Adult height and the risk of cardiovascular diseases: investigating causality using genetic associations in two ethnically diverse populations



Andrew Brent Linden

Brasenose College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2018

Abstract

Adult height and the risk of cardiovascular diseases: investigating causality using genetic associations in two ethnically diverse populations

Andrew Brent Linden, Brasenose College, University of Oxford
A thesis submitted for the degree of Doctor of Philosophy, Michaelmas 2018

Background Taller adult height has been associated with lower risk of most vascular and respiratory diseases, but higher risk of other diseases (cancer and some vascular diseases) in studies predominantly from European populations. The causal relevance of height for different cardiovascular diseases remains unclear.

Methods Instrumental variables for height were constructed from up to 3200 genetic variants having genome-wide significant association with height in published European ancestry based sources. Conventional observational and causal genetic associations (Mendelian randomisation analyses) of 1 SD (~6.5 cm) taller height with cardiovascular diseases and risk factors were assessed in 1M participants from the UK Biobank (UKB) and the China Kadoorie Biobank (CKB) studies, supplemented with data from CARDIoGRAMplusC4D and METASTROKE consortia.

Results The genetic instruments selected explained 16.9% of the variance of height in UKB but only 10.5% in CKB. The association of genetically elevated height was inversely associated with coronary artery disease (odds ratio [OR] per 1 SD taller height: 0.87; 95% CI: 0.84, 0.89; n=78 960 cases), systolic blood pressure and LDL cholesterol, positively with atrial fibrillation (1.32; 95% CI: 1.27, 1.38; n=13 365) and lean body mass, but null with ischaemic stroke overall. However, for ischaemic strokes subtypes available in METASTROKE, the association was inverse with large-vessel stroke (OR: 0.91; 95% CI: 0.84, 0.99; n=2167) and positive with cardioembolic stroke (1.13; 95% CI: 1.05, 1.22; n=2365). Genetically elevated height was also positively associated with all ischaemic stroke among UKB and CKB participants with atrial fibrillation. Observational and genetic associations with diseases were directionally consistent, and were inverse with haemorrhagic stroke (n=7819, mostly from CKB).

Conclusions Genetically elevated height is inversely associated with occlusive cardiovascular diseases, potentially mediated by blood lipids and blood pressure, but also positively associated with atrial fibrillation and cardioembolic stroke, potentially through its association with lean body mass.

Table of contents

Abstract ..... **i**

Table of contents ..... **ii**

List of tables ..... **v**

List of figures ..... **viii**

List of appendices ..... **xi**

Abbreviations ..... **xii**

1. Introduction **1**

1.1 Height and disease 1

1.2 The need for further research 8

1.3 Objectives of the thesis 9

1.4 Tables 11

2. Systematic review **14**

2.1 Objectives of the systematic review 14

2.2 Methods 16

2.3 Results 19

2.4 Discussion 28

2.5 Tables 32

3. Study design and general methods **43**

3.1 UK Biobank study design 43

3.2 China Kadoorie Biobank study design 45

3.3 Height phenotypes 48

3.4 Disease outcomes 49

3.5 Exclusions 50

3.6 General statistical methods 50

3.7 Tables 52
3.8 Figures 54
4. Baseline correlates of height 55
4.1 Introduction 55
4.2 Methods 56
4.3 Results 58
4.4 Discussion 62
4.5 Tables 65
4.6 Figures 72
5. Prospective associations of height with cardiovascular diseases 76
5.1 Introduction 76
5.2 Methods 78
5.3 Results 81
5.4 Discussion 86
5.5 Tables 92
5.6 Figures 93
6. Genetic instruments for height in UKB and CKB 100
6.1 Introduction 100
6.2 Methods 106
6.3 Results 117
6.4 Discussion 123
6.5 Tables 128
6.6 Figures 139
7. Associations of genetically instrumented height with cardiovascular diseases 147
7.1 Introduction 147
7.2 Methods 151

7.3 Results 158
7.4 Discussion 167
7.5 Tables 178
7.6 Figures 183
8. Conclusions 191
8.1 Summary of findings 191
8.2 Strength and limitations of the study 193
8.3 Further research 194
8.4 Tables 197
Appendix 198
Supplementary tables for chapter 2 198
Supplementary figures for chapter 2 201
Supplementary tables for chapter 3 205
Supplementary figures for chapter 3 207
Supplementary tables for chapter 4 208
Supplementary figures for chapter 4 212
Supplementary tables for chapter 5 213
Supplementary figures for chapter 5 215
Supplementary appendix for chapter 5 217
Supplementary tables for chapter 7 219
Supplementary figures for chapter 7 227
Acknowledgements 228
References 229

List of tables

Title	Page no.
Table 1.1 Associations of taller standing height with selected major diseases in Europeans (from the Emerging Risk Factors Collaboration [ERFC] meta-analysis and Rosenberg et al.) and Asians (from the Lee et al. meta-analysis).	11
Table 1.2 Previously published observational and genetic associations of taller height with cardiovascular diseases, and a comparison with the number of events available in UKB and CKB.	13
Table 2.1 Search criteria used for the systematic review.	32
Table 2.2 Summary of search 1 results: published studies of the associations of the components of adult height (standing, sitting, and leg height) with major diseases.	33
Table 2.3 Summary of search 2 results: published studies of the associations of adult height with major diseases in East Asian populations.	36
Table 2.4 Summary of search 3 results: published studies of the causal relevance of the associations of adult height with major diseases.	39
Table 2.5 Summary of search 4 results: published studies of the genetics of height in East Asian populations.	41
Table 3.1 Standard deviation (SD) of measured height, by sex.	52
Table 3.2 International Classification of Diseases (10th revision) (ICD-10) and Office of Population Censuses and Surveys Classification of Interventions and Procedures (version 4) (OPCS-4) codes of investigated outcomes.	53
Table 4.1 Baseline characteristics of included study participants (excluding participants with prior major cardiovascular diseases), shown separately for men and women.	65
Table 4.2 Cross-sectional associations of 1 SD taller standing height with binary baseline characteristics, shown separately for men and women, and adjusted for age, year of birth, and region in CKB.	66
Table 4.3 Cross-sectional associations of 1 SD taller standing height with continuous baseline characteristics, shown separately for men and women, and adjusted for age, year of birth, and region in CKB.	68
Table 4.4 Pearson correlation coefficients of selected continuous traits with standing, sitting, and leg height at baseline.	70
Table 4.5 Pearson correlation coefficients of selected continuous traits with standing height at baseline and resurvey.	71
Table 5.1 Associations of taller standing height with prevalent or incident atrial fibrillation, shown for different adiposity adjustments.	92
Table 6.1 Size and ancestry of genome-wide association studies (GWAS) for height.	128

Table 6.2 Size and ancestry of large genome-wide association studies (GWAS) consortia for various traits. 129

Table 6.3 SNPs associated with height that reached genome-wide significance (identified in the literature) available in UKB and CKB. 130

Table 6.4 SNPs associated with height that reached genome-wide significance (identified in GIANT+UKB and by clumping) available in UKB and CKB. 131

Table 6.5 Number of available and missing height-associated SNPs in UKB and CKB, subdivided by minor allele frequency (MAF) ranges. 132

Table 6.6 Percentage of the variance of height explained by genetic risk scores constructed using literature effect sizes in UKB and CKB. 133

Table 6.7 Estimated attribution of the lower R^2 of GIANT+UKB SNP based genetic instrument in CKB compared to in European ancestry populations. 134

Table 6.8 Power of genetic instruments to detect odds ratios per 1 SD genetically elevated height, assuming 6.5K cases (approximate number of haemorrhagic stroke cases in the CKB genotyped dataset), a sample of 95K, and a Type I error rate of 5%. 136

Table 6.9 Power of genetic instruments to detect odds ratios for ischaemic stroke per 1 SD genetically elevated height, in METASTROKE, UKB and CKB, assuming a Type I error rate of 5%. 137

Table 6.10 Associations of the genetic risk scores with height in UKB and CKB. 138

Table 7.1 Associations of height-associated SNPs with ischaemic stroke and coronary artery disease using data from the CARDIoGRAMplusC4D (C+C4D) and METASTROKE consortia. 178

Table 7.2 Associations of genetically instrumented height with continuous traits in UKB and CKB. 179

Table 7.3 Associations of genetically instrumented height with atrial fibrillation in UKB and CKB, shown for different adiposity and blood pressure adjustments. 181

Table 7.4 Sensitivity analyses of standardised lean body mass and standardised systolic blood pressure for UKB (336433 participants). 182

Table 8.1 Number of selected events for phenome-wide analyses of height in UKB and CKB participants with genotyping. 197

Supplementary Table 2.1 Associations of genetically elevated height and cardiovascular risk factors, extracted from results of systematic search 3 (published studies of the causal relationship between the genetics of height and major diseases). 198

Supplementary Table 3.1 Additional information about UKB outcome variables. 205

Supplementary Table 3.2 Additional information about the CKB outcome variables. 206

Supplementary Table 4.1 Mean standing height across 10 regions in China, shown separately for men and women, after adjusting for age and year of birth. 208

Supplementary Table 4.2 Number of participants by individual year of birth and 5-year age group in UKB and CKB. 209

Supplementary Table 5.1 Pearson self-correlations between baseline and resurvey for exposure variables and covariates in UKB and CKB, shown for resurvey participants and for non-hypertensive participants at baseline. 213

Supplementary Table 7.1 Associations of genetically instrumented height with different blood pressure phenotypes in UKB and CKB. 219

Supplementary Table 7.2 Associations of genetically instrumented height with coronary artery disease, stroke types and atrial fibrillation in CKB before and after exclusion of related individuals. 221

Supplementary Table 7.3 Associations of genetically instrumented height with coronary artery disease, stroke types and atrial fibrillation before and after the inclusion of the height phenotype as a covariate in the model. 222

Supplementary Table 7.4 List of height-associated SNPs from Wood et al. showing heterogeneity with the common lean body mass effect or identified as outliers by the MR-PRESSO method. 223

Supplementary Table 7.5 List of height-associated SNPs from Wood et al. showing heterogeneity with the common systolic blood pressure effect or identified as outliers by the MR-PRESSO method. 225

List of figures

Title	Page no.
Figure 3.1 Measurement of sitting height in CKB using the Stadiometer device. (*Source: CKB Standard Operating Procedure on "Measurement of standing and sitting heights using Stadiometer", 2013*)	54
Figure 4.1 Distribution of standing height, shown separately for men and women in UKB and CKB.	72
Figure 4.2 Standing, sitting and leg height by year of birth, shown separately for men and women, and adjusted for region in CKB.	73
Figure 4.3 Standing, sitting and leg height by age at baseline, shown separately for men and women, and adjusted for region in CKB.	74
Figure 4.4 Systolic blood pressure and diastolic blood pressure by standing height, shown separately for men and women, and adjusted for age, year of birth, and region in CKB.	75
Figure 5.1 Associations of taller standing, sitting, and leg height with incident major coronary events and stroke types, shown for increasing levels of adjustment.	93
Figure 5.2 Combined associations of taller standing height with incident major coronary events and stroke types in UKB (n=496228) and CKB (n=513214), after basic adjustments and adjustment for major CVD and socio-economic risk factors.	94
Figure 5.3 Associations of taller standing height with incident major coronary events, by various baseline characteristics, after basic adjustments and adjustment for major CVD and socio-economic risk factors.	95
Figure 5.4 Associations of taller standing height with incident ischaemic stroke, by various baseline characteristics, after basic adjustments and adjustment for major CVD and socio-economic risk factors.	96
Figure 5.5 Associations of taller standing height with incident haemorrhagic stroke, by various baseline characteristics, after basic adjustments and adjustment for major CVD and socio-economic risk factors.	97
Figure 5.6 Associations of taller standing, sitting, and leg height with prevalent or incident atrial fibrillation, shown for increasing levels of adjustment.	98
Figure 5.7 Associations of taller standing height with prevalent or incident atrial fibrillation, by various baseline characteristics, after basic adjustments, adjustment for major CVD and socio-economic risk factors, and adjustment for lean body mass.	99
Figure 6.1 Effect allele frequency in UKB (336433 participants) versus Wood et al. and CKB (96046 participants) versus GIANT+UKB.	139
Figure 6.2 The distributions of minor allele frequencies (MAFs) of Wood SNPs in UKB, CKB, and the literature, and the corresponding percentages of the variance of height	140

explained by genetic risk scores (constructed using Wood effect sizes), shown for 5% MAF ranges.

Figure 6.3 The distributions of minor allele frequencies (MAFs) of clumped GIANT+UKB SNPs in UKB, CKB, and the literature, and the corresponding percentages of the variance of height explained by genetic risk scores (constructed using GIANT+UKB effect sizes), shown for 5% MAF ranges. 141

Figure 6.4 Imputation quality and minor allele frequency for Wood SNPs in UKB (336433 participants) and clumped GIANT+UKB SNPs in CKB (96046 participants). 142

Figure 6.5 Univariate effect sizes of available Wood SNPs on standardised standing height (336433 UKB participants and 96046 CKB participants). 143

Figure 6.6 Univariate effect sizes of clumped GIANT+UKB SNPs on standardised standing height (96046 CKB participants). 144

Figure 6.7 Univariate effect sizes of clumped GIANT+UKB SNPs on standardised standing height (96046 CKB participants), classified by SNPs with good and weak information. 145

Figure 6.8 Univariate effect sizes of clumped GIANT+UKB SNPs on standardised standing height (96046 CKB participants), shown for SNPs with good information and classified by imputation quality ≥0.95 and <0.95. 146

Figure 7.1 Associations of genetically instrumented height with coronary artery disease and stroke types in UKB (336433 participants) and CKB (77419 controls). 183

Figure 7.2 Comparisons of observation and genetic associations of taller height with coronary disease and stroke types. 184

Figure 7.3 Associations of genetically instrumented height with atrial fibrillation in UKB (336433 participants) and CKB (77419 controls). 185

Figure 7.4 Associations of genetically instrumented height with ischaemic stroke subtypes in METASTROKE, and with related outcomes in UKB (336433 participants). 186

Figure 7.5 Univariate effect sizes of Wood height-associated SNPs on standardised lean body mass and standardised standing height in 336433 UKB participants. 187

Figure 7.6 Univariate effect sizes of Wood height-associated SNPs on standardised systolic blood pressure (SBP) and standardised standing height in 336433 UKB participants. 188

Figure 7.7 Funnel plots of instrumental variable strength against instrumental variable estimates for standardised lean body mass and standardised systolic blood pressure for Wood height-associated SNPs in 336433 UKB participants. 189

Figure 7.8 Diagrammatic representation of the associations of genetically instrumented height with coronary artery disease, ischaemic stroke subtypes, haemorrhagic stroke, and atrial fibrillation. 190

Supplementary Figure 2.1 Flow diagram illustrating search 1: published studies of the associations of the components of adult height (standing, sitting, and leg height) with major diseases. 201

Supplementary Figure 2.2 Flow diagram illustrating search 2: published studies of the associations of adult height with major diseases in East Asian populations. 202

Supplementary Figure 2.3 Flow diagram illustrating search 3: published studies of the causal relevance of the associations of adult height with major diseases. 203

Supplementary Figure 2.4 Flow diagram illustrating search 4: published studies of the genetics of height in East Asian populations. 204

Supplementary Figure 3.1 Data available in the UKB and CKB baseline surveys and resurveys. 207

Supplementary Figure 4.1 Distribution of sitting and leg height, shown separately for men and women in UKB and CKB. 212

Supplementary Figure 5.1 The shape of the associations of height with cardiovascular diseases (major coronary events, stroke types, and atrial fibrillation), after basic adjustments. 215

Supplementary Figure 5.2 Associations of taller standing height with incident major coronary events and stroke types, shown for further levels of adjustment. 216

Supplementary Figure 7.1 Associations of increasing quartiles of the height genetic risk score (GRS) with coronary artery disease, stroke types, and atrial fibrillation in UKB (n=336433) and CKB (77419 controls). 227

List of appendices

Title	Page no.
Appendix	198
Supplementary appendix for chapter 5	217

Abbreviations

The following abbreviations were used in the thesis without explanation:

Abbreviation	Description
BMI	Body mass index
CI	Confidence interval
FEV_1	Forced expiratory volume in one second
FVC	Forced vital capacity
HR	Hazard ratio
LDL	Low density lipoprotein
MAF	Minor allele frequency
OR	Odds ratio
SD	Standard deviation
SE	Standard error
SNP	Single nucleotide polymorphism

1. Introduction

1.1 Height and disease

Overview

Adult height is a widely used biomarker that differs across birth cohorts and populations. Adult height reflects the combined effects of genetic predisposition, environmental factors experienced in childhood, and shrinkage with aging. In addition, height is clustered with multiple traits, physiological systems, and diseases, the extent of which is not fully understood.

Taller adult height has been associated with a lower risk of several diseases (most vascular and respiratory diseases), but a higher risk of other diseases (most types of cancer and some vascular diseases, such as atrial fibrillation and pulmonary heart disease).[1–3] Genetic studies have suggested that the inverse association of height with coronary artery disease is causal, and that shared biological pathways that affect both attained adult height and the development of atherosclerosis may be the mechanism underlying such associations.[4] However, the causal relevance of the associations of height with other diseases remains unclear.

