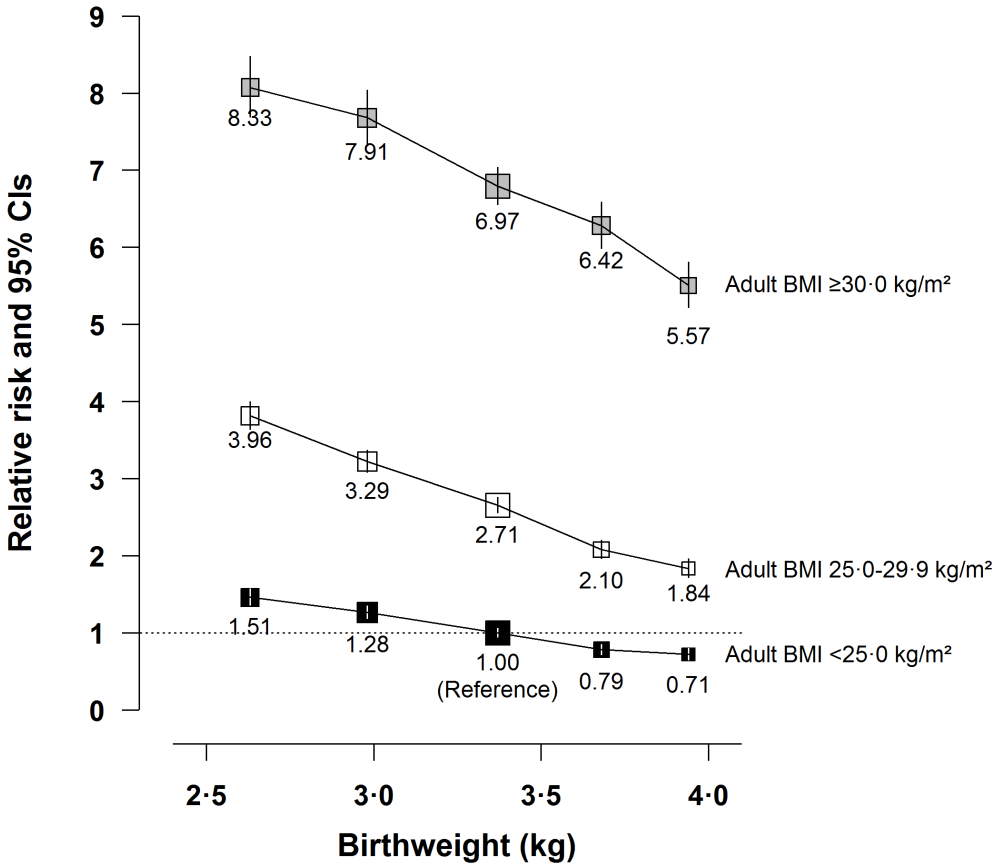


Figure 7.2: Relative risks (RRs) for diabetes per 1 kg increase in birthweight: effect of adjustment by various factors

Analysis restricted to 338,512 women (including 19,256 cases of adult-onset diabetes) with information on all adjustment variables. Likelihood ratio tests showed that additional adjustments for each factor contributed significantly to the model fit (P<0.001).

7.4.2 Association of birthweight and risk of adult-onset diabetes among women with different levels of adult adiposity

The inverse association between birthweight and risk of adult-onset diabetes was strongly evident separately among women who were of normal BMI, overweight and obese in adulthood ($p<0.0001$ for each group) (Figure 7.3. When compared to those with birthweight 3.0-3.4 kg and BMI <25 kg/m^2 (the common baseline risk group), the RRs for diabetes decreased almost linearly from 1.51 for those with birthweight <2.5 kg to 0.71 for those with birthweight ≥4.0 kg among those who had BMI <25 kg/m^2. Among women with BMI 25.0-29.9 kg/m^2, the RRs were 3.96 and 1.84 for birthweight <2.5 kg and ≥4.0 kg, respectively. As expected, the RRs of diabetes for all birthweight groups were much higher among women who were obese as adults, and the highest risk (RR 8.33) was observed among those who were born small and later became obese.



Number of cases					
Adult BMI <25·0 kg/m²	831	1088	1513	624	421
Adult BMI 25·0-29·9 kg/m²	1858	2136	3250	1392	1015
Adult BMI ≥30·0 kg/m²	1764	2005	3328	1818	1485

Figure 7.3: Relative risks (RRs) for diabetes by birthweight and adult body mass index (BMI)

Analyses are adjusted for age, year of birth, year at baseline, region of residence, social deprivation index, smoking, alcohol consumption, strenuous exercise, and maternal history of diabetes. Women with birthweight of 3.0-3.4 kg and adult BMI <25 kg/m^2 are the reference group. RRs are plotted against mean measured birthweight in each category.

7.4.3 Association between high birthweight and adult-onset diabetes risk

Some recently published reports [9, 129, 131] suggested a small increase in risk for diabetes for women with high birthweight (≥4.5 kg), making the overall association non-linear. Consistent with those reports, I also found women who were born with birthweight of ≥4.5 kg had slightly higher risk of diabetes in compared to women who had birthweight of 4.0-4.4 kg (RR 0.87 [0.82-0.92] and 0.79 [0.75-0.83], respectively when compared with a reference group of women with 3.0-3.4 kg birthweight). This was attenuated with further adjustment for adult body mass index (Table 7.2).

Table 7.2: Relative risks (RRs) for diabetes, by birth weight including a separate category for ≥4.5 kg

			Adjusted RR (95% CIs)[a]	
Birthweight (kg)	**No. of women/-cases**	**Mean reported birthweight**	**Without adjustment for adult BMI**	**With adjustment for adult BMI**
<2.5	54187/4453	2.0	1.34 (1.30-1.38)	1.34 (1.30-1.38)
2.5 - 2.9	78583/5229	2.8	1.15 (1.12-1.18)	1.20 (1.17-1.23)
3.0 - 3.4	145386/8091	3.2	1.00 (0.98-1.02)	1.00 (0.98-1.02)
3.5 - 3.9	78359/3834	3.7	0.87 (0.84-0.89)	0.84 (0.81-0.86)
4.0-4.4	34501/1629	4.2	0.79 (0.75-0.82)	0.74 (0.70-0.78)
≥4.5	22500/1292	4.7	0.86 (0.81-0.91)	0.74 (0.70-0.78)

[a] RRs are adjusted for age, year of birth, year at baseline, region of residence, deprivation index, age at menarche, adult height, smoking, alcohol intake, strenuous exercise, and maternal diabetes.

The shape of the association between birthweight and risk of adult-onset diabetes was also examined by piece-wise linear spline regressions. Figure 7.4 and Figure 7.5 show the association between birthweight and risk of diabetes in adulthood, using piece-wise linear spline regressions with cubic knots observed a significant non-linear trend (p<0.0001), which was not significant after further adjustment for adult BMI (p = 0.451). Without adjustment for adult BMI, significant non-linearity ($\chi^2 = 23.45$, p<0.0001) With adjustment for adult BMI, no significant non-linearity ($\chi^2 = 0.57$,)

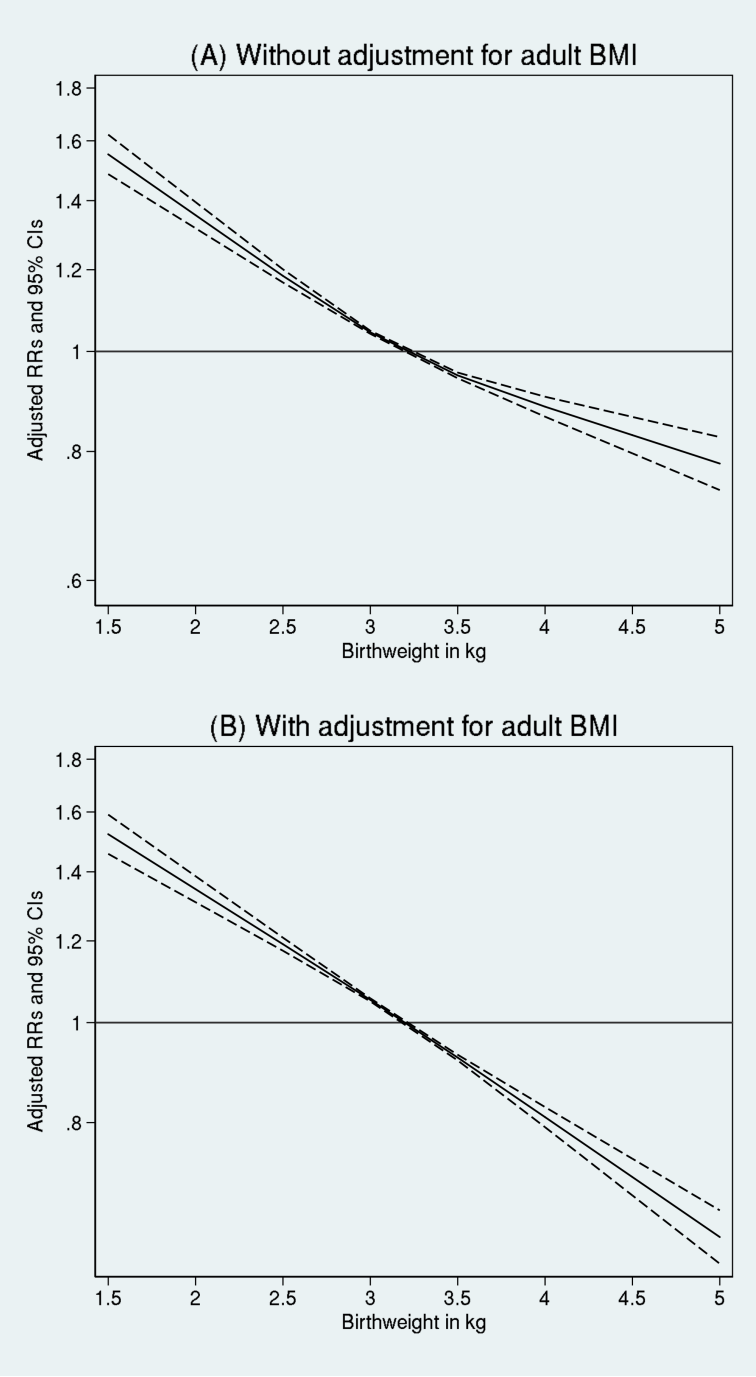


Figure 7.4: Association between birthweight and risk of diabetes in adulthood, using piece-wise linear spline regressions with cubic knots at 10^{th} 50^{th} and 90^{th} percentiles

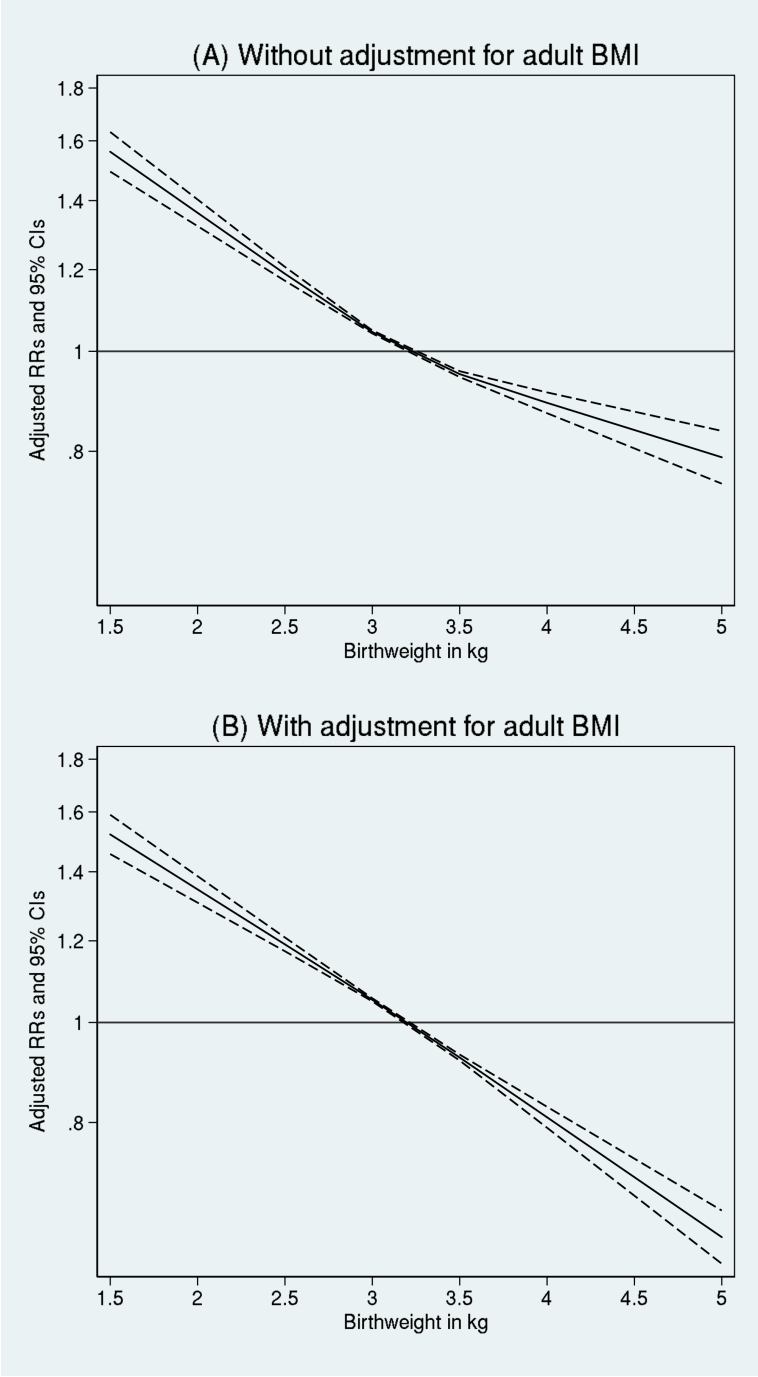


Figure 7.5: Association between birthweight and risk of diabetes in adulthood, using piece-wise linear spline regressions with cubic knots at 2.5, 3.2, and 3.9 kg

7.4.4 Effect modification by other factors

Figure 7.6 shows the RR for diabetes per 1 kg increase in birthweight, in subgroups of women defined by their year of birth, age at follow-up, socioeconomic deprivation, adult height, age at menarche, smoking, alcohol intake, strenuous exercise, and maternal history of diabetes. As opposed to what was seen for the effects of adult BMI on diabetes risk, the birthweight-associated RRs for diabetes did not vary by age at diagnosis. Evidence of heterogeneity in the effect of birthweight on risk of developing diabetes in adulthood was observed only for alcohol drinking. The associations between birthweight and adult-onset diabetes were not modified by other factors studied here (Figure 7.6).

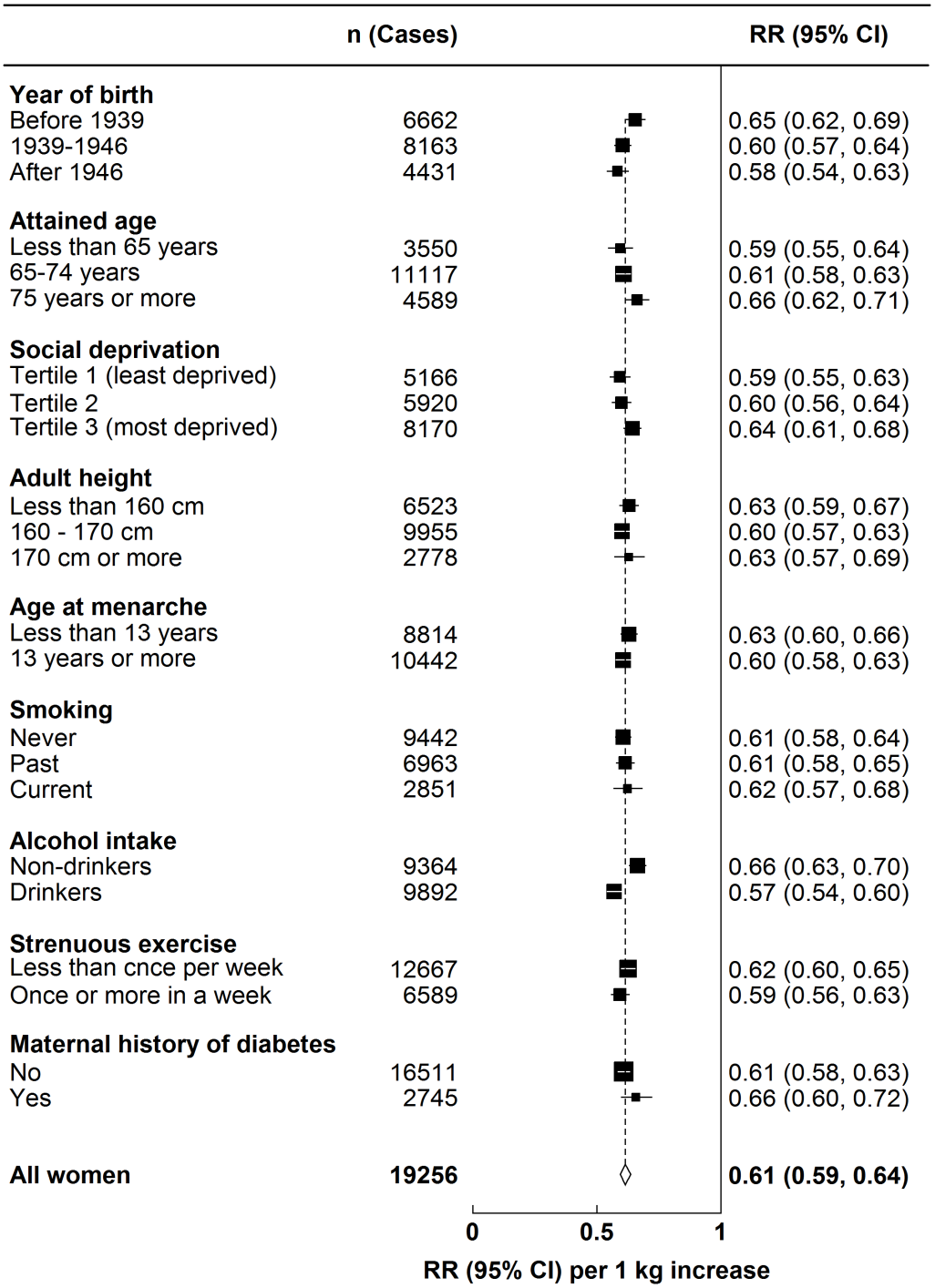


Figure 7.6: Relative risks (RRs) for diabetes per 1 kg increase in birthweight, by various characteristics at baseline

The dotted line represents the RR per 1 kg increase in birthweight for all women. RRs are adjusted for age, year of birth, year at baseline, region of residence, social deprivation, smoking, alcohol intake, strenuous exercise, body mass index, and maternal history of diabetes.

7.4.5 Additive interaction between the effects of lower birthweight and higher adult BMI

I found a significant additive interaction between low birthweight and high adult BMI (Table 7.3). The multivariable-adjusted RR of diabetes was 1.57 (1.52-1.62) per kg lowering of birthweight, 2.15 (2.12-2.17) per 5 kg/m^2 increase in adult BMI, and 3.38 (3.27-3.49) for their joint effect, with a relative excess risk due to interaction 0.66 (0.60-0.71). The attributable proportions of joint effect were 26.5% (95% CI: 25.5%-27.6%) to lower birthweight alone, 45.7% (43.7%-47.7%) to higher adult BMI alone, and 27.7% (26.7%-28.8%) to their interaction (Table 7.3).

Table 7.3: Attributable effects to additive interaction between birthweight and adult BMI on the risk of adult-onset diabetes

	Estimates (95% CI) [a]
Main effects	
Lower birthweight (per kg decrease)	1.57 (1.52-1.62)
Higher BMI (per 5 kg/m^2 increase)	2.15 (2.12-2.17)
Joint effect	3.38 (3.27-3.49)
Relative excess risk due to interaction	
Relative excess risk due to interaction	0.66 (0.60-0.71)
Attributable proportion, %	
Lower birthweight	26.5 (25.5-27.6)
Higher BMI	45.7 (43.7-47.7)
Additive interaction	27.7 (26.7-28.8)

[a] Based on multivariable adjusted RRs estimated from Cox proportional hazards models
Adjusted for age, year of birth, year at baseline, region of residence, deprivation index, age at menarche, adult height, smoking, alcohol intake, strenuous exercise, and maternal diabetes.

7.4.6 Sensitivity analyses

Table 7.4 shows the birthweight-associated RRs for diabetes in various sensitivity analyses: diabetes recorded in primary care (i.e. CPRD) data, restricting to diabetes coded specifically as type-2, excluding initial years of follow-up, excluding women with poor or fair health, and excluding women who changed their diet due to illness.

In all sensitivity analyses, when compared to women with birthweight 3.0-3.4 kg, the higher risks of adult-onset diabetes among those with birthweight <2.5 kg did not change, whereas women with birthweight ≥4.0 kg consistently had lower risks of diabetes. In the previous chapter, RRs for diabetes by adult BMI was much lower in analysis with diabetes recorded in primary care when compared with RRs using hospital admissions data. However, the corresponding RRs for birthweight categories were almost similar for analyses using hospital admissions data and primary care data.

Table 7.4: Sensitivity analyses for the association between birthweight and risk of diabetes in adulthood

		Multivariable-adjusted RR (95% CI)	
Birthweight categories	**No. of women/ cases**	**Adjusted for socioeconomic and lifestyle factors**	**Additionally adjusted for adult BMI**
Restricting to diabetes coded specifically as type-2 (ICD-10 code: E11)			
<2.5 kg	53875/4141	1.34 (1.30-1.38)	1.35 (1.30-1.39)
2.5-2.9 kg	78223/4869	1.15 (1.12-1.19)	1.20 (1.17-1.24)
3.0-3.4 kg	144822/7527	1.00 (0.98-1.02)	1.00 (0.98-1.02)
3.5-3.9 kg	78078/3553	0.86 (0.84-0.89)	0.82 (0.80-0.85)
≥4.0 kg	56766/2686	0.81 (0.78-0.84)	0.72 (0.69-0.74)
Diabetes recorded in primary care (i.e. CPRD) data			
<2.5 kg	3384/324	1.31 (1.17-1.47)	1.31 (1.17-1.47)
2.5-2.9 kg	4928/380	1.14 (1.03-1.27)	1.16 (1.04-1.29)
3.0-3.4 kg	9095/612	1.00 (0.92-1.08)	1.00 (0.92-1.09)
3.5-3.9 kg	4840/324	1.00 (0.89-1.12)	0.96 (0.86-1.07)
≥4.0 kg	3396/205	0.81 (0.71-0.94)	0.72 (0.62-0.83)
Excluding the first 5 years of follow-up			
<2.5 kg	52502/3878	1.34 (1.29-1.38)	1.34 (1.30-1.39)
2.5-2.9 kg	76498/4578	1.15 (1.11-1.18)	1.20 (1.16-1.23)
3.0-3.4 kg	141877/7134	1.00 (0.98-1.02)	1.00 (0.98-1.02)
3.5-3.9 kg	76463/3363	0.86 (0.83-0.89)	0.83 (0.80-0.86)
≥4.0 kg	55446/2560	0.82 (0.79-0.85)	0.73 (0.70-0.76)
Excluding the first 10 years of follow-up			
<2.5 kg	48984/2242	1.31 (1.25-1.36)	1.32 (1.26-1.37)
2.5-2.9 kg	72224/2778	1.17 (1.12-1.21)	1.22 (1.17-1.26)
3.0-3.4 kg	134558/4243	1.00 (0.97-1.03)	1.00 (0.97-1.03)
3.5-3.9 kg	72737/2080	0.90 (0.86-0.94)	0.86 (0.82-0.90)
≥4.0 kg	52483/1531	0.83 (0.79-0.88)	0.75 (0.71-0.79)
Excluding women who rated their health as poor or fair at baseline			
<2.5 kg	39086/2495	1.34 (1.29-1.40)	1.35 (1.29-1.40)
2.5-2.9 kg	59440/3132	1.17 (1.13-1.21)	1.22 (1.17-1.26)
3.0-3.4 kg	112317/4886	1.00 (0.97-1.03)	1.00 (0.97-1.03)
3.5-3.9 kg	60245/2247	0.84 (0.81-0.88)	0.81 (0.77-0.84)
≥4.0 kg	42478/1627	0.78 (0.74-0.82)	0.69 (0.66-0.73)
Excluding women who changed their diet due to illness			
<2.5 kg	45683/3404	1.37 (1.32-1.41)	1.37 (1.33-1.42)
2.5-2.9 kg	67866/4053	1.16 (1.12-1.19)	1.21 (1.17-1.25)
3.0-3.4 kg	126709/6305	1.00 (0.98-1.03)	1.00 (0.98-1.03)
3.5-3.9 kg	68379/2971	0.86 (0.83-0.89)	0.82 (0.79-0.85)
≥4.0 kg	49486/2268	0.81 (0.78-0.85)	0.73 (0.70-0.76)

Analyses were adjusted for age, year of birth, year of baseline, region,
deprivation index, smoking,alcohol intake, strenuous exercise, and maternal diabetes

7.5 Discussion

7.5.1 Main findings

This is the largest prospective study so far with more than 24,000 adult-onset diabetes cases, and it demonstrates strong, independent and persistently increased risks of adult onset-diabetes associated with decreasing birthweight, with no evidence for effect modification by other socioeconomic and lifestyle characteristics. The inverse associations between birthweight and risk of adult-onset diabetes were strong when looked separately among women who were of normal weight, and who were overweight and obese. It seems unlikely that I have overlooked any major effects of high birthweight on adult-onset diabetes risk, and in case there were some additional risks, it was largely attributable to adult adiposity. There was a significant additive interaction between low birthweight and high adult BMI and its interpretation needs further discussion.

7.5.2 Results in relation to other studies

The inverse relationship between birthweight and risk of adult-onset diabetes observed in this chapter is consistent with previously published results from large-scale prospective cohorts. As shown in the systematic review, increased risk of adult diabetes associated with lower birthweight has been reported consistently [129, 131–133]. The association between birthweight and risk of adult-onset diabetes remained strong after adjustment for adult BMI, which was also observed in studies which could account for adult adiposity in the analysis [131–133]. The risk estimates for this study and previous studies are slightly different, which might be due to differences in methods of birthweight assessment and identification of diabetes cases.

Higher risks of diabetes in adulthood among those with lower birthweight have been shown in Mendelian Randomisation (MR) studies, twin studies, and famine studies. A recent MR study based on 237,631 adults from the UK Biobank also reported that birthweight was causally related to type 2 diabetes in adulthood, and this relationship was not mediated by adult obesity [118]. A Swedish twin

study reported that lower birthweight was associated with increased risk of type 2 diabetes among dizygotic twin pairs but not among monozygotic twin pairs, suggesting that genetic factors contribute to the association between birthweight and risk of diabetes later in life [256]. The Dutch and Chinese famine studies showed that prenatal exposure to famine was linked to increased risk of glucose intolerance and diabetes among adults [257, 258].

Only two previous studies [132, 133] published results on birthweight-associated risks of diabetes separately for those who were of normal weight and who were overweight or obese as adults. Although statistical power was limited with few significant differences between the groups, it appeared that low birthweight was associated with an increased risk of diabetes separately in those who were of normal BMI and who were overweight or obese in adulthood. This is the first study to reliably show that lower birthweight is persistently associated with higher risks of diabetes in adulthood when examined separately among adults who are of normal weight, overweight, and obese. This finding suggests that there might be at least one pathway between low birthweight and adult-onset diabetes that is not be mediated by adult adiposity.

In 2007, Harder *et al.* [9] first suggested that high birthweight of $\geq$4.0 kg could increase the risk of adult-onset diabetes, making the overall shape of the association non-linear. But, there were not enough cases of diabetes in high birthweight category at the time of the review to show reliable estimates. In later years, a study [129] based on the Copenhagen School Health Records Register Study with more than 800 cases of diabetes in the high birthweight category showed that high birthweight could be associated with a higher risk of adult diabetes. However, due to the lack of data on adult adiposity, it was not clear whether the increased diabetes risk associated with high birthweight was independent. With approximately 1300 incident cases of diabetes in high birthweight group in this analysis, it was evident that the slight additional risk associated with higher birthweight was largely attributable to adult adiposity. Another recent study in the US reported similar findings for high birthweight-associated risk of diabetes with and without adjustment

for adult BMI [131]. High birthweight is often a consequence of maternal obesity and gestational diabetes, both of which can lead to a higher risk of diabetes in offspring [58, 259, 260]. In the Million Women Study, data on maternal obesity and gestational diabetes are not available.

Analysis in this chapter also shows that low birthweight and high adult BMI were jointly related to an increased risk of adult-onset diabetes which was more than the addition of the risk associated with each individual factor, indicating a significant interaction on an additive scale. It is important to remember that a statistically significant interaction between two exposures can be observed either on a multiplicative or additive scale if both of them have an effect on the outcome [255]. A significant additive interaction between low birthweight and unhealthy lifestyles including overweight was reported in a recent study based on pooled analyses of Nurses' Health Study, Nurses' Health Study II, and Health Professional Follow-up Study [131]. Additive interactions can be relevant to understanding the public health impacts - such interactions indicate that the risk of diabetes in adulthood would be larger among those with lower birthweights, but whether such an interaction is biological in nature cannot be ascertained from such analysis.

Increased risk of diabetes among adults who were born with a lower birthweight can be a consequence of growth restriction in the uterine environment, prematurity (i.e. born before the 37^{th} week of pregnancy), or both. I could not investigate whether the association between birthweight and adult diabetes is independent of premature birth because data on gestational age is not available in this cohort. However, previous studies using data from birth registers in Sweden and Finland reported that low birthweight could increase the risk of diabetes in later life irrespective of gestational age [130, 261, 262].

Some studies [129, 131] reported potential sex differences in the association between birthweight and adult diabetes risk, for example, stronger inverse associations in women than in men, and an increased risk of diabetes associated with high birthweights in women only. I could not examine the possible sex differences in this cohort of middle-aged women. It is unclear how sex could modify the association

between birthweight and diabetes risk in adulthood, but some researchers speculated that sex hormones in utero might have a role [129, 263].

7.5.3 Potential mechanisms

Although the evidence for increased risk of diabetes among those who were born smaller is consistent in epidemiological studies, the mechanisms underlying such associations are unclear. As mentioned earlier, it has been hypothesised that nutritional thriftiness due to poor foetal growth may increase the susceptibility of diabetes in adult life [140], or those genetic variants affecting foetal pancreas development may lead to both reduced foetal growth and a higher risk of diabetes in adulthood [251].

Animal studies [264, 265] show that animals with intrauterine growth restriction have decreased β-cell mass and insulin content in adulthood, even though they regained their body and pancreatic weights with regular postnatal feeding. Famine studies [257, 258] showed that exposure to prenatal under-nutrition could increase the risk of impaired glucose metabolism and subsequent type 2 diabetes in adult life, lending support to the thrifty phenotype hypothesis. Reduction in -cell mass due to lower birth weights has also been hypothesized to contribute to type 2 diabetes over the years [266]. On the other hand, type 2 diabetes genetic loci *CDKAL1* and *HHEX-IDE* are strongly associated with low birth weight, and both are involved in pancreatic β-cell dysfunction [252, 254]. Because the underlying defect is in the pancreas development and β-cell dysfunction, one could argue that the increased adiposity in later life would not mediate the association between birthweight and adult diabetes.

7.5.4 Strengths and limitations

Strengths and limitations concerning the prospective investigations in the Million Women Study are described in details in chapter 9. The strengths and limitations that are specific to investigating the effects of birthweight are discussed below.