The associations of height with different diseases have been well-studied in Western countries, but less so in East Asian countries (see systematic review in Chapter 2). However, the associations of height with different diseases may differ in low- and middle-income countries due to different nutritional exposures, disease patterns, and confounding factors.[5,6] For example, China experienced a

great famine between the years of 1958 and 1961, which is believed to have had an impact on the attainment of adult height among children living or born during this period, and the disease profile of the Chinese population in general.[7,8] Understanding the mechanisms underlying the different associations of height with cardiovascular diseases could lead to interventions or novel therapeutic targets for disease prevention in the general population.

Observational evidence of the associations of height with disease

The inverse association of adult height with coronary heart disease was first reported in 1951.[9] Adult height has since been significantly associated with a wide range of diseases.[1,4,10–14] In a meta-analysis of prospective studies, involving more than 1M European individuals, the Emerging Risk Factors Collaboration (ERFC) reported inverse associations of taller adult height with major cardiovascular diseases. Table 1.1 shows the inverse associations of taller adult height with coronary heart disease (hazard ratio [HR] per 1 standard deviation [SD 6.5 cm] taller height 0.93; 95% CI: 0.91, 0.94; n=43 204), ischaemic stroke (0.94; 95% CI: 0.90, 0.97; n=5720) and haemorrhagic stroke (0.90; 95% CI: 0.85, 0.95; n=2169).[1] Some diseases, such as haemorrhagic stroke, are much more common in low-income countries.[5] A meta-analysis of prospective studies, involving 0.5M mostly Asian adults contributed few additional cases of coronary heart disease (3933 cases) but contributed a substantial number of stroke cases (1983 ischaemic strokes and 1534 haemorrhagic strokes).[2] Despite differences in mean adult height between European and Asian populations the associations of adult height with vascular disease in Asian populations were largely similar to those observed in European populations (Table 1.1).[1,2] Inverse associations of

taller height with chronic obstructive pulmonary disease have also been reported in European and Asian populations (3958 and 1597 deaths, respectively) (Table 1.1).[1,2]

By contrast, positive associations of taller height have been reported in Europeans for most types of cancers (almost 50K cancer deaths)[1] and for some types of vascular disease, including atrial fibrillation (1252 cases; odds ratio [OR] per 6.5 cm taller height [scaled to 6.5 cm and combined across men and women]: 1.18; 95% CI: 1.11, 1.25)[3], ruptured aortic aneurysm (1324 deaths)[1] and pulmonary embolism (714 deaths)[1] (Table 1.1). Additionally, in a prospective study of 1.3M British women in the Million Women Study, Green et al. (2011) reported a positive association of taller height with cancer for most sites based on almost 100K incident cancers.[14]

The large meta-analyses of height and diseases have focussed solely on standing height. No meta-analyses of large scale studies have yet been published for sitting and leg height with major diseases (see systematic review in Chapter 2). The components of height (standing, sitting or leg height) may have different associations with diseases as they are impacted differently during growth and aging, and this might be informative about the underlying mechanisms of disease. In addition, over the last century, mean adult height has been increasing in all populations,[15] suggesting that the rates of diseases that are causally associated with height may also be changing.

Observational associations are susceptible to distortion by confounding factors which may differ in different populations. For example, taller height has been associated with higher social status[16–18] which, in low-income countries may be associated with higher levels of obesity, but this may not be the case in high-income countries.[17,19] In addition, height is associated not only with disease but also with many other traits,[20,21] which may lead to residual confounding in observational analyses. Genetic analyses, such as Mendelian randomisation, can help overcome some of the limitations of traditional observational analyses.[22,23]

Mendelian randomisation analysis

Mendelian randomisation is a powerful tool that is widely used in population studies to investigate the causal relevance of associations of risk factors with disease or other outcomes. In Mendelian randomisation, genetic variants associated with a risk factor are used to form an instrumental variable (such as a genetic risk score) or variables to assess the associations of the genetically instrumented risk factor with the outcome, thereby avoiding the problems of residual confounding and reverse causation which are often present in observational associations.[22,23] For example, Mendelian randomisation studies of coronary artery disease have clarified the causal relevance of some risk factors (such as LDL cholesterol and lipoprotein[a]) and refuted the causal relevance of other risk factors (such as HDL cholesterol and C-reactive protein).[24,25]

Genetics of height

The heritability of height has been estimated in twin studies to be 70-90%.[26–29] A major genome-wide association study (GWAS) was performed on 0.25M participants of European ancestry by the Genetic Investigation of Anthropometric Traits (GIANT) consortium, an international collaboration set up to discover the genetic loci that modulate human body size and shape (Wood et al., 2014).[20] Wood discovered 697 SNPs (single nucleotide polymorphisms) that were significantly associated with height. These 697 SNPS explained about 16% of the variance of height (estimated using a genomic relatedness matrix restricted maximum likelihood [GREML] approach) and had an R^2 (squared correlation coefficient between the genetic predictors and height) of about 14% (estimated using genetic predictors in 5 validation populations).[20] These genetic loci for height were also associated with several different physiological systems, including cardiovascular, digestive, muscular skeletal, respiratory, and urogenital systems.[20]

Very recently, the GIANT consortium combined its data with 0.45M UKB participants and increased the number of SNPs that were significantly associated with height from 697 to 3290.[30] These 3290 height-associated SNPs now explained 24.6% of the variance of height (estimated using GREML) and had an R^2 of 19.7% (estimated using a genetic predictor in a validation population) in European populations.

Height has been estimated to have similar heritability in Europeans and East Asians (see Chapter 6). However, the height-associated variants have been

discovered in European populations and, hence, would not be expected to perform as well in non-Europeans (because the allele frequencies, linkage disequilibrium structures, and lead SNPs at loci may differ across ethnic populations).[31] In addition, there may be ethnic-specific height-associated SNPs that remain to be discovered.

Genetic evidence of the associations of height with cardiovascular diseases

The causal relevance of the observational associations of height with several vascular diseases has been supported by Mendelian randomisation studies (see systematic review in Chapter 2). A Mendelian randomisation analysis of height and major coronary events conducted by Nelson et al. (2015), involving 65 066 cases in CARDIoGRAMplusC4D (C+C4D) and 180 height-associated SNPs, demonstrated that 1 SD (6.5 cm) taller genetically instrumented height was associated with a 12% lower risk of coronary artery disease (OR: 0.88; 95% CI: 0.82, 0.95).[4] Thus, the strength of association estimated from genetically elevated height was almost twice as strong as those reported in traditional observational analyses in the ERFC meta-analysis.[1,4] In a Mendelian randomisation study based on 4848 cases of stroke from the C+C4D Consortium and 69 height-associated SNPs, no statistically significant association of genetically elevated height with stroke was found.[32] However, the associations of genetically elevated height with stroke types have not been previously reported and may differ between ischaemic and haemorrhagic stroke types.

The causal relevance of height with atrial fibrillation has been confirmed in a Mendelian randomisation study based on 209 height-associated SNPs and 879

atrial fibrillation cases from the Cardiovascular Health Study, where Rosenberg et al. (2014) reported a similar odds ratio to those reported in traditional observational analyses.[3,10]

Possible mediators of the associations of height with cardiovascular diseases

Possible mechanisms by which adult height is associated with cardiovascular diseases were investigated in a systematic review (Chapter 2). To identify potential mechanisms, Nelson et al. investigated the extent to which the associations of genetically elevated height were mediated by established cardiovascular risk factors using summary data on the associations of each SNP in a genome-wide panel with individual traits, such as blood pressure, cholesterol, and body mass index (BMI).[4] Nelson et al. concluded that about 30% of the inverse associations of genetically elevated height with coronary artery disease could be explained by LDL cholesterol and triglycerides (quantified using previously reported predictions of lipid-related markers and cardiovascular disease[33]), but that height was unrelated to blood pressure.[4] Nüesch et al. investigated the associations of genetically elevated height with cardiovascular risk factors and also reported inverse associations with LDL cholesterol and triglycerides and a null association with blood pressure.[4] In a recent large GWAS consortium meta-analysis of blood pressure, the International Consortium of Blood Pressure reported that some blood pressure SNPs were strongly positively associated and some were strongly negatively associated with height.[34]

Mendelian randomisation studies of height can also be used to explain environmental or social effects of height that can influence the development of

disease. For example, Tyrrell et al. (2016) in a Mendelian randomisation study based on 396 height-associated SNPs in 120K participants of British ancestry in UKB reported that shorter height was associated with lower educational attainment, lower job status, and less income, particularly in men.[17]

1.2 The need for further research

Value of further research

The causal relevance of height for other major diseases warrants further study. Previous studies of height and cardiovascular diseases have been largely limited to European populations. It is expected that the causal relationship of height with different cardiovascular diseases should be similar in different ethnic populations, including East Asian populations. However, if different associations are found then this might bring into question some deductions from European populations and lead to revised inferences. In addition, probably due to osteoporosis, sitting height is impacted more by aging than leg height. If some diseases were more strongly associated with one component of height (sitting or leg height), this might be informative about the underlying mechanisms of disease.

Opportunities in the UK Biobank and China Kadoorie Biobank studies

The China Kadoorie Biobank (CKB) and the UK Biobank (UKB) studies are ongoing, population-based, cohort studies, each involving 0.5M participants, that were designed to examine both the environmental and genetic determinants of disease.[35–38] All UKB participants and a random population sample of 77K CKB participants (enriched with an additional ~20K cases of cardiovascular and respiratory diseases) have been genotyped. The combined number of cases in

the UKB and CKB genotyped datasets (>16K ischaemic stroke cases, >8K haemorrhagic stroke cases, and >13K cases of atrial fibrillation) are much larger than those previously examined in Mendelian randomisation studies. (The potential contribution of UKB and CKB to the current literature is summarised in Table 1.2.) Since the publication of the most recent Mendelian randomisation studies of height and heart disease and stroke, the number of known genome-wide significant height-associated SNPs has increased substantially (from ~200 to ~3000 SNPs), which should result in much more powerful genetic instruments for height. Moreover, UKB and CKB also have individual participant data on a wide range of anthropometric and cardiovascular risk factors and, therefore, permit reliable analyses of the extent to which the associations of height with different types of cardiovascular diseases are mediated by established cardiovascular risk factors.

1.3 Objectives of the thesis

The aims of this thesis are to use the UKB and CKB resources and publicly available consortia data to:

(1) To determine the observational associations of adult standing height and its components (sitting and leg height) with major coronary events, stroke types and atrial fibrillation in two large prospective studies of Europeans and East Asians, each involving 0.5M participants.

(2) To use previously identified height-associated genetic variants to construct genetic instruments for height for application in European and East Asian populations.

(3) To use Mendelian randomisation approaches to assess the extent to which the observational associations of height with cardiovascular diseases are causal.

Note: As UKB and the consortia are publicly available resources, new reports overlapping the thesis work may emerge at any time, altering the rationale for the thesis. Therefore, one such report[39] which was published during the final stages of the thesis (25 October 2018) has been ignored except in the final chapter.

1.4 Tables

Table 1.1 Associations of taller standing height with selected major diseases in Europeans (from the Emerging Risk Factors Collaboration [ERFC][1] meta-analysis and Rosenberg et al.[3]) and Asians (from the Lee et al.[2] meta-analysis).*

	Europeans			Asians		
Disease outcome*	**No. of events**	**Hazard ratio (95% CI) per 1 SD (6.5 cm) taller standing height†**	**Source**	**No. of events**	**Hazard ratio (95% CI) per 1 SD (6 cm) taller standing height¶**	**Source**
All vascular death	46846	0.94 (0.93, 0.96)	ERFC[1]	-	-	-
Coronary heart disease	43204	0.93 (0.91, 0.94)	ERFC[1]	3933	0.93 (0.90, 0.96)	Lee[2]
Non-fatal myocardial infarction	16774	0.91 (0.89, 0.93)	ERFC[1]	-	-	-
Coronary death	7491	0.93 (0.90, 0.96)	ERFC[1]	-	-	-
Stroke	18502	0.93 (0.91, 0.96)	ERFC[1]	5389	0.93 (0.91, 0.96)	Lee[2]
Ischaemic stroke	5720	0.94 (0.90, 0.97)	ERFC[1]	1983	0.96 (0.92, 1.00)	Lee[2]
Haemorrhagic stroke	2169	0.90 (0.85, 0.95)	ERFC[1]	1534	0.90 (0.85, 0.95)	Lee[2]
Subarachnoid haemorrhage	1315	0.91 (0.84, 0.98)	ERFC[1]	-	-	-
Atrial fibrillation	1252	1.18 (1.11, 1.25)	Rosenberg[3]	577	2.07 (1.70, 2.52)	Suzuki[40]
Ruptured aortic aneurysm death	1324	1.12 (1.05, 1.20)	ERFC[1]	-	-	-
Pulmonary embolism death	714	1.12 (1.03, 1.21)	ERFC[1]	-	-	-
Diabetes death	1421	0.94 (0.87, 1.01)	ERFC[1]	504	0.85 (0.77, 0.95)	Song[41]
COPD and related conditions death	3958	0.84 (0.80, 0.89)	ERFC[1]	1597	0.94 (0.90, 0.99)	Lee[2]
Cancer death	47502	1.04 (1.03, 1.06)	ERFC[1]	7497	1.06 (1.04, 1.09)	Lee[2]

*Outcomes not specified as fatal ("death") included both fatal and non-fatal events. Atrial fibrillation included both prevalent and incident cases. COPD= chronic obstructive pulmonary disease.

†The hazard ratios from the ERFC meta-analysis (1M individuals of European ancestry) were adjusted for age, sex, year of birth, and smoking status.[1] The log hazard ratios from Rosenberg at al. (5860 individuals of European ancestry from the Cardiovascular Health Study) were scaled pro rata to 6.5 cm taller height and combined across men and women (the hazard ratios were adjusted for age, BMI, clinic site, race, and education).[3]

¶The log hazard ratios from the Lee et al. meta-analysis (0.5M individuals mostly of Asian descent) were combined across men and women (the hazard ratios were adjusted for age, year of birth, and study).[2] The relative risks from Suzuki et al. (4719 individuals of East Asian ancestry) were reported for the highest tertile of height relative to the lowest tertile (the relative risks were adjusted for age, sex, and left atrial dimension).[40] The log hazard ratio from Song et al. (345K women of East Asian ancestry) were scaled pro rata to 6 cm taller standing height (the hazard ratios were adjusted for age, systolic blood pressure, serum total cholesterol level, fasting blood glucose level, body mass index, cigarette smoking, alcohol consumption, regular exercise, monthly salary, occupation, and area of residence).[41]

Table 1.2 Previously published observational and genetic associations of taller height with cardiovascular diseases, and a comparison with the number of events available in UKB and CKB.

Outcome and type of association	No. of events			Relative risk (95% CI) per 1 SD taller height from the literature	Relative risk source
	Literature	UKB*	CKB*		
Coronary disease					
Observational (major coronary event)	24265	6632	8219	0.92 (0.90, 0.93)†¶	Emerging Risk Factor Collaboration[1]
Genetic (coronary artery disease)	65066	14072	4645	0.88 (0.82, 0.95)‡	Nelson et al. (CARDIoGRAMplusC4D consortium)[4]
Ischaemic stroke					
Observational	5720	2545	38269	0.94 (0.90, 0.97)†	Emerging Risk Factor Collaboration[1]
Genetic	-	3254	12791	-	
Haemorrhagic stroke					
Observational	2169	874	9789	0.90 (0.85, 0.95)†	Emerging Risk Factor Collaboration[1]
Genetic	-	1447	6757	-	
Atrial fibrillation					
Observational	1252	14912	672	1.18 (1.11, 1.25)§	Rosenberg et al. (Cardiovascular Health Study)[3]
Genetic	879	13225	191	1.19 (1.02, 1.38)**	Rosenberg et al. (Cardiovascular Health Study)[10]

*Observational number of events provided after exclusions of participants with prior major cardiovascular diseases.
†Adjusted for age, sex, smoking status, and year of birth.
¶Relative risks for non-fatal myocardial infarction and coronary death were combined to produce a major coronary event estimate.
‡The genetic association was based on 180 height-associated SNPs.
§The log relative risks were scaled pro rata to 6.5 cm taller height and combined across men and women. Relative risks were adjusted for age, BMI, clinic site, race, and education.
**The genetic association was based on 209 height-associated SNPs. The log relative risks were scaled pro rata to 6.5 cm genetically elevated height.

2. Systematic review

2.1 Objectives of the systematic review

The ERFC meta-analysis summarised the available evidence from prospective studies of the associations of standing height with major diseases.[1] However, the ERFC meta-analysis was largely limited to participants from European populations and focused exclusively on disease associations with standing height.

Less is known about the associations of sitting or leg height with major diseases. If some diseases were more strongly associated with one component of height (standing, sitting or leg height), this might provide pointers to mechanisms of disease. In addition, sitting and leg height may be affected differently by osteoporosis and the aging process.

In different ethnic populations, observational associations of height with cardiovascular diseases may differ due to different confounding factors. However, the causal effects, assessed using genetic instruments and Mendelian randomisation, would be expected to yield comparable results in different ethnic groups.

The wide-spread availability of global genetic consortia summary data (the associations of genome-wide variants with risk factors and disease outcomes) has greatly increased the power and utility of Mendelian randomisation studies to

address causality of risk factors with disease. Such consortia summary data are often used to generate a raft of genetic associations. For example, Bulik-Sullivan et al. used consortia summary data to estimate 276 genetic correlations among 24 traits and diseases.[21] These associations are often included in supplementary appendices of GWAS or Mendelian randomisation studies. Hence, the present thesis sought to identify all relevant genetic studies of the associations of height with cardiovascular diseases and risk factors, in order to elucidate the mechanisms underlying these associations.