One of the major limitations of the analyses presented in this chapter is the systematic and random measurement errors in reported birthweight. A previous report [96] from the Million Women Study compared reported birthweight and birthweight recorded around the time of birth in a subsample of participants. Although reported and recorded birthweights were highly correlated, there was systematic under-reporting of birthweight and the under-reporting was substantial in the highest quintile of recorded birthweight. However, the use of mean recorded birthweights for each reported category in the analysis and presentation of results is expected to minimise the effects of measurement errors and associated regression dilution bias [241]. The mean recorded birthweight for the high birthweight category (i.e. $\geq$4.5 kg) was not available in the validation study [96]; therefore findings for high birthweight should be interpreted cautiously.

Another important limitation is that new cases of diabetes were identified from hospital admissions record. There are several potential issues with such measures including but not limited to: missing diabetes cases who did not have hospital admission during follow-up, under- and delayed-reporting of diabetes in hospital admissions, differential recording of diabetes by birthweight and other factors. Diabetes recorded in hospital admissions data was compared against diabetes recorded in primary care data in chapter 4 to understand the degree and likely effects of these issues. Half of those with a diabetes record in primary care had a hospital admission mentioning diabetes in the next 4.2 (95% CI: 3.3-5.5) years if they had birthweight of less than 2.5 kg. The corresponding numbers for those with birthweight 2.5-3.4, and $\geq$3.5 kg were 4.1 (3.4-4.6), and 3.2 (2.4-3.9) years, respectively. These numbers indicate that even if differential diabetes recording by birthweight can bias the risk estimates, the direction of such bias should be in opposition to the direction of the association between birthweight and risk of adult-onset diabetes. Thereby, the observed association is highly unlikely to be explained by such outcome misclassification.

When compared to individuals with a healthy weight, those with overweight and obesity are more likely to under-report their birthweight while their diabetes

status is recorded earlier in hospital admissions. Therefore, adult BMI is an important confounder of the observed association. When examined separately among women with different levels of adult BMI, birthweight-associated risks of diabetes were strong and consistent. Given the observational nature of this study, there are also possibilities of residual confounding (due to unmeasured or imperfect assessment of measured factors) and reverse causation. A wide range of sensitivity analyses were conducted to deal with the issues related to using hospital data and to minimise any reverse causation bias and potential confounding by general ill-health or changes in behaviour, and similar results in those analyses indicate the robustness of the findings. However, information on important maternal and pre-natal factors (e.g. gestational age, gestational diabetes, weight gain during pregnancy) was not available in the Million Women Study; hence there might be some residual confounding by those factors.

The large size of the Million Women Study allowed me to subdivide the association of birthweight and adult-onset diabetes among women with different levels of adult BMI, and still get reliable estimates for all subgroups. The prospective follow-up for diabetes ensured that any errors in reported birthweight were unlikely to be substantially different between those who developed and did not develop diabetes during follow-up. Record-linkage allowed a completed and unbiased follow-up for the whole cohort, with only 1.5% loss to follow-up and at least one hospital admission for 89% of women.

7.6 Conclusion

In summary, this chapter showed that being small at birth increases the risk of diabetes in adult life, which is unlikely to be mediated by the association between birthweight and adult BMI. To help understand the paradoxical effects of body size at birth, and in adulthood on adult-onset diabetes risk, the effects of body size in childhood and early adulthood need to be assessed.

8

Body size at ages 10 and 20 years and adult-onset diabetes

Summary points

- Increased risks of adult-onset diabetes associated with larger body size at ages 10 and 20 years were largely due to their associations with adult adiposity.
- At every level of birthweight and adult BMI, there were additional risks of diabetes associated with being comparatively thin at age 10 years.
- After allowing for birthweight, size at age 10, and adult BMI, being small at age 20 years had a little additional effect on adult-onset diabetes risk.

8.1 Introduction

So far in this thesis, I have shown that larger body size (i.e. higher adult body mass index [BMI]) in adulthood increases the risk of developing diabetes, whereas paradoxically, smaller size at birth (i.e. lower birthweight) increases the risk of the disease. To help understand the apparently contradictory effects of body size at birth, and in adulthood on adult-onset diabetes risk, separate effects of body size in childhood and early adulthood need to be assessed.

The systematic reviews in chapter 3 highlight several key questions that remain unanswered in the current evidence. Because larger body size tracks from early life to later life [11], it is important to understand whether any increased risks of adult-onset diabetes for larger body size in childhood or in early adulthood are driven by the associations between body sizes in early life and adulthood. Additionally, it is possible that thinness in early life can lead to an increased risk of developing diabetes in adulthood, signifying different underlying mechanisms for the development of diabetes in adults.

Even if there are some effects of body sizes in childhood and early adulthood on adult-onset diabetes risk that are not mediated through adult body size, the question remains whether these observed effects are reflections of body size at an earlier stage. That is why it is important to account for birthweight while examining separate effects of childhood body size, or for both birthweight and childhood body size while examining the effects of body size in early adulthood. Examining simultaneous effects of birthweight and childhood body size also helps with understanding whether increased risks of diabetes among low birthweight babies can be explained by faster growth in postnatal period, as suggested by some studies [141–143].

For these reasons, a good way to investigate the risk of adult-onset diabetes would be by simultaneously subdividing by levels of body size at different ages - from birth to late adulthood. Nevertheless, large numbers are needed to do so. One previous study [133] tried to report the risks of adult-onset diabetes for childhood body size separately among women with different levels of birthweight and adult overweight, but it did not have enough number of cases in each group to have reliable results.

8.2 Aim of this chapter

This chapter aims to explain the associations of body size in childhood (at age 10 years) and in early adulthood (at age 20 years) with the risk of adult-onset diabetes, after accounting for body size at birth and late adulthood. To achieve this aim, I will systematically perform the following tasks:

1. Investigate the associations of body size at ages 10 and 20 years with the risk of adult-onset diabetes by estimating category-specific risks

2. Explore whether there is any confounding or effect modification by various socioeconomic and lifestyle factors on the associations of body size at ages 10 and 20 years with adult-onset diabetes

3. Investigate the associations of body size at ages 10 and 20 years with adult-onset diabetes separately among women with different levels of adult adiposity

4. Investigate the associations of body size at ages 10 and 20 years with adult-onset diabetes separately among women with different levels of birthweight

5. Investigate the risk of adult-onset diabetes by simultaneously subdividing by three categories of body size at four different ages: birthweight, body size at age 10, body size at age 20, and adult BMI (i.e. into 3x3x3x3=81 subgroups)

8.3 Methods

The details about the Million Women Study and the framework of analysis in this thesis have been described in chapter 4 and chapter 5, and are summarised below.

8.3.1 Study design and participants

The Million Women Study is a population-based prospective cohort of 1.3 million women aged 50 years and over in the UK [18, 19]. In brief, between 1996 and 2001 the Million Women Study recruitment questionnaire was included together with a postal invitation letter for routine breast cancer screening, sent to women aged 50-64 years in 66 National Health Services (NHS) screening centres. Half of the invited women completed the recruitment questionnaire and gave written consent for re-contact and for follow-up through other health records. All participants have a unique National Health Service (NHS) number, and using NHS number and date of birth, they are linked electronically to routinely collected national databases on hospital admissions, cancers, and deaths provided by NHS Digital [154] in England and the Information Services Division [155] in Scotland. The study was approved by the Oxford and Anglia Multi-Centre Research Ethics Committee (MREC 97/5/001). The study questionnaires can be found on the study website (www.millionwomenstudy.org).

8.3.2 Body size at ages 10 and 20 years

About three years after recruitment, in median year 2001 (IQR 2000-03), participants reported their body sizes at age 10 (thin, average, plump) and at age 20 years (UK clothes size $\leq$ 12 [small], 14 [medium], $\geq$16 [large]). Previously, these two ordinal body size indices were compared respectively against BMI measured at age 11 and reported at age 20 years from the MRC National Survey of Health and Development (NSHD) study in a subsample of the Million Women Study [96] (details can be found in subsection 4.3.3 in chapter 4). There were significant trends in recorded anthropometry across categories of reported body size at ages 10 and 20 years ($P < 0.001$) [96]. The mean measured BMIs for thin, average, and plump categories at

age 10 were 16.2, 17.5, and 20.4 kg/m^2, respectively. The average recorded BMIs for being small, average, and large at age 20 was 19.9, 21.3, and 24.3 kg/m^2, respectively.

8.3.3 Outcome

New cases of adult-onset diabetes were identified if women with no prior record of diabetes had a hospital admission with International Classification of Diseases 10$^{\text{th}}$ revision (ICD-10) codes for diabetes (E10-E14). The date of the earliest hospital admission with diabetes record was taken as the event date for incident diabetes. To validate hospital diagnoses of diabetes, I compared information on diabetes recorded in primary care for 7% of the cohort linked to the Clinical Practice Research Datalink (CPRD) [169] and examined the time between the first record of diabetes in primary care and in hospital data (more details in chapter 4).

8.3.4 Statistical analysis

Analyses included women with information on all four indices of body size, who had no prior record of diabetes (either self-reported or hospital recorded) or of cancer (other than non-melanoma skin cancer) at study baseline. Person-years were calculated from the date the baseline questionnaire was completed to the earliest of the following: first hospital admission with diabetes, emigration, death, loss to follow-up, or end of follow-up (31 December 2017).

I used Cox proportional hazards regression models with attained age as the underlying time variable to estimate relative risks (RRs) of diabetes recorded in hospital admissions. Group-specific confidence intervals (g-sCIs) were estimated because comparisons were made for body size indices with more than two categories [244], but they are referred to simply as CIs (details on why the terms "CIs" and "g-sCIs" can be used interchangeably is given in subsection 5.6.7 in chapter 5). To allow for measurement errors and regression dilution bias [241] for reported body size indices, mean measured values within each self-reported category were used to plot the RRs in figures.

To ensure that comparisons were made within women who were as similar as possible, all analyses were stratified by year of birth (before 1939, 1939-1945, after 1495), single year of completing the baseline questionnaire (2000 or earlier, 2001, 2002, 2003, 2004, 2005), ten regions of residence in the UK, and social deprivation (according to the Townsend index [222] in quintiles), and were adjusted for adult height (<155, 155-159, 160-164, 165-169, 170-174, and ≥175 cm), age at menarche (11 or less, 12, 13, 14, 15 or more years), smoking (never, past, current), alcohol intake (non-drinkers, <7, 7-14, and ≥15 drinks per week), strenuous exercise (none, once or more per week), and maternal history of diabetes (yes, no). Where appropriate, analyses were also adjusted for body size at the other ages. So that the same women were compared in all analyses, those with missing or non-applicable values (usually less than 5%) for each adjustment variable were treated as a separate category. All adjustment variables were reported at baseline, except social deprivation and strenuous exercise reported at the recruitment.

I explored the effects of adjustment for potential confounding factors on category-specific RRs for body size at ages 10 and 20 years in an analysis restricted to 343,251 women with full information on all adjustment variables.

I cross-classified all women into 3 X 3 categories based on the level of their body size at either age 10 or 20, and adult BMI (<25.0, 25.0-29.9, and ≥30.0 kg/m^2). Women with average or medium body size and adult BMI <25 kg/m^2 were considered as the reference category in respective analysis. I also cross-classified women into 3 X 5 categories based on level of their body size at either age 10 or 20, and birthweight (<2.5, 2.5-2.9, 3.0-3.4, 3.5-3.9, and ≥4.0 kg). Women with average or medium body size and birthweight 3.0-3.4 kg were considered as the reference category in respective analysis. This approach of analysis estimated the RRs for adult diabetes for different combinations of body sizes with reference to common baseline risk.

To investigate the risk of adult-onset diabetes across categories of body size indices at different ages, I also cross-classified women according to their birthweight (<3.0, 3.0-3.4, ≥3.5 kg), body size at age 10 (thin, average, plump), body size at age 20 (small, medium, large) and adult BMI (<25.0, 25.0-29.9, and ≥30.0

kg/m^2). For this analysis, women with birthweight of 3.0-3.4 kg, average body size at age 10, medium body size at age 20 and adult BMI <25.0 kg/m^2 were considered as the reference group.

To assess any further potential for effect modification by other factors, I also assessed the association of body size at ages 10 and 20 years with adult-onset diabetes separately in subgroups defined by various individual characteristics, including year of birth, age at follow-up, socioeconomic status, height, age at menarche, parity, smoking, alcohol intake, strenuous exercise, and maternal history of diabetes. I also looked at the association by dividing women according to the age at diagnosis (<65, 65-74, and ≥75 years).

Sensitivity analyses were done to deal with the issues related to using hospital data for identifying adult-onset diabetes and minimising any reverse causation bias and potential confounding by general ill-health or changes in behaviour.

I used Stata version 15.1 for all analyses. All statistical tests were two-sided.

8.4 Results

Among 413,516 women without previous diabetes who reported their body size at different ages, 129,753 (31%) reported they were thin, 219,481 (53%) reported of average body size, and 64,282 (16%) reported they were plump at age 10 years. For each of these body size levels, the mean measured BMI at a similar age was 16.2, 17.5 and 20.4 kg/m^2, respectively. When these women were asked about their body size at age 20 years, 128,066 (31%) reported they were small, 130,917 (32%) reported of medium body size, and 154,533 (37%) reported they were large. The mean measured BMIs at the same age for these body size levels were 19.9, 21.3, and 24.3 kg/m^2, respectively. As expected, higher body size at ages 10 and 20 years were associated both with higher mean birthweight and with higher mean BMI at baseline. Those with thin and plump body size were more likely to have mothers who had diabetes, whereas the likelihood of having a mother who had diabetes increased with increasing body size at age 20 years. The higher the body size at ages 10 and 20 years, the lower the age at menarche on average. Characteristics at study baseline were not strongly related to body size at ages 10 and 20 years, although those who were thin and who were plump were more likely to be of lowest socioeconomic status, non-drinkers, and physically inactive than those with average body size. Mean adult height increased with increasing body size at age 20, but not with body size at age 10. Those with large body size at age 20 were more likely to be of lowest socioeconomic status, current smokers, non-drinkers, and physically inactive compared to their counterparts (Table 8.1).

During 6.1 million person-years of follow-up, a mean of 15 (SD 3) years per woman, 88% (n=364,526) had at least one hospital admission, and among them, 6.7% (n=24,528) had a hospital record of diabetes, first recorded at mean age 70 (SD 6) years. The large majority (93%, 22,784/24,528) had a hospital diagnosis of type 2 diabetes (ICD-10 code E11).

Table 8.1: Characteristics of the study population by body size at ages 10 and 20 years

	Body size at age 10			Body size at age 20			
	Thin	**Average**	**Plump**	**Small**	**Average**	**Large**	**All women**
Mean measured BMI at similar age (SE) [a]	16.2 (0.3)	17.5 (0.2)	20.4 (0.4)	19.9 (0.2)	21.3 (0.2)	24.3 (0.4)	
Number of women (%)	129753 (31.4)	219481 (53.1)	64282 (15.5)	128066 (31.0)	130917 (31.7)	154533 (37.4)	413516 (100.0)
Median year of birth (IQR)	1943 (1938-1946)	1943 (1938-1946)	1944 (1939-1947)	1943 (1939-1946)	1943 (1938-1946)	1942 (1938-1946)	1938 (1946-1943)
Mean age at baseline (SD)	59.7 (4.8)	59.6 (4.8)	58.9 (4.7)	59.1 (4.6)	59.6 (4.8)	59.9 (4.9)	59.5 (4.8)
Maternal history of diabetes (%)	7.4	6.9	8.3	6.7	7.0	8.0	7.3
Mean birthweight in kg (SD)	3.1 (0.7)	3.3 (0.7)	3.4 (0.7)	3.1 (0.7)	3.2 (0.7)	3.3 (0.7)	3.2 (0.7)
Mean age at menarche (SD)	13.3 (1.6)	12.8 (1.5)	12.4 (1.6)	13.1 (1.6)	12.9 (1.5)	12.7 (1.6)	12.9 (1.6)
Characteristics at baseline							
Mean adult height in cm (SD)	162.8 (6.8)	162.4 (6.4)	162.3 (6.5)	160.1 (6.0)	162.5 (6.2)	164.6 (6.6)	162.5 (6.6)
Lowest socioeconomic quintile (%)	19.2	17.6	19.6	17.9	17.9	19.2	18.4
Current smoker (%)	10.1	11.7	14.2	12.3	11.6	11.1	11.6
Nondrinker (%)	30.0	29.8	31.8	27.9	28.2	33.8	30.2
Mean alcohol drinks per week, in drinkers (SD)	7.2 (6.0)	7.2 (6.0)	7.4 (6.4)	7.5 (6.1)	7.3 (6.0)	7.0 (6.1)	7.2 (6.1)
Strenuous exercise at least once per week (%)	42.6	44.5	41.9	44.8	44.3	41.7	43.5
Mean BMI in kg/m^2 (SD)	25.3 (4.2)	25.8 (4.2)	27.9 (5.5)	24.4 (3.6)	25.4 (3.8)	27.9 (5.1)	26.0 (4.5)
Follow-up							
Person-years (in thousands)	1916.3	3266.1	951.1	1904.9	1952.3	2276.2	6133.5
Mean years of follow-up per woman (SD)	14.8 (3.1)	14.9 (3.0)	14.8 (3.2)	14.9 (3.0)	14.9 (3.0)	14.7 (3.3)	14.8 (3.1)
Number of women with newly recorded diabetes (%)	8946 (6.9)	11337 (5.2)	4245 (6.6)	6208 (4.8)	6624 (5.1)	11696 (7.6)	24528 (5.9)

All values except social deprivation and strenuous exercise were assessed when information about birthweight was reported.

[a] For body size at age 10, BMI measured at age 11 years; and for body size at age 20, BMI measured at age 20 years in a subsample

Abbreviations: BMI = body mass index; SD = standard deviation; SE = standard error

8.4.1 Risks of adult-onset diabetes associated with body size at ages 10 and 20 years

After adjustment for socioeconomic and lifestyle factors, both thin vs average (RR 1.36, 95% CI: 1.33-1.38) and plump vs average (RR 1.21, 95% CI: 1.18-1.25) at age 10 were associated with increased risk of adult-onset diabetes (Figure 8.1A). The increased risk of diabetes associated with being thin vs average at age 10 remained strong when the analysis was additionally adjusted for adult BMI (RR 1.40, 95% CI: 1.36-1.43). However, RR of diabetes for women who were plump at age 10 decreased from 1.21 to 0.92 (95% CI: 0.89-0.95) after additional adjustment for adult BMI.

When compared to medium size at age 20, being large at age 20 was associated with an increased risk of adult-onset diabetes (RR 1.39, 95% CI: 1.37-1.42) (Figure 8.1B). However, the higher risk of diabetes among women who were large at age 20 was almost attenuated after adjustment for adult BMI (RR 1.05, 95% CI: 1.02-1.07). There was little separate effect for being small at age 20 on adult-onset diabetes risk (adult BMI adjusted-RR 1.07, 95% CI: 1.04-1.10).

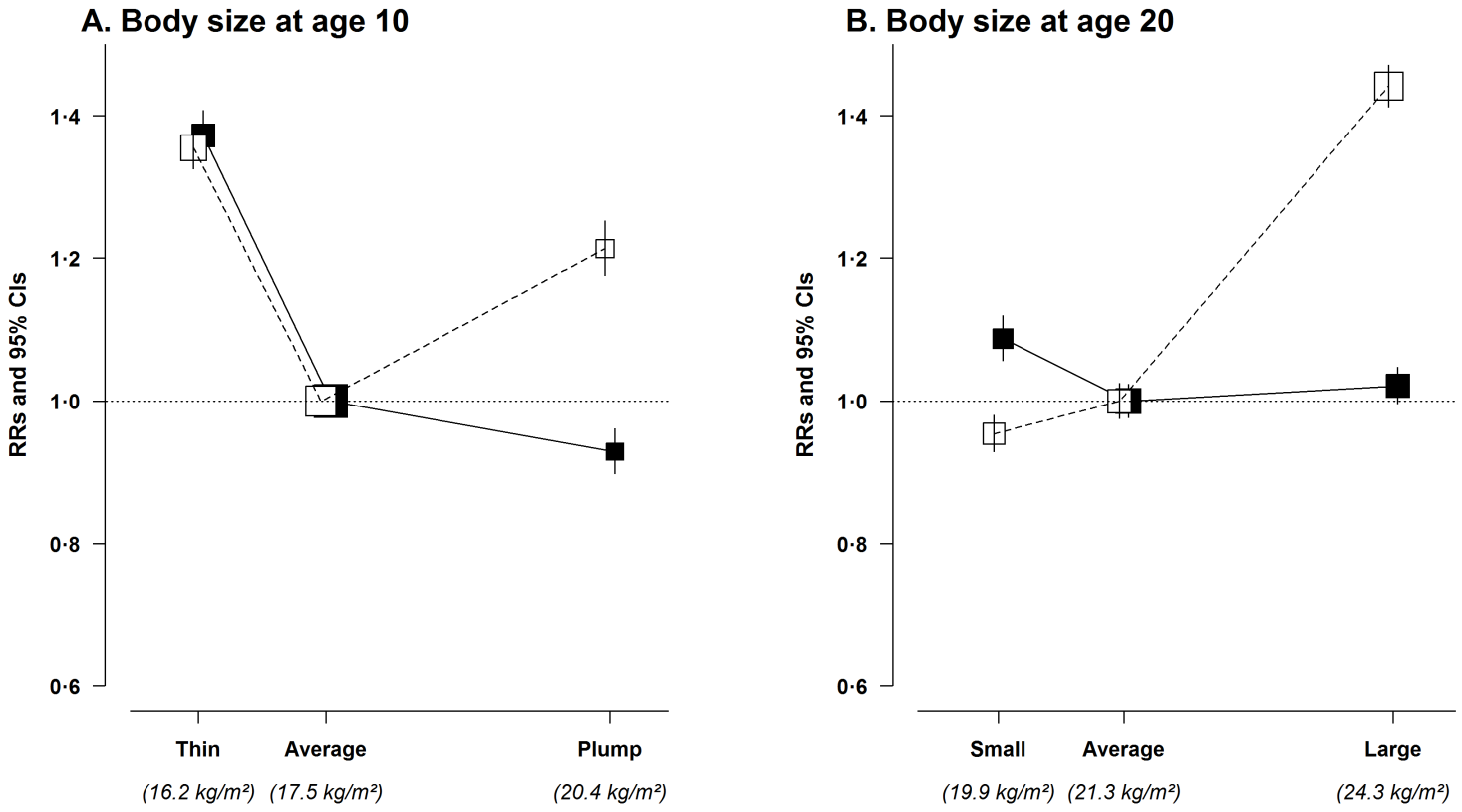


Figure 8.1: Relative risks (RRs) for diabetes, by body size at ages 10 and 20 years

RRs are adjusted for age, year of birth, year at baseline, region of residence, deprivation index, age at menarche, adult height, smoking, alcohol intake, strenuous exercise, and maternal history of diabetes. Category-specific RRs are plotted against the mean recorded BMIs at around relevant ages. To avoid overlap, RRs are moved apart slightly from the actual points. Black boxes and solid lines indicate RRs adjusted for adult BMI, whereas white boxes and dashed lines indicate RRs with no adjustment for adult BMI.

Figure 8.2 shows the effect of adjustment by potential confounding variables on the RR for adult-onset diabetes for thin vs average and plump vs average size at age 10 in an analysis restricted to 338,512 women who had valid information on all variables. Factors like socioeconomic status, maternal history of diabetes, age at menarche, adult height, smoking, alcohol intake, and strenuous physical activity had little effect individually on the RRs for thin vs average and for plump vs average. Compared with the RR for thin vs. average with adjustment for age, year of birth, year of baseline and region (RR 1.31, 95% CI: 1.27-1.35), additional adjustment for birthweight reduced the RR slightly to 1.26 (1.22-1.30), whereas additional adjustment for body size at age 20 and adult BMI increased the RR to 1.48 (1.43-1.53) and 1.46 (1.41-1.50), respectively. The RRs for plump vs. average decreased from 1.36 (1.31-1.42) to 1.18 (1.13-1.23) after adjustment for body size at age 20 and to 0.93 (0.89-0.97) after adjustment for adult BMI (Figure 8.2).

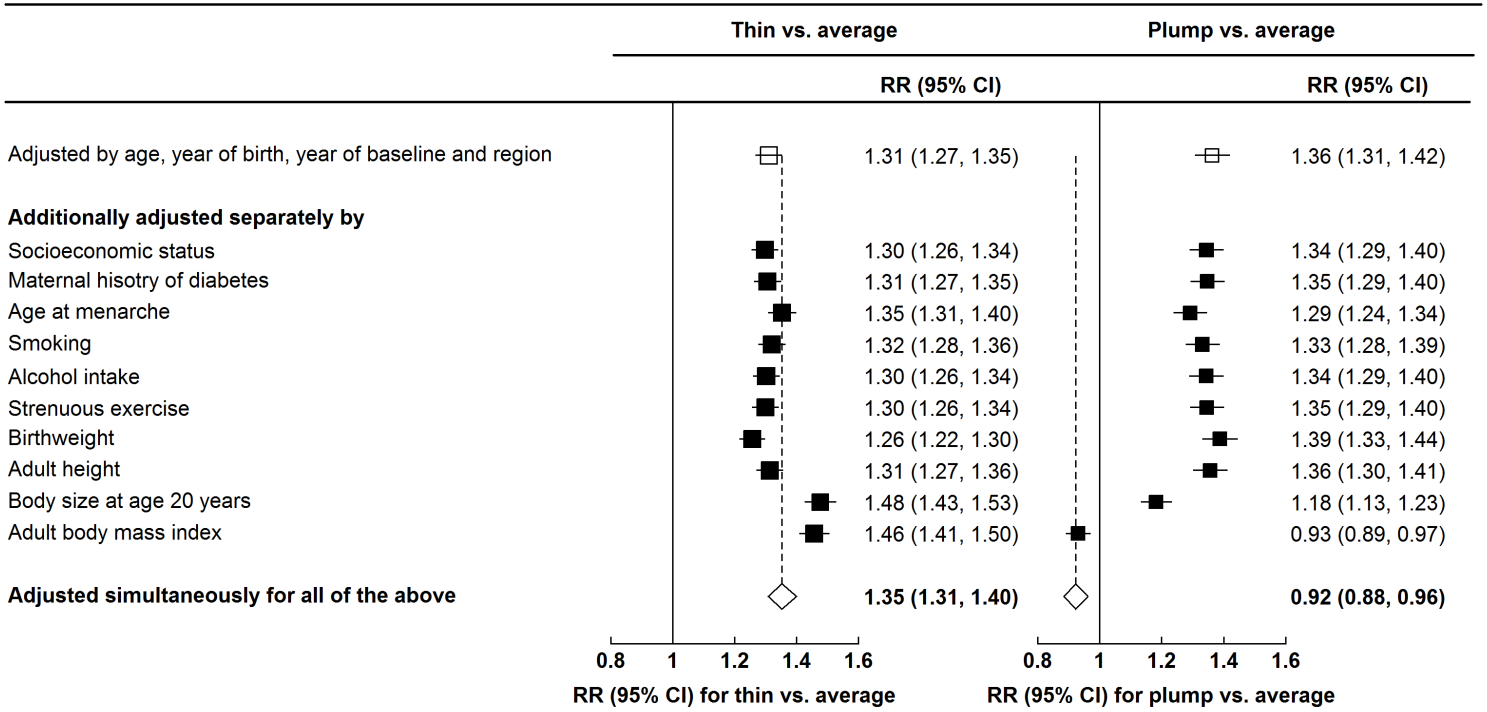


Figure 8.2: Relative risks (RRs) for diabetes by body size at age 10 years: effect of adjustment by various factors

Analysis restricted to 338,512 women (including 19,256 cases of adult-onset diabetes) with valid information on all adjustment variables. Likelihood ratio tests showed that additional adjustments for each factor contributed significantly to the model fit (P<0.0001).

Figure 8.3 shows the effect of adjustment by potential confounding variables on the RR for adult-onset diabetes for small vs medium and large vs medium size at age 20 in an analysis restricted to 338,512 women who had valid information on all variables. Factors like socioeconomic status, maternal history of diabetes, age at menarche, adult height, smoking, alcohol intake, and strenuous physical activity had little effect individually on the RRs for small vs medium and for large vs medium. When birthweight, adult height and body size at age 10 years were adjusted individually to the minimally-adjusted model, small body size was associated with reduced risk of adult-onset diabetes. When adult BMI was adjusted in the model, there was an increased risk of diabetes associated with being small. The effect of large body size at age 20 was completely attenuated after adjustment for adult BMI (Figure 8.3).

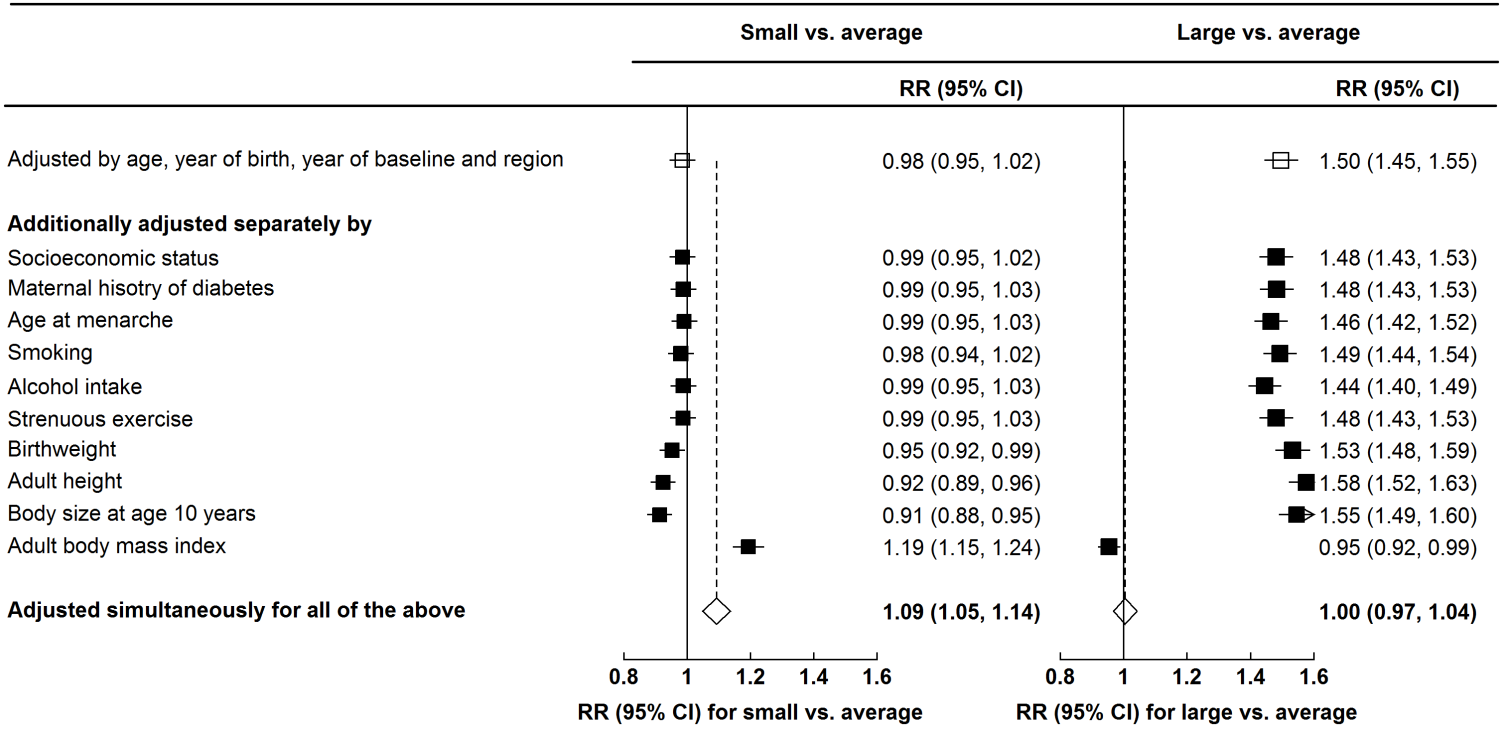


Figure 8.3: Relative risks (RRs) for diabetes by body size at age 20 years: effect of adjustment by various factors

Analysis restricted to 338,512 women (including 19,256 cases of adult-onset diabetes) with information on all adjustment variables. Likelihood ratio tests showed that additional adjustments for each factor contributed significantly to the model fit (P<0.0001).