For standing height, the major GWASs were largely performed on data collected by the GIANT consortium, which is an international collaboration that was set up to discover genetic loci that modulate human body size and shape.[20,42,43] The participants in these studies were mostly of European ancestry. SNPs associated with height, discovered in European populations, are likely to also be associated with height in other populations but may explain a lower proportion of variance of height due to differences in allele frequencies and linkage disequilibrium patterns between ethnic groups.[31] Analyses in the present thesis were based on UK and Chinese populations. Hence, it was necessary to identify additional relevant genetic studies that were conducted in East Asian populations.

Therefore, in this thesis, four systematic searches were conducted to identify previous studies that have addressed the following research questions:

(1) The associations of the components of adult height (standing, sitting, and leg height) with major diseases.

(2) The associations of adult height with major diseases in East Asian populations.

(3) The associations of genetically elevated height with major diseases and cardiovascular disease risk factors.

(4) The genetics of height in East Asian populations.

2.2 Methods

PubMed and Embase databases were used to perform four systematic searches:

- Search 1: The associations of the components of adult height (standing, sitting, and leg height) with major diseases. The following search criteria (defined in Table 2.1) and operators were used: *Height* AND (*Sitting* OR *Leg*) AND *Disease*.
- Search 2: The associations of adult height with major diseases in East Asian populations. The following search criteria and operators were used: *Height* AND *Disease* AND *East Asian*.
- Search 3: The causal relevance of the associations of adult height with major diseases. The following search criteria were used: *Height* AND *Genetics* AND *Disease*. In addition, the results of search 3 were used to extract associations of genetically elevated height with cardiovascular risk factors.
- Search 4: The genetics of height in East Asian populations. The following search criteria were used: *Height* AND (*Genetics* OR *SNP*) AND *East Asian*.

The searches excluded animal studies and studies involving children, and were restricted to full-length articles in published peer reviewed English language journals. Free text search terms were searched for in the title or abstract of articles in the databases. In searches 1-3, where the *Disease* criteria were used, studies were only included if the number of events was greater than 5% of the events in the ERFC meta-analysis[1] (listed in Table 1.1). Assistance with the systematic review was sought from Mrs Nia Wyn Roberts (Outreach Librarian for Cancer services, Gastroenterology, Population Health and Primary Care [University of Oxford Bodleian Health Care Libraries]) who advised on search platforms, search criteria, and other nuances of the systematic search process. The searches were conducted on the databases on 26 March 2017.

After the initial database search, the reference for each article was recorded and duplicate articles were removed. Articles were then screened for eligibility by title and abstract. Irrelevant articles were excluded, and the full-texts of the remaining articles were assessed for eligibility. If full-text articles were excluded, details of the reasons for such exclusions were recorded.

For included studies in searches 1-3, the following information was extracted: author, year of publication, a description of the study (including the genetic ancestry of the participants), the disease outcomes, the number of events of each outcome, details of the height exposure variable, relative risks of the disease outcomes, relative risk adjustments, and the number of height-associated SNPs (search 3 only). For included studies in search 4, the following information was extracted: author, year of publication, a description of the study, genetic ancestry

of the participants, the size of the discovery study, main findings (including the association significance level), and the novel SNPs that were identified.

The results of the associations of height with disease were presented using several different approaches. The relative risks for standing height were presented per SD, per 5 cm, per 6 cm, per 6.5 cm, per 10 cm height difference, or per height category (highest versus lowest height quartile, highest versus lowest height tertile, or height >166 cm versus <157 cm). In order for these relative risks to be easily compared, the relative risks were converted to per 6.5 cm taller standing height (the log relative risks were scaled pro rata to 6.5 cm). By expressing the relative risks per 6.5 cm standing height, the relative risks were consistent with the ERFC[1] and Nelson et al.[4] (where relative risks were reported per 6.5 cm standing height [~1 SD]). For articles where relative risks were given per height category, the relative risks were first converted to relative risks per linear increase in height. This was done using the mean and SD for height provided in the article and assuming a normal distribution to calculate the means for each height category by conditional tail expectation. The conditional tail expectation, assuming a normal distribution, was defined as:

$$\int_{Z_1}^{Z_2} \frac{x e^{-\frac{x^2}{2}}}{\sqrt{2\,\pi}}\,dx$$

where Z_1 is the Z-value of the lower bound of the quantile (or point of interest), and Z_2 is Z-value of the upper bound. For example, the conditional tail expectation for the lowest quartile, in standard units, was -0.318 (Z_1=-∞, and Z_2=-0.674). Where articles did not provide the mean and SD for height, the original relative risk per height category was reported in the systematic search results and a

footnote provided. Where articles only provided sex-specific relative risks, these were reported in the systematic search results separately for men and women.

In various studies the relative risks for sitting height, trunk length and leg height were presented per SD, per 4 cm, per 4.2 cm, and per 5 cm height difference, or per height category (highest versus lowest height quartile, sitting height >87.4 cm versus <82.2 cm, sitting height ≥88 cm versus < 86 cm, or leg height ≥78 cm versus <75 cm). These relative risks were converted to per 4.0 cm (~1 SD[44,45]) taller sitting height, trunk length, or leg height. For articles where relative risks were given per height category, the relative risks were first converted to relative risks per linear increase in height (using the same method of conditional tail expectation that was used for standing height).

2.3 Results

Search 1: the associations of the components of adult height (standing, sitting, and leg height) with major diseases

The search procedure for examining the associations of the components of adult height (standing, sitting, and leg height) with major diseases was summarised in a PRISMA[46] flow diagram (Supplementary Figure 2.1). A total of 689 articles were identified through database searches, of which 134 were duplicates. A total of 555 articles were screened by title and abstract of which 18 were screened by full-text for eligibility. Overall, a total of 10 studies were included in the systematic search results (Table 2.2).

The search identified studies involving a total of 1520 ischaemic heart disease deaths, 852 myocardial infarction deaths, 844 cerebrovascular deaths, 319 haemorrhagic stroke deaths, 97 ischaemic stroke deaths, and 2730 diabetes cases (Table 2.2). Sawada et al. (2017)[47], the only study with cardiovascular disease events, reported relative risks for standing height and sitting height separately for men and women. For standing height, the relative risk of ischaemic heart disease death in men (1099 events) was 0.82 (95% CI: 0.78, 0.86) per 6.5 cm taller standing height; and relative risk in women (421 events) was 0.85 (95% CI: 0.78, 0.91) per 6.5 cm taller standing height.[47] Consistent inverse associations of taller standing height were reported for myocardial infarction death (852 events), cerebrovascular disease death (844 events), haemorrhagic stroke death (319 events), and ischaemic stroke death (97 events) for men and women. For sitting height, the relative risk of ischaemic heart disease death in men (1099 events) was 0.81 (95% CI: 0.75, 0.89) per 4 cm taller sitting height; and relative risks in women (421 events) was 0.78 (95% CI: 0.69, 0.89) per 4 cm taller sitting height.[47] Slightly stronger sitting height associations (relative to standing height) were also reported for myocardial infarction death (852 events), cerebrovascular disease death (844 events), and haemorrhagic stroke death (319 events) in men and women, and ischaemic stroke death (48 events) in women.

For diabetes the evidence was unclear. Weitzman et al. (2010)[48] reported inverse associations of taller leg height with diabetes in white men (569 events), white women (439 events) and black women (324 events) (no statistically significant association was found in black men [184 events]), but did not examine the associations with standing or sitting height (Table 2.2). Consistent results were

also reported by Asao et al. (2006)[44], who reported no statistically significant associations of leg height or standing height with diabetes (839 events), and Lawlor et al. (2002)[45], who reported no statistically significant associations of sitting height or standing height with diabetes (375 events), but reported an inverse association with taller leg height.

The search identified studies with 9371 cancer deaths, 3274 breast cancer cases, and 1531 prostate cancer cases (Table 2.2). Most studies reported positive associations of taller standing, sitting, and leg height with cancer types. Sawada et al. (2017)[47] reported relative risks for standing height and sitting height separately for men and women. For standing height, the relative risks per 6.5 cm taller standing height were 1.04 (95% CI: 1.00, 1.07) for cancer death in men (3247 deaths) and 1.07 (95% CI: 1.03, 1.09) for cancer death in women (3662 deaths). For sitting height, the relative risk of cancer death in both men (3247 deaths) and women (3662 deaths) was 1.04 (95% CI: 1.00, 1.08) per 4 cm taller sitting height. In general, the associations of taller sitting and leg height with cancers were of a similar magnitude as the associations of standing height with cancers (Table 2.2).

Search 2: the associations of adult height with major diseases in East Asian populations

The search procedure for examining the associations of adult height with major diseases in East Asian populations was summarised in a PRISMA[46] flow diagram (Supplementary Figure 2.2). A total of 3167 articles were identified through database searches, of which 697 were duplicates. A total of 2470 articles were

screened by title and abstract of which 32 were screened by full-text for eligibility. Overall, a total of 6 studies were included in the systematic search results (Table 2.3).

The search identified studies with 3933 coronary heart disease cases, 2776 ischaemic stroke cases/deaths, 2131 haemorrhagic stroke cases/deaths, 577 atrial fibrillation cases, and 504 diabetes deaths (Table 2.3). In a meta-analysis of prospective cohort studies in Asian and Australasian populations (413 748 Asians and 97 052 Australasians), Lee et al. (2009),[2] the largest study of cardiovascular diseases identified by the search, reported an inverse association of taller standing height with coronary heart disease in men (2780 events; relative risk of 0.91 per 6.5 cm taller standing height; 95% CI: 0.88, 0.96) but no statistically significant association in women (1153 events; relative risk of 0.96; 95% CI: 0.89, 1.01). Lee et al. reported no statistically significant association of taller standing height with ischaemic stroke in men (1334 events; relative risk of 0.95 per 6.5 cm taller standing height; 95% CI: 0.89, 1.01) or women (649 events; relative risk of 0.98; 95% CI: 0.89, 1.07), but reported inverse associations of taller standing height with haemorrhagic stroke in men (1050 events; relative risk of 0.90; 95% CI: 0.84, 0.96) and women (484 events; relative risk of 0.86; 95% CI: 0.77, 0.96). However, in a prospective cohort study of Japanese adults, Suzuki et al. (2010)[40] reported a strong positive association of taller standing height with atrial fibrillation (577 events; relative risk of 2.00 per 6.5 cm taller standing height; 95% CI: 1.66, 2.41). For diabetes, in a prospective cohort of Korean women, Song et al. (2008)[41] reported an inverse association of taller

standing height with diabetes death (504 deaths; relative risk of 0.83 per 6.5 cm taller standing height; 95% CI: 0.75, 0.95).

The search identified studies with 50 212 cancer cases or deaths (Table 2.3). Positive associations of taller standing height with cancers were generally found. In a prospective cohort study of Koreans, Sung et al. (2009)[49] reported a positive association of taller standing height with all cancers in men (23 725 events; relative risk of 1.07 per 6.5 cm taller standing height; 95% CI: 1.04, 1.08), and in women (9443 events; relative risk of 1.09; 95% CI: 1.07, 1.12). In a meta-analysis of prospective cohort studies in Asian and Australasian populations (413 748 Asians and 97 052 Australasians), Lee et al. (2009)[2] reported a positive association of taller standing height with cancer death in men (5232 events; relative risk of 1.05 per 6.5 cm taller standing height; 95% CI: 0.88, 0.96) and in women (2265 events; relative risk of 1.10; 95% CI: 1.05, 1.15). Positive associations of taller standing height with cancer death was also reported by Song et al. (2003)[50] in a prospective cohort of Korean men (5769 deaths; relative risk of 1.03 per 6.5 cm taller standing height; 95% CI: 1.00, 1.07) and by Song et al. (2008)[41] in a prospective cohort of Korean women (3778 deaths; relative risk of 1.07; 95% CI: 1.03, 1.12).

Search 3: the causal relevance of the associations of adult height with major diseases

The search procedure for examining the causal relevance of the associations of adult height with major diseases was summarised in a PRISMA[46] flow diagram (Supplementary Figure 2.3). A total of 2208 articles were identified through

database searches, of which 482 were duplicates. A total of 1727 articles were screened by title and abstract of which 16 were screened by full-text for eligibility. Overall, a total of 11 studies were included in the systematic search results, of which 9 were Mendelian randomisation studies, 1 was a twin study, and 1 was based on genetic correlations (Table 2.4).

The search identified studies with up to 65 066 coronary artery disease cases from the C+C4D[4,32,51] consortium, 4848 stroke cases (from studies forming part of the C+C4D[32] consortium dataset), 879 atrial fibrillation cases, and 35 566 diabetes cases (34 840 from the DIAGRAM[4] consortium) (Table 2.4). In a Mendelian randomisation study based on 180 height-associated SNPs and 65 066 cases of coronary artery disease from the C+C4D consortium, Nelson et al. (2015)[4] reported that 1 SD (6.5 cm) genetically elevated height was associated with a 12% lower risk of coronary artery disease (OR: 0.88; 95% CI: 0.82, 0.95) (Table 2.4). Similar results reported by Nüesch et al. (2016)[32] and van Iperen et al. (2016)[51] greatly overlap Nelson et al. (2015)[4] as they were based on the same consortium data. In a Mendelian randomisation study based on 69 height-associated SNPs and 4848 cases of stroke from studies forming part of the C+C4D consortium dataset, Nüesch et al. (2016)[32] reported no statistically significantly association of genetically instrumented height with stroke (relative risk per unit [6.5 cm] genetically elevated height: 0.97; 95% CI: 0.79, 1.19) (Table 2.4). By contrast, based on 879 cases of atrial fibrillation and 209 height-associated SNPs, Rosenberg et al. (2014)[10] reported a positive association of genetically elevated height with atrial fibrillation (relative risk per unit [scaled to 6.5 cm] genetically elevated height: 1.19; 95% CI: 1.02, 1.38).

Based on 180 height-associated SNPs and 34 840 diabetes cases (from the DIAGRAM consortium), Nelson et al. (2015)[4] reported no statistically significant association of genetically instrumented height with type 2 diabetes (relative risk of 0.95 per unit [6.5 cm] genetically elevated height; 95% CI: 0.83, 1.07) (Table 2.4). Based on the same dataset and 697 height-associated SNPs, Bulik-Sullivan et al. (2015)[21] confirmed that there was no statistically significant genetic correlation between height and diabetes (genetic correlation of -0.02; 95% CI: -0.11, 0.07). Davies et al. (2015),[52] based on 179 height-associated SNPs and 726 diabetes cases (from the ProtecT population-based study) reported no statistically significant association of genetically instrumented height with diabetes (relative risk of 0.94 per unit [6.5 cm] genetically elevated height; 95% CI: 0.70, 1.25).

Of the 11 studies identified in the search, 4 Mendelian randomisation studies also reported associations of genetically elevated height with cardiovascular risk factors, which included BMI (249 796 participants in the largest study), lung function (11 131 participants), systolic blood pressure (69 899 participants), LDL cholesterol (95 454 participants), glucose (96 496 participants), and smoking (41 150 participants) (Supplementary Table 2.1). In addition, Bulik-Sullivan et al. (2015)[21] reported genetic correlations between height and cardiovascular risk factors which included obesity measures (236 231 participants), LDL cholesterol (95 454 participants), and education (101 069 participants) (Supplementary Table 2.10.

Based on 180 height-associated SNPs and 95 454 participants (from 46 participating studies compiled by Teslovich [2010][53]), Nelson et al. (2015)[4] reported an inverse association of genetically elevated height with LDL cholesterol (-0.06 SD units of LDL cholesterol per unit [6.5 cm] genetically elevated height; 95% CI: -0.09, -0.04) (Supplementary Table 2.1). Consistent results, based on the same dataset and 697 height-associated SNPs, were reported by Bulik-Sullivan et al. (2015)[21] (an inverse genetic correlation between height and LDL cholesterol of -0.09; 95% CI: -0.16, -0.02). Likewise, based on 69 height-associated SNPs and 41 477 participants (from 21 included studies), Nüesch et al. (2016)[32] reported an inverse association of genetically elevated height with non-HDL cholesterol (-0.12 SD units of non-HDL cholesterol per unit [6.5 cm] genetically elevated height; 95% CI: -0.17, -0.06).

Based on 180 height-associated SNPs and 69 899 participants (from International Consortium for Blood Pressure[54]), Nelson et al. (2015)[4] reported no statistically significant association of genetically instrumented height with systolic blood pressure, adjusted for BMI (0.34 mm Hg of systolic blood pressure unit [per 6.5 cm] genetically elevated height; 95% CI: -0.31, 1.00) (Supplementary Table 2.1). Likewise, based on 69 height-associated SNPs and 52 345 participants, Nüesch et al. (2016)[32] also reported no statistically significant association of genetically instrumented height with measured systolic blood pressure.

The search identified studies with 62 328 breast cancer cases (46 325 cases from the BCAC[55] consortium and 16 003 from the DRIVE[55,56] consortium), 35 008 prostate cancer cases (20 848 cases from the PRACTICAL[52] consortium and 14

160 from the ELLIPSE[56] consortium), 12 537 lung cancer cases (from the TRICL[56] consortium), 15 326 colorectal cancer cases (10 226 cases from the GECCO[57] and C-CFR[57] consortia, and 5100 cases from the CORECT[56] consortium), and 999 oesophageal adenocarcinoma (from the BEACON[58] consortium) (Table 2.4). Mendelian randomisation analyses of cancer consortia data reported positive associations of genetically elevated height with breast cancer (based on 168[55] and 423[56] height-associated SNPs), lung cancer (423 SNPs)[56] and colorectal cancer (346[56] and 696[58] SNPs); and no statistically significant associations were observed for prostate cancer (179[52] and 419[56] SNPs) or oesophageal adenocarcinoma (243 SNPs)[58] (Table 2.4).

Search 4: the genetics of height in East Asian populations

The search procedure for examining the genetics of height in East Asian populations was summarised in a PRISMA[46] flow diagram (Supplementary Figure 2.4). A total of 1403 articles were identified through database searches, of which 302 were duplicates. A total of 1101 articles were screened by title and abstract of which 12 were screened by full-text for eligibility. Overall, a total of 8 studies were included in the systematic search results (Table 2.5).

The studies reported 21 novel SNPs associated with height reaching genome-wide significance (P-value <5.0E-8) and 13 SNPs reaching candidate gene significance (5.0E-8≤ P-value <1.0E-5, or P-value corrected for multiple testing) (Table 2.5). He et al. (2015)[59], in a GWAS meta-analysis of 36 227 East Asians (Japanese, Chinese, Korean, Malay, and Filipino) based on ~2.7M genotyped or imputed SNPs, reported 17 novel and 81 previously identified genome-wide

significant height-associated SNPs. Hao et al. (2013)[60], in a GWAS meta-analysis of 6534 Chinese based on ~1.5M SNPs in one study and ~2M SNPs in the other, reported 3 novel and 2 previously identified genome-wide significant height-associated SNPs. Okada et al. (2010)[61], in a GWAS of 19633 Japanese based on ~0.4M SNPs, reported 1 novel and 7 previously identified genome-wide significant SNPs.

2.4 Discussion

The systematic search of the relationship between the components of height (sitting and leg height) and major diseases in adults identified studies with much lower numbers of events than have accumulated in UKB and CKB (Tables 1.2 and 2.2). For example, the identified studies reported 319 haemorrhagic stroke deaths and 97 ischaemic stroke deaths, while the combined results of UKB and CKB, included more than 10K haemorrhagic stroke cases and more than 40K ischaemic stroke cases. The results of the search showed little difference between the associations of standing height and the components of height (sitting and leg height) with major diseases.

The systematic search of the relationship between height and major diseases in East Asian adult populations (Table 2.3) identified generally concordant associations of height with major diseases in East Asians and Europeans (as reported by the ERFC[1] meta-analysis). However, the numbers of events reported in the studies in East Asian populations were much lower (approximately one tenth for coronary heart disease and one half for ischaemic stroke) than in European populations, as reported by the ERFC[1] meta-analysis. The number of

events in CKB greatly increases the number of events available in East Asian populations (more than 13-fold increase in the number of ischaemic stroke events, 4-fold increase in haemorrhagic stroke events, and 2-fold increase in the number of major coronary events) (Table 1.2).

The systematic search of the causal relationship between the genetics of height and major diseases in adults identified, mostly, Mendelian randomisation studies, the largest of which were conducted by consortia, particularly C+C4D for coronary heart/artery disease (Table 2.4).[4,32,51] No Mendelian randomisation studies of stroke types (ischaemic stroke or haemorrhagic stroke) were found, and one study of stroke (all types) reported no statistically significant association.[32] Most Mendelian randomisation studies were based on 180 or fewer height-associated SNPs, which were identified by Lango et al. (2010)[62] and explained 10.5% of height variance (R^2). Since then, Wood et al. (2014)[20] identified 697 height-associated SNPs which explained ~14% of height variance (R^2), and Yengo et al. (2018)[43] identified 3290 height-associated SNPs which explained 19.7% of height variance (R^2).

Some studies have reported an inverse association of genetically elevated height with blood lipids (Supplementary Table 2.1).[4,21,32] Nelson et al. (2015)[4], from C+C4D consortium, concluded that the inverse associations of genetically elevated height with LDL cholesterol and triglycerides could explain ~30% of the inverse association of genetically elevated height with coronary artery disease. No statistically significant association of genetically elevated height with blood pressure has been reported (Supplementary Table 2.1).[4,21,32] However, the

present study has much more power than these studies to detect the association with blood pressure (particularly in UKB where there were 336K genotyped participants [Chapter 6.2]).

The systematic search of the genetics of height in East Asian populations identified studies which reported 21 novel SNPs associated with height reaching genome-wide significance in East Asian populations (Table 2.5). He et al. (2015)[59], the largest study (36K East Asian participants), reported 17 of these SNPs, Hao et al. (2013)[60] (6.5K Chinese participants) reported 3 of these SNPs, and Okada et al. (2010)[61] (20K Japanese participants) reported 1 of these SNPs. The largest GWAS of height available at the time these studies were published was Lango et al. (2010)[62], a GWAS meta-analysis of 134K Europeans which identified 180 height-associated SNPs. However, subsequent, larger GWAS of height have been published, namely Wood et al. (2014)[20] (253K European participants which identified 697 height-associated SNPs) and Yengo et al. (2018)[43] (694K European participants which identified 3290 height-associated SNPs). He et al. (2015)[59], although published after Wood et al. (2014)[20], did not include the SNPs identified by Wood in their comparisons as Wood et al. (2014) was published while their paper was under review. Therefore, the height-associated SNPs reported as novel in the studies included in the systematic search may not be independent of those identified in subsequent large GWAS of European populations. The CKB genotyped data contains ~100K genotyped Chinese participants (Chapter 6.2), which could be used to perform the largest GWAS of height in East Asian populations.

A limitation of the systematic reviews is their recency (conducted in March 2017), particularly since genetics research is such a rapidly evolving field.

2.5 Tables

Table 2.1 Search criteria used for the systematic review.

Criteria	MeSH terms for PubMed search	Subject headings for Embase search	Free text for PubMed and Embase searches
Height	body height	height OR body height	height
Sitting	trunk length	trunk	sitting OR trunk
Leg	leg length	leg length	leg
Disease	stroke OR coronary artery disease OR cardiovascular diseases OR atrial fibrillation OR pulmonary disease, chronic obstructive OR diabetes mellitus OR neoplasms	cerebrovascular disease OR occlusive cerebrovascular disease OR coronary artery disease OR ischemic heart disease OR cardiovascular disease OR heart atrial fibrillation OR chronic obstructive lung disease OR diabetes mellitus OR neoplasm	stroke OR artery disease OR cardiovascular disease* OR atrial fibrillation OR pulmonary disease* OR diabetes OR cancer
East Asian	Asia OR China OR Japan OR Asian continental ancestry group	Asia OR Asian OR East Asian OR China OR Chinese OR Japan OR Japanese	Asia* OR Chin* OR Japan*
Genetics	Mendelian randomisation analysis OR genetic association studies OR genome wide association study OR genetic variation	Mendelian randomisation analysis OR human genetics OR genetic variability OR genetic association OR genome-wide association study	Mendelian randomisation OR genetic* OR genom*
SNP	polymorphism, single nucleotide	single nucleotide polymorphism	single nucleotide polymorphism OR SNP

*Indicates wildcard search in free text search terms.

Table 2.2 Summary of search 1 results: published studies of the associations of the components of adult height (standing, sitting, and leg height) with major diseases.

Author, year	Study description	Disease outcome	No. of events	Sex (M/F/M&F)	Relative risk (95% CI) per 4 cm taller sitting height/ trunk length	Relative risk (95% CI) per 4 cm taller leg height	Relative risk (95% CI) per 6.5 cm taller standing height	Adjustments
Cardiometabolic diseases								
Sawada, 2017[47]	Prospective cohort study of 409748 Europeans (144962 men and 264786 women). Sitting height was measured in a subset of 253427 participants*	Ischaemic heart disease death	1099†	M	0.81 (0.75, 0.89)‡		0.82 (0.78, 0.86)‡	Recruitment centre, age, birth cohort, education level, smoking status, physical activity, alcohol consumption, weight, energy intake, menopausal status, and menopausal hormone use
			421	F	0.78 (0.69, 0.89)		0.85 (0.78, 0.91)	
		Myocardial infarction death	592	M	0.81 (0.75, 0.92)		0.82 (0.77, 0.87)	
			260	F	0.78 (0.66, 0.92)		0.86 (0.77, 0.95)	
		Cerebrovascular disease death	396	M	0.85 (0.75, 0.96)		0.88 (0.82, 0.96)	
			448	F	0.85 (0.78, 0.96)		0.95 (0.87, 1.03)	
		Haemorrhagic stroke death	133	M	0.75 (0.60, 0.92)		0.86 (0.74, 0.99)	
			186	F	0.89 (0.75, 1.08)		0.96 (0.85, 1.09)	
		Ischaemic stroke death	49	M	0.96 (0.69, 1.41)		0.81 (0.65, 1.01)	
			48	F	0.72 (0.50, 1.04)		0.85 (0.78, 0.91)	
Weitzman, 2010[48]	Prospective cohort study of 12801 US adults (4606 white men, 5298 white women, 1089 black men, and 1808 black women)	Diabetes mellitus	569	M		0.88 (0.79, 0.96) (white M)‡		Age, parental history of diabetes, parental socioeconomic status, and weight at age 25
			439	F		0.83 (0.74, 0.92) (white F)		
			184	M		0.99 (0.81, 1.21) (black M)		
			324	F		0.92 (0.79, 1.06) (black F)		
Asao, 2006[44]	Cross-sectional study of 7424 US adults	Type 2 diabetes	839	M&F		0.86 (0.73, 1.02)‡	0.91 (0.72, 1.02)	Sex, race, age, parental history of diabetes, physical activity, education, income, smoking, age at menarche, and percent body fat and its interaction with sex
Lawlor, 2002[45]	Cross sectional study of 3660 British women	Type 2 diabetes	375	F	1.00 (0.87, 1.16)‡	0.88 (0.78, 0.97)‡	0.91 (0.80, 1.03)‡	Systolic blood pressure, diastolic blood pressure, HDL, LDL, log triglycerides, weight, waist-to-hip ratio, smoking, adult social class, and childhood social class

Cancer and respiratory diseases

Sawada, 2017[47]	Prospective cohort study of 409748 Europeans (144962 men and 264786 women). Sitting height was measured in a subset of 253427 participants*	Cancer death	3247†	M	1.04 (1.00, 1.08)‡		1.04 (1.00, 1.07)‡	Recruitment centre, age, birth cohort, education level, smoking status, physical activity, alcohol consumption, weight, energy intake, menopausal status, and menopausal hormone use
			3662	F	1.04 (1.00, 1.08)		1.07 (1.03, 1.09)	
		Respiratory disease death	296	M	0.75 (0.66, 0.85)		0.96 (0.87, 1.05)	
			321	F	0.81 (0.72, 0.92)		0.95 (0.86, 1.04)	
Wang, 2011[63]	Prospective cohort study of 136202 Chinese adults (61333 men and 74869 women)	Cancer death	808†	M	1.09 (1.00, 1.19)§	1.10 (1.03, 1.16)§	1.13 (1.05,1.22)§	Birth calendar year, age, education, income, occupation, BMI, WHR, cigarette smoking, alcohol consumption, total physical activity, daily intake of energy, red meat, fruit and vegetables, and menopause status
			1654	F	1.07 (1.01, 1.12)	1.03 (0.98, 1.08)	1.06 (1.01, 1.12)	
Fagherazzi, 2012[64]	Prospective cohort study of 50704 French women	Breast cancer	2065	F	1.03 (0.84, 1.24)¶	1.21 (1.04, 1.42)¶	1.27 (1.02, 1.59)¶	Age, education level, total physical activity, age at menarche, nulliparity, age at first full-term pregnancy, breastfeeding, age at menopause, birth height, birth weight, menopausal status, use of oral contraceptives, family history of breast cancer, mammography during the previous follow-up period, history of benign breast disease, use of oral progestogens alone before menopause, and use of menopausal hormone therapy
Zuccolo, 2008[65]	Case-control study (1357 cases and 7990 controls) of British men	Prostate cancer	1357	M	1.03 (0.98, 1.08)‡	1.02 (0.97, 1.08)‡	1.04 (0.98, 1.10)‡	Age
Mellemkjær, 2012[66]	Prospective cohort study of 23864 Danish women	Breast cancer	1209	F	1.14 (1.06, 1.23)‡	1.09 (1.03, 1.16)‡	1.15 (1.08, 1.21)‡	Age, parity, age at birth of first child, age at menarche, use and duration of HRT, alcohol intake, length of schooling, and BMI

Rinaldi, 2012[67]	Prospective cohort study of 520000 Europeans (150000 men and 370000 women)	Differentiated thyroid carcinoma	58 508	M F		1.39 (1.09, 1.77)‡ 1.02 (0.92, 1.13)	1.43 (1.12, 1.87)‡ 1.16 (1.05, 1.28)	Recruitment centre, age, and smoking
Gunnel, 2001[68]	Systematic review of the following studies:**							
	Case-control study of 700 US women	Endometrial cancer	403	F	0.5 (0.3, 0.8)††		1.0 (0.6, 1.7)††	Age, education, parity, hormone replacement therapy, and smoking
	Prospective cohort study of 8006 Japanese men living in Hawaii	Prostate cancer	174	M	0.9 (0.7, 1.3)‡‡	1.2 (0.9, 1.8)‡‡		Age

*Number of men and women for sitting height not provided.
†Number of events were given for sitting height/ trunk length.
‡The log relative risks were scaled pro rata to 4 cm taller sitting height/ trunk length/ leg height or to 6.5 cm taller standing height.
§Relative risks given per SD taller trunk/ leg/ standing height. SD values not provided in article.
¶Relative risks have been expressed per linear increase in height (relative risks were given per height category) by assuming a normal distribution and using conditional tail expectations to estimate means of the height categories. The log relative risks were then scaled pro rata to 4 cm taller sitting height/ trunk length/ leg height or to 6.5 cm taller standing height.
**Studies in the review involving children, where number of events was less than 5% of the events in the ERFC meta-analysis[1], or where no relative risk was given, were excluded.
††Relative risks given for sitting height >87.4 cm versus <82.2 cm, and for standing height >166 cm versus <157 cm. Conversion of the relative risks per linear increase in height was not possible as mean and SD values for each height were not provided in the article.
‡‡Relative risks given for sitting height ≥88 cm versus <86 cm, and for leg height ≥78 cm versus <75 cm. Conversion of the relative risks per linear increase in height was not possible as mean and SD values for each height were not provided in the article.

Table 2.3 Summary of search 2 results: published studies of the associations of adult height with major diseases in East Asian populations.

Author, year	Study description	Ancestry	Disease outcome	No. of events	Sex (M/F/M&F)	Relative risk (95% CI) per 6.5 cm taller standing height	Adjustments
Cardiometabolic diseases							
Lee, 2009[2]	Meta-analysis of prospective cohort studies of 510800 adults (332020 men and 178780 women)	Asian (n=413748) and Australasian (n=97052)	Coronary heart disease	2780	M	0.91 (0.88, 0.96)*	Age, study, and year of birth
				1153	F	0.96 (0.89, 1.01)	
			Ischaemic stroke	1334	M	0.95 (0.89, 1.01)	
				649	F	0.98 (0.89, 1.07)	
			Haemorrhagic stroke	1050	M	0.90 (0.84, 0.96)	
				484	F	0.86 (0.77, 0.96)	
Suzuki, 2010[40]	Prospective cohort study of 4719 adults	Japanese	Atrial fibrillation	577	M&F	2.00 (1.66, 2.41)†	Age, sex, and left atrial dimension
Song, 2008[41]	Prospective cohort study of 344519 women	Korean	Diabetes mellitus death	504	F	0.83 (0.75, 0.95)*	Age, systolic blood pressure, serum total cholesterol level, fasting blood glucose level, body mass index, cigarette smoking, alcohol consumption, regular exercise, monthly salary, occupation, and area of residence
			Ischaemic stroke death	502	F	0.77 (0.66, 0.88)	
			Haemorrhagic stroke death	323	F	0.78 (0.70, 0.88)	
Honjo, 2011[69]	Prospective cohort study of 15564 adults	Japanese	Ischaemic stroke	291	M&F	0.82 (0.71, 0.95)*	Age, gender, area, education level, occupation, BMI, smoking status, alcohol intake, physical activity, perceived psychological stress, hypertension, diabetes, hyperlipidaemia, family history of stroke/coronary heart disease, marital status and, for women, menopausal status
			Haemorrhagic stroke	274	M&F	0.79 (0.68, 0.91)	

Cancer and respiratory diseases

Sung, 2009[49]	Prospective cohort study of 788789 adults (449214 men and 339575 women)	Korean	All cancers	23725	M	1.07 (1.04, 1.08)*	Age, body mass index, cigarette smoking, alcohol consumption, exercise, monthly salary, occupation, and area of residence
				9443	F	1.09 (1.07, 1.12)	
			Stomach cancer	8777	M	1.01 (0.99, 1.04)	
				2274	F	1.00 (0.94, 1.05)	
			Liver cancer	5809	M	1.07 (1.04, 1.11)	
				940	F	1.12 (1.03, 1.21)	
			Lung cancer	4453	M	1.09 (1.05, 1.13)	
				943	F	1.07 (0.99, 1.17)	
			Colon cancer	2499	M	1.05 (1.00, 1.11)	
				1007	F	1.11 (1.01, 1.20)	
			Rectum cancer	2281	M	1.08 (1.01, 1.13)	
				892	F	1.00 (0.92, 1.11)	
			Prostate cancer	1612	M	1.11 (1.04, 1.17)	
			Breast cancer	2102	F	1.24 (1.15, 1.34)	
Song, 2003[50]	Prospective cohort study of 386627 men	Korean	Cancer death	5769	M	1.03 (1.00, 1.07)*	Age, diastolic blood pressure, glucose, cholesterol, body mass index, alcohol drinking habits, smoking habits, exercise, monthly salary, occupation, and area of residence
Song, 2008[41]	Prospective cohort study of 344519 women	Korean	Cancer death	3778	F	1.07 (1.03, 1.12)*	Age, systolic blood pressure, serum total cholesterol level, fasting blood glucose level, body mass index, cigarette smoking, alcohol consumption, regular exercise, monthly salary, occupation, and area of residence

Lee, 2009[2]	Meta-analysis of prospective cohort studies of 510800 adults (332020 men and 178780 women)	Asian (n=413748) and Australasian (n=97052)	Cancer death	5232 2265	M F	1.05 (1.02, 1.09)* 1.10 (1.05, 1.15)	Age, study, and year of birth

*The log relative risks were scaled pro rata to 6.5 cm taller standing height.
†Relative risks have been expressed per linear increase in height (relative risks were given per height category) by assuming a normal distribution and using conditional tail expectations to estimate means of the height categories. The log relative risks were then scaled pro rata to 6.5 cm taller standing height.