8.4.2 Risks of adult-onset diabetes associated with body size at ages 10 and 20 years among women with different levels of adult adiposity

I investigated the risk of adult-onset diabetes by body size at age 10 within subgroups of women by adult BMI (Figure 8.4A). There was strong evidence of additional risks of diabetes associated with being thin at age 10 among women who were of normal BMI, overweight and obese ($p<0.0001$ for each group). Women who were plump vs average at age 10 had slightly lower risks of adult-onset diabetes at each level of adult BMI.

For body size at age 20, there were slight additional increases in the risk of adult-onset diabetes for small body size when compared to medium body size among women who were of normal BMI, overweight and obese (Figure 8.4B). However, being large at age 20 had no additional effect on adult-onset diabetes risk at each level of adult BMI.

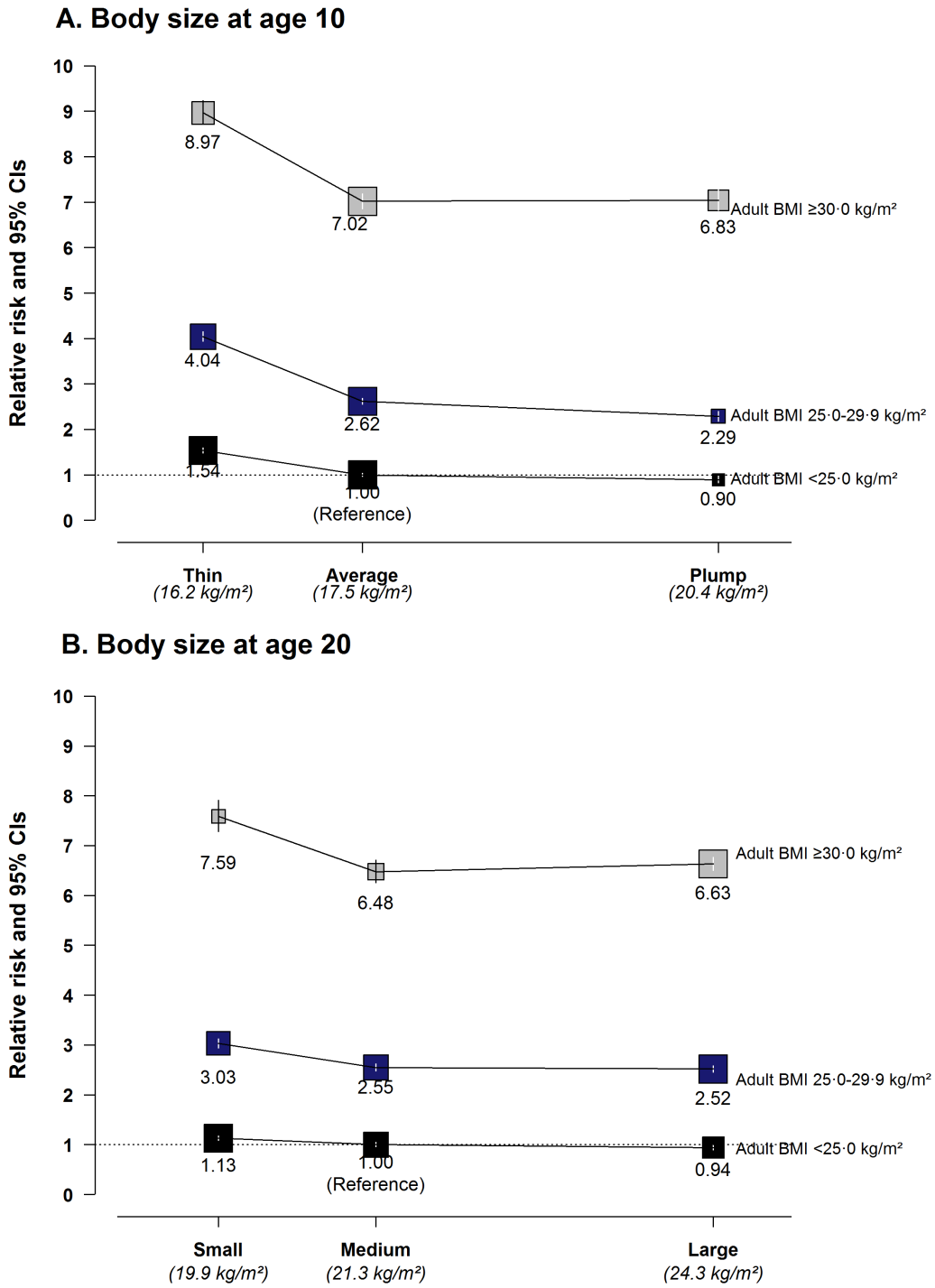


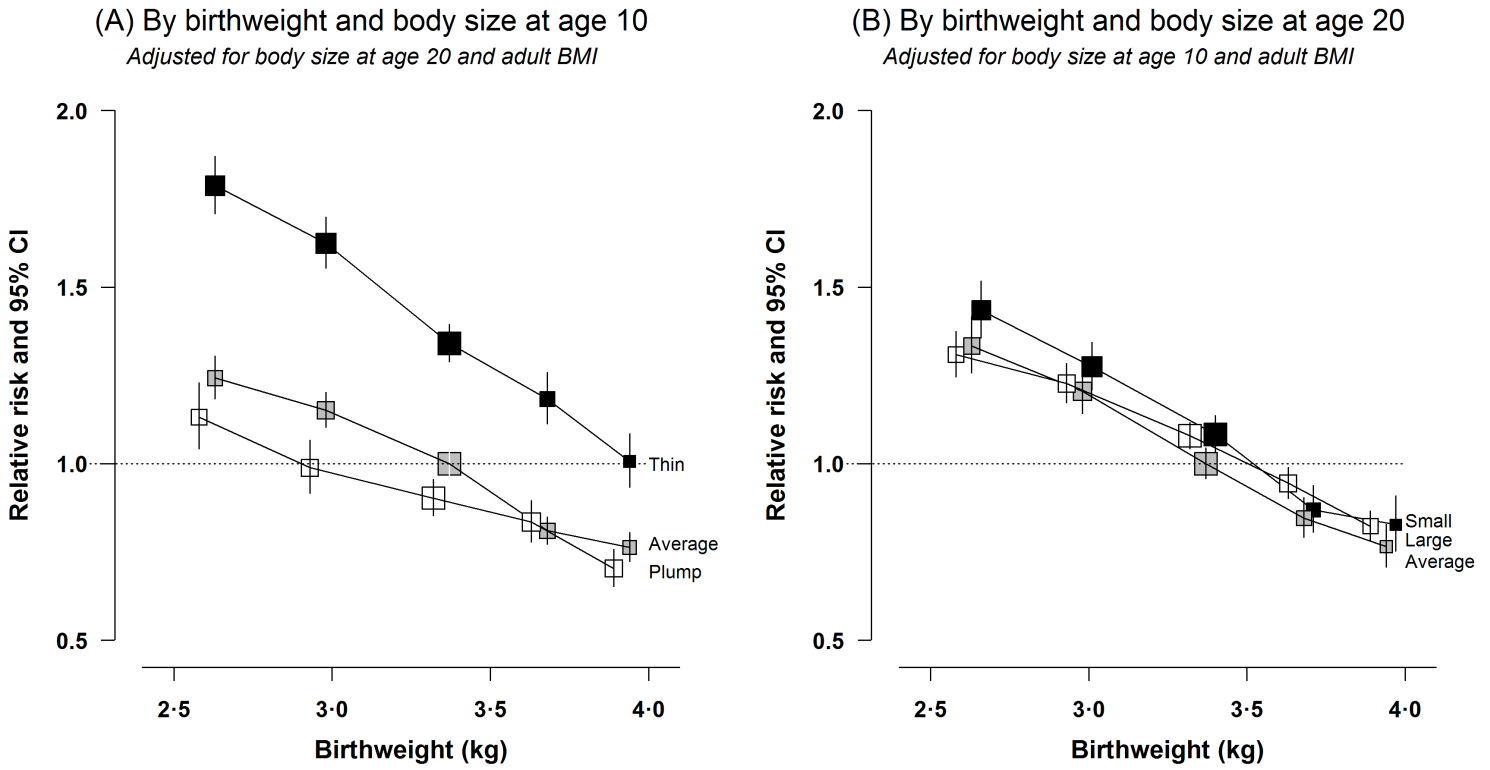
Figure 8.4: Relative risks (RRs) for diabetes by body size at ages 10 and 20 among women with different levels of adult body mass index (BMI)

RRs were adjusted for age, year of birth, year at baseline, region of residence, social deprivation, maternal history of diabetes, age at menarche, smoking, alcohol intake, and strenuous exercise. Women with average or medium body size at ages 10 or 20, and adult BMI <25.0 kg/m^2 were considered as the reference groups

8.4.3 Risks of adult-onset diabetes associated with body size at ages 10 and 20 years among women with different levels of birthweight

Figure 8.5 shows the RRs for diabetes by body size at ages 10 and 20 among women with different levels of birthweight. The inverse associations between birthweight and risk of diabetes were strongly evident at each level of body size at age 10. It is also evident that after allowing for body size at age 20 and adult BMI, at every level of birthweight, diabetes risk tended to be greater for women who were thin compared to those who were average or plump - indicating additional risks of adult-onset diabetes associated with being thin at age 10 (Figure 8.5A).

The inverse associations between birthweight and risk of diabetes were also strong at each level of body size at age 20 (Figure 8.5B). However, at every level of birthweight there was a little, if any, additional effect of body size at age 20 on the risk of adult-onset diabetes.



RR (95% CI)					
Thin	1.8 (1.7-1.9)	1.6 (1.6-1.7)	1.3 (1.3-1.4)	1.2 (1.1-1.3)	1.0 (0.9-1.1)
Average	1.3 (1.2-1.3)	1.2 (1.1-1.2)	1.0 (1.0-1.0)	0.8 (0.8-0.8)	0.8 (0.7-0.8)
Plump	1.1 (1.0-1.2)	1.0 (0.9-1.1)	0.9 (0.9-1.0)	0.8 (0.8-0.9)	0.7 (0.7-0.8)

RR (95% CI)					
Small	1.5 (1.4-1.5)	1.3 (1.2-1.3)	1.1 (1.1-1.2)	0.9 (0.8-1.0)	0.8 (0.7-0.9)
Average	1.4 (1.3-1.4)	1.2 (1.2-1.3)	1.0 (1.0-1.0)	0.9 (0.8-0.9)	0.8 (0.7-0.8)
Large	1.3 (1.2-1.4)	1.2 (1.2-1.3)	1.1 (1.0-1.1)	0.9 (0.9-1.0)	0.8 (0.8-0.9)

Figure 8.5: Relative risks (RRs) for diabetes by (A) birthweight and body size at age 10 and (B) birthweight and body size at age 20

Analyses are adjusted for age, year of birth, year of baseline, region of residence, social deprivation index, smoking, alcohol intake, strenuous exercise, age at menarche, and maternal history of diabetes. Women with birthweight of 3.0-3.4 kg and average body size at age 10 or average body size at age 20 are the reference group. RRs are plotted against mean recorded birthweight in each category.

8.4.4 Effect modification by other factors

The RRs of adult-onset diabetes for body size at ages 10 and 20 years were assessed in subgroups of women defined by their year of birth, age at follow-up, social deprivation, adult height, age at menarche, smoking, alcohol intake, strenuous exercise, and maternal history of diabetes. There was weak evidence for variations in RRs for thin vs average and for plump vs average by most of these factors (Figure 8.6).

There was weak evidence for variations in RRs for small vs medium and for large vs medium at age 20 years by these factors (Figure 8.7). Increased risks of diabetes for women who were large at age 20 years were observed only in a more contemporary cohort (i.e. those born after 1946) and for younger ages at diabetes diagnosis (i.e. <65 years).

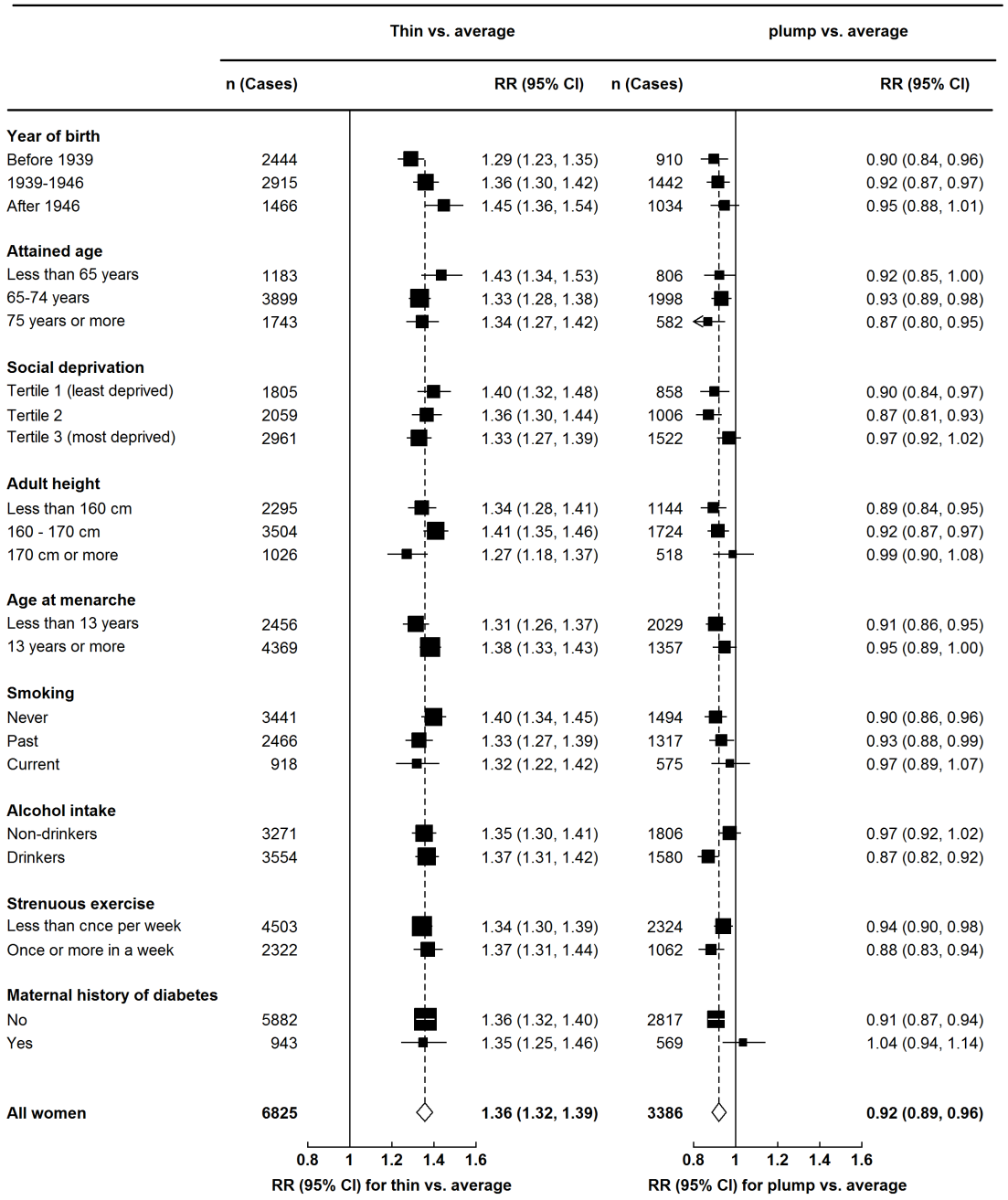


Figure 8.6: Relative risks (RRs) for diabetes by body size at age 10 years, by various characteristics at baseline

The dotted lines represent the RRs for thin vs. average and plump vs. average for all women. RRs are adjusted as appropriate for age, year of birth, year at baseline, region of residence, social deprivation, age at menarche, adult height, smoking, alcohol intake, strenuous exercise, body mass index, and maternal history of diabetes.

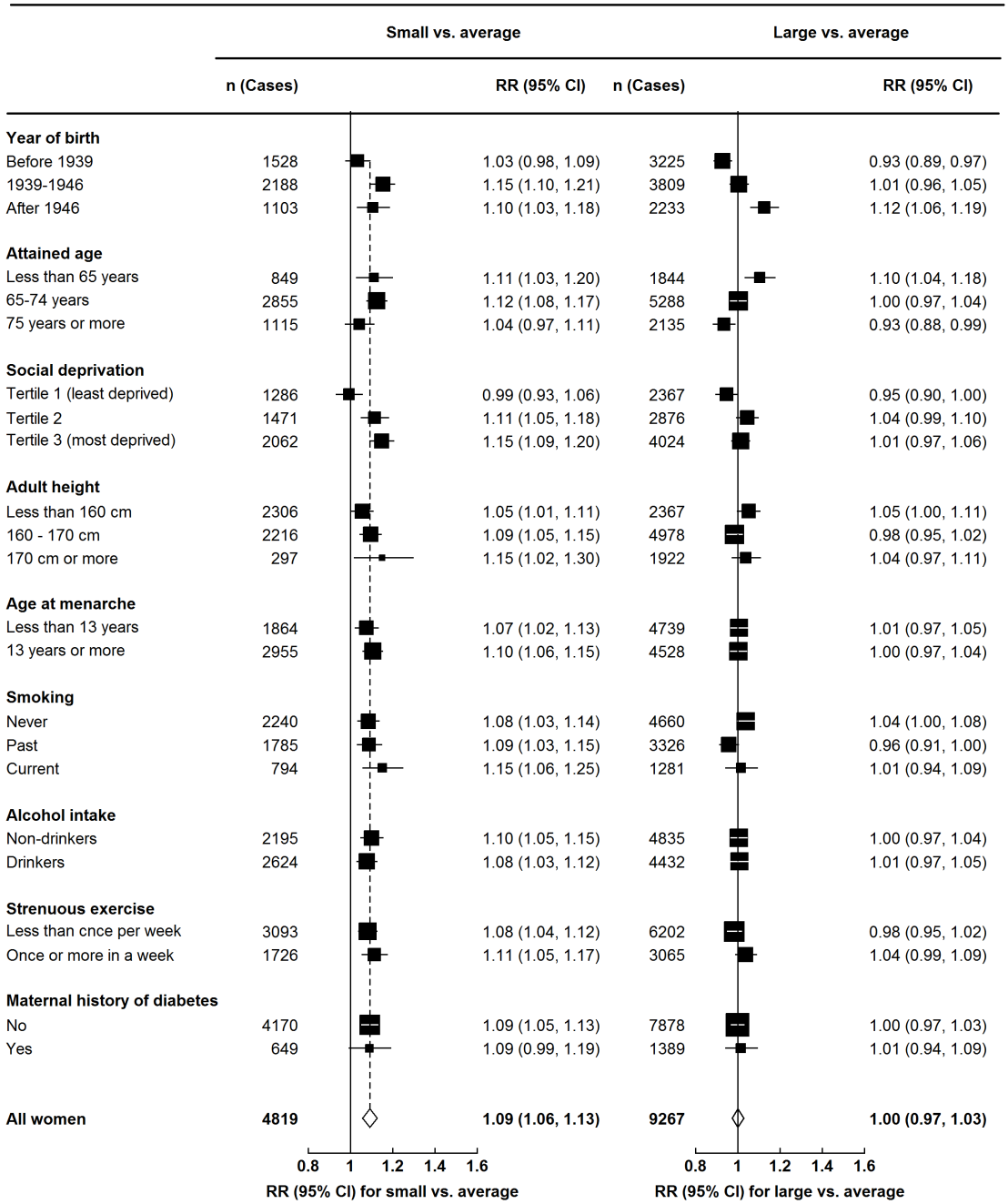


Figure 8.7: Relative risks (RRs) for diabetes by body size at age 20 years, by various characteristics at baseline

The dotted lines represent the RRs for small vs. average and large vs. average for all women. RRs are adjusted as appropriate for age, year of birth, year at baseline, region of residence, social deprivation, age at menarche, adult height, smoking, alcohol intake, strenuous exercise, body mass index, and maternal history of diabetes.

8.4.5 Sensitivity analysis

Table 8.2 and 8.3 show the RRs for diabetes by body size at ages 10 and 20 years from various sensitivity analyses: diabetes recorded in primary care (i.e. CPRD) data, restricting to diabetes coded specially as type-2, excluding initial years of follow-up, excluding women with poor or fair health, and excluding women who changed their diet due to illness.

RRs of diabetes for thin vs average body size at age 10 changed slightly in analysis with diabetes recorded in primary care when compared with RRs using hospital admissions data 8.2. The RRs for body size at age 20 years categories were almost similar for analyses using hospital admissions data and primary care data 8.3. Additional analyses to minimise biases associated with general ill-health or changes in health-related behavioural factors also yielded similar results for body size at ages 10 and 20 years. There was also little evidence that reverse causation could explain the observed associations.

Table 8.2: Sensitivity analyses for the association between body size at age 10 and risk of diabetes in adulthood

Body size at age 10	No. of women/ cases	Multivariable-adjusted RR (95% CI)	
		Adjusted for socioeconomic and lifestyle factors	Additionally adjusted for adult BMI
Restricting to diabetes coded specifically as type-2 (ICD-10 code: E11)			
Thin	129116/8309	1.49 (1.45-1.52)	1.37 (1.34-1.41)
Average	218672/10528	1.00 (0.98-1.02)	1.00 (0.98-1.02)
Plump	63976/3939	1.07 (1.03-1.10)	0.92 (0.89-0.96)
Diabetes recorded in primary care (i.e. CPRD) data			
Thin	13805/1109	1.36 (1.27-1.46)	1.28 (1.19-1.37)
Average	23795/1585	1.00 (0.96-1.05)	1.00 (0.96-1.05)
Plump	6691/588	1.17 (1.07-1.28)	1.07 (0.98-1.17)
Excluding the first 5 years of follow-up			
Thin	126294/7844	1.47 (1.44-1.51)	1.37 (1.33-1.40)
Average	214036/9990	1.00 (0.98-1.02)	1.00 (0.98-1.02)
Plump	62456/3679	1.05 (1.02-1.09)	0.92 (0.88-0.95)
Excluding the first 10 years of follow-up			
Thin	119018/4594	1.42 (1.37-1.47)	1.33 (1.29-1.38)
Average	203081/6086	1.00 (0.98-1.02)	1.00 (0.98-1.02)
Plump	58887/2194	1.04 (0.99-1.08)	0.91 (0.87-0.96)
Excluding women who rated their health as poor or fair at baseline			
Thin	95393/5005	1.44 (1.39-1.49)	1.36 (1.31-1.40)
Average	170994/7016	1.00 (0.98-1.02)	1.00 (0.98-1.02)
Plump	47179/2366	1.04 (1.00-1.09)	0.91 (0.87-0.95)
Excluding women who changed their diet due to illness			
Thin	109971/6735	1.47 (1.43-1.51)	1.37 (1.33-1.41)
Average	192905/9011	1.00 (0.98-1.02)	1.00 (0.98-1.02)
Plump	55247/3255	1.06 (1.02-1.10)	0.92 (0.88-0.95)

Analyses were adjusted for age, year of birth, year of baseline, region,
deprivation index, smoking,alcohol intake, strenuous exercise, and maternal diabetes

Table 8.3: Sensitivity analyses for the association between body size at age 20 and risk of diabetes in adulthood

		Multivariable-adjusted RR (95% CI)	
Body size at age 20	**No. of women/ cases**	**Adjusted for socioeconomic and lifestyle factors**	**Additionally adjusted for adult BMI**
Restricting to diabetes coded specifically as type-2 (ICD-10 code: E11)			
Small	127618/5760	0.86 (0.84-0.89)	1.09 (1.06-1.13)
Medium	130408/6115	1.00 (0.98-1.02)	1.00 (0.98-1.02)
Large	153738/10901	1.56 (1.53-1.60)	1.03 (1.00-1.05)
Diabetes recorded in primary care (i.e. CPRD) data			
Small	13408/843	0.91 (0.84-0.98)	1.10 (1.02-1.19)
Medium	14159/898	1.00 (0.94-1.07)	1.00 (0.94-1.06)
Large	16724/1541	1.45 (1.36-1.54)	1.02 (0.95-1.10)
Excluding the first 5 years of follow-up			
Small	125088/5496	0.86 (0.83-0.88)	1.08 (1.05-1.12)
Medium	127770/5869	1.00 (0.98-1.02)	1.00 (0.98-1.02)
Large	149928/10148	1.53 (1.49-1.56)	1.02 (0.99-1.04)
Excluding the first 10 years of follow-up			
Small	118989/3301	0.85 (0.82-0.88)	1.06 (1.02-1.11)
Medium	121505/3590	1.00 (0.97-1.03)	1.00 (0.97-1.03)
Large	140492/5983	1.48 (1.44-1.53)	1.01 (0.98-1.05)
Excluding women who rated their health as poor or fair at baseline			
Small	98110/3634	0.86 (0.83-0.89)	1.09 (1.05-1.13)
Medium	101209/3985	1.00 (0.97-1.03)	1.00 (0.97-1.03)
Large	114247/6768	1.55 (1.50-1.59)	1.03 (1.00-1.07)
Excluding women who changed their diet due to illness			
Small	110982/4738	0.85 (0.83-0.88)	1.08 (1.05-1.12)
Medium	114290/5177	1.00 (0.97-1.03)	1.00 (0.98-1.03)
Large	132851/9086	1.56 (1.52-1.60)	1.02 (0.99-1.05)

Analyses were adjusted for age, year of birth, year of baseline, region,
deprivation index, smoking,alcohol intake, strenuous exercise, and maternal diabetes

8.5 Global analysis of body size over the life course

A global analysis of body sizes over the life course can help us to understand whether the observed effects for one body size on diabetes risk are similar at every level of other body size variables. If similar effects are observed, they can suggest whether the associations are separate and independent.

The large size of the Million Women Study allowed me to cross-classify women by their birthweight, body size at age 10, body size at age 20, and BMI at age 60, with a sufficient number of incident diabetes cases in most of the cells. Table 8.4 shows the RRs for diabetes, cross-classified by birthweight, body size at age 10, body size at age 20, and BMI at age 60. The clear increases in diabetes risk associated with increasing BMI at age 60 and with decreasing birthweight were evident in all cells of the table. It can also be seen that diabetes risk tended to be greater for women who were thin, compared to those with an average size or plump at age 10, irrespective of the levels of birthweight, body size at age 20 and BMI at age 60. There was a little separate effect of body size at age 20 on adult-onset diabetes risk after allowing for other body size variables.

Figure 8.8 shows the RRs for diabetes per 1 kg decrease in birthweight, thin vs average at 10 years and per 5 kg/m^2 increase in BMI at age 60 years, by other body size indices. As expected, the RRs remained consistent and strong at every level of other body size indices. It is important to understand that even though there are some noticeable differences in the risk estimates at relative scale, they might not indicate effect modification in terms of absolute risks.

Table 8.4: Relative risks (RRs) for diabetes by birthweight, body size at age 10, body size at age 20, and BMI at age 60

RRs (95% CIs) with *Number of cases*									
	Birthweight <3.0 kg			**Birthweight 3.0-3.5 kg**			**Birthweight ≥3.5 kg**		
	Thin at age 10	Average at age 10	Plump at age 10	Thin at age 10	Average at age 10	Plump at age 10	Thin at age 10	Average at age 10	Plump at age 10
BMI <25.0 kg/m^2 at age 60									
Small at age 20	1.85 (1.72-2.00) *649*	1.10 (0.98-1.25) *267*	1.28 (0.89-1.83) *30*	1.36 (1.22-1.50) *366*	0.98 (0.86-1.11) *245*	0.96 (0.63-1.46) *22*	1.10 (0.95-1.26) *194*	0.71 (0.60-0.83) *143*	0.76 (0.49-1.16) *21*
Medium at age 20	1.78 (1.58-2.01) *263*	1.22 (1.09-1.37) *291*	1.27 (0.94-1.72) *42*	1.27 (1.10-1.47) *190*	1.00 (0.89-1.12) *293*	1.07 (0.78-1.46) *39*	1.03 (0.86-1.24) *117*	0.75 (0.65-0.86) *212*	0.57 (0.39-0.83) *28*
Large at age 20	1.97 (1.64-2.38) *109*	1.38 (1.20-1.58) *211*	0.98 (0.76-1.27) *57*	1.25 (1.00-1.55) *79*	1.00 (0.87-1.14) *212*	0.88 (0.69-1.12) *67*	1.23 (0.97-1.56) *68*	0.79 (0.69-0.91) *189*	0.68 (0.54-0.85) *73*
BMI 25.0-29.9 kg/m^2 at age 60									
Small at age 20	4.88 (4.57-5.21) *919*	3.28 (2.97-3.63) *386*	3.63 (2.82-4.69) *61*	3.75 (3.44-4.08) *523*	2.70 (2.43-3.01) *330*	1.85 (1.29-2.65) *30*	2.95 (2.63-3.32) *282*	1.96 (1.71-2.25) *203*	1.84 (1.30-2.61) *32*
Medium at age 20	4.89 (4.52-5.30) *615*	2.92 (2.69-3.17) *563*	2.84 (2.31-3.49) *90*	3.79 (3.44-4.17) *414*	2.32 (2.13-2.53) *537*	2.54 (2.07-3.11) *92*	3.02 (2.67-3.40) *265*	1.86 (1.69-2.05) *399*	1.75 (1.40-2.20) *74*
Large at age 20	5.26 (4.76-5.80) *395*	3.40 (3.16-3.66) *732*	2.55 (2.25-2.90) *233*	4.24 (3.80-4.73) *323*	2.68 (2.49-2.88) *731*	2.34 (2.08-2.64) *270*	2.88 (2.52-3.29) *217*	2.04 (1.89-2.20) *648*	1.74 (1.55-1.96) *287*
BMI ≥30 kg/m^2 at age 60									
Small at age 20	9.99 (9.08-10.9) *435*	8.09 (7.03-9.30) *199*	6.85 (4.72-9.94) *28*	9.16 (8.16-10.28) *294*	7.45 (6.43-8.64) *178*	6.32 (4.19-9.53) *23*	8.21 (7.10-9.49) *185*	6.20 (5.26-7.30) *144*	4.36 (2.78-6.85) *19*
Medium at age 20	9.49 (8.58-10.5) *376*	7.61 (6.88-8.41) *388*	7.07 (5.63-8.88) *75*	7.43 (6.56-8.43) *247*	6.49 (5.86-7.19) *371*	5.92 (4.66-7.53) *67*	7.33 (6.37-8.44) *197*	5.74 (5.13-6.41) *313*	4.80 (3.77-6.12) *66*
Large at age 20	10.1 (9.24-11.0) *514*	7.98 (7.50-8.49) *1031*	7.92 (7.35-8.52) *723*	8.90 (8.05-9.85) *384*	7.32 (6.89-7.78) *1069*	6.81 (6.31-7.34) *695*	7.72 (6.92-8.61) *326*	6.03 (5.68-6.41) *1052*	6.45 (6.06-6.86) *1001*

Analyses are adjusted for age, year of birth, year at baseline, region of residence, social deprivation index,age at menarche, smoking, alcohol consumption, strenuous exercise, and maternal history diabetes.