Table 2.4 Summary of search 3 results: published studies of the causal relevance of the associations of adult height with major diseases.

Author, year	Study description*	No. of height-associated SNPs	Disease outcome	No. of events	Relative risk (95% CI) per unit (6.5 cm) genetically elevated height
Cardiometabolic diseases					
Nelson, 2015[4]	Mendelian randomisation analysis of CARDIoGRAMplusC4D consortium data	180 180	Coronary artery disease Type 2 diabetes	65066 34840	0.88 (0.82, 0.95) 0.95 (0.83, 1.07)
Nüesch, 2016[32]	Mendelian randomisation analysis of studies included in the CARDIoGRAMplusC4D consortium data†	69 180 69	Coronary heart disease Coronary heart disease Stroke	63746 10848 4848	0.90 (0.85, 0.95) 0.90 (0.78, 1.03) 0.97 (0.79, 1.19)
Bulik-Sullivan, 2015[21]	Genetic correlations across human diseases and traits, calculated from consortia data (DIAGRAM for diabetes)	697	Type 2 diabetes	34840	-0.02 (-0.11, 0.07) (genetic correlation)
van Iperen, 2016[51]	Mendelian randomisation analysis of CARDIoGRAM consortium data†	173	Coronary artery disease	22233	0.87 (0.80, 0.93)‡
Silventoinen, 2006[27]	Prospective cohort study of 35042 twin pairs from Scandinavia	N/A	Coronary heart disease death	5943	0.79 (0.69, 0.89)§ 0.93 (0.86, 1.02)¶
Rosenberg, 2014[10]	Mendelian randomisation analysis of a prospective cohort study of 3309 Europeans	209	Atrial fibrillation	879	1.19 (1.02, 1.38)‖
Davies, 2015[52]	Mendelian randomisation analysis of PRACTICAL consortium data	179	Diabetes	726	0.94 (0.70, 1.25)††

Cancer and respiratory diseases					
Zhang, 2015[55]	Mendelian randomisation analysis of BCAC and DRIVE consortia data	168	Breast cancer (BCAC) Breast cancer (DRIVE)	46325 16003	1.14 (1.08, 1.20)‖** 1.13 (1.03, 1.24)
Khankari, 2016[56]	Mendelian randomisation analysis of DRIVE, ELLIPSE, TRICL and CORECT consortia data (collectively forming the GAME-ON consortium)	423 419 423 346	Breast cancer Prostate cancer Lung cancer Colorectal cancer	16003 14160 12537 5100	1.12 (1.04, 1.20)‖ 1.02 (0.95, 1.10) 1.06 (1.00, 1.14) 1.35 (1.09, 1.66)
Davies, 2015[52]	Mendelian randomisation analysis of PRACTICAL consortium data	179	Prostate cancer	20848	0.99 (0.97, 1.01)††
Thrift, 2015[57]	Mendelian randomisation analysis of GECCO and C-CFR consortia data	696	Colorectal cancer	10226	1.55 (1.07, 2.34)‖**
Thrift, 2014[58]	Mendelian randomisation analysis of BEACON consortium data	243	Oesophageal adenocarcinoma	999	0.82 (0.60, 1.10)‖

*CARDIoGRAMplusC4D=Coronary ARtery DIsease Genome-wide Replication and Meta-analysis (CARDIoGRAM) plus The Coronary Artery Disease (C4D) Genetics; DIAGRAM=DIAbetes Genetics Replication And Meta-analysis; PRACTICAL=Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome; BCAC=Breast Cancer Association Consortium; DRIVE=Discovery, Biology, and Risk of Inherited Variants in Breast Cancer; ELLIPSE=Elucidating Loci Involved in Prostate Cancer Susceptibility; TRICL=Transdisciplinary Research for Cancer of Lung; CORECT=ColoRectal Transdisciplinary Study; GAME-ON=Genetic Associations and Mechanisms in Oncology; GECCO=Genetics and Epidemiology of Colorectal Cancer Consortium; C-CFR=Colorectal Cancer Family Registry; BEACON=International Barrett's and Esophageal Adenocarcinoma Consortium.
†Study greatly overlaps Nelson using CARDIoGRAM or C4D consortia cases.
‡Relative risks given per SD taller height. SD values not provided in article.
§Relative risk given for death of the taller monozygotic twin.
¶Relative risk given for death of the taller dizygotic twin.
‖The log relative risks were scaled pro rata to 6.5 cm genetically elevated height.
**Adjusted for age and principal components.
††Adjusted for principal components.

Table 2.5 Summary of search 4 results: published studies of the genetics of height in East Asian populations.

Author, year	Type of study	Study size (GWAS discovery)	Ancestry	Relevant findings	Identified novel SNPs
Studies reaching genome-wide significance (P-value <5.0E-8)					
He, 2015[59]	GWAS meta-analysis	36227	East Asian*	98 loci associated with height at P<5.0E-8 explaining 8.89% of height variance (17 loci were novel and 81 were previously reported)	rs4472734, rs3755206, rs3771381, rs1865760, rs11970475, rs2251830, rs10448080, rs4733789, rs174547, rs1938679, rs3809128, rs7184046, rs258324, rs2270518, rs7250071, rs12459943, and rs6030712
Okada, 2010[61]	GWAS	19633	Japanese	8 loci associated with height at P<5.0E-8 (1 locus was novel and 7 were previously reported)	rs12338076
Hao, 2013[60]	GWAS meta-analysis	6534	Chinese	5 loci associated with height at P<5.0E-8 (3 loci were novel and 2 were previously reported)	rs12612930, rs11021504, and rs11082671
Studies reaching candidate gene significance (5.0E-8≤ P-value <1.0E-5, or P-value corrected for multiple testing)					
Shen, 2016[70]	Candidate gene study	20426	Chinese	COL11A1 identified as a candidate gene associated with height at P<4.5E-3 (i.e. the multiple corrected threshold of P=0.05/11)	rs992100, and rs111980222

Kim, 2010[71]	GWAS	8842	Koreans	15 loci associated with height at P<1.0E-5 (8 loci were novel and 7 were previously reported)	rs10948197, rs11989122, rs10961780, rs7032940, rs2292303, rs1520223, rs2315504, and rs4811971
Hong, 2008[72]	Linkage analysis	7113	Chinese	Chromosome 17q24 identified a strong linkage to height at P=1.0E-5	No individual SNPs identified
Takeuchi, 2009[73]	Replication study	1530	Japanese	Replication of locus (ZFAT) previously reported as showing a suggestive association with eight in a Korean population. P=0.015 in the replication study (Japanese population).	rs1036821†
Lei, 2009[74]	GWAS	618	Chinese	5 loci associated with height at P<1.0E-5 (2 loci were novel and 3 were previously reported)	rs10816533, and rs7859940

*Japanese, Chinese, Korean, Malay, and Filipino.
†SNP reported in Wood et al. (2014)[20].

3. Study design and general methods

The present thesis was based on analyses of both the UK Biobank (UKB) and the China Kadoorie Biobank (CKB) studies.

3.1 UK Biobank study design

Recruitment

UKB is an ongoing, population-based, prospective cohort study, of 496 228 participants.[36] The UKB cohort is comprised of UK men and women, aged 40-69 years, who were recruited between 2006 and 2010. Standard operating procedures were used to collect participant information (demographic, socio-economic, medical, dietary, and lifestyle information), physical measurements, urine and blood samples (for genetic and biomarker analyses). During an assessment centre visit, a participant completed a self-administered touchscreen questionnaire (which was used to collect most of the study information) followed by a nurse-administered personal interview, physical measurements (height, waist and hip circumference, bio-impedance, spirometry, blood pressure, heart rate hand-grip strength, and heel bone ultrasound), and sample collection (urine and blood). Weight and percentage body fat were measured using a bio-impedance device (Tanita BC418MA body composition analyser). Fat body mass was calculated as weight multiplied by percentage body fat, and lean body mass was weight less fat body mass. Details of the survey methods and baseline characteristics have been described elsewhere.[36,37]

Measurement of height

Standing and sitting height were measured using a Seca 202 device.[75] Staff read measurements off analogue rulers and manually entered the readings into the IT system, which automatically flagged impossible or implausible values.[75] Participants were asked to remove their shoes prior to measuring standing height. Sitting height, the distance from rump to crown when sitting, was calculated as the difference between seated height and box size (the custom-made box on which the participant was seated during the seat height measurement). Leg height was calculated as standing height minus sitting height

Follow-up and resurvey

All participants are subsequently monitored for death and disease outcomes through electronic record linkage. The UKB study has electronic record linkage to electronic health records including cancer registries, death registries, and hospital episodes (including OPCS-4 hospital procedure codes). UKB participants have been followed for ~6 years. The analyses presented in this thesis were based on UKB data available at 31 March 2015. Details of the resurvey methods and long-term follow-up have been described elsewhere.[36]

In addition, data from two resurveys have been used (Supplementary Figure 3.1). The first UKB resurvey, conducted between 2012 and 2013, included ~20K randomly selected participants (~4 years after the baseline survey). The second resurvey, primarily for the collection of imaging data, was conducted between 2014 and 2015, included ~22K randomly selected participants (~7 years after the baseline survey). Approximately 17% of participants in the second resurvey were

also in the first resurvey. The second resurvey included measurements of carotid intima media thickness (cIMT), which was not measured at baseline.

Genotyping

All participants were genotyped using an 0.8M SNP panel in 106 batches (0.05M participants were genotyped using Affymetrix UK BiLEVE Axiom array and 0.44M were genotyped using an Affymetrix UK Biobank Axiom array). Further details of the genotyping methods are described in Chapter 6.2.

Research ethics and permissions

The most recent ethics approval for the UKB study was granted by the North West Multi-centre Research Ethics Committee (MREC) on 10 May 2016 (REC reference: 16/NW/0274). No additional research ethics approval was requested for this project. The UKB application number for this research is 10 061.

3.2 China Kadoorie Biobank study design

Recruitment

CKB is an ongoing, population-based, prospective cohort study, of 513 214 participants.[35] The CKB cohort is comprised of men and women, aged 30-79 years, who were recruited between 2004 and 2008 from 10 diverse regions of China (5 urban and 5 rural). Interviewer-administered electronic questionnaires were used to collect information on the participants, which included, inter alia, demographic, socio-economic, medical (including pre-existing diseases and related medications), dietary, and lifestyle information.[38] A range of physical measurements (height, waist and hip circumference, bio-impedance, spirometry,

blood pressure, and heart rate) were recorded using standard protocols and instruments with regular calibrations. For measurements such as waist and hip circumference, some women could only be measured over their underclothes. Weight and percentage body fat were measured using a bio-impedance device (Tanita BC-418MA body composition analyser). Fat body mass was calculated as weight multiplied by percentage body fat, and lean body mass was weight less fat body mass. In addition, blood samples (for genetic and biomarker analyses) were collected for long term storage. Details of the survey methods and baseline characteristics have been described elsewhere.[35,38]

Measurement of height

Standing and sitting height were measured using a Stadiometer device.[76] During the measurement process, participants were asked to remove shoes, hats and hairpins that would affect the measurements. For participants who were: in a wheelchair, unable to stand upright, had neck surgery, or had severe curvature of the spine, arm span was measured with a soft tape measure instead of height. Participants sat on the seat plate (part of the measurement device) for the sitting height measurements (Figure 3.1). Measurements were recorded to the nearest 0.1 cm. Standing and sitting height were measured directly, and leg height was calculated as standing height minus sitting height.

Follow-up and resurvey

All participants are subsequently monitored for death and disease outcomes through electronic record linkage. The CKB study has electronic record linkage to disease registries, death registries, and the national health insurance claim

database, which includes information on disease events and procedures. CKB participants have been followed for ~10 years. The analyses presented in this report were based on CKB data available at 31 December 2016. Details of the resurvey methods and long-term follow-up have been described elsewhere.[35,38]

In addition, data from two resurveys have been used (Supplementary Figure 3.1). The first CKB resurvey, conducted in 2008, included ~20K randomly selected participants (~3 years after the baseline survey). The second resurvey, conducted between 2013 and 2014, included ~25K randomly selected participants (~8 years after the baseline survey). Approximately 50% of participants in the second resurvey were also in the first resurvey. The second resurvey included carotid measurements (measures of cIMT and carotid plaque) and hand grip strength measurements, which were not measured at baseline.

Genotyping

A total of ~0.1M participants were genotyped as part of a nested case-control study. This study included a random population sample of 77K participants (which included the 25K participants who attended the second resurvey ~8 years after the baseline survey) and an additional ~20K cases (stroke types, ischaemic heart disease, and chronic obstructive pulmonary disease). Participants were genotyped using a 0.8M SNP panel in batches of ~50 plates (Affymetrix custom-designed biobank array for CKB). Further details of the genotyping methods are described in Chapter 6.2.

Research ethics and permissions

Ethics approval for the CKB study was granted by the Oxford Tropical Research Ethics Committee (OxTREC) on 3 February 2005 (reference: 025-04), and by the Ethics Review Committee of the Chinese Center for Disease Control and Prevention on 8 July 2004 (approval notice: 005/2004). No additional research ethics approval was requested for this project.

3.3 Height phenotypes

The standing, sitting, and leg height phenotypes were prepared following the methodology proposed by the Global Lipids Genetics Consortium and GIANT consortium to account for height differences by sex and age (Global Lipids Genetics/GIANT Consortium HRC and 1000 Genome Phase 3 Imputation and Analysis plan [unpublished analysis plan circulated by Doctor Robin Walters, April 2017]). For each biobank and component of height (standing, sitting, and leg height), the entire cohort (UKB or CKB) was stratified by sex (and also by region in CKB). Separately for each stratum, height was adjusted for age and age squared, and the residuals were transformed using inverse normal transformation. This gave the height phenotype in standard deviation units.

For UKB and CKB, respectively, the standard deviations of standing height were 6.8 cm versus 5.7 cm for men, and 6.2 cm versus 5.3 cm for women (Table 3.1). For UKB and CKB, respectively, the standard deviations of sitting height were 3.9 cm versus 3.3 cm for men, and 3.7 cm versus 3.1 cm for women, and the standard deviations of leg height were 4.6 cm versus 3.8 cm for men, and 4.3 cm versus 3.5 cm for women (Table 3.1).

3.4 Disease outcomes

Several major cardiovascular diseases were examined including: myocardial infarction and coronary heart disease death, coronary artery disease (which included myocardial infarction and coronary heart disease death, and coronary revascularisation procedures [percutaneous transluminal coronary angioplasty or coronary artery bypass grafting]), stroke types (ischaemic stroke and haemorrhagic stroke), and atrial fibrillation. The International Classification of Diseases (10th revision) (ICD-10) and the Office of Population Censuses and Surveys Classification of Interventions and Procedures (version 4) (OPCS-4) codes of these outcomes are presented in Table 3.2.

The UKB outcome adjudication group developed algorithms to classify selected disease outcomes (cross-sectional and prospective myocardial infarction and stroke types) with high positive predictive value (i.e. a high probability that a recorded event was in fact a true event).[77] The algorithms make use of participant self-reported data collected at the baseline survey (self-reported myocardial infarction and self-reported stroke types), linked hospital admissions data (hospital episode statistics), and death registry data. Where algorithmically defined outcomes were available in UKB, these were used. (The following outcomes were not algorithmically defined: coronary heart disease death, atrial fibrillation, and coronary revascularisation procedures [a component of coronary artery disease].)

The CKB outcomes examined in this thesis were unadjudicated. However, the process of outcome adjudication in CKB is ongoing and has reported that stroke

phenotyping has a positive predictive value of >90% and myocardial infarction has a positive predictive value of 70%.

Additional information about the UKB and CKB outcome variables, including data sources, links to the UKB data showcase website, CKB data request form codes are presented in Supplementary Tables 3.1 and 3.2.

3.5 Exclusions

Participants with prior major cardiovascular diseases were excluded from prospective analyses and most cross-sectional analyses (except cross-sectional analyses of prior major cardiovascular diseases). In UKB, prior major cardiovascular diseases included 14 561 participants with algorithmically defined or self-reported myocardial infarction or stroke and, in CKB, prior major cardiovascular diseases included 23 147 participants with self-reported coronary heart disease or stroke (including transient ischaemic attack).

3.6 General statistical methods

Linear regression was used for continuous outcomes, and logistic regression was used for cross-sectional analyses of disease outcomes, producing relative risks expressed as odds ratios (ORs). Cox regression was used for prospective analyses of disease outcomes, producing relative risks expressed as hazard ratios (HRs). Adjustments were made for potential confounders and well-established cardiovascular risk factors. Basic adjustments were defined as adjustment for age (divided into 5-year age groups), sex, year of birth, and region in CKB (10 regions).

In this thesis, all P-values were two-sided and all confidence intervals (CIs) were at 95%. Where appropriate, estimates of the relative risks for categorical variables were presented as floating absolute risks (whereby standard errors [SEs] were assigned nearly independently to each category so as avoid comparison to an arbitrary reference group).[78] Regression estimates were presented as points with area sized in proportion to the amount of statistical information they represent.

Analyses were performed using SAS version 9.4, R version 3.3.3, and PLINK version 1.90. Figures, particularly line and forest plots, were plotted using Jasper, an R plotting package developed by Doctor Matthew Arnold (in Professor Sarah Parish's team) and available across the Nuffield Department of Population Health. Additional methods are discussed in each chapter.

3.7 Tables

Table 3.1 Standard deviation (SD) of measured height, by sex.

	UKB		CKB	
Height	**Men**	**Women**	**Men**	**Women**
Standing height, cm	6.8	6.2	5.7	5.3
Sitting height, cm	3.9	3.7	3.3	3.1
Leg height, cm	4.6	4.3	3.8	3.5

Table 3.2 International Classification of Diseases (10th revision) (ICD-10) and Office of Population Censuses and Surveys Classification of Interventions and Procedures (version 4) (OPCS-4) codes of investigated outcomes.

Outcome	ICD-10 code	OPCS-4 code (UKB only)
Myocardial infarction	I21-I22	-
Coronary heart disease death	I20-I25 (fatal)	-
Major coronary event	I21-I22, I20-I25 (fatal)	-
Coronary artery bypass grafting	-	K40-K46
Percutaneous transluminal coronary angioplasty	-	K49, K50.1, K75
Coronary artery disease	I21-I22, I20-I25 (fatal)	K40-K46, K49, K50.1, K75
Ischaemic stroke	I63	-
Haemorrhagic stroke	I60-I61	-
Atrial fibrillation	I48	-

3.8 Figures



Figure 3.1 Measurement of sitting height in CKB using the Stadiometer device. (*Source: CKB Standard Operating Procedure on "Measurement of standing and sitting heights using Stadiometer", 2013*)

4. Baseline correlates of height

4.1 Introduction

In both the UK and China, mean height has increased by similar amounts over the last century, ~10.8 cm and ~10.2 cm, respectively (estimated from a study involving over 18M adults from 200 countries born between 1896 and 1996).[15] However, UK men and women remain ~5.7 cm and ~4.7 cm taller, respectively, than their Chinese counterparts.[15] The years between 1958 and 1961 were a time of great famine and social upheaval in China. This famine was the worst in human history and resulted in an estimated total of 15-30M premature deaths and a decrease in births by ~30M.[7,8] The physical development and adult height attainment of children living or born around this time may have been impacted by the famine. Using data from the China Health and Nutrition Surveys, which was a series of 5 cross-sectional surveys conducted over a 10-year period (1989-2000) (the first survey included 15917 participants), Chen et al. (2007) estimated that individuals of the 1959 birth cohort would have grown ~3 cm taller in the absence of famine.[8]

Adult height is influenced by the combined effects of age on height (the age effect) and the cohort (calendar year of birth) effects on height. In a meta-analysis of prospective studies involving more than 1M people, the ERFC reported that height was inversely associated with older age, decreasing, on average, by 1.4 cm per decade older age in adulthood.[1] In contrast, average height, adjusted to individuals aged 50 years, increased by ~0.5-1.0 cm per decade higher year of birth.[1] The components of height (sitting and leg height) may be influenced

differently by such effects. For example, it is expected that the effects of aging will have a more extreme impact on sitting height than on leg height due to the effects of osteoporosis.

Understanding the associations of height with confounding factors, such as age, year of birth, sex, and regional differences, is essential for reliable analysis and interpretation of the associations of height with disease. Investigating the cross-sectional associations of height with various factors measured at baseline (prior disease and use of medication, anthropometric, lifestyle and social factors) can also provide useful additional insights into potential mediating and confounding factors of the associations of height with disease.

This chapter explores the characteristics of the included study participants in addition to the cross-sectional associations of height with various factors measured at baseline.

4.2 Methods

Study populations

Details of the study designs of UKB and CKB have already been provided in Chapter 3. Briefly, a total of 496 228 participants in UKB and 513 214 participants in CKB were included in the cross-sectional analyses of height with prior major cardiovascular diseases (myocardial infarction [UKB only], stroke [UKB and CKB], and coronary heart disease [CKB only]). All other cross-sectional analyses were performed in participants with no prior major cardiovascular disease at baseline (481 667 participants in UKB and 490 067 participants in CKB).

The standing, sitting, and leg height phenotypes were normalised within strata of sex (and also region in CKB), which accounted for strata specific height differences (described in Chapter 3.3). Consistent with previous published studies,[1,4] an exposure variable of 1 SD taller height was used to quantify the cross-sectional associations of height with baseline characteristics. For UKB and CKB, respectively, 1 SD taller standing height was 6.8 cm versus 5.7 cm for men, and 6.2 cm versus 5.3 cm for women (SD for sitting and leg height are provided in Table 3.1). Deciles of measured height were used to investigate the shape of the associations of height with selected baseline characteristics. (For presentation purposes, when deciles were plotted, consecutive pairs of the middle 6 deciles were combined, with results presented for 7 groups.)

Statistical methods

The shape of the associations of calendar year of birth (and age at baseline) with measured standing, sitting, and leg height were investigated, in both men and women, by estimating standing height (and 95% CIs) for each calendar year of birth (and age at baseline), adjusted for region in CKB. (For presentation purposes, the years of birth, and ages, at the extremes of the distributions [for CKB, years of birth ≤1924 or ≥1983 and ages at baseline ≤24 or ≥79; and for UKB, years of birth ≤1937 or ≥1970 and ages at baseline ≤40 and ≥70] were combined to obtain reliable estimates.)

For binary and continuous baseline characteristics, logistic and linear regression, respectively, was used to quantify the associations with 1 SD taller standing

height, separately for men and women after basic adjustments. Binary baseline characteristics included prior disease and medications, lifestyle factors (smoking, alcohol consumption, and marital status), social factors (income and education), and self-rated frailty measures; and continuous baseline characteristics included anthropometric measures, blood pressure and heart rate, carotid measures and lipids, and lifestyle factors (smoking, alcohol consumption, and physical activity). The associations with binary and continuous characteristics were presented as ORs (and 95% CIs) and regression coefficients (and SEs), respectively, per 1 SD taller height. In addition, the shape of the associations of deciles of measured standing height with systolic and diastolic blood pressure were investigated separately for both men and women, after basic adjustments (age [5-year age bands], year of birth, and region in CKB).

Pearson correlation coefficients of each continuous trait with measured standing, sitting, and leg height were estimated for participants at baseline, using residuals after adjustment for age, sex and region in CKB. Correlations with standing height also estimated for participants with resurvey data. For resurvey participants, measurements of continuous traits were averaged over the baseline and resurvey measurements to provide a more accurate representation of their long-term usual values.

4.3 Results

Characteristics of included study participants (481 667 UKB and 490 067 CKB participants) are presented separately for men and women in Table 4.1. Participants in UKB were, on average, older than those in CKB (56.4 [SD 8.1]

years versus 51.6 [SD 10.6] years). CKB had a higher proportion of women than men (59.1% versus 55.4%). The mean levels of height differed substantially between UKB and CKB. Men in UKB were taller than those in CKB (175.7 [SD 6.8] cm versus 165.2 [SD 6.5] cm). Likewise, women in UKB were taller than those in CKB (162.5 [SD 6.3] cm versus 154.2 [SD 6.0] cm). There were also substantial differences in smoking behaviour between men and women in both studies, with a much higher prevalence of smoking in men than in women in CKB.

Standing, sitting, and leg height were normally distributed in men and women in both studies (Figure 4.1 and Supplementary Figure 4.1). Mean height (adjusted for age and year of birth) for men and women differed between regions in China (Supplementary Table 4.1). On average, participants living in urban regions were taller than those living in rural regions (166.8 cm versus 164.1 cm for men, and 155.5 cm versus 153.1 cm for women). Individuals living in Qingdao (an urban region) had the tallest mean height (169.6 cm for men and 158.3 cm for women) in contrast with individuals living in Sichuan (a rural region) who had the shortest mean height (160.6 cm for men and 149.8 cm for women).

The number of participants by individual year of birth and 5-year age group within each study are presented in Supplementary Table 4.2. Standing height increased linearly with increasing year of birth (Figure 4.2) and decreased linearly with increasing age (Figure 4.3). In both studies, the change in standing height was mainly driven by changes in sitting height as changes in leg height were less pronounced (Figures 4.2 and 4.3). For UKB and CKB, respectively, differences in standing height at baseline between the youngest and oldest participants were

3.8 cm versus 7.4 cm for men, and 4.2 cm versus 8.0 cm for women (Figure 4.3). For UKB and CKB, respectively, standing height declined by 1.3 cm versus 1.6 cm per decade older age for men, and by 1.4 cm versus 1.8 cm per decade older age for women (Figure 4.3).

Cross-sectional associations of binary baseline characteristics (prior disease and medications, lifestyle factors, social factors, and self-rated frailty measures) with 1 SD taller standing height, after basic adjustments, are presented in Table 4.2. In both populations, taller height was inversely associated with reported medical history of stroke and chronic obstructive pulmonary disease, and positively associated with history of cancer, higher household income, higher education level, and higher self-rated health. However, current smoking status, history of hypertension and diabetes each showed opposite associations in UKB and CKB (inverse in UKB and positive in CKB). Taller height was inversely associated with myocardial infarction (data only available in UKB), but was positively associated with coronary heart disease (data only available in CKB).

Cross-sectional associations of continuous baseline characteristics (anthropometric measures, blood pressure and heart rate, carotid measures and lipids, and lifestyle factors) with 1 SD taller standing height, after basic adjustments, are presented in Table 4.3. In both populations, taller height was positively associated with most anthropometric measures (weight, waist circumference, lean body mass, fat body mass, hand grip strength and lung function) and carotid intima media thickness, and inversely associated with heart rate and physical activity. However, baseline blood pressure and cigarette

equivalents per day each showed opposite associations in UKB and CKB (inverse in UKB and positive in CKB). The opposite associations of systolic and diastolic blood pressure in UKB and CKB are shown for measured height deciles in Figure 4.4. Adiposity measures (waist-to-hip ratio and BMI) were inversely associated with taller height for men and women in UKB, and positively associated for men in CKB (there were no statistically significant associations for women in CKB). LDL cholesterol and triglycerides were positively associated with taller height for men in CKB and there were no statistically significant associations for women in CKB (data for lipids was only available in a subset of 17 122 CKB participants).

Among the anthropometric and blood pressure measures, standing height was most strongly correlated with lean body mass (correlation coefficient: r=0.55 in UKB and r=0.68 in CKB) and weight (r=0.34 in UKB and r=0.46 in CKB), and weakly correlated with systolic blood pressure (r=-0.07 in UKB and r=0.02 in CKB) (Table 4.4). The correlations with sitting and leg height were similar to the correlations with standing height, except for systolic and diastolic blood pressure in CKB, which were inversely correlated with leg height but positively correlated with sitting and standing height. The correlations with standing height were largely unaltered when measurements of these traits were averaged between the baseline and resurvey measurements to provide a more accurate representation of their long-term usual values (Table 4.5). For lean body mass, the correlation in UKB was unaltered at 0.57 (when calculated using average measurements) and decreased in CKB from 0.68 to 0.66.

4.4 Discussion

The difference in standing height at baseline between the youngest and oldest participants (~4.0 cm in UKB and ~7.9 cm in CKB) was a combination of a cohort effect on height and an age effect (the effect of age on height). Standing height increased with increasing year of birth (the cohort effect) (Figure 4.2) and decreased with increasing age (the age effect) (Figure 4.3). On average, standing height increased by ~1.4 cm and ~1.7 cm per decade higher year of birth in UKB and CKB, respectively (Figure 4.2); and decreased by ~1.4 cm and ~1.7 cm per decade older age in UKB and CKB, respectively (Figure 4.3). The ERFC reported the same decrease in height per decade older age as found in UKB.[1] The decrease in standing height with older age in both UKB and CKB was mainly driven by a decrease in sitting height (which decreased 2-3 times more per decade than leg height), suggesting that the decrease in height with age was due to osteoporosis. In CKB, geographic region had a substantial impact on height (Supplementary Table 4.1). Participants from urban regions were on average 2.4 cm taller than those from rural regions. Hence, all subsequent analyses of the associations of height and disease were adjusted for region. No clear discontinuity in height was observed for participants born close to the time of the Chinese famine (1958-1961) (Figure 4.2), certainly not to the extent that has previously been estimated by Chen et al. (2007)[8] (that the famine reduced the adult height of the 1959 birth cohort by ~3 cm).

The inverse associations of taller standing height with stroke, myocardial infarction (data only available in UKB), chronic obstructive pulmonary disease, and diabetes, and the positive associations of cancer in both studies (Table 4.2)

were consistent with the direction of the associations reported by the ERFC.[1] (The associations reported by the ERFC were prospective associations [hazard ratios estimated by Cox regression] of height with disease.) In contrast, the positive associations of standing taller height with coronary heart disease (data only available in CKB) differed substantially from the findings reported in the ERFC[1]. The difference in the direction of the associations of adult height with coronary heart disease between UKB and CKB could reflect the reduced quality of self-reported outcomes (baseline self-reported outcomes were not adjudicated in CKB) and large differences in the severity of diagnosed coronary artery disease.

The associations of standing height with blood pressure measured at baseline differed between UKB and CKB, being strongly inverse in UKB, but weakly positive in CKB (Figure 4.4). This discrepancy highlighted the need to adequately adjust for blood pressure when examining the associations between height and disease. Overall, taller standing height was associated with more favourable cardiovascular disease risk profiles for participants in UKB, including lower levels of BMI, waist-to-hip ratio, blood pressure and smoking, and a lower proportion with self-reported doctor diagnoses of hypertension and diabetes (Tables 4.3 and 4.4). However, this was not the case for taller participants in CKB who had worse cardiovascular disease risk profiles, including higher levels of BMI, waist-to-hip ratio, blood pressure and smoking, and a higher proportion with self-reported doctor diagnoses of hypertension and diabetes. Taller height was associated with higher income in UKB and CKB, which may lead to unhealthier lifestyles in developing countries (where people with higher income are more likely to smoke,

drink and eat more) and healthier lifestyles in developed countries (where people with higher income tend to be more health conscious). Correlations provided little evidence for different associations of baseline characteristics with sitting and leg height (relative to standing height) (Table 4.4).

4.5 Tables

Table 4.1 Baseline characteristics of included study participants (excluding participants with prior major cardiovascular diseases), shown separately for men and women.

		UKB		CKB		UKB	CKB
		Men	Women	Men	Women	All	All
Baseline characteristics Mean (SD) or percentage		**n=215039 (44.6%)**	**n=266628 (55.4%)**	**n=200364 (40.9%)**	**n=289703 (59.1%)**	**n=481667**	**n=490067**
Age (years)		56.5 (8.2)	56.3 (8.0)	52.3 (10.8)	51.0 (10.4)	56.4 (8.1)	51.6 (10.6)
Age groups (%)	<40	0.0%	0.0%	14.8%	16.6%	0.0%	15.9%
	40–49	24.1%	23.8%	29.1%	31.9%	23.9%	30.8%
	50–59	32.3%	34.5%	30.4%	31.0%	33.5%	30.7%
	60–69	43.1%	41.3%	18.6%	15.6%	42.1%	16.8%
	≥70	0.5%	0.4%	7.0%	5.0%	0.5%	5.8%
Smoking status (%)	Never	49.7%	59.7%	14.2%	95.1%	55.2%	62.0%
	Ex-smoker	37.5%	31.2%	12.5%	0.8%	34.0%	5.6%
	Current smoker	12.3%	8.8%	73.3%	4.1%	10.4%	32.4%
	Prefer not to answer	0.4%	0.4%	-	-	0.4%	-
Cigarette equivalents (cigarettes/day)		1.3 (5.1)	0.9 (4.0)	13.7 (12.3)	0.3 (2.1)	1.1 (4.5)	5.8 (10.4)
Systolic blood pressure (mm Hg)		141.1 (17.4)	135.3 (19.2)	132.3 (19.8)	129.4 (21.8)	137.9 (18.6)	130.6 (21.1)
Diastolic blood pressure (mm Hg)		84.3 (9.9)	80.8 (10.0)	79.1 (11.3)	76.7 (10.8)	82.4 (10.1)	77.6 (11.1)
Hypertension (%)		27.2%	20.4%	9.5%	10.2%	23.4%	9.9%
Diabetes (%)		6.3%	3.6%	5.1%	5.6%	4.8%	5.4%
Standing height (cm)		175.7 (6.8)	162.5 (6.3)	165.2 (6.5)	154.2(6.0)	168.4 (9.3)	158.7 (8.3)
Sitting height (cm)		92.5 (4.0)	86.4 (3.8)	88.4 (3.9)	83.2 (3.8)	89.1 (4.9)	85.3 (4.6)
Leg height (cm)		83.3 (4.6)	76.0 (4.3)	76.9 (4.0)	71.0 (3.7)	79.3 (5.7)	73.4 (4.8)

Table 4.2 Cross-sectional associations of 1 SD taller standing height with binary baseline characteristics, shown separately for men and women, and adjusted for age, year of birth, and region in CKB.