Women with birthweight of 3.0-3.4 kg, average body size at age 10, average body size at age 20 and adult BMI <25.0 kg/m^2 are the reference group.

Cells with red case numbers had less than 50 new cases of diabetes

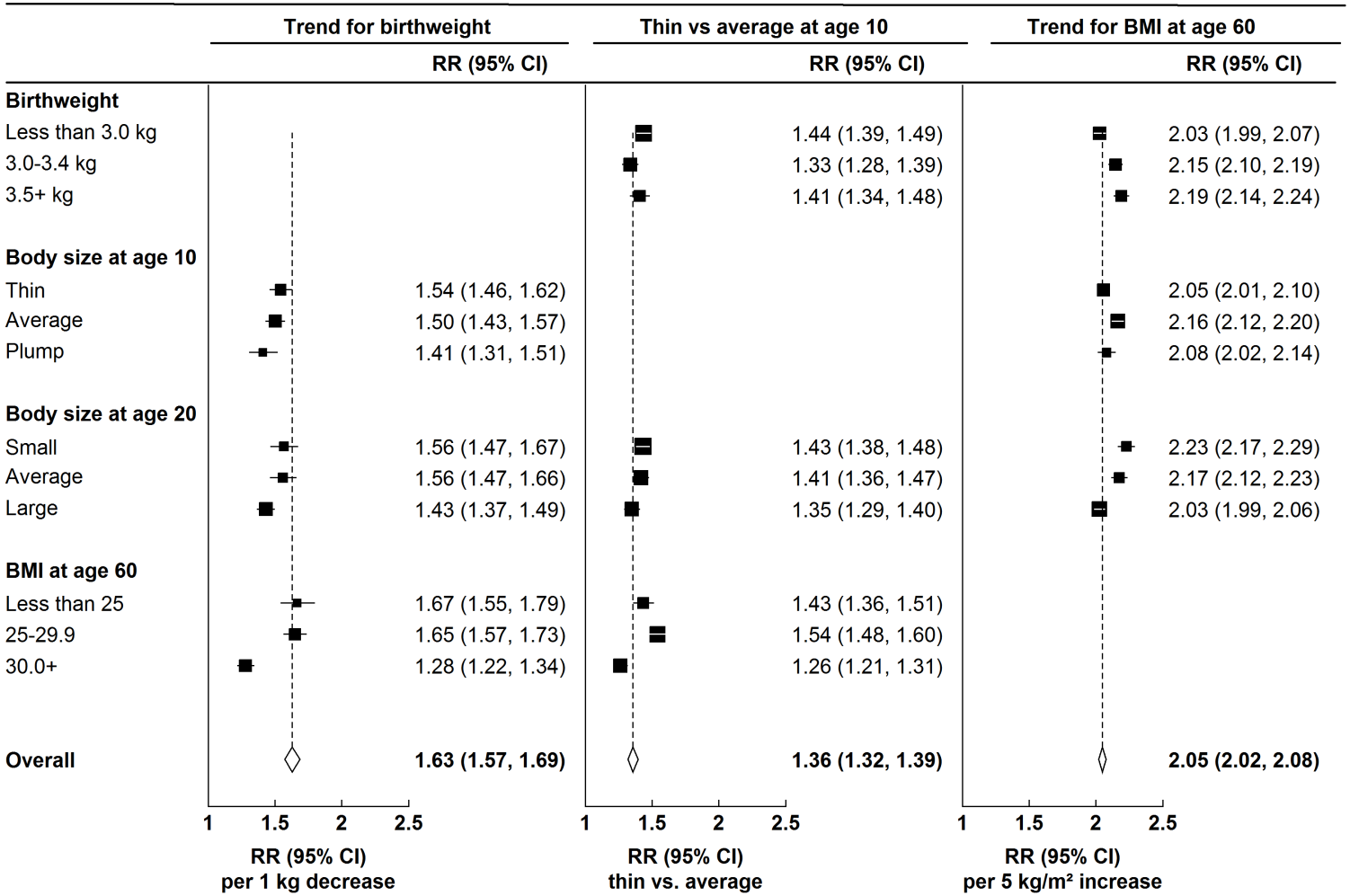


Figure 8.8: Relative risks (RRs) for diabetes for per kg decrease in birthweight, thin vs average at age 10 years and per 5 kg/m^2 increase in BMI at age 60 years, by other body size indices

Analyses are adjusted for age, year of birth, year of baseline, region of residence, social deprivation index, smoking, alcohol intake, strenuous exercise, age at menarche, and maternal history of diabetes.

8.6 Discussion

8.6.1 Main findings

Analyses in this chapter used information on body size at ages 10 and 20 years which were validated against measured BMI at round the relevant ages and shown that at every level of birthweight and adult BMI, there was a strong evidence of increased risks of adult-onset diabetes associated with being thin at age 10 (validated BMI of 16.2 kg/m^2 at age 11). Being large at ages 10 and 20 years was relevant to adult-onset diabetes risk largely as a predictor of adult adiposity. However, being plump at age 10 (validated BMI of 20.4 kg/m^2 at age 11) slightly decreased the risk of adult-onset diabetes, after adult adiposity was taken into account. It is also found that low birthweight could increase the risk of adult-onset diabetes in women with different levels of body size at ages 10 and 20 years. After allowing for body sizes at other ages, it seems body size at age 20 had little, if any, independent effect on adult-onset diabetes risk.

8.6.2 Results in relation to other studies

Examining the associations of adult-onset diabetes with different body size indices over the life course offered insights into the separate effects of these body sizes

The large number of incident cases (n>24,000) in the Million Women Study allowed us to subdivide the results separated by birthweight, body size at age 10 and age 20, and adult BMI, with sufficient numbers in almost every cell for reliable estimation of the risk of adult-onset diabetes. By doing so, this chapter offers substantial insights into the separate effects of body size at ages 10 and 20 years on the risk of adult-onset diabetes. Body sizes in early life are correlated with each other, and also with adult adiposity - the strongest risk factor for diabetes in adulthood. Comprehensive investigations by cross-classifying on different body size indices, therefore, help us to understand their separate effects on adult-onset diabetes risk, after allowing for their correlatedness; which previous studies have been unable to provide.

Given the tracking nature of adiposity from early life to later life, it is not clear whether the increased risks of adult diabetes associated with larger body size in childhood and adolescence are driven by their associations with adult BMI. Many prospective studies [13, 15, 16, 141, 143–145, 147, 148, 267, 268] have shown that large body size in childhood and in early adulthood are associated with higher risk of diabetes in adulthood, but only a few of them [13, 16, 148] had results with adjustment for adult BMI. From the analyses in this chapter, it is evident that the additional risks associated with being plump at age 10 years and with being large at age 20 years were attenuated when adult BMI was taken into account. These results are in line with findings from recent prospective investigations on early life body size and risk of adult diabetes. The Copenhagen School Health Record Register Study [144] reported that overweight in childhood ages (between age 7 and 13 years) was associated with increased risk of diabetes only if it continued until early adulthood. Studies based on the China Kadoorie Biobank [16], E3N [133] cohort and MELANY cohort [148] reported that higher BMI in early adulthood increased the risk of type 2 diabetes in later life, but adjustment for adult BMI either attenuated or reversed the associations.

This chapter clearly shows that at every level of birthweight, there was an additional increased risk of developing diabetes for being thin at age 10 years. Additionally, this increased risk of diabetes appears to be persistent in women who were of normal weight, overweight and obese as adults. Increased risk of adult-onset diabetes for childhood thinness was reported in some studies [133, 145] but it was not observed in all subgroups of birthweight and adult adiposity [133], or in all age groups during follow-up [145]. It has been suggested previously that childhood adiposity can be related to earlier pubertal development, hence can modify the risk of diabetes [236, 237]. Analyses here showed that it is unlikely that pubertal development can modify the observed associations between childhood body size and diabetes risk as adults. Results in this chapter also suggested that the increased risks of adult-onset diabetes associated with low birthweight could be seen in women with any levels of body size at childhood and early adulthood, suggesting low birthweight

might influence diabetes risk through pathways other than rapid postnatal growth which was suggested by some previous studies [141, 143, 267].

The evidence on whether a smaller body size in early adulthood ages can increase diabetes risk is not clear. A recent investigation based on the Health Professional Follow-up Study [15] reported that those with BMI <18.5 kg/m^2 had a higher risk of diabetes in adulthood when compared to those with BMI 18.5-22.9 kg/m^2. When adjusted for adult BMI, higher risks of diabetes were observed among those in the lowest decile for early adulthood BMI than those in the higher deciles in a large prospective study based on Israeli men, but the results were not statistically significant in all groups [148]. However, with the statistical power of the Million Women Study and the comprehensive nature of the investigation here in this chapter, it seems being small at age 20 years can only have a little additional effect on adult-onset diabetes risk.

8.6.3 Potential mechanisms

The mechanisms that underlie the associations observed here are unclear. For the effect of thinness at age 10 years, one possible explanation is that development of islet cells in the pancreas can continue postnatally [269, 270], and both childhood thinness and a higher risk of adult-onset diabetes could be shared consequences of poor development of endocrine pancreas [133]. Another explanation could be that even among individuals with similar levels of adult adiposity, those who are small in childhood have more fat mass than those who are large over their life course [271, 272]. A recent study examining trajectories of body size across early and middle life also suggested that being small in early life followed by adult adiposity had more deleterious effects on diabetes risk than being large from early life to adult life [150], supporting the differential effects of adult adiposity by early life growth. Although the associations of being small at birth and age 10 years with increased risk of adult-onset diabetes were statistically independent of each other, it is unclear whether they share the same underlying biological mechanisms.

8.6.4 Strengths and limitations

Strengths and limitations concerning the prospective investigations in the Million Women Study are described in detail in chapter 9. The strengths and limitations that are specific to investigating the associations of body size at ages 10 and 20 years with the risk of adult-onset diabetes in the Million Women Study are discussed below.

The major limitations of using reported body size indices for ages 10 and 20 years involve systematic and random measurement errors in those indices and difficulties in their clinical and public health interpretations. Although body size indices based on relative size and clothes size are considered to be less precise measures than asking about actual height and weight, they are particularly helpful when asking participants to recall their body size in the distant past [97, 98]. A previous report [96] from the Million Women Study compared these reported body size indices against recorded BMI at around relevant ages in a subsample of participants. Although clear associations between categories of body size and mean measured BMIs were observed, it is likely that recalled body sizes in early life are generally under-reported, with more significant under-reporting among those who are overweight or obese as adults [210, 211]. However, the use of mean recorded BMIs for each level of body size at age 10 and age 20 in relevant figures help to interpret the results better. It is also important to acknowledge that the results for reported body size at ages 10 and 20 years in the Million Women Study can only be interpreted for moderate levels of thinness and overweight at relevant ages.

Another important limitation is that new cases of diabetes were identified from hospital admissions record. There are several potential issues with such measure including but not limited to: missing diabetes cases who did not have hospital admission during follow-up, under- and delayed-reporting of diabetes in hospital admissions, differential recording of diabetes by body size in early life and other factors. Diabetes recorded in hospital admissions data was compared against diabetes recorded in primary care data in chapter 4 to understand the degree and likely effects of these issues. Half of those with a diabetes record in primary care had a hospital admission with diabetes recorded in the next 3.8 (95% CI: 3.4-4.3) years

if they had average body size at age 10 years. The corresponding numbers for those who were thin and plump were 3.9 (3.4-4.4) and 3.6 (3.0-4.6) years, respectively. The corresponding numbers for body size at age 20 years were: small (4.0 [3.4-4.9]), medium (4.2 [3.5-4.9)], and large (3.5 [3.2-4.0]) years. These numbers indicate that the observed associations were highly unlikely to be explained by differential diabetes recording in hospital admissions according to the level of body size in early life.

When compared to individuals with a healthy weight, those with overweight and obesity are more likely to under-report their body size in early life while their diabetes status is recorded earlier in hospital admissions. Therefore, adult BMI is an important confounder of the observed association. When examined separately among women with different levels of adult BMI, risks of diabetes by body size at ages 10 and 20 were consistent. Given the observational nature of this study, there are also possibilities of residual confounding (due to unmeasured or imperfect assessment of measured factors) and reverse causation. A wide range of sensitivity analyses was conducted to deal with the issues related to using hospital data and minimising any reverse causation bias and potential confounding by general ill-health or changes in behaviour, and similar results in those analyses indicate the robustness of the findings. However, information on important childhood factors (e.g. childhood socioeconomic status) was not available in the Million Women Study; hence, there might be some residual confounding by those factors.

The large size of the Million Women Study allowed me to subdivide the RRs of diabetes among women with different levels of birthweight, body size at age 10, body size at age 20 and adult BMI, and still get reliable estimates for most of the subgroups. The prospective follow-up for diabetes ensured that any errors in reported birthweight were unlikely to be substantially different between those who developed and did not develop diabetes during follow-up. Record-linkage allowed a completed and unbiased follow-up for the whole cohort, with only 1.5% loss to follow-up and at least one hospital admission for 89% of women. The participants of this cohort also have detailed information on many factors, allowing for statistical adjustment for many potential confounding factors.

8.7 Conclusion

In summary, this largest prospective investigation to date on body size in childhood and early adulthood suggests that the increased risks of developing diabetes in adulthood associated with larger body size at these ages are largely explained by the fact that they are predictors of adult adiposity. On the other hand, at every level of birthweight and adult BMI, being thin at age 10 years is associated with a higher risk of diabetes in adulthood. After allowing for birthweight, size at age 10, and adult BMI, body size at age 20 years had a little residual effect on adult-onset diabetes risk.

9
Strengths and limitations

9.1 Introduction

This chapter discusses the strengths and limitations of analyses in this thesis which are essential for understanding the validity and appropriateness of observed findings. Like any other observational epidemiological studies, there are conditions in the Million Women Study, which could have undermined the validity of the findings to a significant extent. Discussing these conditions and their influence on the conclusions are crucial for assessing the robustness of the findings, placing the current findings in context, and improving future research in this field.

In the following sections, I am going to discuss issues related to specific conditions in the Million Women Study and the likely biases associated with those conditions to better understand the validity of the findings in this thesis.

9.2 Limitations of analyses in this thesis

In epidemiology, various issues with the selection of participants, measurements of factors including exposure and outcome, and confounding might provide alternative explanations for a specific research finding. Acknowledgement of such issues should consist of interpretation of their meaning, assessment of their likely influences on the observed findings, and explaining whether any measures (either in design phase or analysis phase) have been taken to account for them. The findings of a study are considered as "internally valid" when it is evident that they are unlikely to be explained by any sources of errors. External validity or generalisability looks at how the conclusions can be generalised to other circumstances.

9.2.1 Issues with reported body size indices at different ages

Women in the Million Women Study reported their body size at different ages over the life course at baseline, many years after the relevant ages. These reported body size variables were previously compared against recorded values at similar ages and found to have good overall agreements [96, 181]. However, as discussed

in the previous chapters, reported body size indices can have various forms of measurement errors. The utility of reported body size indices can be determined by examining the magnitudes of errors, their likely effects on the risk estimates, and whether such biases can be corrected.

To understand the validity of observed findings in this thesis, the following issues with reported body size indices at different ages need careful considerations:

9.2.1.1 Reporting errors in body size indices

In the Million Women Study, when compared with recorded values at relevant ages, reported birthweight and adult body mass index (BMI), on average, tended to be lower than measured values (i.e. systematic under-reporting) [96, 181]. Usually, under-reporting was bigger for higher categories of birthweight and BMI, although such differences were not statistically significant. Errors in reported body size indices can result in systematic underestimation of their associations with risk of developing diabetes in adulthood, also known as regression dilution. Previously, regression dilution ratios (RDRs) for reported birthweight and adult BMI were estimated to quantify the apparent attenuation of estimates due to reporting errors [96, 181]. RDR is a non-parametric estimate of the ratio of such a biased estimate to the coefficient that would be found if analyses were conducted using true values of the variable of interest [186, 187]. RDR for adult BMI (0.96, 95% CI: 0.95-0.98) suggested little or no attenuation of any apparent disease-exposure association due to under-reporting, while RDR for birthweight (0.59, 95% CI: 0.50-0.67) suggested substantial attenuation.

It was not possible to assess any under-reporting or over-reporting between reported categorical body size in early life (ages 10 and 20 years), and BMI recorded at the relevant times. However, there were significant trends in recorded BMI across categories of body size at ages 10 and 20 years ($P < 0.001$) [96, 181].

For all body size indices, I used the mean recorded values for reported categories to plot category-specific relative risks or to estimate log-linear trends. Such methods

helped with the interpretation the results and minimising the effects of reporting errors and associated regression dilution bias.

9.2.1.2 Differential reporting errors in body size indices by the outcome or covariates

If errors in reported body size indices occur more frequently in one of the groups being compared, then differential misclassification will occur, and the estimate of association can be overestimated or underestimated. The prospective nature of this investigation ensures that any errors in reported body size indices are unlikely to be different for those with and without diabetes.

As discussed in chapter 3, the reporting errors can also contribute to confounding depending on their correlations with other covariates and their errors too. When recalled, body sizes in early life are generally under-reported, with more significant under-reporting among overweight or obese adults [97, 98]. In this thesis, effects of body size in early life on diabetes risk were consistent when examined separately among women with different levels of adult BMI, indicating that the observed associations were unlikely due to any differential reporting errors by adult BMI.

9.2.1.3 Clinical and public health interpretations

Another important limitation of reported body size indices in the Million Women Study is their clinical and public health interpretations. In particular, interpretations for categorical body size at ages 10 and 20 years are challenging, because they are somewhat subjective measures and can vary widely between populations and over periods. In the Million Women Study, categories of reported body size indices were characterised by distinct means of recorded values of corresponding anthropometry. The mean recorded values within each category of these indices helped with the interpretation of relevant results on a more objective scale.

9.2.2 Issues with identifying diabetes cases from hospital admissions data

9.2.2.1 Under-reporting of diabetes in hospital admissions

In the Million Women Study, new cases of adult-onset diabetes were identified in hospital admissions record. As expected, diabetes is generally under-reported in hospital admissions, and therefore, estimates of the incidence of the disease are lower in hospital admissions data than in primary care data (for details see Figure 4.9). There are several factors responsible for lower incidence rate in hospital admissions including but not limited to: missing diabetes cases who did not have hospital admission during follow-up, under-reporting of diabetes in hospital admissions, and the possible time lag between diabetes recording in primary care and in hospital admissions.

In chapter 4, diabetes recorded in hospital admissions was compared against diabetes recorded in primary care. By the end of 10 years of follow-up, about 80% of those with diabetes recorded in primary care also had a hospital admission mentioning diabetes and the median time between diabetes first recorded in CPRD and first recorded in HES was 3.7 (95% CI 3.4-4.1) years (Figure 4.11). They provide some evidence that most of the diabetes cases can be identified from hospital admissions data, with some delay in cohort studies over at least a decade of follow-up.

9.2.2.2 Differential diabetes recording in hospital admissions by body size indices

When compared with women who have a healthy weight, those with higher BMI are more likely to be admitted to hospital, and if admitted, their diabetes is more likely to be recorded because these women are generally admitted for obesity-related conditions. For body size in early life, the similar bias in diabetes recording can happen, the degree of which depends on the correlations between the cause for admission and adult adiposity. As shown in Figure 4.12 in chapter 4, cumulative probabilities of diabetes recorded in hospital admissions among those who had a record of diabetes in primary care were not substantially different by levels of early

life body size indices. Therefore, the observed associations of adult-onset diabetes with body size in early life are highly unlikely to be explained by differential diabetes recording by the exposure.

However, when compared with women who have a healthy weight, those with higher BMI are more likely to be admitted to hospital, and if admitted, their diabetes is more likely to be recorded because these women are generally admitted for obesity-related conditions. For body size in early life, a similar bias in diabetes recording can happen, the degree of which depends on the correlations between the cause for admission and adult adiposity. The consistent and strong associations of low birthweight and being thin at age 10 years with increased risks of adult-onset diabetes when investigated separately in women with different levels of adult BMI suggest little, if any, the effect of such issues.

9.2.2.3 Differential diabetes recording in hospital admissions by other variables

Diabetes recording in hospital admissions may vary by many factors, including region and socioeconomic status. Diabetes recorded in hospital admissions data was compared against diabetes recorded in primary care data in chapter 4 by socioeconomic and lifestyle variables. There was differential diabetes recording in hospital admissions by levels of these variables. These factors should be taken into account in the analysis; otherwise, they might confound observed risk estimates. Therefore, these factors were adjusted in all relevant analyses for this thesis.

9.2.2.4 Recording of diabetes in primary care among those with diabetes in hospital admissions

Information on biomarkers related to diabetes (e.g. blood sugar level, HbA1C) is not available in hospital admissions data, and codes for diabetes are generated from clinical notes relevant to hospital admissions. So, it is difficult to answer whether having a diabetes code in HES accurately indicates diabetes in an individual. Comparing hospital and primary care data for diabetes recording in a subsample, 95% of those with a hospital record of diabetes also had a primary care record

of diabetes. This indicates that diabetes codes recorded in hospital data actually reflect diabetes.

9.2.2.5 Assigning type to diabetes

As mentioned previously, most cases of adult-onset diabetes are of type 2, but type 1 diabetes can also be seen. In this thesis, the large majority (93%) of women had a hospital diagnosis of type 2 diabetes, and sensitivity analyses restricting to diabetes coded specifically as type 2 showed similar results.

Moreover, diagnostic confusion can be seen due to those who are diagnosed with LADA. Patients with LADA present like type 2 diabetes at diagnosis but then require treatment with insulin within the first few years after diagnosis and further tests show that they have GADA antibodies [21]. So, their diagnosis is later changed to type 1 diabetes in routine health data. According to the UK Prospective Diabetes Study, LADA cases are found in between 6% and 10% of all diabetes cases among men and women aged 25–65 years [27]. For a small proportion (<2%) of women in this study, different ICD-10 codes for diabetes were recorded in subsequent hospital admissions. Results from analysis excluding those cases were similar to the main results (data not shown).

9.2.2.6 Number of hospital admission during follow-up

A woman's probability of diabetes recording at hospital admissions depends on her number of hospital admissions during the follow-up. Sensitivity analysis with adjustment for the number of hospital admission did not influence the associations substantially (data not shown), indicating that such an issue did not have a significant impact in this study.

Overall, the issues with diabetes recording in hospital admissions discussed above can bias the estimates of associations in a complex manner. Even though I tried to estimate the likely effects of those issues by comparing diabetes recording between primary care and hospital admissions, it is impossible to disentangle or separate their individual effects. Sensitivity analyses were conducted using diabetes recorded in primary care for a subsample of women, which would indirectly address these issues related to diabetes recording in hospital admissions data. I found that the RRs of diabetes associated with birthweight, body size at age 10 and body size at age 20 were similar to the main results reported in this thesis. However, the RRs for diabetes by adult BMI were greatly underestimated, which highlights that adult adiposity is likely to increase the rate of hospital admission and subsequently, the probabilities of diabetes recording when hospitalised.

9.2.3 Issues related to reverse causation

In observational epidemiological studies, reverse causation can occur when the outcome itself influences the exposure being studied. Short-term associations of adult-onset diabetes risk with body size variables examined here could be due to reverse causation, whereby the preclinical stage of the disease could affect not only women's adult body size and also how they report their body size in early life. There are also possibilities that diabetes cases occurring in the initial period of follow-up are actually pre-existing cases which are either undiagnosed or unrecorded cases.

The Million Women Study is a large prospective study with a long period of follow-up, which allows me to exclude the incident cases of the disease in the early period of follow-up. I excluded the first five years and then the first ten years of follow-up, which increased the length of time between the reporting of body size and recording of a new diabetes case. By doing so, the likelihood that reporting of body size indices were affected by preclinical or undiagnosed diabetes was reduced. When compared to the main results, these results were similar for all body size variables, indicating that the risk estimates observed in this thesis

are unlikely to be influenced by reverse causation. Large sample sizes even after progressive exclusions to the follow-up period reduced any possibilities of statistical artefact due to random variability.

9.2.4 Issues related to residual confounding

Although I adjusted for many important factors related to body size and risk of adult-onset diabetes, I cannot rule out the possibility of residual confounding due to imperfect assessment, unmeasured factors, or both. Information on important factors, for example, gestational age, gestational diabetes, childhood height, childhood socioeconomic status was not available in the Million Women Study; hence the possibilities of some residual confounding cannot be excluded. Women who reported their health as fair or poor (instead of excellent or good) and who reported a change in their diet due to illness at baseline can indicate general ill-health and changes in health-related behavioural factors. Therefore, sensitivity analyses excluding these women can help to minimise potential biases associated with residual confounding. Consistent findings in analyses excluding these women suggested that residual confounding had little, if any, effect on the observed associations.

9.2.5 Issues related to missing data

As mentioned previously in chapter 5, only 56% of those who completed the 3-year re-survey questionnaire reported their birthweight. Women who did not report their birthweight were excluded from the analysis which could give rise to biased estimates if those with missing data differed in relation to both exposure and outcome than those with complete data [223, 224]. Table 5.2 shows that women who reported birthweight had similar adult BMI on average when compared with women who did not report birthweight, although there were differences in other socioeconomic and lifestyle variables. To understand whether the exclusion of women with missing data on birthweight can give rise to biased estimates, I did sensitivity analyses by using a separate category for the missing value in birthweight (data not shown), and the results for body size at ages 10 and 20 and adult BMI did not change.

So that the same women were compared in all analyses, the small numbers with missing data (usually <5%) for each adjustment variable were included as a separate category. Similar results were observed when I conducted separate analyses on 338,512 women (including 19,256 cases of adult-onset diabetes) who had valid information on all adjustment variables.

9.2.6 Baseline prevalence of diabetes and "healthy volunteer effect"

As mentioned earlier, the 3-year re-survey (median year 2001 [IQR: 2000-03]) is the study baseline for the prospective investigations in this thesis. Up until the baseline, 33,510 women (3.9% of those who completed 3-year re-survey) reported that they had diabetes, and were excluded from the analysis. According to the Health Survey England (HSE) 2003 [273], the diabetes prevalence was 7.0% among women aged 55 years or more. The possible reasons for the lower prevalence of diabetes within the cohort than the general UK population could be that only known cases of diabetes were captured in the study questionnaires while the HSE used blood sugar tests to identify diabetes cases. Due to the long and asymptomatic latent period, a significant proportion of diabetes cases remain undiagnosed [274]. A recent study [275] based on five annual cross-sectional health examination surveys in England found that 33% of diabetes cases are undiagnosed at the community level. Even if we adjust for undiagnosed cases, the estimated prevalence of diabetes at the baseline of Million Women Study population (5.2%) was lower than that in general population at around same period (7.0%).

The lower prevalence of diabetes at baseline maybe also because cohort studies like the Million Women Study generally have a lower risk population than the underlying eligible source population due to selective non-participation of individuals with a poorer health status (also known as "healthy volunteer effect"). While such a phenomenon is unlikely to introduce a bias in the risk estimates, other factors like age, education, region, and socioeconomic status can influence such participation in differential manners to cause bias or to limit generalisability, and thereby these

factors were adjusted during the analyses. Excluding the first few years of follow-up can also deal with any possible methodological issues because the effects of "healthy volunteer", if any, tend to disappear after the initial years of follow-up [276].

9.2.7 Issues with analysing correlated body size data

As explained in chapter 3, adjustment for body size in later life that is likely to be on the causal path from body size in early life and adult-onset diabetes (i.e. mediator) can effectively take account of the effect of the body size which is mediated through the body size in later life. This often pulls the effect estimates towards the null [240]. However, the large size of the Million Women Study allowed me to analyse the risk of diabetes in different combinations of body size over the life course, which offers substantial insights into the separate effects of body sizes at different ages, after allowing for their correlations. I also presented results both before and after adjustment for other body size variables.

9.2.8 Issues with generalising the findings

The Million Women Study recruited about one in every four UK women born in 1935–1950 (aged 50–64 years). Nevertheless, only 64% of those at recruitment completed the 3-year re-survey questionnaire, i.e. the study baseline questionnaire. Again, 44% of those who completed the baseline questionnaire did not report their birthweight, and therefore were excluded from the analysis. Table 5.1 and Table 5.2 in chapter 5 show the differences in key characteristics between those who were included and excluded from this study. However, these factors are unlikely to cause selection bias because, in order for bias to occur, selection into the study has to be related to both exposure and outcome. However, they are essential for understanding how to generalise the findings to the background population. Although the study population is different from the underlying population of UK women at the time of recruitment, there are sufficiently large numbers of women across the levels of various socioeconomic and lifestyle factors to study the associations reliably.

The great majority of the Million Women Study participants are of White ethnicity (96%), and therefore the observed findings should be generalised cautiously to other ethnic populations. The cohort has no information on men, so any possible sex differences in the observed associations cannot be understood from this thesis. Some previous studies [129, 131] reported stronger associations of birthweight and risk of diabetes in adulthood, but for other body size variables, possible sex differences were not evident. This cohort consists of middle-aged and elderly women, and their nutritional status in early ages might have been affected by World War II (1935-1945). Although analyses among women grouped according to their year of birth showed that the results remained strong and consistent, generalising the findings to a more recent cohort of young individuals needs caution.

9.3 Strengths of this thesis

The strengths of the Million Women Study include its large size, which limits the random errors due to chance. It also allowed examining the separate effects of body size at different ages on the risk of adult-onset diabetes in a large number of subgroups of women, with reliable estimates for almost all subgroups. The prospective follow-up for diabetes ensured that any errors in reported body size indices were unlikely to be different between those who developed and did not develop diabetes during follow-up. Record-linkage allowed a completed and unbiased follow-up for the whole cohort, with only 1.5% loss to follow-up and at least one hospital admission for 89% of women. Many other middle-aged cohorts recontacted their participants for follow-up, which could lead to high attrition rate and subsequent selection bias [13, 132].

Relevant comparisons between self-reported and measured or recorded body sizes indices at different ages and between diabetes recording in primary care and hospital admissions data helped to better understand the validity of findings in this thesis. The participants of this cohort also have detailed information on many factors, allowing for statistical adjustment for many potential confounding factors, which many record-linkage studies with measured values cannot do. A wide range of sensitivity analyses was conducted to deal with various methodological issues, and similar results in those analyses indicated the robustness of the findings.

Also, the systematic review of the evidence before this thesis helped to contextualise the findings of this thesis for better understanding of all available evidence on the observed associations for body size in early life and risk of diabetes in adulthood.

10
Discussion

10.1 Aim of my DPhil research

In my DPhil research, I set out to provide a detailed understanding of body size in early life and the risk of diabetes in adulthood, in relation to the already established strong association between adult body mass index (BMI) and diabetes. I specifically examined the roles of birthweight, body size at ages 10 and 20 years and adult BMI on the risk of adult-onset diabetes. I conducted advanced statistical analyses on data from a large-scale prospective cohort of middle-aged and elderly women in the UK (The Million Women Study) and also synthesised relevant findings from previous studies by doing systematic reviews. They helped me to interpret the totality of the evidence on the relative importance of body sizes from birth to late adulthood in relation to the risk of adult-onset diabetes. Examining the risk of adult-onset diabetes separately in subgroups defined by these body size indices allowed me to understand their separate effects, recognising tracking of body size from early life to late adulthood.

10.2 Main findings from this thesis

The main findings of this thesis, as detailed in Chapters 6 to 8 are summarised below.

- **Diabetes risk increased strongly with increasing adult adiposity:** Consistent with previous reports, the analyses in chapter 6 showed that the risk of adult-onset diabetes increased strongly with increasing adult BMI. When compared with women with a healthy weight, being overweight and being obese was associated with 2.6 times and 6.6 times higher risk of developing diabetes in adult life, respectively. The associations between adult adiposity and diabetes risk remained strong even after accounting for various socioeconomic and lifestyle factors.