	UKB (n=481667)			CKB (n=490067)		
		Odds ratio (95% CI) per 1 SD taller height			Odds ratio (95% CI) per 1 SD taller height	
Baseline characteristics	**Prevalence (%) Men/Women**	**Men (n=215039)**	**Women (n=266628)**	**Prevalence (%) Men/Women**	**Men (n=200364)**	**Women (n=289703)**
Prior disease and medications						
Coronary heart disease*	-/-	-	-	2.9/3.4	1.10 (1.07, 1.13)	1.10 (1.07, 1.12)
Myocardial infarction*	4.4/0.9	0.79 (0.78, 0.81)	0.78 (0.75, 0.81)	-/-	-	-
Angina	2.9/1.4	0.81 (0.79, 0.83)	0.76 (0.73, 0.78)	-/-	-	-
Atrial fibrillation	2.1/1.0	1.35 (1.31, 1.39)	1.23 (1.18, 1.28)	-/-	-	-
Stroke*†	0.8/0.5	0.82 (0.78, 0.86)	0.85 (0.81, 0.90)	2.5/1.4	0.97 (0.94, 0.99)	0.94 (0.91, 0.97)
Ischaemic stroke*	0.6/0.2	0.79 (0.75, 0.84)	0.78 (0.72, 0.85)	-/-	-	-
Haemorrhagic stroke*	0.3/0.2	0.89 (0.82, 0.97)	0.92 (0.85, 1.00)	-/-	-	-
Self-reported hypertension	27.2/20.4	0.91 (0.90, 0.92)	0.88 (0.87, 0.89)	9.5/10.2	1.14 (1.12, 1.16)	1.12 (1.11, 1.14)
Self-reported diabetes	6.3/3.6	0.87 (0.86, 0.89)	0.82 (0.80, 0.84)	5.1/5.6	1.06 (1.04, 1.08)	1.06 (1.04, 1.07)
Chronic obstructive pulmonary disease	0.3/0.3	0.80 (0.75, 0.87)	0.88 (0.82, 0.94)	8.6/6.0	0.96 (0.94, 0.97)	0.94 (0.92, 0.95)
Cancer	6.0/9.0	1.07 (1.05, 1.09)	1.08 (1.06, 1.09)	0.4/0.5	1.11 (1.03, 1.18)	1.11 (1.05, 1.17)
Cancer (excluding skin cancer)	5.0/8.4	1.06 (1.04, 1.08)	1.07 (1.06, 1.09)	-/-	-	-
Fracture	8.5/10.1	1.01 (0.99, 1.02)	1.06 (1.04, 1.07)	8.9/5.5	1.03 (1.02, 1.05)	1.05 (1.03, 1.07)
Current use of blood pressure medication	22.2/17.1	0.87 (0.86, 0.88)	0.84 (0.83, 0.85)	3.6/3.6	1.17 (1.14, 1.20)	1.17 (1.15, 1.20)
Current use of cholesterol medication	19.8/12.0	0.86 (0.85, 0.87)	0.81 (0.80, 0.82)	0.2/0.2	1.16 (1.04, 1.30)	1.24 (1.13, 1.35)
Current use of diabetes medication	1.3/0.8	0.91 (0.87, 0.94)	0.91 (0.87, 0.95)	1.6/1.8	1.15 (1.11, 1.19)	1.13 (1.10, 1.17)

Lifestyle factors						
Being an ever smoker	49.8/40.0	1.00 (0.99, 1.01)	1.08 (1.07, 1.09)	74.5/3.1	1.09 (1.08, 1.10)	1.04 (1.02, 1.06)
Being a current smoker	12.3/8.8	0.90 (0.88, 0.91)	0.95 (0.93, 0.96)	62.0/2.3	1.05 (1.04, 1.06)	1.03 (1.00, 1.05)
Being a weekly alcohol drinker	-/-	-	-	33.9/2.1	1.07 (1.06, 1.08)	1.10 (1.07, 1.12)
Social factors						
Income >£100000 (UKB) or ¥35000 (CKB)	5.6/4.0	1.41 (1.39, 1.44)	1.33 (1.31, 1.36)	20.4/16.5	1.27 (1.26, 1.29)	1.26 (1.25, 1.27)
Income >£52000 (UKB) or ¥20000 (CKB)	25.8/19.5	1.37 (1.35, 1.38)	1.28 (1.26, 1.29)	45.7/40.7	1.27 (1.26, 1.29)	1.25 (1.24, 1.26)
Income >£31000 (UKB) or ¥10000 (CKB)	50.0/40.6	1.37 (1.36, 1.38)	1.26 (1.25, 1.27)	73.9/70.1	1.27 (1.26, 1.29)	1.20 (1.19, 1.21)
Income >£18000 (UKB) or ¥5000 (CKB)	71.6/62.3	1.37 (1.36, 1.38)	1.26 (1.25, 1.28)	90.8/89.8	1.21 (1.19, 1.23)	1.16 (1.15, 1.18)
Income >¥2500 (CKB)	-/-	-	-	97.1/96.9	1.22 (1.19, 1.26)	1.13 (1.11, 1.16)
Education to degree level	34.4/31.4	1.27 (1.26, 1.28)	1.28 (1.27, 1.29)	7.6/4.4	1.33 (1.31, 1.36)	1.43 (1.40, 1.46)
Ever married	-/-	-	-	98.5/99.8	1.73 (1.67, 1.79)	1.10 (1.02, 1.19)
Self-rated frailty measures						
Self-rated health: good	72.9/76.6	1.17 (1.16, 1.19)	1.22 (1.21, 1.23)	50.6/44.5	1.05 (1.04, 1.06)	1.06 (1.05, 1.06)
Self-rated walking pace: brisk	40.1/38.7	1.25 (1.24, 1.26)	1.24 (1.23, 1.25)	-/-	-	-
Self-rated satisfaction level: very satisfied	-/-	-	-	17.5/17.6	1.09 (1.08, 1.11)	1.09 (1.08, 1.10)

*Associations estimated before exclusions of participants with prior myocardial infarction (UKB only), stroke (UKB and CKB), or coronary heart disease (CKB only).
†CKB included stroke and transient ischaemic attack.

Table 4.3 Cross-sectional associations of 1 SD taller standing height with continuous baseline characteristics, shown separately for men and women, and adjusted for age, year of birth, and region in CKB.

	UKB (n=481667)			CKB (n=490067)		
		Beta (SE) for 1 SD taller height			Beta (SE) for 1 SD taller height	
Baseline characteristic	**Mean (Men/Women)**	**Men (n=215039)**	**Women (n=266628)**	**Mean (Men/Women)**	**Men (n=200364)**	**Women (n=289703)**
Anthropometric measures						
Weight (kg)	86/71	5.8 (0.03)	3.8 (0.03)	64/57	4.8 (0.02)	3.8 (0.01)
Waist circumference (cm)	97/85	1.7 (0.02)	0.5 (0.02)	82/79	2.2 (0.02)	1.6 (0.02)
Waist-to-hip ratio (ratio)	0.93/0.82	-0.003 (0.0001)	-0.005 (0.0001)	0.90/0.87	0.003 (0.0001)	0.000 (0.0001)
BMI (kg/m^2)	28/27	-0.2 (0.01)	-0.6 (0.01)	23/24	0.1 (0.01)	0.0 (0.01)
Lean body mass (kg)	64/45	4.7 (0.01)	2.3 (0.01)	50/38	3.6 (0.01)	2.7 (0.00)
Fat body mass (kg)	22/27	1.1 (0.02)	1.5 (0.02)	15/19	1.2 (0.01)	1.2 (0.01)
Hand grip strength (kg)*	40/23	2.5 (0.02)	1.6 (0.01)	32/20	2.5 (0.06)	1.4 (0.04)
FEV_1 (L)	3.2/2.3	0.29 (0.001)	0.20 (0.001)	2.7/2.0	0.19 (0.001)	0.13 (0.001)
FVC (L)	4.3/3.0	0.41 (0.002)	0.28 (0.001)	3.2/2.4	0.23 (0.001)	0.16 (0.001)
Blood pressure and heart rate						
Systolic blood pressure (mm Hg)	141/135	-0.93 (0.03)	-1.38 (0.04)	132/129	0.50 (0.05)	0.28 (0.04)
Diastolic blood pressure (mm Hg)	84/81	-0.21 (0.02)	-0.49 (0.02)	79/77	0.73 (0.03)	0.37 (0.02)
Heart rate (bpm)	69/70	-0.64 (0.03)	-0.55 (0.02)	78/80	-0.18 (0.03)	-0.51 (0.02)
Carotid measures and lipids						
Carotid intima media thickness (mm)*†	0.71/0.65	0.014 (0.004)	0.006 (0.003)	0.74/0.67	0.006 (0.002)	0.005 (0.001)
Carotid plaque score (mm)*	-/-	-	-	0.94/0.62	-0.010 (0.010)	-0.004 (0.007)
LDL cholesterol (mmol/L)¶	-/-	-	-	2.3/2.4	0.021 (0.007)	-0.006 (0.007)

Triglycerides (mmol/L)¶	-/-	-	-	2.0/2.0	0.071 (0.017)	0.025 (0.015)
Lifestyle factors						
Cigarette equivalents (cigarettes/day)	1.3/0.9	-0.2 (0.01)	-0.1 (0.01)	13.7/0.3	0.5 (0.03)	0.0 (0.00)
Alcohol consumption (units)	15/6	0.3 (0.04)	0.5 (0.01)	-/-	-	-
Physical activity (MET hours/day)	38/32	-2.0 (0.11)	-0.1 (0.07)	23/21	-0.7 (0.03)	-0.2 (0.02)

*Associations based on CKB second resurvey (measurements were available in 21713 participants).
†Associations based on UKB imaging data (measurements were available in 1836 participants).
¶Associations based on lipids data available in 17122 CKB participants.

Table 4.4 Pearson correlation coefficients of selected continuous traits with standing, sitting, and leg height at baseline.*

	Pearson correlation coefficients of baseline continuous traits					
	UKB (n=481667)			CKB (n=490067)		
Continuous trait	**Standing height**	**Sitting height**	**Leg height**	**Standing height**	**Sitting height**	**Leg height**
Standing height	1.00	0.74	0.82	1.00	0.76	0.83
Sitting height	0.74	1.00	0.23	0.76	1.00	0.27
Leg height	0.82	0.23	1.00	0.83	0.27	1.00
Weight	0.34	0.33	0.21	0.46	0.46	0.30
Waist circumference	0.09	0.10	0.05	0.21	0.22	0.12
Waist-to-hip ratio	-0.06	-0.09	-0.02	0.02	0.00	0.03
BMI	-0.09	0.02	-0.15	0.02	0.13	-0.09
Lean body mass	0.55	0.51	0.37	0.68	0.61	0.48
Fat body mass	0.14	0.16	0.06	0.19	0.23	0.08
Hand grip strength¶	0.29	0.25	0.21	-	-	-
FEV_1	0.40	0.36	0.28	0.36	0.31	0.27
FVC	0.44	0.38	0.32	0.39	0.33	0.29
Systolic blood pressure	-0.07	-0.02	-0.08	0.02	0.06	-0.03
Diastolic blood pressure	-0.04	0.01	-0.06	0.04	0.08	-0.01
Heart rate	-0.05	-0.02	-0.06	-0.03	-0.01	-0.04

*Correlation coefficients were estimated using residuals after adjustment for age, sex and region in CKB.

Table 4.5 Pearson correlation coefficients of selected continuous traits with standing height at baseline and resurvey.*

	Pearson correlation coefficients					
	UKB			CKB		
	Baseline participants (n=481667) using:	1st and 2nd resurvey participants (n=36681) using:		Baseline participants (n=490067) using:	1st and 2nd resurvey participants (n=29878) using:	
Continuous trait	Baseline measurements	Baseline measurements	Average measurements†	Baseline measurements	Baseline measurements	Average measurements†
Weight	0.34	0.36	0.38	0.46	0.46	0.46
Waist circumference	0.09	0.11	0.12	0.21	0.20	0.21
Waist-to-hip ratio	-0.06	-0.05	-0.07	0.02	0.02	0.02
BMI	-0.09	-0.09	-0.08	0.02	0.01	0.01
Lean body mass	0.55	0.57	0.57	0.68	0.68	0.66
Fat body mass	0.14	0.15	0.17	0.19	0.18	0.20
Hand grip strength¶	0.29	0.27	0.35	-	-	0.34
FEV_1	0.40	0.40	0.41	0.36	0.37	0.36
FVC	0.44	0.45	0.42	0.39	0.39	0.39
Systolic blood pressure	-0.07	-0.07	-0.06	0.02	0.01	0.00
Diastolic blood pressure	-0.04	-0.04	-0.03	0.04	0.04	0.04
Heart rate	-0.05	-0.04	-0.07	-0.03	-0.03	-0.02
cIMT¶‡	-	-	0.08	-	-	0.04
Carotid plaque score¶	-	-	-	-	-	-0.01

*Correlation coefficients were estimated using residuals after adjustment for age, sex and region in CKB.
†Average measurements were the average of the baseline measurements and the earliest available resurvey measurements.
¶Associations based on CKB second resurvey (measurements were available in 21713 participants).
‡Associations based on UKB imaging data (measurements were available in 1836 participants). cIMT=carotid intima media thickness.

4.6 Figures

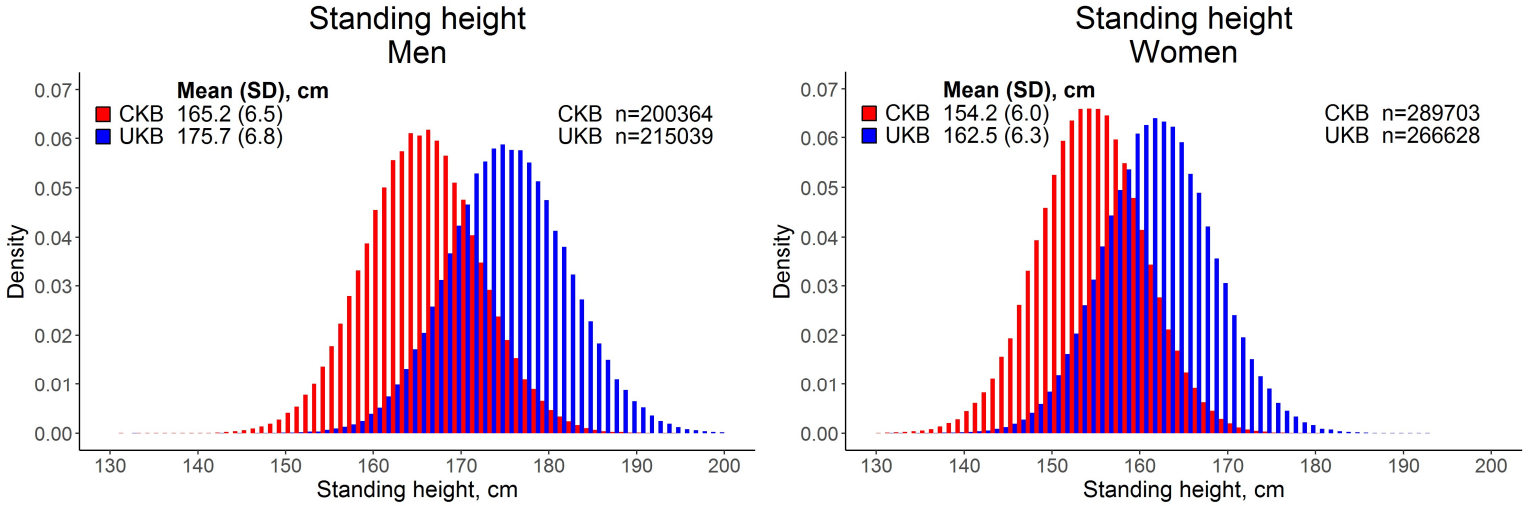


Figure 4.1 Distribution of standing height, shown separately for men and women in UKB and CKB.