- **Being small at birth increased risk of adult-onset diabetes:** Paradoxically, but also as expected, being born with a lower birthweight increased the risk of diabetes as adults, with no evidence for effect modification

by other socioeconomic and lifestyle characteristics. The inverse relationship between birthweight and risk of diabetes was strongly evident when examined separately among women who were of healthy weight, overweight and obese as adults. Also, when body size in childhood and early adulthood were taken into account, the association between birthweight and risk of adult-onset diabetes remained robust, independent and persistent.

- **Larger body size in early life was largely relevant to adult-onset diabetes risk as a predictor of adult adiposity:**
  The analyses in chapter 8 showed that the increased diabetes risks associated with large body size at ages 10 and 20 years were largely explained by their associations with adult adiposity. Women who became lean in adulthood did not have an increased risk of diabetes for being large in childhood and in early adulthood. Also, chapter 8 showed that the slight additional risk associated with extremely high birthweight was largely attributable to adult adiposity.

- **Being thin in childhood ages was associated with additional risks of adult-onset diabetes:**
  Chapter 8 also suggested that at every level of birthweight and adult BMI, there was an additional risk of diabetes associated with being thin at age 10 years. However, after allowing for birthweight, size at age 10, and adult BMI, being small at age 20 years had a little additional effect on adult-onset diabetes risk.

10.3 Body size in early life and diabetes in adulthood: the totality of evidence

This thesis prospectively investigated the associations between body size in early life and risk of developing diabetes in adulthood. In the beginning part of this thesis, in chapter 3, I outlined the topics that needed careful reflections to understand the research questions involving body size in early life and risk of adult-onset diabetes. Systematic reviews of the evidence before this thesis have added richness and depth to the research inquiries in this thesis. To understand the totality of evidence, I now look at questions that have been answered or remained unanswered.

10.3.1 Questions answered in this thesis

Analyses in this thesis demonstrated that in addition to the strong risk of diabetes for higher BMI in adulthood, there were independent and persistent increased risks of adult onset-diabetes associated with lower birthweight and being thin at age 10 years. For the effects of low birthweight on adult-onset diabetes risk, growth in postnatal period or increased life-long susceptibility to obesity-related diabetogenesis did not have any large influences. On the other hand, the adverse effects of large body size in early life on the risk of developing diabetes in adulthood were largely due to the tendency to gain adiposity in adulthood. The associations between body size in early life and risk of adult-onset diabetes remained unchanged even after accounting for various socioeconomic and lifestyle factors.

10.3.2 Questions remaining unanswered in this thesis

Several questions remain unanswered in this thesis. I investigated the separate effects of birthweight and body size at ages 10, 20 and 60 years on the risk of diabetes in adulthood. They generally reflect body size at around birth, mid-childhood, early adulthood, and late adulthood, respectively. However, it is not clear whether body size in these stages has distinct effects of diabetes development, or they are just manifestations of some immeasurable constructs that determine the risk of diabetes in adulthood. The underlying biological pathways explaining

the associations in this thesis are also unknown. Information on many factors that might influence body size at different ages is not available for this cohort, therefore testing a new hypothesis for multiple mechanisms was not possible. The body size indices used in this thesis could represent either adiposity or muscularity, which makes it more challenging to understand possible mechanisms of observed associations. For example, when compared with direct measures of adiposity, for example, dual-energy X-ray absorptiometry (DEXA), BMI has high specificity (0.90) but low sensitivity (0.50) for defining obesity [277].

10.4 Added value of this thesis

10.4.1 Validity of evidence

It is highly unlikely that alternative explanations can threaten the validity of the findings in this thesis. Epidemiological investigations in this thesis utilise validated measures of body size at different ages from the Million Women Study which has other advantages too: prospective design, large size, and virtually complete, long-term follow-up through linkage to routinely collected NHS registers. The analysis framework is conceived based on the understandings of current literature, and sophisticated statistical techniques have been used to analyse the data. Issues related to reporting errors in body size, diabetes recording in hospital admissions, confounding, and reserve causation have been scrutinised through a wide range of sensitivity analyses. From the relevant discussions in chapter 9, it is evident that these issues do not have any significant implications that could have undermined the validity of the findings to a large extent. Detailed discussions on the representativeness of the findings also help to contextualise the conclusions.

10.4.2 Relevance to the field

In this thesis, I tried to answer several important research questions in relation to body size at different ages and the risk of adult-onset diabetes.

Many previous studies in this field were unable to show the presence of long term adverse effects of body size in early life, acting independently of their relationships with adult BMI. This was mainly due to limitations in statistical analysis attributable to standard adjustments for adult BMI, study design, and sample size. A recent approach identifies subgroups of individuals who share similar trajectories and compare the risks of developing diabetes in adulthood directly across different trajectory groups [278]. As discussed previously, one way of dealing with the difficulties in separating the effects of early-versus late-life body size on risks of developing diabetes in adulthood is to directly compare the effects for different combinations of body size throughout life.

The large size of this study offers substantial insights into the separate effects of body sizes at different ages since there were sufficient numbers to subdivide women simultaneously by their body size at birth, and at ages 10, 20 and 60 years. Previous studies were either too small to yield reliable estimates for most subgroups [12, 132, 133, 146] or could not allow for adiposity in later life while examining the diabetes risks associated with body size in early life [144, 267]. Examining the risk of adult-onset diabetes in different combinations of body size across the life course is essential because it helps us to understand whether the observed effects for one body size on diabetes risk are similar at every level of other body size variables. Such analysis also takes the correlations between body size indices into account. By studying the effects of body size from birth to late adulthood, the thesis also provides a life course perspective on body size and adult-onset diabetes.

This thesis also provides insights into research questions relevant to body size at specific ages, for example, whether there is an increased risk of diabetes associated with extremely high birthweight, and whether birthweight can modify the associations of childhood body size and diabetes in adulthood. Analyses in

this thesis also showed that the effects of body size on adult-onset diabetes risk were not modified by other socioeconomic and lifestyle variables.

Overall, this is the first study which can reliably report separate effects of body size at a particular age after taking account of body size in three other stages.

10.4.3 Providing new hypotheses for underlying pathophysiology

This thesis provides convincing evidence that for a given BMI in adulthood, being small at birth and thin at age 10 years can increase the risk of adult-onset diabetes. However, the mechanisms that underlie these associations are still unclear. As mentioned earlier, according to the thrifty phenotype hypothesis [139, 140], adult BMI would be a possible mediator of the association between low birthweight and adult diabetes. However, the increased risks of adult-onset diabetes associated with low birthweight were consistent and strong among women who were and were not overweight and obese, suggesting that such association is unlikely to be mediated by adult BMI.

Although there is some evidence from animal studies on links between being small in childhood and poor development of endocrine pancreas [269, 270], more basic research on thinness in early life may provide more understanding on such relationships. It is also essential to explore whether the risks of diabetes associated with being small at birth and at age 10 years share the same underlying biological mechanisms.

10.4.4 Public health implications

The results of this thesis will provide a better understanding of healthy growth and development in early life and their roles in aetiology of diabetes in later life. This thesis also provides a stronger literature base to inform relevant public health policies and practices, particularly to characterise the groups of individuals who should be targeted for public health interventions to maximise protection against diabetes. There are higher risks of diabetes in later life for those who are small at

birth and thin at age 10 years, and therefore, as expected, their risks of developing diabetes are even more if they develop overweight or obesity as adults.

For assessing public health implications of body size at different stages, it is important to remember that the effect size for the association between adult BMI and diabetes is multi-fold higher than those for body size at other ages. Findings from this thesis suggest that larger body size in early life was mostly relevant to adult-onset diabetes risk as a predictor of adult adiposity. However, given the fact that it is challenging to reduce overweight and obesity once it is established [8], children and adolescents must be encouraged to maintain optimal growth in early years of life to minimise the long-term risk of diabetes.

10.4.5 Insights on diabetes recording in hospital admissions data

In order to understand the possible limitations of identifying diabetes cases from hospital admissions record, I compared diabetes recording between primary care and hospital admissions, overall and then by various factors. Results from such comparisons will not only help to understand the validity of findings in this thesis but also inform other researchers who work with diabetes in routinely-collected healthcare data about possible issues and their likely effects epidemiological investigations.

10.5 Recommendations for future research

Although the investigations in this thesis could greatly expand our understanding of the roles of body size on adult-onset diabetes risk, several issues can be improved or extended in future research. Recommendations for future research on this topic can be broadly grouped into two categories: those related to methodological limitations of this thesis, and those related to questions that could not be answered in this thesis.

The major limitations in this research on body size in early life and risk of diabetes in adulthood are the use of reported body size information and identifying new diabetes cases from hospital admissions data. Therefore, priorities for future research would be to use measured body size indices in a large cohort with long-term

prospective follow-up in primary care data for incident diabetes. Several record-linkage studies [143–145, 267] so far have measured data on birthweight and early life BMI, but BMI at a particular age does not capture the dynamic processes of growth in early life [8], and BMI can reflect either adiposity or muscularity [146]. Indices of body fat percentage and distribution can be more accurate markers of adiposity, but getting such measures in large population studies is quite challenging. Like the Million Women Study, most of the current record-linkage studies [143–145, 267] are linked to national hospital admissions registers to identify diabetes, but the linkage to primary care and prescriptions registers can be more accurate. Moreover, record-linkage studies often do not have information on important factors that might be related to both body size and diabetes risk; therefore analyses cannot take those factors into account. This information can be obtained from other data sources or by re-contacting participants.

Since standard statistical adjustment for adult BMI could be problematic, advanced statistical techniques like structural equation modelling, multi-level modelling or analysis of different combinations of body size over the life course should be done [8]. Randomised studies may not be feasible for many issues, but Mendelian Randomisation study can be informative to understand the associations.

The underlying biological pathways explaining the associations observed in this thesis are mostly unknown. Understanding the relationships of body size at different ages with intermediary bio-markers, for example, beta-cell development, insulin secretion, insulin sensitivity are also needed to delineate potential biological pathways. Future research should also investigate whether body size at different ages over the life course has distinct effects of diabetes development or whether they are just manifestations of some immeasurable constructs that determine the risk of diabetes in adulthood. I envisage that the findings of this thesis will inform on-going research and also stimulate new lines of investigation in basic sciences and animal studies.

10.6 Conclusions

This large prospective study of UK women, with more than 24,000 cases of diabetes, demonstrates as well as the known increased risk with adult adiposity, there are clear, independent, and persistently increased risks of adult-onset diabetes with decreasing birthweight, and separately with being comparatively thin at age 10 years. Allowing for these associations and the adverse effect of adiposity in middle-age, body size at age 20 appears to have little, if any, independent effect on adult-onset diabetes risk.

Appendices

A

Systematic review

A.1 Association between birthweight and risk of diabetes in adulthood

A.1.1 Search terms

Table A.1: Search terms for studies investigating the association between birthweight and risk of diabetes in adulthood

	Search terms
EMBASE	
Search interface: Ovid	
Embase 1980 to 2018 week 14	
1	birth weight.ti,ab.
2	size at birth.ti,ab.
3	weight at birth.ti,ab.
4	birth size.ti,ab.
5	1 or 2 or 3 or 4
6	diabetes.ti,ab.
7	diabetes mellitus.ti,ab.
8	type 2 diabetes.ti,ab.
9	non-insulin dependent diabetes.ti,ab.
10	niddm.ti,ab.
11	adult onset diabetes.ti,ab.
12	6 or 7 or 8 or 9 or 10 or 11
13	5 and 12
PubMed	
Search interface: NLM	
Embase 1980 to 2018 week 14	
1	birth weight [tiab]
2	size at birth [tiab]
3	weight at birth [tiab]
4	birth weight [tiab]
5	((("birth weight" [tiab]) OR "size at birth" [tiab]) OR "weight at birth" [tiab]) OR "birth size" [tiab]
6	diabetes mellitus [tiab]
7	adult onset diabetes [tiab]
8	type 2 diabetes [tiab]
9	non-insulin dependent diabetes [tiab]
10	niddm [tiab]
11	((((("diabetes mellitus" [tiab]) OR "adult onset diabetes" [tiab]) OR "type 2 diabetes" [tiab]) OR "non-insulin dependent diabetes" [tiab]) OR "niddm" [tiab]
12	(((((("birth weight" [tiab]) OR "size at birth" [tiab]) OR "weight at birth" [tiab]) OR "birth size" [tiab])) AND (((((("diabetes mellitus" [tiab]) OR "adult onset diabetes" [tiab]) OR "type 2 diabetes" [tiab]) OR "non-insulin dependent diabetes" [tiab]) OR "niddm" [tiab])

Table A.2: Summary information of all included prospective studies investigating the association between birthweight and risk of diabetes in adulthood

Year	Study, country [Reference]	Study population at baseline	Follow-up	Assessment of birthweight	Assessment of diabetes
2018	Shanghai Women's Health Study and the Shanghai Men's Health Study, China [134]	11,515 men and 13,569 women aged 40–70 years were recruited in 1996–2000 and in 2002-06, respectively	Follow-up surveys were conducted every 2–3 years since recruitment	Birthweight was self-reported at recruitment	1027 incident cases of T2DM were identified by self-reporting
2018	National Child Development Study, UK [279]	15,043 men and women born during the week of March 9, 1958 were included	Followed-up in 2000 (aged 42 years)	Birthweight was extracted from medical records	341 incident diabetes cases were identified by self-reporting
2017	Saskatchewan cohort of First Nations, Canada [136]	2003 women born in 1950-1984 were included	Followed-up from 1980 to 2009	Birthweight was collected from birth registration information	384 diabetes cases were identified using health administrative data
2015	Copenhagen School Health Records Register, Denmark [129]	113,801 men and 109,298 women born in 1936-1983 were included in 1977 or at age 30 years (median age 30 years)	Followed-up in 1977-2013 (mean age 59 years)	Birthweight was reported by mothers at school entry (age 6-7 years) in school health records	7750 men and 4736 women developed adult T2DM (aged $\geq$30 years) identified from hospital admission data
2015	Health Professional Follow-up Study, USA [131]	18,305 men were recruited in 1986, mean age 51.3 years	Followed-up in 1986-2010	Birthweight was self-reported on the 1994 follow-up questionnaire	1603 T2DM cases were identified by self-reporting and confirmed by validated supplementary questionnaire

Table A.2 continued: Information of prospective studies investigating the association between birthweight and adult-onset diabetes

Year	Study, Country [Reference]	Study population at baseline	Follow-up	Assessment of birthweight	Assessment of diabetes
2015	Nurses' Health Study, USA [131]	49,757 women were recruited in 1980, mean age 45.2 years	Followed-up in 1980-2010	Birthweight was self-reported on the 1992 follow-up questionnaire	5381 T2DM cases were identified by self-reporting and confirmed by validated supplementary questionnaire
2015	Nurses' Health Study II, USA [131]	81,732 women were recruited in 1991, mean age 36.0 years	Followed-up in 1991-2011	Birthweight was self-reported on the 1991 follow-up questionnaire	4725 T2DM cases were identified by self-reporting and confirmed by validated supplementary questionnaire
2014	Swedish Medical Birth Register [280], Sweden	759,999 men and women born during 1973-1982 were included	Followed-up from 2004 to 2010 (aged 27.5-37.5 years)	Birthweight was obtained from Swedish Medical Birth Register	2092 T2DM cases were identified from Swedish Medical Birth register
2014	Black Women's Health Study, USA [132]	21,624 black women aged 21-69 years were included in 1995	Followed-up between 1997 and 2013 by biennial questionnaire.	Birthweight was self-reported in 1997 follow-up questionnaire and validated	2388 self-reported diabetes cases were identified
2013	AGES-Reykjavik Study, Iceland [281]	1682 men and women born in the greater Reykjavik area in 1914-1935 and participated in Reykjavik study in 1967 were included	Followed-up at mean age 75.5 years	Birthweight was recorded in hospital data	249 T2DM cases were identified based on plasma sugar level at follow-up study

Table A.2 continued: Information of prospective studies investigating the association between birthweight and adult-onset diabetes

Year	Study, Country [Reference]	Study population at baseline	Follow-up	Assessment of birthweight	Assessment of diabetes
2010	E3N cohort, France [133]	91,453 women born in 1925-1950 and recruited in 1990 (aged 40-65 years)	Followed-up in 1990-2005. Mean years of follow-up 15.1 years	Birthweight was self-reported at study recruitment	2534 diabetes cases were identified by self-reporting or drug reimbursement record linkage
2009	Stockholm Birth Cohort 1925-49, Sweden [130]	6425 men and women born in 1925-1949 in Stockholm and lived in Sweden after 1987 were included	Followed-up from 1987 to 2006 (aged 38 – 81 years)	Birthweight was recorded in birth records	508 diabetes cases were identified in hospital registers
2006	Aberdeen Birth Cohort 1950-56, Scotland [282]	5793 men and women born between 1950 and 1956 in Aberdeen, Scotland were included	Followed-up in 2000 (aged 44-50 years)	Birthweight was recorded in hospital data	113 T2DM cases were identified by self-reporting
2005	British Birth Cohort 1946, UK [283]	3934 men and women born in March 1946 in the UK (nationally representative)	Followed up between 1977 and 1999 (aged 31-53 years)	Birthweight was extracted from medical records	78 self-reported T2DM cases were identified
2004	Gothenburg 1913 Birth Cohort, Sweden [142]	478 men born in 1913 were recruited in 1963 at the age of 50 years	Followed-up between 1963 and 1993	Birthweight was obtained from hospital deliveries and midwives' record books	54 diabetes cases were identified from questionnaire and checked with medical record
2003	British Birth cohort 1958, UK [14]	10,683 men and women born in 3-9 March 1958 were included	Followed up by a survey questionnaire at 41 years of age (1999-2000)	Birthweight was obtained from midwives' record	88 self-reported cases of T2DM

Table A.2 continued: Information of prospective studies investigating the association between birthweight and adult-onset diabetes

Year	Study, Country [Reference]	Study population at baseline	Follow-up	Assessment of birthweight	Assessment of diabetes
2003	Helsinki Birth Cohort Study 1934-44, Finland [267]	8760 men and women born in 1934-1944 in Helsinki and lived in Finland in 1971 were included	Followed-up between 1964 and 1998 (aged ≥40 years)	Birthweight was obtained from birth record	290 T2DM cases were identified in the national medication register
2002	Helsinki Birth Cohort Study 1934-44, Finland [143]	8760 men and women born in 1934-1944 in Helsinki and lived in Finland in 1971 were included	Followed-up between 1964 and 1998 (aged ≥40 years)	Birthweight was obtained from birth records	290 T2DM cases were identified in the national medication register
2002	Helsinki Birth Cohort Study 1924-44, Finland [141]	13,517 men and women born in 1924-44 in Helsinki and lived in Finland in 1971 were included	Followed-up from 1964 to 1997 in medication register and from 1971 to 1997 in hospital admissions register	Birthweight was obtained from birth records	698 T2DM cases were identified in the medication and hospital admissions registers
2000	Helsinki Birth Cohort Study 1924-33, Finland [268]	7086 men and women born in 1924-1933 in Helsinki and lived in Finland in 1971	Followed-up between 1964 and 1997 (aged 64-73 years)	Birthweight was obtained from birth records	471 T2DM cases were identified in national medication register
2000	Peking Birth Cohort 1948-54, China [284]	627 men and women born in 1948-1954 were included	Followed-up in 1995-1996 (mean age 45 years)	Birthweight was recorded in hospital data	28 T2DM cases were identified by OGTT at the follow-up

Table A.2 continued: Information of prospective studies investigating the association between birthweight and adult-onset diabetes

Year	Study, Country [Reference]	Study population at baseline	Follow-up	Assessment of birthweight	Assessment of diabetes
1999	Nurses' Health Study, USA [12]	69,526 women born between 1921 and 1946 were recruited in 1976	Followed-up from 1976 to 1992 (mean age 58 years)	Birthweight was self-reported on the 1992 follow-up questionnaire	2123 T2DM cases were identified by self-reporting and confirmed by validated supplementary questionnaire
1999	Pieksamaki Birth Cohort 1947-57, Finland [285]	374 men and women born in 1947-1957 in Pieksamaki, Finland and participated in the follow-up study (mean age 48 years)	Followed-up at mean age 48 years	Birthweight was recorded in hospital data	44 T2DM cases were identified by OGTT in the follow-up study
1998	Mysore Birth Cohort 1934-53, India [286]	506 men and women born in 1934-1953 were included	Followed-up at the mean age 47 years	Birthweight was recorded in hospital data	79 T2DM cases were identified by OGTT in the follow-up study
1998	Sheffield Birth Cohort 1939-40, England [287]	337 men and women born in 1939-1940 in Sheffield were included	Interviewed and examined at mean age of 52 years	Birthweight was recorded in hospital data	18 T2DM cases were identified based on OGTT (mean age 52 years)
1998	Rancho Bernardo study, USA [288]	303 postmenopausal women born in 1907-1941 and completed the survey in 1972-1974	Followed-up in 1984-1987, aged 50-84 years (mean age 67 years)	Birthweight was self-reported on the 1984-87 follow-up questionnaire	30 T2DM cases were identified by OGTT
1996	Uppsala Birth Cohort, Sweden [289]	1093 men born in 1920-1924 in Uppsala were included	Followed-up in 1970-1973 (aged 50 – 60 years)	Birthweight was extracted from medical records	61 T2DM cases were identified by OGTT

Table A.2 continued: Information of prospective studies investigating the association between birthweight and adult-onset diabetes

Year	Study, Country [Reference]	Study population at baseline	Follow-up	Assessment of birthweight	Assessment of diabetes
1994	Pima Indian community study, USA [135]	1179 men and women born in 1940-1972 in Gila River Indian community, Arizona and had at least one OGTT were included	Followed-up in ages 20-39 years	Birthweight was extracted from medical records	210 T2DM were identified by OGTT test
1991	Hertfordshire Birth Cohort 1920-30, England [290]	765 men born in 1920-30 in Hertfordshire who attended follow-up study were included	Followed up between age of 59-70 years (mean age 64 years)	Birthweight was obtained from health visitors' records	48 diabetes cases were identified by OGTT

Abbreviations: FBS: fasting blood sugar, IGT: impaired glucose tolerance, OGTT: oral glucose tolerance test, T2DM: type 2 diabetes mellitus

A.2 Meta-analysis on the association between birthweight and risk of adult-onset diabetes

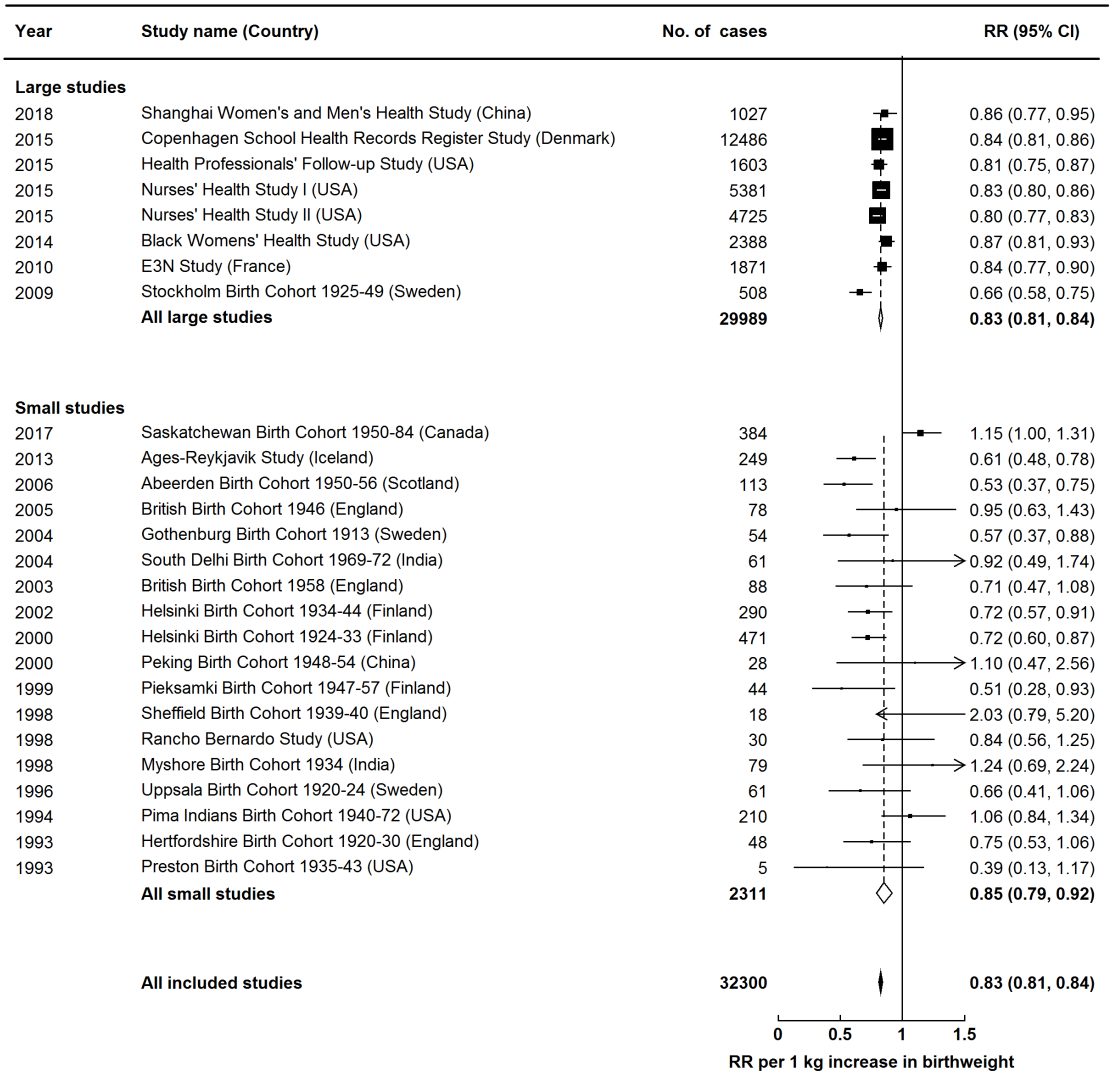


Figure A.1: Meta-analysis of the association between birthweight and risk of adult-onset diabetes in all published prospective cohort studies

Study-specific linear trends (per kg increase) for birthweight were extracted from the published reports or estimated using the generalized least square method proposed by Orsini [126]. Two-stage fixed-effects dose-response models were used for combining results. (see subsection 3.2.1 for more details)

Table A.3: Relative risks (RRs) of adult-onset diabetes per 1 kg increase in birthweight, meta-analysis results in subgroups by various factors

	RR per 1 kg increase in birthweight
Sex	
Men	0.80 (0.75- 0.86)
Women	0.83 (0.81- 0.85)
Birthweight information	
Self-reported	0.82 (0.80- 0.84)
From birth records	0.83 (0.81- 0.85)
Region	
North America	0.83 (0.81- 0.86)
Europe	0.81 (0.79- 0.83)
Others	0.87 (0.79- 0.96)
Mean age at follow-up	
Less than 50 years	0.83 (0.81- 0.86)
50 years or more	0.82 (0.80- 0.84)

A.3 Associations of body size in childhood and early adulthood with adult-onset diabetes risk

A.3.1 Search terms

Table A.4: Search terms in EMBASE for studies investigating the association between body size in early life and risk of diabetes in adulthood

	Search terms
EMBASE	
Search interface: Ovid	
Embase 1980 to 2018 week 14	
1	childhood obesity.ti,ab.
2	childhood overweight.ti,ab.
3	childhood adiposity.ti,ab.
4	childhood body mass index.ti,ab.
5	childhood bmi.ti,ab.
6	childhood body size.ti,ab.
7	adolescen* obesity.ti,ab.
8	adolescen* overweight.ti,ab.
9	adolescen* adiposity.ti,ab.
10	adolescen* body mass index.ti,ab.
11	adolescen* bmi.ti,ab.
12	adolescen* body size.ti,ab.
13	early life body size.ti,ab.
14	early life weight.ti,ab.
15	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14
16	diabetes.ti,ab.
17	diabetes mellitus.ti,ab.
18	type 2 diabetes.ti,ab.
19	non-insulin dependent diabetes.ti,ab.
20	niddm.ti,ab.
21	adult onset diabetes.ti,ab.
22	16 or 17 or 18 or 19 or 20 or 21
23	15 and 22

Table A.5: Search terms in PubMed for studies investigating the association between body size in early life and risk of diabetes in adulthood

	Search terms
PubMed	
Search interface: NLM	
Embase 1980 to 2018 week 14	
1	"childhood obesity" [tiab]
2	"childhood overweight" [tiab]
3	"childhood adiposity" [tiab]
4	"childhood body mass index" [tiab]
5	"childhood bmi" [tiab]
6	"childhood body size" [tiab]
7	"adolescen* obesity" [tiab]
8	"adolescen* overweight" [tiab]
9	"adolescen* adiposity" [tiab]
10	"adolescen* body mass index" [tiab]
11	"adolescen* bmi" [tiab]
12	"adolescen* body size" [tiab]
13	"early life body size" [tiab]
14	"early life weight" [tiab]
15	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14
16	diabetes [tiab]
17	diabetes mellitus [tiab]
18	type 2 diabetes [tiab]
19	non-insulin dependent diabetes [tiab]
20	niddm [tiab]
21	adult onset diabetes [tiab]
22	16 or 17 or 18 or 19 or 20 or 21
23	15 and 22

Table A.6: Summary information of all included prospective studies investigating the association between body size in childhood and risk of diabetes in adulthood

Year	Study, country [Reference]	Study population at baseline	Follow-up	Assessment of body size	Assessment of diabetes
2018	Copenhagen School Health Records Register, Denmark [144]	62,565 men born between 1930 and 1989 and had linked data in the Danish Conscription Database were included	Followed-up from either 1977 or age 30 years to 2015	BMI was measured at age 7 and 13 years and in early adulthood (17 to 26 years)	6710 T2DM cases were identified from hospital admissions data
2017	Copenhagen School Health Records Register, Denmark [145]	128,843 women and 163,983 men born between 1930 and 1989 were included	Followed up from either 1977 or 30 years of age, whichever came later until 2013	BMI was measured at age 7, 10, and 13 years by school doctors or nurses	11,548 in men and 7,472 in women T2DM cases were identified from hospital admissions record
2015	Young Finns Study, Finland [291]	2478 children and adolescents (aged 3-18 years & female 55%) were randomly selected from population register in 1980	Followed-up between 2001 and 2011 (aged 24-49 years)	BMI was measured at 3-6 years and 9-18 years	84 T2DM cases were identified (mean age 39.6 years) based on FBS or HbA1C
2015	Kaunas Cardiovascular Risk cohort, Lithuania [292]	506 random school-children were included in the baseline survey (1977; aged 12-13 years)	Followed up in 2012 (age 48-49 years)	BMI was measured at the baseline (age 12-13 years)	15 T2DM cases were identified by self-reporting
2014	Young Finns Study, Finland [293]	1617 randomly selected children aged 9-24 years in 1986 were included	Followed-up in 2007 (aged 30-45 years)	BMI was measured at the time of recruitment (age 9-24 years)	43 T2DM cases were identified in 2007 using blood glucose measurement

Table A.6 continued: Information of studies on associations between body size in childhood and adult-onset diabetes

Year	Study, country [Reference]	Study population at baseline	Follow-up	Assessment of body size	Assessment of diabetes
2011	1958 British Birth Cohort, UK [294]	7,855 participants born in one week in March 1958 in England, Scotland and Wales were included	Followed-up at 41 years of age (in 1999-2000)	BMI was measured at age 7, 11 and 16 years and self-reported at age 23 years	88 self-reported cases of T2DM
2011	Bogalusa, Muscatine, Childhood Determinants of Adult Health, and Young Finns Study; USA and Finland [146]	6328 participants (women 53%) mean age at baseline 11.4 years were included	Followed-up for mean duration of 23 years (mean age 34.5 years)	BMI was measured at the baseline	129 T2DM cases were identified through FBS
2010	E3N cohort, France [133]	91,453 women born in 1925-1950 and recruited in 1990 (aged 40-65 years)	Followed-up in 1990-2005. Mean years of follow-up 15.1 years	Recalled body silhouettes (1-8 scale) for age 8 years and for menarche age; adult BMI was self-reported	2534 diabetes cases were identified by self-reporting or drug reimbursement record linkage
2010	Nurses' Health Study II, USA [13]	109,172 women aged 25-42 years (mean age: 34.3 years) were recruited in 1989	Followed-up over 16 years between 1989-2005	Recalled somatotypes by 9-level pictogram for age 5, 10 and 20 years, and validated against measured BMI	3307 T2DM cases were self-reported in biennial questionnaires and validated by supplementary questionnaire
2010	Princeton Follow-up study, USA [295]	822 school children aged 6-18 years (mean age 12.4 years) at baseline (1973-1978) were included	Followed-up for median 26 years (in 1998-2003), mean age 38.6 years	BMI was obtained at baseline	38 T2DM cases were identified based on FBS

Table A.6 continued: Information of studies on associations between body size in childhood and adult-onset diabetes

Year	Study, country [Reference]	Study population at baseline	Follow-up	Assessment of body size	Assessment of diabetes
2009	New Delhi Birth Cohort 1969-72, India [296]	1492 men and women (42%) born during 1969-72 were included	Followed-up in 1998-2002 (aged 26-32 years)	BMI was measured every 3-6 months between age 2-14 years	219 IGT or T2DM cases were identified by OGTT
2004	New Delhi Birth Cohort 1969-72, India [297]	1364 men and women born in 1969-1972 in South Delhi were included	Followed-up between 1998 and 2002; aged 26 to 32 years (mean age 29 years)	Birthweight was recorded within 72 hours after birth by trained personnel	61 T2DM cases were identified by OGTT at follow-up
2003	British Birth cohort 1958, UK [14]	10,683 men and women born in 3-9 March 1958 were included	Followed up by a survey questionnaire at 41 years of age (1999-2000)	Birthweight was obtained from midwives' record	88 self-reported cases of T2DM
2003	Helsinki Birth Cohort Study 1934-44, Finland [267]	8760 men and women born in 1934-1944 in Helsinki and lived in Finland in 1971 were included	Followed-up between 1964 and 1998 (aged $\geq$40 years)	BMI was measured 18 times between birth and 12 years of age	290 T2DM cases were identified in the national medication register
2002	Helsinki Birth Cohort Study 1924-44, Finland [141]	13,517 men and women born in 1924-44 in Helsinki and lived in Finland in 1971 were included	Followed-up from 1964 to 1997 in medication register and from 1971 to 1997 in hospital admissions register	BMI was measured 18 times between birth and 12 years of age	698 T2DM cases were identified in the medication and hospital admissions registers
2000	Helsinki Birth Cohort Study 1924-33, Finland [268]	7086 men and women born in 1924-1933 in Helsinki and lived in Finland in 1971	Followed-up between 1964 and 1997 (aged 64-73 years)	BMIs between age 6 and 16 years were taken from school health records	471 T2DM cases were identified in national medication register

Table A.6 continued: Information of studies on associations between body size in childhood and adult-onset diabetes

Year	Study, country [Reference]	Study population at baseline	Follow-up	Assessment of body size	Assessment of diabetes
1992	Third Harvard Growth Study, USA [298]	508 adolescents (aged 13 to 18 years; female 49.6%) were recruited between 1922 and 1935	Followed-up in 1988 (mean age 73 years)	BMI was measured at the baseline	32 diabetes cases were identified by self-reporting

Abbreviations: BMI: body mass index, FBS: fasting blood sugar, HbA1C: glycated haemoglobin,

IGT: impaired glucose tolerance, OGTT: oral glucose tolerance test,

T2DM: type 2 diabetes mellitus

Table A.7: Summary information of all included prospective studies investigating the association of body size in adolescence and in early adulthood with risk of diabetes in adulthood

Year	Study, country [Reference]	Study population at baseline	Follow-up	Assessment of body size	Assessment of diabetes
2018	Copenhagen School Health Records Register, Denmark [144]	62,565 men born between 1930 and 1989 and had linked data in the Danish Conscription Database were included	Followed-up from either 1977 or age 30 years to 2015	BMI was measured at age 7 and 13 years and in early adulthood (17 to 26 years)	6710 T2DM cases were identified from hospital admissions data
2018	China Kadoorie Biobank, China [16]	482,589 men and women aged 30-79 years were recruited during 2004-08	Mean during of follow-up was 9 years	BMI at age 25 years was calculated from self-reported weight at age 25 and measured height at study baseline	13,416 incident cases of diabetes were identified from bespoke disease surveillance system and health insurance databases
2015	Young Finns Study, Finland [291]	2478 children and adolescents (aged 3-18 years & female 55%) were randomly selected from population register in 1980	Followed-up between 2001 and 2011 (aged 24-49 years)	BMI was measured at 3-6 years and 9-18 years	84 T2DM cases were identified (mean age 39.6 years) based on FBS or HbA1C
2015	British Regional Heart study, UK [147]	4846 men born between 1919 and 1939 were recruited in 1978-80 (aged 40-59 years)	Followed-up until 2010	Measured BMI at age 21 years was taken from either Military record or recalled in 1996 questionnaire	627 diabetes cases were identified from medical records

Table A.7 continued: Information of studies on body size in early adulthood and adult-onset diabetes

Year	Study, country [Reference]	Study population at baseline	Follow-up	Assessment of body size	Assessment of diabetes
2014	Health Professional Follow-up Study, USA [15]	39,909 male participants (aged 40-75 years) were recruited in 1986	Followed-up for 22 years (from 1986-2008).	BMI was calculated from self-reported recall of weight at 21 years and height reported in 1986 questionnaire.	2970 diabetes cases were self-reported in biennial questionnaires and validated by supplementary questionnaire
2013	Fifth Military Conscription District Study, Denmark [299]	6502 men born in 1955 and eligible for conscription in Northern Denmark were included	Follow-up began at participants' 22^nd birthday in Danish National Registry of Patients and in Aarhus University Prescription Database between 1989-2010	BMI was measured at age 22 years	316 incident cases of diabetes was identified
2011	1958 British Birth Cohort, UK [294]	7,855 participants born in one week in March 1958 in England, Scotland and Wales were included	Followed-up at 41 years of age (in 1999-2000)	BMI was measured at age 7, 11 and 16 years and self-reported at age 23 years	88 self-reported cases of T2DM
2011	MELANY study, Israel [148]	37,674 healthy men from Israeli Army were included (mean age 17.4 years)	Followed-up at mean age of 30.6 years (mean duration 17.4 years)	BMI was measured at the baseline	1173 cases of T2DM were identified based on FBS
2010	E3N cohort, France [133]	91,453 women born in 1925-1950 and recruited in 1990 (aged 40-65 years)	Followed-up in 1990-2005. Mean years of follow-up 15.1 years	Recalled body silhouettes (1-8 scale) for age 8 years, for menarche age and for age 20-25 years; adult BMI was self-reported	2534 diabetes cases were identified by self-reporting or drug reimbursement record linkage

Table A.7 continued: Information of studies on body size in early adulthood and adult-onset diabetes

Year	Study, country [Reference]	Study population at baseline	Follow-up	Assessment of body size	Assessment of diabetes
2010	Nurses' Health Study II, USA [13]	109,172 women aged 25-42 years (mean age: 34.3 years) were recruited in 1989	Followed-up over 16 years between 1989-2005	Recalled somatotypes by 9-level pictogram for age 5, 10 and 20 years, and validated against measured BMI	3307 T2DM cases were self-reported in biennial questionnaires and validated by supplementary questionnaire
2006	British Women's Heart and Health study, UK [300]	4286 women aged 60-79 years (median age 69 years) were randomly selected from GP practitioners in 23 British towns in 1999-2001	Followed-up by questionnaire after 3 years from baseline and reviewing of primary care medical record	BMI at age 21 was calculated from self-reported recalled weight and height at baseline (age 60-79 years).	314 adult-onset diabetes cases were identified either from primary care medical record or from self-reporting
2006	Glasgow Alumni study, UK [300]	5,571 men and women who attended University of Glasgow between 1948 and 1968 (median age 19 years)	Followed-up in NHS central register in 1998 and filled a postal questionnaire in 2001 (median age 64 years)	BMI was measured at median age of 19 years	207 adult onset diabetes cases were identified through self-reporting

Abbreviations: FBS: fasting blood sugar, IGT: impaired glucose tolerance, OGTT: oral glucose tolerance test, T2DM: type 2 diabetes mellitus

B

Relevant study information

B.1 Million Women Study 3-year re-survey questionnaire

This questionnaire was completed by study participants in 2000-2003. I have no role in developing the questionnaire.

THE MILLION WOMEN STUDY

Confidential National Study of Women's Health

The Million Women Study is an important national study of women's health. A few years ago you received the first questionnaire with your invitation to the National Health Service Breast Screening Programme. Your help is needed again. Can you find time to complete this second questionnaire? Some of the questions may seem familiar and others are new, but all will provide vital up to date information for the study. Your answers are valuable and important - the enclosed leaflet explains how the study will benefit women and improve medical knowledge world-wide. We very much hope you are still willing to be one of the Million Women in the study.

We guarantee that all information provided will be treated with absolute confidentiality and used for medical research only.

To help us read your answers please write as clearly as possible and complete the questionnaire as shown:

Please put a cross in the appropriate box(es) ☒ ☐

OR put numbers in the appropriate places

eg 54 *(age)* 0 2 | 0 8 | 2 0 0 0
2nd August 2000

Any questions? Ring us on Freephone 0800 262 872

QUESTIONS ABOUT YOU AND YOUR HEALTH

Please answer every question as best you can as all the information that you give us is very useful. If you are not sure about exact dates or ages an approximate answer is better than none. Please use a black pen, if possible.

1. In the last 5 years has a doctor told you that you have had any of the following conditions?

If YES please cross the box ☒ *and write your age when the condition was first diagnosed (eg* 57 *(age))*

Yes			Yes			Yes	
☐ High blood pressure	____ (age)		☐ Asthma	____ (age)		☐ Gallstones/gall bladder problems	____ (age)
☐ High blood cholesterol	____ (age)		☐ Osteoporosis	____ (age)		☐ Blood clot in leg	____ (age)
☐ Diabetes	____ (age)		☐ Thyroid problem	____ (age)		☐ Blood clot elsewhere (please describe below)	____ (age)
☐ Heart problem (please describe below)	____ (age)		☐ Breast cancer	____ (age)		☐ Inflammatory bowel disease (please describe below)	____ (age)
☐ Stroke/TIA (please describe below)	____ (age)		☐ Other cancer (please describe below)	____ (age)		☐ Any other serious illness (please describe below)	____ (age)

Please give as many details as possible about any illness mentioned above

2. Are you NOW being treated for: *If YES please cross the box* ☒ *and write your age when the condition was first treated (eg* 60 *(age))*

Yes			Yes			Yes	
☐ High blood pressure	____ (age)		☐ Asthma	____ (age)		☐ Osteoporosis	____ (age)
☐ Diabetes	____ (age)		☐ Rheumatoid arthritis	____ (age)		☐ Depression/anxiety	____ (age)
☐ Heart problem (please describe below)	____ (age)		☐ Osteoarthritis	____ (age)		☐ Other serious illness or disability (please describe below)	____ (age)

Please give further details of any serious illness or disability you are now being treated for

3. Have you had any broken/fractured bones in the last 5 years? *(please cross)* ☐ No ☐ Yes

If Yes, which bones were broken? ☐ wrist ☐ arm ☐ ankle ☐ hip ☐ spine ☐ other _____ *(please describe)*

If Yes, how did the fracture occur? ☐ after a fall ☐ in a car accident ☐ some other way _____ *(please describe)*

If Yes, when did it occur? month | year *(give month/year of the most recent fracture, if you have had more than one)*

4. Have you had any major operations in the last 5 years? ☐ No ☐ Yes - ***If Yes***, when? month | year

If Yes, please describe the operation and why it was done. (If you have had more than one operation please give the dates and details of each)

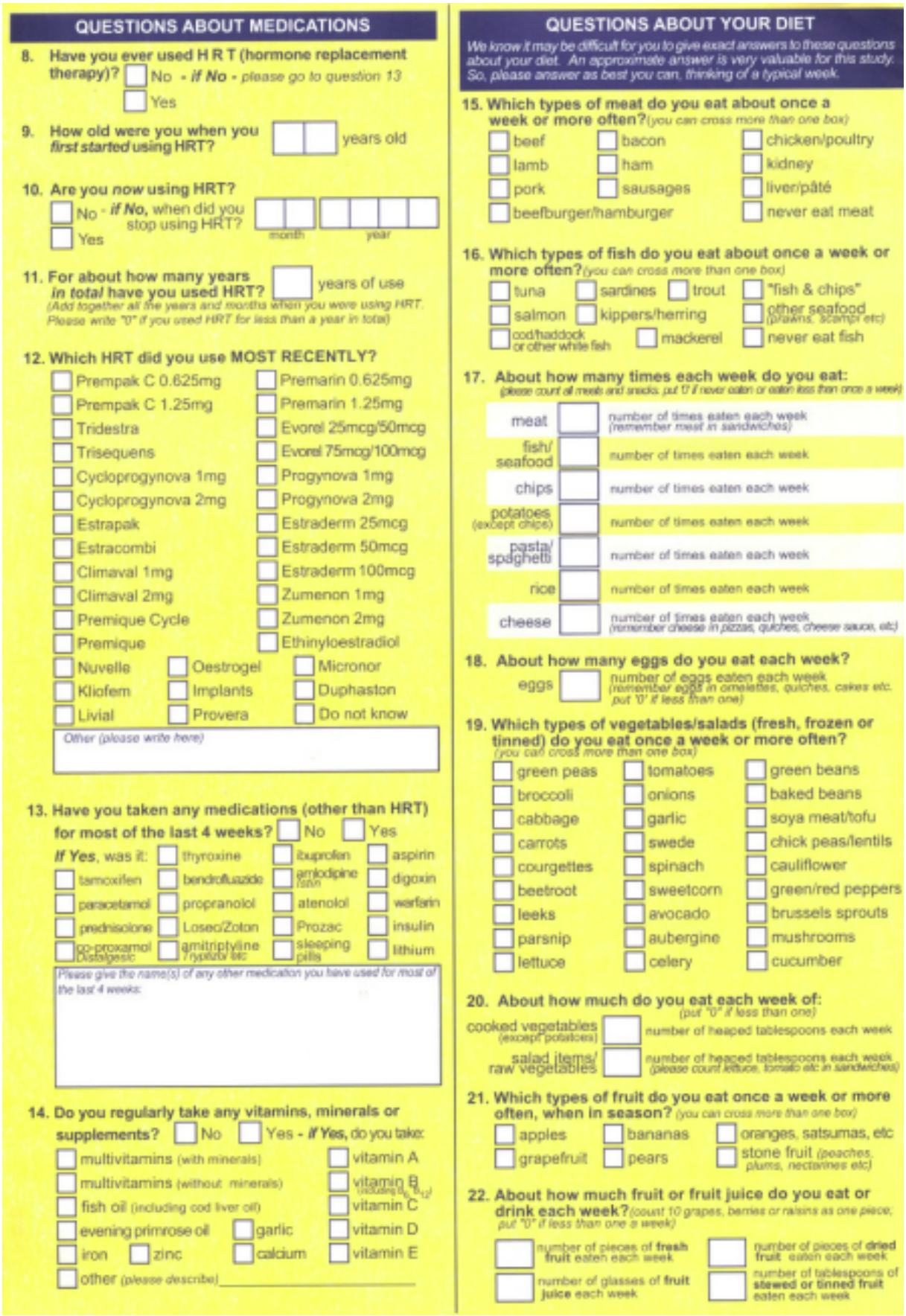
5. When did you last go for breast screening? month | year

6. Have you ever had a cervical smear test? ☐ No ☐ Yes - ***If Yes***, when was your last test? month | year

About how many cervical smear tests have you had in the last 10 years? ____ number of tests

Were you told that any of the cervical smear tests (in the last 10 years) were abnormal? ☐ No ☐ Yes

7. How would you describe your health now? ☐ excellent ☐ good ☐ fair ☐ poor



QUESTIONS ABOUT MEDICATIONS

8. **Have you ever used H R T (hormone replacement therapy)?** ☐ No - *if No - please go to question 13* ☐ Yes

9. **How old were you when you *first started* using HRT?** ☐☐ years old

10. **Are you *now* using HRT?** ☐ No - *if No,* when did you stop using HRT? ☐☐ month ☐☐☐☐ year ☐ Yes

11. **For about how many years *in total* have you used HRT?** ☐☐ years of use
(Add together all the years and months when you were using HRT. Please write "0" if you used HRT for less than a year in total)

12. **Which HRT did you use MOST RECENTLY?**

☐ Prempak C 0.625mg	☐ Premarin 0.625mg	
☐ Prempak C 1.25mg	☐ Premarin 1.25mg	
☐ Tridestra	☐ Evorel 25mcg/50mcg	
☐ Trisequens	☐ Evorel 75mcg/100mcg	
☐ Cycloprogynova 1mg	☐ Progynova 1mg	
☐ Cycloprogynova 2mg	☐ Progynova 2mg	
☐ Estrapak	☐ Estraderm 25mcg	
☐ Estracombi	☐ Estraderm 50mcg	
☐ Climaval 1mg	☐ Estraderm 100mcg	
☐ Climaval 2mg	☐ Zumenon 1mg	
☐ Premique Cycle	☐ Zumenon 2mg	
☐ Premique	☐ Ethinyloestradiol	
☐ Nuvelle	☐ Oestrogel	☐ Micronor
☐ Kliofem	☐ Implants	☐ Duphaston
☐ Livial	☐ Provera	☐ Do not know

Other (please write here)

13. **Have you taken any medications (other than HRT) for most of the last 4 weeks?** ☐ No ☐ Yes
If Yes, was it:

☐ thyroxine	☐ ibuprofen	☐ aspirin	
☐ tamoxifen	☐ bendrofluazide	☐ amlodipine (Istin)	☐ digoxin
☐ paracetamol	☐ propranolol	☐ atenolol	☐ warfarin
☐ prednisolone	☐ Losec/Zoton	☐ Prozac	☐ insulin
☐ co-proxamol Distalgesic	☐ amitriptyline Tryptizol etc	☐ sleeping pills	☐ lithium

Please give the name(s) of any other medication you have used for most of the last 4 weeks:

14. **Do you regularly take any vitamins, minerals or supplements?** ☐ No ☐ Yes - ***If Yes,*** do you take:

☐ multivitamins (with minerals)	☐ vitamin A	
☐ multivitamins (without minerals)	☐ vitamin B (including B_6, B_{12})	
☐ fish oil (including cod liver oil)	☐ vitamin C	
☐ evening primrose oil	☐ garlic	☐ vitamin D
☐ iron	☐ zinc	☐ calcium ☐ vitamin E

☐ other *(please describe)*_____

QUESTIONS ABOUT YOUR DIET

We know it may be difficult for you to give exact answers to these questions about your diet. An approximate answer is very valuable for this study. So, please answer as best you can, thinking of a typical week.

15. **Which types of meat do you eat about once a week or more often?** *(you can cross more than one box)*

☐ beef	☐ bacon	☐ chicken/poultry
☐ lamb	☐ ham	☐ kidney
☐ pork	☐ sausages	☐ liver/pâté
☐ beefburger/hamburger		☐ never eat meat

16. **Which types of fish do you eat about once a week or more often?** *(you can cross more than one box)*

☐ tuna	☐ sardines	☐ trout	☐ "fish & chips"
☐ salmon	☐ kippers/herring		☐ other seafood *(prawns, scampi etc)*
☐ cod/haddock or other white fish	☐ mackerel		☐ never eat fish

17. **About how many times each week do you eat:**
(please count all meals and snacks, put '0' if never eaten or eaten less than once a week)

meat	☐	number of times eaten each week *(remember meat in sandwiches)*
fish/ seafood	☐	number of times eaten each week
chips	☐	number of times eaten each week
potatoes (except chips)	☐	number of times eaten each week
pasta/ spaghetti	☐	number of times eaten each week
rice	☐	number of times eaten each week
cheese	☐	number of times eaten each week *(remember cheese in pizzas, quiches, cheese sauce, etc)*

18. **About how many eggs do you eat each week?**
eggs ☐ number of eggs eaten each week *(remember eggs in omelettes, quiches, cakes etc. put '0' if less than one)*

19. **Which types of vegetables/salads (fresh, frozen or tinned) do you eat once a week or more often?**
(you can cross more than one box)

☐ green peas	☐ tomatoes	☐ green beans
☐ broccoli	☐ onions	☐ baked beans
☐ cabbage	☐ garlic	☐ soya meat/tofu
☐ carrots	☐ swede	☐ chick peas/lentils
☐ courgettes	☐ spinach	☐ cauliflower
☐ beetroot	☐ sweetcorn	☐ green/red peppers
☐ leeks	☐ avocado	☐ brussels sprouts
☐ parsnip	☐ aubergine	☐ mushrooms
☐ lettuce	☐ celery	☐ cucumber

20. **About how much do you eat each week of:** *(put "0" if less than one)*
cooked vegetables (except potatoes) ☐ number of heaped tablespoons each week
salad items/ raw vegetables ☐ number of heaped tablespoons each week *(please count lettuce, tomato etc in sandwiches)*

21. **Which types of fruit do you eat once a week or more often, when in season?** *(you can cross more than one box)*

☐ apples	☐ bananas	☐ oranges, satsumas, etc
☐ grapefruit	☐ pears	☐ stone fruit *(peaches, plums, nectarines etc)*

22. **About how much fruit or fruit juice do you eat or drink each week?** *(count 10 grapes, berries or raisins as one piece; put "0" if less than one a week)*

☐ number of pieces of **fresh fruit** eaten each week	☐ number of pieces of **dried fruit** eaten each week
☐ number of glasses of **fruit juice** each week	☐ number of tablespoons of **stewed or tinned fruit** eaten each week

MORE ABOUT YOUR DIET

23. About how many of the following do you eat: *(put "0" if none or less than one)*

slices/pieces of white bread	slices each week
slices/pieces of brown/wholemeal bread *(also include granary, rye bread etc)*	slices each week
crackers, crispbread etc *(Ryvita, water biscuits etc)*	number each week
sweet biscuits	number each week
dairy desserts (yoghurts etc)	number each week
cakes, puddings, pies, buns etc	number each week
chocolate (in any food or drink)	approx. number of pieces each week
nuts (including peanut butter)	tablespoons each week
soup	bowls/cups each week
gravy, cream/cheese sauces etc	tablespoons each week
breakfast type cereal	bowls each week

If you eat breakfast cereal is it usually: *(please cross)*

- ☐ bran cereal *(allbran, branflakes etc)*
- ☐ biscuit cereal *(weetabix, shreddies etc)*
- ☐ oat cereal *(porridge, ready brek etc)*
- ☐ muesli
- ☐ other *(e.g. cornflakes, rice crispies etc)*

24. Which type of spread do you use on bread, crispbreads etc, once a week or more often? *(you can cross more than one box)*

- ☐ butter
- ☐ margarine
- ☐ soft cheese
- ☐ low fat spread
- ☐ mayonnaise
- ☐ salad cream
- ☐ olive oil spread
- ☐ marmite etc
- ☐ rarely use spread

Do you spread it: ☐ thick? ☐ medium? ☐ thin? *(please cross)*

Do you add butter etc to: ☐ potatoes? ☐ other vegetables?

25. Which types of fats or oils do you use for cooking or salad dressing once a week or more often? *(you can cross more than one box)*

- ☐ butter
- ☐ soft (tub) margarine
- ☐ white flora
- ☐ olive oil
- ☐ hard (block) margarine
- ☐ lard/dripping
- ☐ corn oil
- ☐ sunflower oil
- ☐ mayonnaise
- ☐ soya oil
- ☐ other vegetable oil
- ☐ salad cream

Please put a cross in the box if you RARELY OR NEVER:

☐ use fats or oils for cooking ☐ use salad dressing/cream

26. Please put a cross in the box if you NEVER eat:

- ☐ beef
- ☐ pork/ham
- ☐ lamb
- ☐ dairy products
- ☐ kidney
- ☐ liver/pâté
- ☐ sugar
- ☐ wheat products
- ☐ salami
- ☐ sausages
- ☐ eggs
- ☐ beefburgers

27. Which type of milk or cream do you drink or use once a week or more often? *(you can cross more than one box)*

milk: ☐ full cream ☐ semi-skimmed ☐ skimmed ☐ soya

cream: ☐ single ☐ double ☐ other

other: ☐ dairy ice cream ☐ never have milk/cream

28. Do you:

	never	sometimes	usually	always
add milk to your tea?	☐	☐	☐	☐
add milk to your coffee?	☐	☐	☐	☐
add salt to your food?	☐	☐	☐	☐
remove fat from meat? *(cross "never" if vegetarian)*	☐	☐	☐	☐
eat organic food?	☐	☐	☐	☐

29. Have you made any major changes to your diet in the last 5 years? ☐ No ☐ Yes- because of illness ☐ Yes- for some other reason

30. About how much alcohol do you drink each week?

☐ number of drinks of alcohol each week
one drink = a glass of wine, half pint of lager, or tot of spirits
(put "0" if you do not drink, or have less than one drink each week)

If you have more than one drink of alcohol each week:

is it usually with meals? ☐ No ☐ Yes ☐ it varies

on how many days each week do you usually drink? ☐ days each week

31. About how much do you drink EACH DAY of:

tea? *cups daily*	☐	milk? *(include hot chocolate etc) cups daily*	☐	fizzy/soft drink? *glasses daily*	☐
coffee? *cups daily*	☐	water? *glasses daily*	☐	fruit squash? *glasses daily*	☐

32. How many teaspoons of sugar do you add to tea, coffee, cereal, fruit etc EACH DAY? ☐ teaspoons of sugar each day

33. What size clothes do you wear now? *(you can cross more than one box if the size varies)*

Clothes	☐ 10 or less	☐ 12	☐ 14	☐ 16	☐ 18	☐ 20+
Bra	☐ 32	☐ 34	☐ 36	☐ 38	☐ 40	☐ 42+
Cup	☐ A/AA	☐ B	☐ C	☐ D	☐ DD/E +	

34. What is your: *(please put "0" if you do not know)*

waist measurement? ☐☐ inches hip measurement? ☐☐ inches

35. About how much do you weigh now? ☐ stone ☐ lbs *(Put "0" if you do not know)*

36. About how many hours each week do you spend doing:

	summer	winter	
housework? *(include cooking, cleaning etc)*		☐	hours per week
gardening?	☐	☐	hours per week
walking?	☐	☐	hours per week
cycling?	☐	☐	hours per week
any work or exercise causing sweating or a fast heartbeat?	☐	☐	hours per week

WHEN YOU WERE YOUNG

37. About how much did you weigh when you were born? ☐ lbs ☐ ozs *(Put "0" if you do not know)*

38. Were you breastfed when you were a baby?

☐ No ☐ Yes ☐ do not know

39. Did your parents smoke:

at around the time that you were born?

Mother	☐ No	☐ Yes	☐ do not know
Father	☐ No	☐ Yes	☐ do not know

at around the time that you were 10 years old?

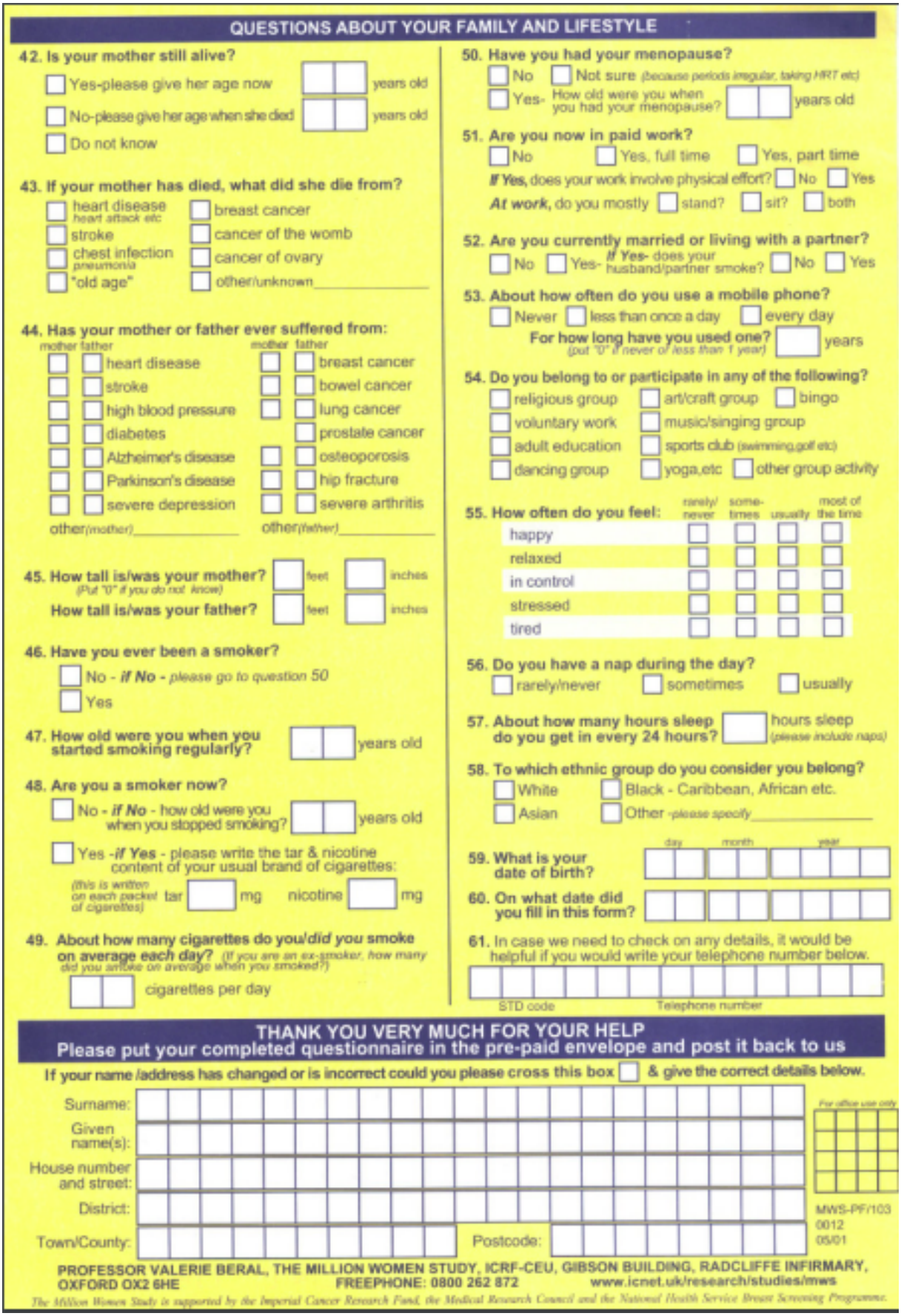
Mother	☐ No	☐ Yes	☐ do not know
Father	☐ No	☐ Yes	☐ do not know

40. When you were about 10 years old, compared to average, would you describe yourself as *(please cross)*:

☐ thinner? ☐ plumper? ☐ about average?

41. What size clothes did you wear when you were about 20 years old? *(you can cross more than one box)*

☐ 8 or less ☐ 10 ☐ 12 ☐ 14 ☐ 16 ☐ 18 +



QUESTIONS ABOUT YOUR FAMILY AND LIFESTYLE

42. Is your mother still alive?
- ☐ Yes-please give her age now ☐☐ years old
- ☐ No-please give her age when she died ☐☐ years old
- ☐ Do not know

43. If your mother has died, what did she die from?
- ☐ heart disease *heart attack etc*
- ☐ stroke
- ☐ chest infection *pneumonia*
- ☐ "old age"
- ☐ breast cancer
- ☐ cancer of the womb
- ☐ cancer of ovary
- ☐ other/unknown_____

44. Has your mother or father ever suffered from:

mother	father		mother	father	
☐	☐	heart disease	☐	☐	breast cancer
☐	☐	stroke	☐	☐	bowel cancer
☐	☐	high blood pressure	☐	☐	lung cancer
☐	☐	diabetes		☐	prostate cancer
☐	☐	Alzheimer's disease	☐	☐	osteoporosis
☐	☐	Parkinson's disease	☐	☐	hip fracture
☐	☐	severe depression	☐	☐	severe arthritis

other*(mother)*_____ other*(father)*_____

45. How tall is/was your mother? *(Put "0" if you do not know)* ☐ feet ☐ inches
How tall is/was your father? ☐ feet ☐ inches

46. Have you ever been a smoker?
- ☐ No - ***if No*** - *please go to question 50*
- ☐ Yes

47. How old were you when you started smoking regularly? ☐☐ years old

48. Are you a smoker now?
- ☐ No - ***if No*** - how old were you when you stopped smoking? ☐☐ years old
- ☐ Yes -***if Yes*** - please write the tar & nicotine content of your usual brand of cigarettes: *(this is written on each packet of cigarettes)* tar ☐ mg nicotine ☐ mg

49. About how many cigarettes do you/*did you* smoke on average *each day*? *(If you are an ex-smoker, how many did you smoke on average when you smoked?)*
☐☐ cigarettes per day

50. Have you had your menopause?
- ☐ No ☐ Not sure *(because periods irregular, taking HRT etc)*
- ☐ Yes- How old were you when you had your menopause? ☐☐ years old

51. Are you now in paid work?
- ☐ No ☐ Yes, full time ☐ Yes, part time
- ***If Yes,*** does your work involve physical effort? ☐ No ☐ Yes
- ***At work,*** do you mostly ☐ stand? ☐ sit? ☐ both

52. Are you currently married or living with a partner?
- ☐ No ☐ Yes- ***If Yes-*** does your husband/partner smoke? ☐ No ☐ Yes

53. About how often do you use a mobile phone?
- ☐ Never ☐ less than once a day ☐ every day
- **For how long have you used one?** *(put "0" if never or less than 1 year)* ☐ years

54. Do you belong to or participate in any of the following?
- ☐ religious group ☐ art/craft group ☐ bingo
- ☐ voluntary work ☐ music/singing group
- ☐ adult education ☐ sports club *(swimming,golf etc)*
- ☐ dancing group ☐ yoga,etc ☐ other group activity

55. How often do you feel:

	rarely/ never	some-times	usually	most of the time
happy	☐	☐	☐	☐
relaxed	☐	☐	☐	☐
in control	☐	☐	☐	☐
stressed	☐	☐	☐	☐
tired	☐	☐	☐	☐

56. Do you have a nap during the day?
☐ rarely/never ☐ sometimes ☐ usually

57. About how many hours sleep do you get in every 24 hours? ☐ hours sleep *(please include naps)*

58. To which ethnic group do you consider you belong?
- ☐ White ☐ Black - Caribbean, African etc.
- ☐ Asian ☐ Other -*please specify*_____

59. What is your date of birth? day ☐☐ month ☐☐ year ☐☐☐☐

60. On what date did you fill in this form? ☐☐ ☐☐ ☐☐☐☐

61. In case we need to check on any details, it would be helpful if you would write your telephone number below.
STD code ☐☐☐☐ Telephone number ☐☐☐☐☐☐☐

THANK YOU VERY MUCH FOR YOUR HELP
Please put your completed questionnaire in the pre-paid envelope and post it back to us

If your name /address has changed or is incorrect could you please cross this box ☐ & give the correct details below.

Surname: ☐☐☐☐☐☐☐

B.2 Letter confirming ethical approval for use of MWS-CPRD linked data

As the Chief Investigator of this data-linkage, Professor Jane Green submitted the application and did all correspondences with the CPRD team. I helped her in preparing the research protocol for this application.

ISAC EVALUATION OF PROTOCOLS FOR RESEARCH INVOLVING CPRD DATA

FEEDBACK TO APPLICANTS

CONFIDENTIAL			***by e-mail***
PROTOCOL NO:	12_070R3A4		
PROTOCOL TITLE:	Comparison of information on prescribing of hormone therapy and oral bisphosphonates, and on diagnosis of dementia and Parkinson's disease, in the Million Women Study cohort and CPRD-GOLD: a record linkage study		
APPLICANT:	Professor Jane Green, co-director, Cancer Epidemiology Unit, University of Oxford and co-PI, Million Women Study: jane.green@ceu.ox.ac.uk		
APPROVED ☒	**APPROVED WITH COMMENTS (resubmission not required)** ☐	**REVISION/ RESUBMISSION REQUESTED** ☐	**REJECTED** ☐
INSTRUCTIONS: ***Protocols with an outcome of 'Approved' or 'Approved with comments' do not require resubmission to the ISAC.*** **REVIEWER COMMENTS:** **APPLICANT FEEDBACK:**			
DATE OF ISAC FEEDBACK:	11/06/18		
DATE OF APPLICANT FEEDBACK:			

For protocols approved from 01 April 2014 onwards, applicants are required to include the ISAC protocol in their journal submission with a statement in the manuscript indicating that it had been approved by the ISAC (with the reference number) and made available to the journal reviewers. If the protocol was subject to any amendments, the last amended version should be the one submitted.

***** Please refer to the ISAC advice about protocol amendments provided below*****

References

[1] International Diabetes Federation. *IDF Diabetes Atlas, 9th edn.* `http://www.diabetesatlas.org`. Online; accessed 06 January 2020. 2019.

[2] NCD Risk Factor Collaboration et al. "Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants". In: *The Lancet* 387.10027 (2016), pp. 1513–1530.

[3] Seamus Kent et al. "Hospital costs in relation to body-mass index in $1 \cdot 1$ million women in England: a prospective cohort study". In: *The Lancet Public Health* 2.5 (2017), e214–e222.

[4] Mika Kivimäki et al. "Overweight, obesity, and risk of cardiometabolic multimorbidity: pooled analysis of individual-level data for 120 813 adults from 16 cohort studies from the USA and Europe". In: *The Lancet Public Health* 2.6 (2017), e277–e285.

[5] Asnawi Abdullah et al. "The magnitude of association between overweight and obesity and the risk of diabetes: a meta-analysis of prospective cohort studies". In: *Diabetes research and clinical practice* 89.3 (2010), pp. 309–319.

[6] GBD 2015 Obesity Collaborators. "Health effects of overweight and obesity in 195 countries over 25 years". In: *New England Journal of Medicine* 377.1 (2017), pp. 13–27.

[7] Peter H Whincup et al. "Birth weight and risk of type 2 diabetes: a systematic review". In: *Jama* 300.24 (2008), pp. 2886–2897.

[8] M. H. Park et al. "The impact of childhood obesity on morbidity and mortality in adulthood: a systematic review". In: *Obesity Reviews* 13.11 (2012), pp. 985–1000.

[9] Thomas Harder et al. "Birth weight and subsequent risk of type 2 diabetes: a meta-analysis". In: *American journal of epidemiology* 165.8 (2007), pp. 849–857.

[10] Bianca L De Stavola et al. "Statistical issues in life course epidemiology". In: *American journal of epidemiology* 163.1 (2005), pp. 84–96.

[11] Mark Simmonds et al. "The use of measures of obesity in childhood for predicting obesity and the development of obesity-related diseases in adulthood: a systematic review and meta-analysis." In: *Health technology assessment (Winchester, England)* 19.43 (2015), pp. 1–336.

[12] Janet W Rich-Edwards et al. "Birthweight and the risk for type 2 diabetes mellitus in adult women". In: *Annals of internal medicine* 130.4_Part_1 (1999), pp. 278–284.

[13] Edwina H Yeung et al. "Childhood size and life course weight characteristics in association with the risk of incident type 2 diabetes". In: *Diabetes care* 33.6 (2010), pp. 1364–1369.

[14] Elina Hyppönen, Chris Power, and George Davey Smith. "Prenatal growth, BMI, and risk of type 2 diabetes by early midlife". In: *Diabetes care* 26.9 (2003), pp. 2512–2517.

[15] Renée de Mutsert et al. "Overweight in early adulthood, adult weight change, and risk of type 2 diabetes, cardiovascular diseases, and certain cancers in men: a cohort study". In: *American journal of epidemiology* 179.11 (2014), pp. 1353–1365.

[16] Fiona Bragg et al. "Associations of general and central adiposity with incident diabetes in Chinese men and women". In: *Diabetes care* 41.3 (2018), pp. 494–502.

[17] Mingyang Song. "Trajectory analysis in obesity epidemiology: a promising life course approach." In: *Current opinion in endocrine and metabolic research* 4 (2019), p. 37.

[18] Jane Green et al. "Cohort Profile: the Million Women Study". In: *International journal of epidemiology* 48.1 (2018), 28–29e.

[19] Million Women Study Collaborative Group et al. "The Million Women Study: design and characteristics of the study population". In: *Breast Cancer Research* 1.1 (1999), p. 73.

[20] Stuart H Ralston et al. *Davidson's Principles and Practice of Medicine E-Book*. Elsevier Health Sciences, 2018.

[21] John Wass and Katharine Owen. *Oxford handbook of endocrinology and diabetes*. OUP Oxford, 2014.

[22] Steven E Kahn, Mark E Cooper, and Stefano Del Prato. "Pathophysiology and treatment of type 2 diabetes: perspectives on the past, present, and future". In: *The Lancet* 383.9922 (2014), pp. 1068–1083.

[23] American Diabetes Association et al. "2. Classification and diagnosis of diabetes: standards of medical care in diabetes—2019". In: *Diabetes Care* 42.Supplement 1 (2019), S13–S28.

[24] NICE. *National Institute for Health and Care Excellence (NICE)*. `https://www.nice.org.uk/guidance`. Online; accessed 24 Septemebr 2019. 2019.

[25] Dana Dabelea et al. "Trends in the prevalence of ketoacidosis at diabetes diagnosis: the SEARCH for diabetes in youth study". In: *Pediatrics* 133.4 (2014), e938–e945.

[26] Gunnar Stenström et al. "Latent autoimmune diabetes in adults: definition, prevalence, β-cell function, and treatment". In: *Diabetes* 54.suppl 2 (2005), S68–S72.

[27] Robert Turner et al. "UKPDS 25: autoantibodies to islet-cell cytoplasm and glutamic acid decarboxylase for prediction of insulin requirement in type 2 diabetes". In: *The Lancet* 350.9087 (1997), pp. 1288–1293.

[28] Nicholas J Thomas et al. "Frequency and phenotype of type 1 diabetes in the first six decades of life: a cross-sectional, genetically stratified survival analysis from UK Biobank". In: *The lancet Diabetes & endocrinology* 6.2 (2018), pp. 122–129.

[29] Francesca Amati et al. "Physical inactivity and obesity underlie the insulin resistance of aging". In: *Diabetes care* 32.8 (2009), pp. 1547–1549.

[30] Mark E Cleasby, Pauline M Jamieson, and Philip J Atherton. "Insulin resistance and sarcopenia: mechanistic links between common co-morbidities". In: *Journal of Endocrinology* 229.2 (2016), R67–R81.

[31] Naomi M Hamburg et al. "Physical inactivity rapidly induces insulin resistance and microvascular dysfunction in healthy volunteers". In: *Arteriosclerosis, thrombosis, and vascular biology* 27.12 (2007), pp. 2650–2656.

[32] Steven E Kahn, Rebecca L Hull, and Kristina M Utzschneider. "Mechanisms linking obesity to insulin resistance and type 2 diabetes". In: *Nature* 444.7121 (2006), p. 840.

[33] Kathrin Maedler et al. "Aging correlates with decreased β-cell proliferative capacity and enhanced sensitivity to apoptosis: a potential role for Fas and pancreatic duodenal homeobox-1". In: *Diabetes* 55.9 (2006), pp. 2455–2462.

[34] Matthew M Rankin and Jake A Kushner. "Adaptive β-cell proliferation is severely restricted with advanced age". In: *Diabetes* 58.6 (2009), pp. 1365–1372.

[35] Annette M Chang and Jeffrey B Halter. "Aging and insulin secretion". In: *American Journal of Physiology-Endocrinology And Metabolism* 284.1 (2003), E7–E12.

[36] Emma Ahlqvist et al. "Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables". In: *The lancet Diabetes & endocrinology* 6.5 (2018), pp. 361–369.

[37] Pouya Saeedi et al. "Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas". In: *Diabetes research and clinical practice* 157 (2019), p. 107843.

[38] Dianna J Magliano et al. "Trends in incidence of total or type 2 diabetes: systematic review". In: *bmj* 366 (2019), p. l5003.

[39] Diabetes UK. *Us, diabetes and a lot of facts and stats.* `https://www.diabetes.org.uk/professionals/position-statements-reports/statistics`. Online; accessed 11 March 2019. 2018.

[40] A Rosemary Tate et al. "Quality of recording of diabetes in the UK: how does the GP's method of coding clinical data affect incidence estimates? Cross-sectional study using the CPRD database". In: *BMJ open* 7.1 (2017), e012905.

[41] Manuj Sharma, Irwin Nazareth, and Irene Petersen. "Trends in incidence, prevalence and prescribing in type 2 diabetes mellitus between 2000 and 2013 in primary care: a retrospective cohort study". In: *BMJ open* 6.1 (2016), e010210.

[42] NHS Digital. *National Diabetes Audit Report 2017.* `bit.ly/1M5R3Op`. Online; accessed 11 March 2019. 2018.

[43] Tapash Roy and Cathy E Lloyd. "Epidemiology of depression and diabetes: a systematic review". In: *Journal of affective disorders* 142 (2012), S8–S21.

[44] Xiyue Jing et al. "Related factors of quality of life of type 2 diabetes patients: a systematic review and meta-analysis". In: *Health and quality of life outcomes* 16.1 (2018), p. 189.

[45] Claudia Langenberg and Luca A Lotta. "Genomic insights into the causes of type 2 diabetes". In: *The Lancet* 391.10138 (2018), pp. 2463–2474.

[46] Anubha Mahajan et al. “Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps”. In: *Nature genetics* 50.11 (2018), p. 1505.

[47] Robert A Scott et al. “An expanded genome-wide association study of type 2 diabetes in Europeans”. In: *Diabetes* 66.11 (2017), pp. 2888–2902.

[48] Wei Zhao et al. “Identification of new susceptibility loci for type 2 diabetes and shared etiological pathways with coronary heart disease”. In: *Nature genetics* 49.10 (2017), p. 1450.

[49] Anubha Mahajan et al. “Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes”. In: *Nature genetics* 50.4 (2018), p. 559.

[50] James B Meigs et al. “Genotype score in addition to common risk factors for prediction of type 2 diabetes”. In: *New England Journal of Medicine* 359.21 (2008), pp. 2208–2219.

[51] Valeriya Lyssenko et al. “Clinical risk factors, DNA variants, and the development of type 2 diabetes”. In: *New England Journal of Medicine* 359.21 (2008), pp. 2220–2232.

[52] Erik Ingelsson and Mark I McCarthy. “Human genetics of obesity and type 2 diabetes mellitus: past, present, and future”. In: *Circulation: Genomic and Precision Medicine* 11.6 (2018), e002090.

[53] Centers for Disease Control, Prevention, et al. *National diabetes statistics report, 2017*. 2017.

[54] Iris Shai et al. “Ethnicity, obesity, and risk of type 2 diabetes in women: a 20-year follow-up study”. In: *Diabetes care* 29.7 (2006), pp. 1585–1590.

[55] L. M. Goff. “Ethnicity and Type 2 diabetes in the UK”. In: *Diabetic Medicine* 36.8 (2019), pp. 927–938. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1111/dme.13895`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/dme.13895`.

[56] Andrew P Hills et al. “Epidemiology and determinants of type 2 diabetes in south Asia”. In: *The Lancet Diabetes & Endocrinology* (2018).

[57] Elias K Spanakis and Sherita Hill Golden. “Race/ethnic difference in diabetes and diabetic complications”. In: *Current diabetes reports* 13.6 (2013), pp. 814–823.

[58] InterAct Consortium et al. “The link between family history and risk of type 2 diabetes is not explained by anthropometric, lifestyle or genetic risk factors: the EPIC-InterAct study”. In: *Diabetologia* 56.1 (2013), pp. 60–69.

[59] Esther van’t Riet et al. “Role of adiposity and lifestyle in the relationship between family history of diabetes and 20-year incidence of type 2 diabetes in US women”. In: *Diabetes care* 33.4 (2010), pp. 763–767.

[60] CHT Tam et al. “Maternal history of diabetes is associated with increased cardiometabolic risk in Chinese”. In: *Nutrition & diabetes* 4.3 (2014), e112.

[61] Emilie Agardh et al. “Type 2 diabetes incidence and socio-economic position: a systematic review and meta-analysis”. In: *International journal of epidemiology* 40.3 (2011), pp. 804–818.

[62] Arleen F Brown et al. "Socioeconomic position and health among persons with diabetes mellitus: a conceptual framework and review of the literature". In: *Epidemiologic reviews* 26.1 (2004), pp. 63–77.

[63] Vincent Connolly et al. "Diabetes prevalence and socioeconomic status: a population based study showing increased prevalence of type 2 diabetes mellitus in deprived areas". In: *Journal of Epidemiology & Community Health* 54.3 (2000), pp. 173–177.

[64] A Espelt et al. "Socioeconomic inequalities in diabetes mellitus across Europe at the beginning of the 21st century". In: *Diabetologia* 51.11 (2008), p. 1971.

[65] Eric B Loucks et al. "Socioeconomic disparities in metabolic syndrome differ by gender: evidence from NHANES III". In: *Annals of epidemiology* 17.1 (2007), pp. 19–26.

[66] Teresa Tamayo, Christian Herder, and Wolfgang Rathmann. "Impact of early psychosocial factors (childhood socioeconomic factors and adversities) on future risk of type 2 diabetes, metabolic disturbances and obesity: a systematic review". In: *BMC public health* 10.1 (2010), p. 525.

[67] Elina Puolakka et al. "Childhood socioeconomic status in predicting metabolic syndrome and glucose abnormalities in adulthood: the cardiovascular risk in young Finns study". In: *Diabetes care* 39.12 (2016), pp. 2311–2317.

[68] Frank B Hu et al. "Diet, lifestyle, and the risk of type 2 diabetes mellitus in women". In: *New England journal of medicine* 345.11 (2001), pp. 790–797.

[69] Amy R Weinstein et al. "Relationship of physical activity vs body mass index with type 2 diabetes in women". In: *Jama* 292.10 (2004), pp. 1188–1194.

[70] Supriya Krishnan et al. "Overall and central obesity and risk of type 2 diabetes in US black women". In: *Obesity* 15.7 (2007), pp. 1860–1866.

[71] Youfa Wang et al. "Comparison of abdominal adiposity and overall obesity in predicting risk of type 2 diabetes among men". In: *The American journal of clinical nutrition* 81.3 (2005), pp. 555–563.

[72] T Nagaya et al. "Increases in body mass index, even within non-obese levels, raise the risk for Type 2 diabetes mellitus: a follow-up study in a Japanese population". In: *Diabetic medicine* 22.8 (2005), pp. 1107–1111.

[73] James B Meigs et al. "Body mass index, metabolic syndrome, and risk of type 2 diabetes or cardiovascular disease". In: *The Journal of Clinical Endocrinology & Metabolism* 91.8 (2006), pp. 2906–2912.

[74] Earl S Ford, David F Williamson, and Simin Liu. "Weight change and diabetes incidence: findings from a national cohort of US adults". In: *American journal of epidemiology* 146.3 (1997), pp. 214–222.

[75] Gang Hu et al. "Physical activity, body mass index, and risk of type 2 diabetes in patients with normal or impaired glucose regulation". In: *Archives of internal medicine* 164.8 (2004), pp. 892–896.

[76] S Goya Wannamethee, A Gerald Shaper, and Mary Walker. "Overweight and obesity and weight change in middle aged men: impact on cardiovascular disease and diabetes". In: *Journal of Epidemiology & Community Health* 59.2 (2005), pp. 134–139.

[77] Robert H Eckel et al. “Obesity and type 2 diabetes: what can be unified and what needs to be individualized?” In: *The Journal of Clinical Endocrinology & Metabolism* 96.6 (2011), pp. 1654–1663.

[78] Steven E Kahn, Rebecca L Hull, and Kristina M Utzschneider. “Mechanisms linking obesity to insulin resistance and type 2 diabetes”. In: *Nature* 444.7121 (2006), p. 840.

[79] An Pan et al. “Relation of active, passive, and quitting smoking with incident type 2 diabetes: a systematic review and meta-analysis”. In: *The lancet Diabetes & endocrinology* 3.12 (2015), pp. 958–967.

[80] Arnaud Chiolero et al. “Consequences of smoking for body weight, body fat distribution, and insulin resistance”. In: *The American journal of clinical nutrition* 87.4 (2008), pp. 801–809.

[81] Henri-Jean Aubin et al. “Weight gain in smokers after quitting cigarettes: meta-analysis”. In: *Bmj* 345 (2012), e4439.

[82] Craig Knott, Steven Bell, and Annie Britton. “Alcohol consumption and the risk of type 2 diabetes: a systematic review and dose-response meta-analysis of more than 1.9 million individuals from 38 observational studies”. In: *Diabetes care* 38.9 (2015), pp. 1804–1812.

[83] S Carlsson, N Hammar, and V Grill. “Alcohol consumption and type 2 diabetes”. In: *Diabetologia* 48.6 (2005), pp. 1051–1054.

[84] Xiao-Hua Li et al. “Association between alcohol consumption and the risk of incident type 2 diabetes: a systematic review and dose-response meta-analysis”. In: *The American journal of clinical nutrition* 103.3 (2016), pp. 818–829.

[85] Mika Kivimäki et al. “Physical inactivity, cardiometabolic disease, and risk of dementia: an individual-participant meta-analysis”. In: *bmj* 365 (2019), p. l1495.

[86] Andrea D Smith et al. *Physical activity and incident type 2 diabetes mellitus: a systematic review and dose–response meta-analysis of prospective cohort studies.* 2016.

[87] Dagfinn Aune et al. *Physical activity and the risk of type 2 diabetes: a systematic review and dose–response meta-analysis.* 2015.

[88] Connor A Emdin et al. “Usual blood pressure and risk of new-onset diabetes: evidence from 4.1 million adults and a meta-analysis of prospective studies”. In: *Journal of the American College of Cardiology* 66.14 (2015), pp. 1552–1562.

[89] Bindu Chamarthi et al. “Inflammation and hypertension: the interplay of interleukin-6, dietary sodium, and the renin–angiotensin system in humans”. In: *American journal of hypertension* 24.10 (2011), pp. 1143–1148.

[90] Stefano Taddei et al. “Antihypertensive drugs and reversing of endothelial dysfunction in hypertension”. In: *Current hypertension reports* 2.1 (2000), pp. 64–70.

[91] Lisa J Moran et al. “Impaired glucose tolerance, type 2 diabetes and metabolic syndrome in polycystic ovary syndrome: a systematic review and meta-analysis”. In: *Human reproduction update* 16.4 (2010), pp. 347–363.

[92] Amy D Anderson, Christine M Burt Solorzano, and Christopher R McCartney. "Childhood obesity and its impact on the development of adolescent PCOS". In: *Seminars in reproductive medicine*. Vol. 32. 03. Thieme Medical Publishers. 2014, pp. 202–213.

[93] Tim J Cole and Tim Lobstein. "Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity". In: *Pediatric obesity* 7.4 (2012), pp. 284–294.

[94] Robert J Kuczmarski. "2000 CDC growth charts for the United States; methods and development". In: (2002).

[95] Mercedes de Onis et al. "Development of a WHO growth reference for school-aged children and adolescents". In: *Bulletin of the World health Organization* 85 (2007), pp. 660–667.

[96] Benjamin J Cairns et al. "Lifetime body size and reproductive factors: comparisons of data recorded prospectively with self reports in middle age". In: *BMC medical research methodology* 11.1 (2011), p. 7.

[97] A Must et al. "Recall of early menstrual history and menarcheal body size: after 30 years, how well do women remember?" In: *American journal of epidemiology* 155.7 (2002), pp. 672–679.

[98] Carol Koprowski, Ralph J Coates, and Leslie Bernstein. "Ability of young women to recall past body size and age at menarche". In: *Obesity research* 9.8 (2001), pp. 478–485.

[99] UNICEF-WHO. *United Nations Children's Fund and World Health Organization. Low birthweight: country, regional and global estimates*. 2004.

[100] NHS Digital. *Health Survey for England 2017 Adult and child overweight and obesity*. `https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england`. Online; accessed 21 March 2019. 2018.

[101] World Health Organisation. *Overweight and obesity: Key facts*. `https://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight`. Online; accessed 18 August 2019. 2018.

[102] Leandra Abarca-Gómez et al. "Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128 · 9 million children, adolescents, and adults". In: *The Lancet* 390.10113 (2017), pp. 2627–2642.

[103] NHS Digital. *Statistics on Obesity, Physical Activity and Diet, England, 2019*. `https://digital.nhs.uk/data-and-information/publications/statistical/statistics-on-obesity-physical-activity-and-diet`. Online; accessed 11 June 2019. 2019.

[104] Mark O Goodarzi. "Genetics of obesity: what genetic association studies have taught us about the biology of obesity and its complications". In: *The lancet Diabetes & endocrinology* 6.3 (2018), pp. 223–236.

[105] Adam E Locke et al. "Genetic studies of body mass index yield new insights for obesity biology". In: *Nature* 518.7538 (2015), p. 197.

[106] Hermine HM Maes, Michael C Neale, and Lindon J Eaves. "Genetic and environmental factors in relative body weight and human adiposity". In: *Behavior genetics* 27.4 (1997), pp. 325–351.

[107] Janine F Felix et al. "Genome-wide association analysis identifies three new susceptibility loci for childhood body mass index". In: *Human molecular genetics* 25.2 (2015), pp. 389–403.

[108] Jonathan P Bradfield et al. "A genome-wide association meta-analysis identifies new childhood obesity loci". In: *Nature genetics* 44.5 (2012), p. 526.

[109] P Magnus. "Causes of variation in birth weight: a study of offspring of twins". In: *Clinical genetics* 25.1 (1984), pp. 15–24.

[110] P Magnus. "Further evidence for a significant effect of fetal genes on variation in birth weight". In: *Clinical genetics* 26.4 (1984), pp. 289–296.

[111] Momoko Horikoshi et al. "Genome-wide associations for birth weight and correlations with adult disease". In: *Nature* 538.7624 (2016), p. 248.

[112] Lindon J Eaves et al. "Resolving the effects of maternal and offspring genotype on dyadic outcomes in genome wide complex trait analysis ("M-GCTA")". In: *Behavior genetics* 44.5 (2014), pp. 445–455.

[113] Astrid Lunde et al. "Genetic and environmental influences on birth weight, birth length, head circumference, and gestational age by use of population-based parent-offspring data". In: *American journal of epidemiology* 165.7 (2007), pp. 734–741.

[114] Robin N Beaumont et al. "Genome-wide association study of offspring birth weight in 86 577 women identifies five novel loci and highlights maternal genetic effects that are independent of fetal genetics". In: *Human molecular genetics* 27.4 (2018), pp. 742–756.

[115] Kirsi H Pietiläinen et al. "Tracking of body size from birth to late adolescence: contributions of birth length, birth weight, duration of gestation, parents' body size, and twinship". In: *American Journal of Epidemiology* 154.1 (2001), pp. 21–29.

[116] Tessa J Parsons, Chris Power, and Orly Manor. "Fetal and early life growth and body mass index from birth to early adulthood in 1958 British cohort: longitudinal study". In: *Bmj* 323.7325 (2001), pp. 1331–1335.

[117] Imogen Rogers. "The influence of birthweight and intrauterine environment on adiposity and fat distribution in later life". In: *International journal of obesity* 27.7 (2003), p. 755.

[118] Daniela Zanetti et al. "Birthweight, type 2 diabetes mellitus, and cardiovascular disease: Addressing the Barker hypothesis with Mendelian randomization". In: *Circulation: Genomic and Precision Medicine* 11.6 (2018), e002054.

[119] Gian-Paolo Ravelli, Zena A Stein, and Mervyn W Susser. "Obesity in young men after famine exposure in utero and early infancy". In: *New England Journal of Medicine* 295.7 (1976), pp. 349–353.

[120] Anita CJ Ravelli et al. "Obesity at the age of 50 y in men and women exposed to famine prenatally". In: *The American journal of clinical nutrition* 70.5 (1999), pp. 811–816.

[121] David J Pettitt et al. "Excessive obesity in offspring of Pima Indian women with diabetes during pregnancy". In: *New England Journal of Medicine* 308.5 (1983), pp. 242–245.

[122] N Spencer and S Logan. *Social influences on birth weight*. 2002.

[123] CRS Dougherty and A David Jones. "The determinants of birth weight". In: *American journal of obstetrics and gynecology* 144.2 (1982), pp. 190–200.

[124] Peymane Adab, Miranda Pallan, and Peter H Whincup. *Is BMI the best measure of obesity?* 2018.

[125] X Song et al. "Comparison of various surrogate obesity indicators as predictors of cardiovascular mortality in four European populations". In: *European Journal of Clinical Nutrition* 67.12 (2013), p. 1298.

[126] Nicola Orsini, Rino Bellocco, and Sander Greenland. "Generalized least squares for trend estimation of summarized dose–response data". In: *The stata journal* 6.1 (2006), pp. 40–57.

[127] Geneviève Chêne and Simon G Thompson. "Methods for summarizing the risk associations of quantitative variables in epidemiologic studies in a consistent form". In: *American Journal of Epidemiology* 144.6 (1996), pp. 610–621.

[128] Sander Greenland and Matthew P Longnecker. "Methods for trend estimation from summarized dose-response data, with applications to meta-analysis". In: *American journal of epidemiology* 135.11 (1992), pp. 1301–1309.

[129] Esther Zimmermann et al. "Sex differences in the association between birth weight and adult type 2 diabetes". In: *Diabetes* 64.12 (2015), pp. 4220–4225.

[130] Magnus Kaijser et al. "Perinatal risk factors for diabetes in later life". In: *Diabetes* 58.3 (2009), pp. 523–526.

[131] Yanping Li et al. "Birth weight and later life adherence to unhealthy lifestyles in predicting type 2 diabetes: prospective cohort study". In: *bmj* 351 (2015), h3672.

[132] Edward A Ruiz-Narváez et al. "Birth weight and risk of type 2 diabetes in the black women's health study: does adult BMI play a mediating role?" In: *Diabetes care* 37.9 (2014), pp. 2572–2578.

[133] Blandine de Lauzon-Guillain et al. "Birth weight, body silhouette over the life course, and incident diabetes in 91,453 middle-aged women from the French Etude Epidemiologique de Femmes de la Mutuelle Generale de l'Education Nationale (E3N) Cohort". In: *Diabetes care* 33.2 (2010), pp. 298–303.

[134] Qinghua Xia et al. "Prospective cohort studies of birth weight and risk of obesity, diabetes, and hypertension in adulthood among the Chinese population". In: *Journal of diabetes* 11.1 (2019), pp. 55–64.

[135] David R McCance et al. "Birth weight and non-insulin dependent diabetes: thrifty genotype, thrifty phenotype, or surviving small baby genotype?" In: *Bmj* 308.6934 (1994), pp. 942–945.

[136] RF Dyck et al. "The hefty fetal phenotype hypothesis revisited: high birth weight, type 2 diabetes and gestational diabetes in a Saskatchewan cohort of First Nations and non-First Nations women". In: *Journal of developmental origins of health and disease* (2017), pp. 1–7.

[137] Paul W Franks et al. “Gestational glucose tolerance and risk of type 2 diabetes in young Pima Indian offspring”. In: *Diabetes* 55.2 (2006), pp. 460–465.

[138] Roland Dyck et al. “A comparison of rates, risk factors, and outcomes of gestational diabetes between aboriginal and non-aboriginal women in the Saskatoon health district”. In: *Diabetes care* 25.3 (2002), pp. 487–493.

[139] C Nicholas Hales and David JP Barker. “The thrifty phenotype hypothesis: Type 2 diabetes”. In: *British medical bulletin* 60.1 (2001), pp. 5–20.

[140] C Nicholas Hales and David JP Barker. “Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis”. In: *Diabetologia* 35.7 (1992), pp. 595–601.

[141] David JP Barker et al. “Fetal origins of adult disease: strength of effects and biological basis”. In: *International journal of epidemiology* 31.6 (2002), pp. 1235–1239.

[142] Margaretha Eriksson et al. “Birth weight and cardiovascular risk factors in a cohort followed until 80 years of age: the study of men born in 1913”. In: *Journal of internal medicine* 255.2 (2004), pp. 236–246.

[143] Johan G Eriksson et al. “Pathways of infant and childhood growth that lead to type 2 diabetes”. In: *Diabetes care* 26.11 (2003), pp. 3006–3010.

[144] Lise G Bjerregaard et al. “Change in overweight from childhood to early adulthood and risk of type 2 diabetes”. In: *New England Journal of Medicine* 378.14 (2018), pp. 1302–1312.

[145] Esther Zimmermann et al. “Childhood body mass index and development of type 2 diabetes throughout adult life—A large-scale danish cohort study”. In: *Obesity* 25.5 (2017), pp. 965–971.

[146] Markus Juonala et al. “Childhood adiposity, adult adiposity, and cardiovascular risk factors”. In: *New England Journal of Medicine* 365.20 (2011), pp. 1876–1885.

[147] Christopher G Owen et al. “Body mass index in early and middle adult life: prospective associations with myocardial infarction, stroke and diabetes over a 30-year period: the British Regional Heart Study”. In: *BMJ open* 5.9 (2015), e008105.

[148] Amir Tirosh et al. “Adolescent BMI trajectory and risk of diabetes versus coronary disease”. In: *New England Journal of Medicine* 364.14 (2011), pp. 1315–1325.

[149] William Johnson, Diana Kuh, and Rebecca Hardy. “A Life Course Perspective on Body Size and Cardio-metabolic Health”. In: *A life course perspective on health trajectories and transitions*. Springer, Cham, 2015, pp. 61–83.

[150] Yan Zheng et al. “Group-based trajectory of body shape from ages 5 to 55 years and cardiometabolic disease risk in 2 US cohorts”. In: *American Journal of Epidemiology* 186.11 (2017), pp. 1246–1255.

[151] Emily Banks et al. “Effect on attendance at breast cancer screening of adding a self administered questionnaire to the usual invitation to breast screening in southern England.” In: *Journal of Epidemiology & Community Health* 52.2 (1998), pp. 116–119.

[152] Patnick J (ed). *NHS Breast Screening Programme Report*. Sheffield. 1996.

[153] Emily Banks et al. "Comparison of various characteristics of women who do and do not attend for breast cancer screening". In: *Breast Cancer Research* 4.1 (2001).

[154] NHS Digital. *Hospital Episode Statistics*. `http://content.digital.nhs.uk/hes`. Online; accessed 11 March 2019. 2019.

[155] NHS National Services Scotland. *Information Services Division*. `https://www.iSDscotland.org/`. Online; accessed 11 March 2019. 2019.

[156] World Health Organization. *International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10)*. `https://icd.who.int/browse10/2016/en`. Online; accessed 01 March 2019. 2019.

[157] UK Biobank. *Hospital Episode Statistics data in Showcase*. `http://biobank.ctsu.ox.ac.uk/showcase/docs/HospitalEpisodeStatistics.pdf`. Online; accessed 01 March 2019. 2013.

[158] Health and Social Care Information Centre. *Hospital Episode Statistics (HES) Analysis Guide*. `http://docs.adrn.ac.uk/888040/mrdoc/pdf/888040_hes-analy-guide-apr13.pdf`. Online; accessed 01 March 2018. 2013.

[159] Office of Population Censuses and Surveys. *Tabular list of the classification of surgical operations and procedures, 4th revision*. London. 1990.

[160] Stephen Andrew Spencer and Mark Price Davies. "Hospital episode statistics: improving the quality and value of hospital data: a national internet e-survey of hospital consultants". In: *BMJ open* 2.6 (2012), e001651.

[161] James Peskett, Howard Davis, and Peter Saunders. "Findings of the national PbR data Assurance Framework: improving the quality of data underpinning payment by results using benchmarking to target clinical coding audits". In: *BMC health services research*. Vol. 8. 1. BioMed Central. 2008, A22.

[162] Sidhartha Sinha et al. "Studies using English administrative data (Hospital Episode Statistics) to assess health-care outcomes—systematic review and recommendations for reporting". In: *The European Journal of Public Health* 23.1 (2012), pp. 86–92.

[163] Annie Britton et al. "Validating self-reported strokes in a longitudinal UK cohort study (Whitehall II): Extracting information from hospital medical records versus the Hospital Episode Statistics database". In: *BMC medical research methodology* 12.1 (2012), p. 83.

[164] Susan E Campbell et al. "A systematic review of discharge coding accuracy". In: *Journal of Public Health* 23.3 (2001), pp. 205–211.

[165] Elaine M Burns et al. "Systematic review of discharge coding accuracy". In: *Journal of public health* 34.1 (2011), pp. 138–148.

[166] NHS Digital. *Payment by Results*. `https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/payment-by-results`. Online; accessed 11 Septemebr 2019. 2019.

[167] ISD Scotland. *SMR Datasets: ISD Scotland Data Dictionary*. `https://www.ndc.scot.nhs.uk/Data-Dictionary/SMR-Datasets/`. Online; accessed 11 Septemebr 2019. 2019.

[168] ISD Scotland. *Secondary Care Information Collection: Information on Standards and Coding for Clinicians.* https://www.isdscotland.org/Products-and-Services/Terminology-Services/Information-for-Clinicians/Secondary-Care-Information/. Online; accessed 11 Septemebr 2019. 2019.

[169] National Institute for Health Research. *Clinical Practice Research Datalink (CPRD).* https://www.cprd.com/. Online; accessed 11 March 2019. 2019.

[170] Emily Herrett et al. "Data resource profile: clinical practice research datalink (CPRD)". In: *International journal of epidemiology* 44.3 (2015), pp. 827–836.

[171] Rohini Mathur et al. "Completeness and usability of ethnicity data in UK-based primary care and hospital databases". In: *Journal of public health* 36.4 (2013), pp. 684–692.

[172] Emily Herrett et al. "Data resource profile: clinical practice research datalink (CPRD)". In: *International journal of epidemiology* 44.3 (2015), pp. 827–836.

[173] John Chisholm. "The Read clinical classification." In: *BMJ: British Medical Journal* 300.6732 (1990), p. 1092.

[174] NHS Digital. *Read Codes.* https://digital.nhs.uk/services/terminology-and-classifications/read-codes. Online; accessed 23 March 2019. 2019.

[175] Shivani Padmanabhan. *CPRD GOLD Data Specification.* https://www.cprd.com/sites/default/files/CPRD%20GOLD%20Full%20Data%20Specification%20v2.0.pdf. Online; accessed 11 March 2018. 2017.

[176] A Rosemary Tate et al. "Quality of recording of diabetes in the UK: how does the GP's method of coding clinical data affect incidence estimates? Cross-sectional study using the CPRD database". In: *BMJ open* 7.1 (2017), e012905.

[177] S De Lusignan et al. "Miscoding, misclassification and misdiagnosis of diabetes in primary care". In: *Diabetic Medicine* 29.2 (2012), pp. 181–189.

[178] NHS Digital. *National Quality Outcomes Framework - 2004/05.* 2005.

[179] Emily Herrett et al. "Validation and validity of diagnoses in the General Practice Research Database: a systematic review". In: *British journal of clinical pharmacology* 69.1 (2010), pp. 4–14.

[180] Nada F Khan, Sian E Harrison, and Peter W Rose. "Validity of diagnostic coding within the General Practice Research Database: a systematic review". In: *Br J Gen Pract* 60.572 (2010), e128–e136.

[181] F Lucy Wright et al. "Validity over time of self-reported anthropometric variables during follow-up of a large cohort of UK women". In: *BMC medical research methodology* 15.1 (2015), p. 81.

[182] Kenneth J Rothman. "BMI-related errors in the measurement of obesity". In: *International journal of obesity* 32.S3 (2008), S56.

[183] Michael L Rowland. "Self-reported weight and height". In: *The American journal of clinical nutrition* 52.6 (1990), pp. 1125–1133.

[184] Guri Skeie et al. "Validity of self-reported body mass index among middle-aged participants in the Norwegian Women and Cancer study". In: *Clinical epidemiology* 7 (2015), p. 313.

[185] Anthony Hill and Julian Roberts. "Body mass index: a comparison between self-reported and measured height and weight". In: *Journal of Public Health* 20.2 (1998), pp. 206–210.

[186] Rudolf Kaaks, Elio Riboli, and Wija van Staveren. "Calibration of dietary intake measurements in prospective cohort studies". In: *American Journal of Epidemiology* 142.5 (1995), pp. 548–556.

[187] Victor Kipnis et al. "Empirical evidence of correlated biases in dietary assessment instruments and its implications". In: *American journal of epidemiology* 153.4 (2001), pp. 394–403.

[188] Emanuele Di Angelantonio et al. "Body-mass index and all-cause mortality: individual-participant-data meta-analysis of 239 prospective studies in four continents". In: *The Lancet* 388.10046 (2016), pp. 776–786.

[189] Dexter Canoy et al. "Body mass index and incident coronary heart disease in women: a population-based prospective study". In: *BMC medicine* 11.1 (2013), p. 87.

[190] Mary E Kroll et al. "Adiposity and ischemic and hemorrhagic stroke: prospective study in women and meta-analysis". In: *Neurology* 87.14 (2016), pp. 1473–1481.

[191] Lianne Parkin et al. "Body mass index, surgery, and risk of venous thromboembolism in middle-aged women: a cohort study". In: *Circulation* 125.15 (2012), pp. 1897–1904.

[192] Bette Liu et al. "Relationship of height, weight and body mass index to the risk of hip and knee replacements in middle-aged women". In: *Rheumatology* 46.5 (2007), pp. 861–867.

[193] Miranda EG Armstrong et al. "Body mass index and physical activity in relation to the incidence of hip fracture in postmenopausal women". In: *Journal of bone and mineral research* 26.6 (2011), pp. 1330–1338.

[194] Jason Lacombe et al. "The Effects of Age, Adiposity, and Physical Activity on the Risk of Seven Site-Specific Fractures in Postmenopausal Women". In: *Journal of Bone and Mineral Research* 31.8 (2016), pp. 1559–1568.

[195] Miranda EG Armstrong et al. "Different effects of age, adiposity and physical activity on the risk of ankle, wrist and hip fractures in postmenopausal women". In: *Bone* 50.6 (2012), pp. 1394–1400.

[196] Bette Liu et al. "Body mass index and risk of liver cirrhosis in middle aged UK women: prospective study". In: *Bmj* 340 (2010), p. c912.

[197] K Coffey et al. "Lifestyle and reproductive risk factors associated with anal cancer in women aged over 50 years". In: *British journal of cancer* 112.9 (2015), p. 1568.

[198] Andrea Burón Pust et al. "Heterogeneity of colorectal cancer risk by tumour characteristics: Large prospective study of UK women". In: *International journal of cancer* 140.5 (2017), pp. 1082–1090.

[199] Collaborative Group on Epidemiological Studies of Ovarian Cancer et al. "Ovarian cancer and body size: individual participant meta-analysis including 25,157 women with ovarian cancer from 47 epidemiological studies". In: *PLoS medicine* 9.4 (2012), e1001200.

[200] F Murphy et al. “Body size in relation to incidence of subtypes of haematological malignancy in the prospective Million Women Study”. In: *British journal of cancer* 108.11 (2013), p. 2390.

[201] Gillian K Reeves et al. “Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study”. In: *Bmj* 335.7630 (2007), p. 1134.

[202] Gillian K Reeves et al. “Hospital admissions in relation to body mass index in UK women: a prospective cohort study”. In: *BMC medicine* 12.1 (2014), p. 45.

[203] Seamus Kent et al. “Hospital costs in relation to body-mass index in $1 \cdot 1$ million women in England: a prospective cohort study”. In: *The Lancet Public Health* 2.5 (2017), e214–e222.

[204] Medical Research Council and UCL. *MRC National Survey of Health and Development (NSHD)*. `https://www.nshd.mrc.ac.uk/`. Online; accessed 15 March 2019. 2019.

[205] Lisa M Troy et al. “Self-reported birthweight and history of having been breastfed among younger women: an assessment of validity”. In: *International journal of epidemiology* 25.1 (1996), pp. 122–127.

[206] Pernille M Wodskou et al. “Validity of self-reported birthweight among middle-aged and elderly women in the Danish Nurse Cohort Study”. In: *Acta obstetricia et gynecologica Scandinavica* 89.9 (2010), pp. 1134–1139.

[207] Parisa Tehranifar et al. “Validity of self-reported birth weight by adult women: sociodemographic influences and implications for life-course studies”. In: *American journal of epidemiology* 170.7 (2009), pp. 910–917.

[208] TO Yang et al. “Birth weight and adult cancer incidence: large prospective study and meta-analysis”. In: *Annals of oncology* 25.9 (2014), pp. 1836–1843.

[209] Bette Liu et al. “Body size from birth to middle age and the risk of hip and knee replacement”. In: *BMC musculoskeletal disorders* 17.1 (2016), p. 260.

[210] Aviva Must, Walter C Willett, and William H Dietz. “Remote recall of childhood height, weight, and body build by elderly subjects”. In: *American journal of epidemiology* 138.1 (1993), pp. 56–64.

[211] Alison E Field et al. “Race differences in accuracy of self-reported childhood body size among white and black women”. In: *Obesity research* 12.7 (2004), pp. 1136–1144.

[212] TYO Yang et al. “Postmenopausal endometrial cancer risk and body size in early life and middle age: prospective cohort study”. In: *British journal of cancer* 107.1 (2012), p. 169.

[213] TienYu Owen Yang et al. “Body size in early life and risk of lymphoid malignancies and histological subtypes in adulthood”. In: *International journal of cancer* 139.1 (2016), pp. 42–49.

[214] Melanie Calvert et al. “Effect of the quality and outcomes framework on diabetes care in the United Kingdom: retrospective cohort study”. In: *Bmj* 338 (2009), b1870.

[215] Institute of Population Health, University of Manchester. *ClinicalCodes.org*. `https://clinicalcodes.rss.mhs.man.ac.uk/`. Online; accessed 23 March 2019. 2019.

[216] Anders Green et al. "Validation of the danish national diabetes register". In: *Clinical epidemiology* 7 (2015), p. 5.

[217] M Whitston et al. "What can be learned about the impact of diabetes on hospital admissions from routinely recorded data?" In: *Diabetic Medicine* 29.9 (2012), pp. 1199–1205.

[218] H Anwar et al. "Assessment of the under-reporting of diabetes in hospital admission data: a study from the Scottish Diabetes Research Network Epidemiology Group". In: *Diabetic Medicine* 28.12 (2011), pp. 1514–1519.

[219] Sophie V Eastwood et al. "Algorithms for the capture and adjudication of prevalent and incident diabetes in UK Biobank". In: *PloS one* 11.9 (2016), e0162388.

[220] Elizabeth Jean Comino et al. "Impact of diabetes on hospital admission and length of stay among a general population aged 45 year or more: a record linkage study". In: *BMC health services research* 15.1 (2015), p. 12.

[221] JM Khalid et al. "Rates and risk of hospitalisation among patients with type 2 diabetes: retrospective cohort study using the UK General Practice Research Database linked to English Hospital Episode Statistics". In: *International journal of clinical practice* 68.1 (2014), pp. 40–48.

[222] P Townsend, P Phillimore, and A Beattie. "Health and deprivation: inequality and the North (Croom Helm, London)". In: (1988).

[223] Amalia Karahalios et al. "A review of the reporting and handling of missing data in cohort studies with repeated assessment of exposure measures". In: *BMC medical research methodology* 12.1 (2012), p. 96.

[224] Helena Jeličić, Erin Phelps, and Richard M Lerner. "Use of missing data methods in longitudinal studies: The persistence of bad practices in developmental psychology." In: *Developmental psychology* 45.4 (2009), p. 1195.

[225] Jonathan AC Sterne et al. "Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls". In: *Bmj* 338 (2009), b2393.

[226] Sander Greenland and William D Finkle. "A critical look at methods for handling missing covariates in epidemiologic regression analyses". In: *American journal of epidemiology* 142.12 (1995), pp. 1255–1264.

[227] A Mackinnon. "The use and reporting of multiple imputation in medical research–a review". In: *Journal of internal medicine* 268.6 (2010), pp. 586–593.

[228] Institute for Digital Research and Education, UCLA. *Testing the proportional hazard assumption in Cox models*. `https://stats.idre.ucla.edu/other/examples/asa2/testing-the-proportional-hazard-assumption-in-cox-models/`. Online; accessed 11 March 2019. 2019.

[229] Health and Social Care Information Centre. *National Diabetes Audit 2011/12: Report 1: Care Processes and Treatment Targets.* `https://www.diabetes.org.uk/resources-s3/2017-11/diabetes-key-stats-guidelines-april2014.pdf`. Online; accessed 01 March 2019. 2014.

[230] Jessica M Robbins et al. "Socioeconomic status and diagnosed diabetes incidence". In: *Diabetes research and clinical practice* 68.3 (2005), pp. 230–236.

[231] Emilie Agardh et al. "Type 2 diabetes incidence and socio-economic position: a systematic review and meta-analysis". In: *International journal of epidemiology* 40.3 (2011), pp. 804–818.

[232] Richard Cookson et al. "Socio-economic inequalities in health care in England". In: *Fiscal Studies* 37.3-4 (2016), pp. 371–403.

[233] Yoshitaka Nishino, Stuart Gilmour, and Kenji Shibuya. "Inequality in diabetes-related hospital admissions in England by socioeconomic deprivation and ethnicity: facility-based cross-sectional analysis". In: *PloS one* 10.2 (2015), e0116689.

[234] Srinivasa Vittal Katikireddi et al. "Assessment of health care, hospital admissions, and mortality by ethnicity: population-based cohort study of health-system performance in Scotland". In: *The Lancet Public Health* 3.5 (2018), e226–e236.

[235] Filippa Juul et al. "Birth weight, early life weight gain and age at menarche: a systematic review of longitudinal studies". In: *Obesity Reviews* 18.11 (2017), pp. 1272–1288.

[236] David S Freedman et al. "The relation of menarcheal age to obesity in childhood and adulthood: the Bogalusa heart study". In: *BMC pediatrics* 3.1 (2003), p. 3.

[237] Susana Bubach et al. "Impact of the age at menarche on body composition in adulthood: results from two birth cohort studies". In: *BMC public health* 16.1 (2016), p. 1007.

[238] Shanshan Li et al. "Offspring risk of obesity in childhood, adolescence and adulthood in relation to gestational diabetes mellitus: a sex-specific association". In: *International journal of epidemiology* 46.5 (2017), pp. 1533–1541.

[239] InterAct Consortium et al. "The link between family history and risk of type 2 diabetes is not explained by anthropometric, lifestyle or genetic risk factors: the EPIC-InterAct study". In: *Diabetologia* 56.1 (2013), pp. 60–69.

[240] Enrique F Schisterman, Stephen R Cole, and Robert W Platt. "Overadjustment bias and unnecessary adjustment in epidemiologic studies". In: *Epidemiology (Cambridge, Mass.)* 20.4 (2009), p. 488.

[241] Stephen MacMahon et al. "Blood pressure, stroke, and coronary heart disease: part 1, prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias". In: *The Lancet* 335.8692 (1990), pp. 765–774.

[242] Robert Clarke et al. "Underestimation of risk associations due to regression dilution in long-term follow-up of prospective studies". In: *American journal of epidemiology* 150.4 (1999), pp. 341–353.

[243] Sture Holm. "A simple sequentially rejective multiple test procedure". In: *Scandinavian journal of statistics* (1979), pp. 65–70.

[244] Martyn Plummer. "Improved estimates of floating absolute risk". In: *Statistics in medicine* 23.1 (2004), pp. 93–104.

[245] StataCorp. *Stata statistical software: Release 15.1.* 2017.

[246] R Development Core Team. *R: A language and environment for statistical computing.* 2015.

[247] Laura J Corbin et al. "BMI as a modifiable risk factor for type 2 diabetes: refining and understanding causal estimates using Mendelian randomization". In: *Diabetes* 65.10 (2016), pp. 3002–3007.

[248] Michael V Holmes et al. "Causal effects of body mass index on cardiometabolic traits and events: a Mendelian randomization analysis". In: *The American Journal of Human Genetics* 94.2 (2014), pp. 198–208.

[249] Shoaib Afzal et al. "Vitamin D concentration, obesity, and risk of diabetes: a mendelian randomisation study". In: *The lancet Diabetes & endocrinology* 2.4 (2014), pp. 298–306.

[250] Tove Fall et al. "The role of adiposity in cardiometabolic traits: a Mendelian randomization analysis". In: *PLoS medicine* 10.6 (2013), e1001474.

[251] Andrew T Hattersley and John E Tooke. "The fetal insulin hypothesis: an alternative explanation of the association of low bir thweight with diabetes and vascular disease". In: *The Lancet* 353.9166 (1999), pp. 1789–1792.

[252] Rachel M Freathy et al. "Type 2 diabetes risk alleles are associated with reduced size at birth". In: *Diabetes* 58.6 (2009), pp. 1428–1433.

[253] Juris J Meier. "Linking the genetics of type 2 diabetes with low birth weight: a role for prenatal islet maldevelopment?" In: *Diabetes* 58.6 (2009), pp. 1255–1256.

[254] Momoko Horikoshi et al. "New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism". In: *Nature genetics* 45.1 (2013), p. 76.

[255] Tyler J VanderWeele and Mirjam J Knol. "A tutorial on interaction". In: *Epidemiologic Methods* 3.1 (2014), pp. 33–72.

[256] Stefan Johansson et al. "The association between low birth weight and type 2 diabetes: contribution of genetic factors". In: *Epidemiology* 19.5 (2008), pp. 659–665.

[257] Anita CJ Ravelli et al. "Glucose tolerance in adults after prenatal exposure to famine". In: *The Lancet* 351.9097 (1998), pp. 173–177.

[258] Yanping Li et al. "Exposure to the Chinese famine in early life and the risk of hyperglycemia and type 2 diabetes in adulthood". In: *Diabetes* 59.10 (2010), pp. 2400–2406.

[259] Dana Dabelea. "The predisposition to obesity and diabetes in offspring of diabetic mothers". In: *Diabetes care* 30.Supplement 2 (2007), S169–S174.

[260] Shanshan Li et al. "Offspring risk of obesity in childhood, adolescence and adulthood in relation to gestational diabetes mellitus: a sex-specific association". In: *International journal of epidemiology* 46.5 (2017), pp. 1533–1541.

[261] Eero Kajantie et al. "Preterm birth—a risk factor for type 2 diabetes?: the Helsinki Birth cohort study". In: *Diabetes care* 33.12 (2010), pp. 2623–2625.

[262] Casey Crump et al. “Risk of diabetes among young adults born preterm in Sweden”. In: *Diabetes Care* 34.5 (2011), pp. 1109–1113.

[263] Eric L Ding et al. “Sex differences of endogenous sex hormones and risk of type 2 diabetes: a systematic review and meta-analysis”. In: *Jama* 295.11 (2006), pp. 1288–1299.

[264] A Garofano, P Czernichow, and B Breant. “In utero undernutrition impairs rat beta-cell development”. In: *Diabetologia* 40.10 (1997), pp. 1231–1234.

[265] Eric Bertin et al. “Development of β-cell mass in fetuses of rats deprived of protein and/or energy in last trimester of pregnancy”. In: *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 283.3 (2002), R623–R630.

[266] JJ Meier. “Beta cell mass in diabetes: a realistic therapeutic target?” In: *Diabetologia* 51.5 (2008), pp. 703–713.

[267] Johan G Eriksson et al. “Early adiposity rebound in childhood and risk of Type 2 diabetes in adult life”. In: *Diabetologia* 46.2 (2003), pp. 190–194.

[268] Tom Forsen et al. “The fetal and childhood growth of persons who develop type 2 diabetes”. In: *Annals of internal medicine* 133.3 (2000), pp. 176–182.

[269] Luc Bouwens and Ilse Rooman. “Regulation of pancreatic beta-cell mass”. In: *Physiological reviews* 85.4 (2005), pp. 1255–1270.

[270] David J Hill. “Development of the endocrine pancreas”. In: *Reviews in Endocrine and Metabolic Disorders* 6.3 (2005), pp. 229–238.

[271] Johan G Eriksson et al. “Trajectories of body mass index amongst children who develop type 2 diabetes as adults”. In: *Journal of internal medicine* 278.2 (2015), pp. 219–226.

[272] W Johnson. “Body size trajectories and cardio-metabolic resilience to obesity”. In: *Nutrition bulletin* 43.4 (2018), pp. 456–462.

[273] University College London, Department of Epidemiology and Public Health, National Centre for Social Research. *Health Survey for England, 2003*. `http://doi.org/10.5255/UKDA-SN-5098-1`. Online; accessed 21 March 2019. 2010.

[274] Nasser Bagheri et al. “Undiagnosed diabetes from cross-sectional GP practice data: an approach to identify communities with high likelihood of undiagnosed diabetes”. In: *BMJ open* 4.7 (2014), e005305.

[275] Alison Moody et al. “Social inequalities in prevalence of diagnosed and undiagnosed diabetes and impaired glucose regulation in participants in the Health Surveys for England series”. In: *BMJ open* 6.2 (2016), e010155.

[276] Kristian D Lindsted et al. “Healthy volunteer effect in a cohort study: temporal resolution in the Adventist Health Study”. In: *Journal of clinical epidemiology* 49.7 (1996), pp. 783–790.

[277] DO Okorodudu et al. “Diagnostic performance of body mass index to identify obesity as defined by body adiposity: a systematic review and meta-analysis”. In: *International journal of obesity* 34.5 (2010), p. 791.

[278] Yan Zheng et al. “Group-based trajectory of body shape from ages 5 to 55 years and cardiometabolic disease risk in 2 US cohorts”. In: *American Journal of Epidemiology* 186.11 (2017), pp. 1246–1255.

[279] Cindy Feng, Nathaniel D Osgood, and Roland F Dyck. “Low birth weight, cumulative obesity dose, and the risk of incident type 2 diabetes”. In: *Journal of diabetes research* 2018 (2018).

[280] Inger Wahlström Johnsson et al. “A high birth weight is associated with increased risk of type 2 diabetes and obesity”. In: *Pediatric obesity* 10.2 (2015), pp. 77–83.

[281] Mikaela B von Bonsdorff et al. “Persistence of the effect of birth size on dysglycaemia and type 2 diabetes in old age: AGES-Reykjavik Study”. In: *Age* 35.4 (2013), pp. 1401–1409.

[282] DA Lawlor et al. “The associations of birthweight, gestational age and childhood BMI with type 2 diabetes: findings from the Aberdeen Children of the 1950s cohort”. In: *Diabetologia* 49.11 (2006), pp. 2614–2617.

[283] M Wadsworth et al. “Early growth and type 2 diabetes: evidence from the 1946 British birth cohort”. In: *Diabetologia* 48.12 (2005), pp. 2505–2510.

[284] Jie Mi et al. “Effects of infant birthweight and maternal body mass index in pregnancy on components of the insulin resistance syndrome in China”. In: *Annals of Internal Medicine* 132.4 (2000), pp. 253–260.

[285] MJ Vanhala et al. “Relative weight gain and obesity as a child predict metabolic syndrome as an adult”. In: *International journal of obesity* 23.6 (1999), p. 656.

[286] CHD Fall et al. “Size at birth, maternal weight, and type 2 diabetes in South India”. In: *Diabetic medicine* 15.3 (1998), pp. 220–227.

[287] CN Martyn et al. “Fetal growth and hyperinsulinaemia in adult life”. In: *Diabetic medicine* 15.8 (1998), pp. 688–694.

[288] Donald E Yarbrough et al. “Birth weight, adult weight, and girth as predictors of the metabolic syndrome in postmenopausal women: the Rancho Bernardo Study”. In: *Diabetes care* 21.10 (1998), pp. 1652–1658.

[289] Hans O Lithell et al. “Relation of size at birth to non-insulin dependent diabetes and insulin concentrations in men aged 50-60 years”. In: *Bmj* 312.7028 (1996), pp. 406–410.

[290] C Nicholas Hales et al. “Fetal and infant growth and impaired glucose tolerance at age 64.” In: *Bmj* 303.6809 (1991), pp. 1019–1022.

[291] Matthew A Sabin et al. “Insulin and BMI as predictors of adult type 2 diabetes mellitus”. In: *Pediatrics* 135.1 (2015), e144–e151.

[292] Janina Petkeviciene et al. “Anthropometric measurements in childhood and prediction of cardiovascular risk factors in adulthood: Kaunas cardiovascular risk cohort study”. In: *BMC Public Health* 15.1 (2015), p. 218.

[293] Juha Koskinen et al. “Youth overweight and metabolic disturbances in predicting carotid intima-media thickness, type 2 diabetes, and metabolic syndrome in adulthood: the Cardiovascular Risk in Young Finns study”. In: *Diabetes Care* 37.7 (2014), pp. 1870–1877.

[294] Chris Power and Claudia Thomas. "Changes in BMI, duration of overweight and obesity, and glucose metabolism: 45 years of follow-up of a birth cohort". In: *Diabetes care* 34.9 (2011), pp. 1986–1991.

[295] John A Morrison et al. "Childhood predictors of adult type 2 diabetes at 9-and 26-year follow-ups". In: *Archives of pediatrics & adolescent medicine* 164.1 (2010), pp. 53–60.

[296] HPS Sachdev et al. "Predicting adult metabolic syndrome from childhood body mass index: follow-up of the New Delhi birth cohort". In: *Archives of disease in childhood* 94.10 (2009), pp. 768–774.

[297] Santosh K Bhargava et al. "Relation of serial changes in childhood body-mass index to impaired glucose tolerance in young adulthood". In: *New England Journal of Medicine* 350.9 (2004), pp. 865–875.

[298] Aviva Must et al. "Long-term morbidity and mortality of overweight adolescents: a follow-up of the Harvard Growth Study of 1922 to 1935". In: *New England journal of medicine* 327.19 (1992), pp. 1350–1355.

[299] Morten Schmidt et al. "Obesity in young men, and individual and combined risks of type 2 diabetes, cardiovascular morbidity and death before 55 years of age: a Danish 33-year follow-up study". In: *BMJ open* 3.4 (2013), e002698.

[300] M Jeffreys et al. "Lifecourse weight patterns and adult-onset diabetes: the Glasgow Alumni and British Women's Heart and Health studies". In: *International journal of obesity* 30.3 (2006), p. 507.