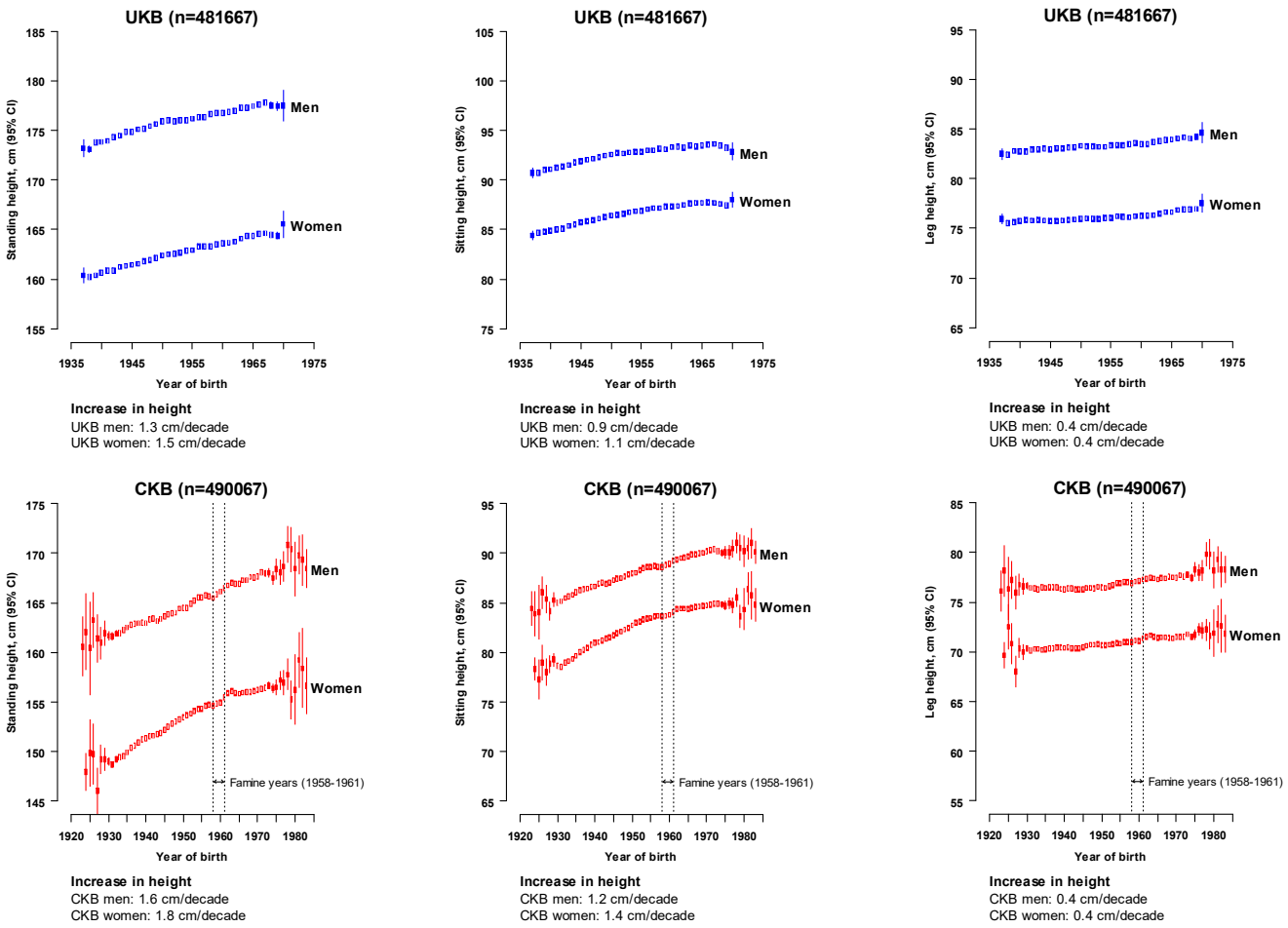


Figure 4.2 Standing, sitting and leg height by year of birth, shown separately for men and women, and adjusted for region in CKB.

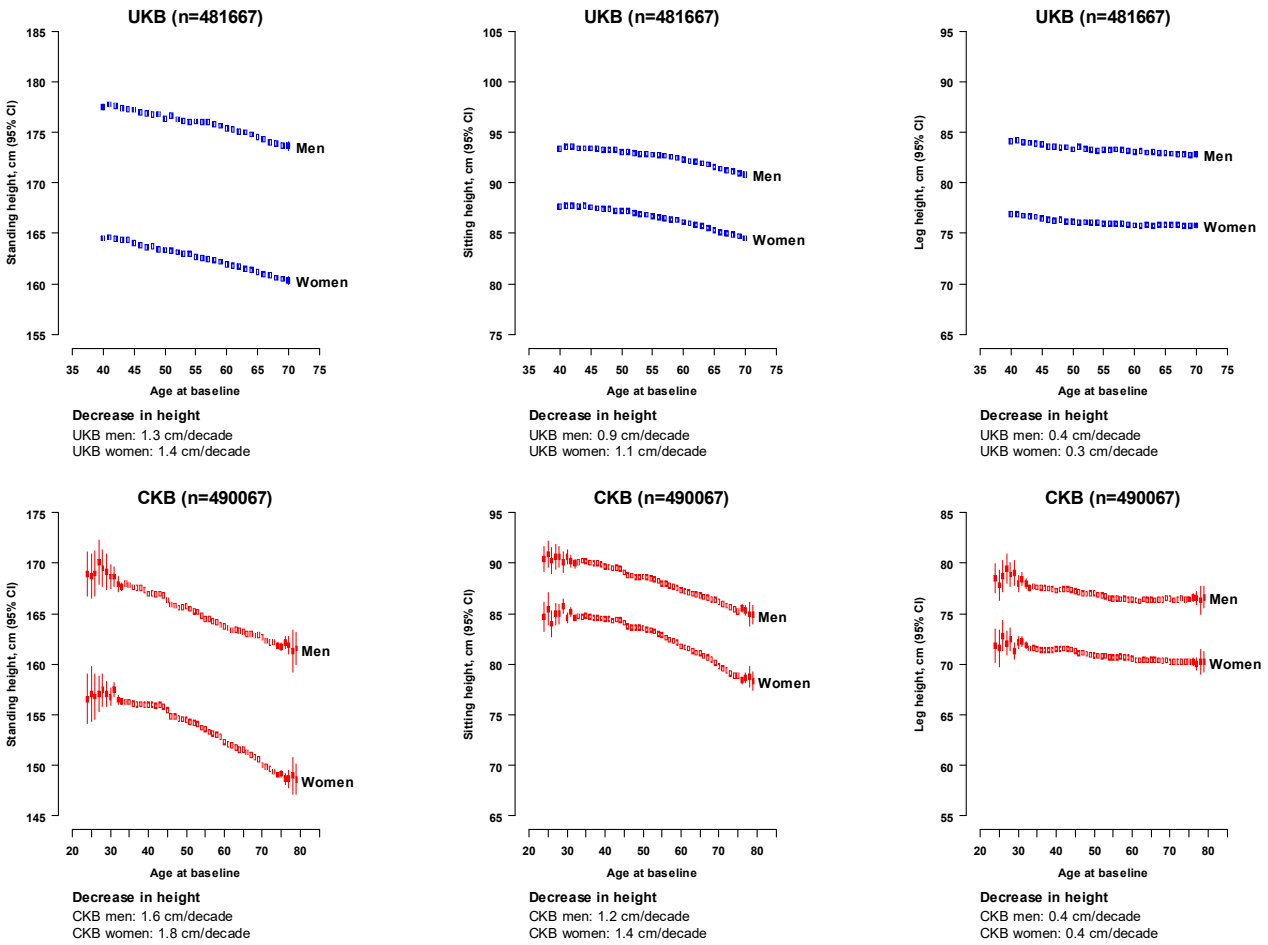


Figure 4.3 Standing, sitting and leg height by age at baseline, shown separately for men and women, and adjusted for region in CKB.

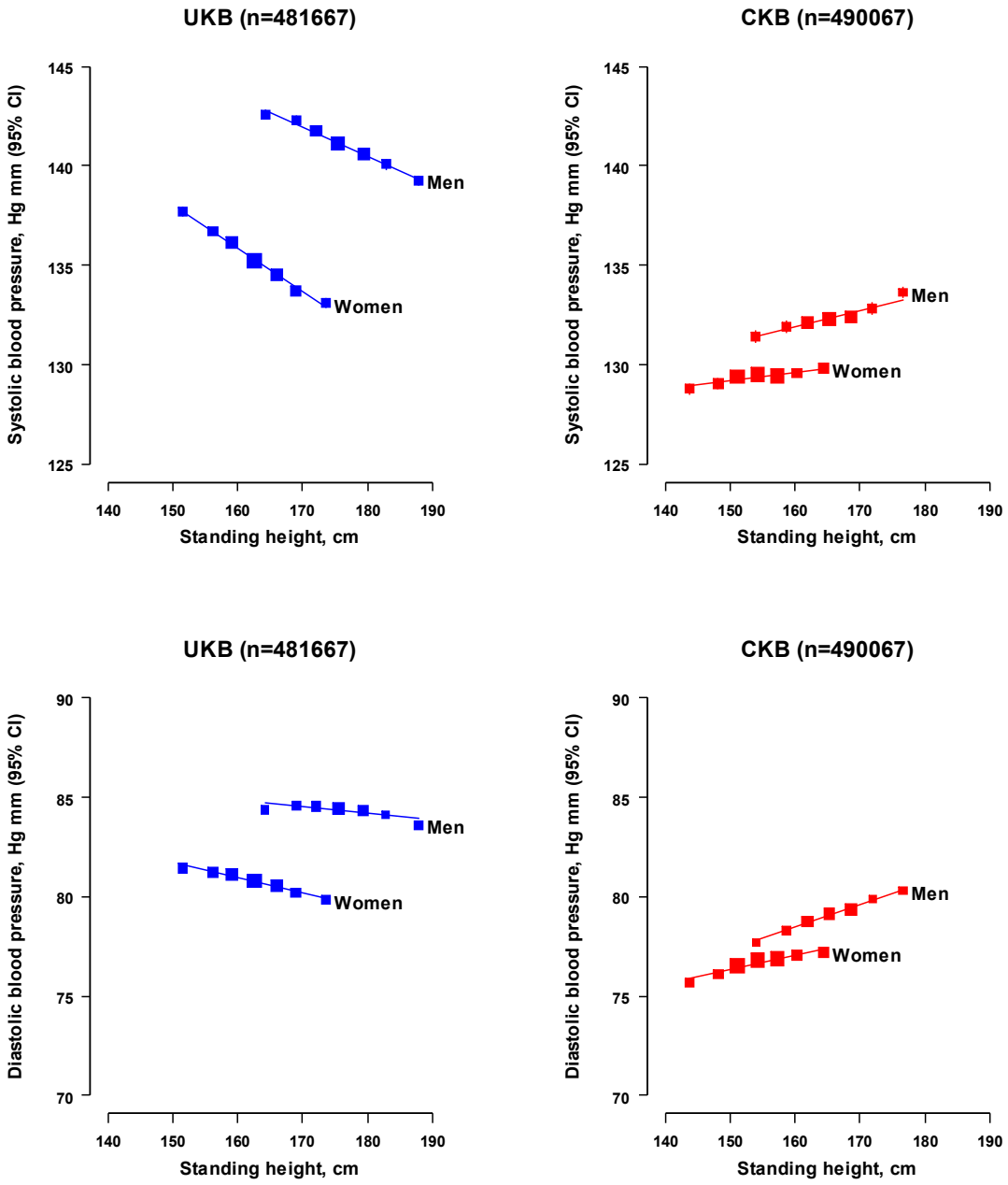


Figure 4.4 Systolic blood pressure and diastolic blood pressure by standing height, shown separately for men and women, and adjusted for age, year of birth, and region in CKB.

5. Prospective associations of height with cardiovascular diseases

5.1 Introduction

The ERFC meta-analysis of prospective studies involving 1M Europeans reported inverse associations of taller standing height with non-fatal myocardial infarction (16 774 cases; HR per 1 SD [6.5 cm] taller height: 0.91; 95% CI: 0.89, 0.93), coronary deaths (7491 deaths; HR: 0.93; 95% CI: 0.90, 0.96), ischaemic stroke (5720 cases; HR: 0.94; 95% CI: 0.90, 0.97), haemorrhagic stroke (2169 cases; HR: 0.90; 95% CI: 0.85, 0.95) and subarachnoid haemorrhage (1315 cases; HR: 0.91; 95% CI: 0.84, 0.98), but positive associations with ruptured aortic aneurysm (1324 deaths; HR: 1.12; 95% CI: 1.05, 1.20) and pulmonary embolism (714 deaths; HR: 1.12; 95% CI: 1.03, 1.21).[1] The Lee et al. (2009) meta-analysis of prospective studies involving 0.5M mostly Asian participants also reported inverse associations of taller standing height with coronary heart disease (3933 cases; HR per 1 SD [6 cm] taller height [combined across men and women]: 0.93; 95% CI: 0.90, 0.96) and haemorrhagic stroke (1534 cases; HR: 0.90; 95% CI: 0.85, 0.95), but no statistically significant association with ischaemic stroke (1983 cases; HR: 0.96; 95% CI: 0.92, 1.00).[2] The clinical relevance, if any, of the associations of height with cardiovascular diseases is not fully understood.

Most studies have been conducted in European populations although some studies have also been conducted in Asian populations (see systematic review in

Chapter 2). Studies of different populations, with different confounding factors, might help to elucidate mechanisms and possibly distinguish causal from non-causal associations.

Little is known about the association of height with major cardiovascular diseases in low- and middle-income countries that have different nutritional exposures and different disease patterns.[5,6] For example, haemorrhagic stroke is much more common in low-income countries, and the associations of height with stroke types (haemorrhagic stroke and ischaemic stroke) remain uncertain.[1,2]

Height has been reported to be positively associated with atrial fibrillation, but the number of cases studied to date has been limited (1252 cases; HR per 6.5 cm SD taller height [scaled to 6.5 cm and combined across men and women]: 1.18; 95% CI: 1.11, 1.25).[3] In addition, lean body mass has been identified as a major anthropometric risk factor for atrial fibrillation.[79]

In addition, there have been few limited studies on the associations of sitting height compared with standing height and risk of cardiovascular diseases (see systematic review in Chapter 2). If some diseases were more strongly associated with one component of height (standing, sitting or leg height), this might be informative about the underlying mechanisms of disease.

This chapter extends the current body of evidence by providing direct comparisons of the associations of height with cardiovascular diseases in East Asian and European populations. In addition, the number of cases of ischaemic

stroke, haemorrhagic stroke and atrial fibrillation were much greater in UKB and CKB than those assessed in previous studies. This chapter examines the shape and strength of the associations of height (standing, sitting, and leg height) with cardiovascular diseases and atrial fibrillation before and after adjustment for confounding factors.

5.2 Methods

Study populations

A total of 481 667 participants in UKB and 490 067 participants in CKB, each with no prior major cardiovascular disease at baseline, were included in the prospective analyses. For further details of the study designs of UKB and CKB and definitions of exclusion, see Chapter 3.

The standing, sitting, and leg height phenotypes were normalised within strata of sex (and also region in CKB) to give height variables in SD units, (described in Chapter 3.3). Lean body mass was also normalised within strata of sex (and region in CKB) and quartiles of this variable were used for subgroup analyses. Deciles of measured height were used to investigate the shape of the associations of height with the outcomes. (For presentation purposes, when deciles were plotted, consecutive pairs of the middle 6 deciles were combined, with results presented for 7 groups.)

Outcomes

Incident cardiovascular disease events included major coronary events (myocardial infarction or coronary heart disease death), ischaemic stroke, and

haemorrhagic stroke. Since atrial fibrillation is a chronic condition, it is appropriate to assess the associations of height with atrial fibrillation using cross-sectional (rather than prospective) methods and prevalent cases of atrial fibrillation at the baseline survey were combined with incident cases. Follow-up data until 31 March 2015 (~6 years of follow-up) for UKB and 31 December 2016 (~6 years of follow-up) for CKB were available. Further information about outcomes is provided in Chapter 3.4.

Statistical methods

The shapes of the associations of height with cardiovascular diseases were investigated by estimating the HRs (and 95% CIs) for each decile of measured height after basic adjustments (age [5-year age bands], sex, year of birth, and region in CKB). Floating absolute risk estimates were calculated relative to the lowest height decile.

The relative risks of taller standing height with incident cardiovascular disease events were assessed using Cox regression after basic adjustments and further adjustment for major cardiovascular disease risk factors, socio-economic risk factors, and additional risk factors. Details of the adjustment variables are provided in Supplementary Appendix 5.1. In addition, the relative risks were estimated for sitting and leg height. The associations with cardiovascular diseases were presented as hazard ratios (HRs) and 95% CIs per 1 SD taller height.

Subgroup analyses for standing height were performed for each of the following groups (after basic adjustments and further adjustment for major cardiovascular disease and socio-economic risk factors): sex, year of birth (<1945, ≥1945), region (urban and rural) (CKB only), smoking status (never, ex-smoker, and current smoker), self-reported diabetes, self-reported hypertension, systolic blood pressure (<120, 120-140, 140-160, ≥160 mm Hg), BMI (<20, 20-24, 25-29, ≥30 kg/m^2), quartiles of sex normalised lean body mass, and atrial fibrillation (prevalent at baseline or incident). The HRs of the linear effects from the UKB and CKB studies were combined using fixed-effects meta-analysis.

The cross-sectional associations of taller height with atrial fibrillation were assessed using logistic regression. The same adjustments, shape, and subgroup analyses, which were examined in the prospective analyses of height and incident cardiovascular diseases, were also examined in the cross-sectional associations of height with atrial fibrillation. To examine potential differences in the aetiology of atrial fibrillation, the associations were also estimated after additional adjustment for adiposity traits (weight, waist circumference, waist-to-hip ratio, BMI, lean body mass, and fat body mass). The subgroup analyses of height and atrial fibrillation were restricted to UKB, which had many more cases than CKB (14 912 versus 621, respectively). The associations with atrial fibrillation were presented as ORs (and 95% CIs) per 1 SD taller height.

Pearson self-correlations (correlations between repeated measurements of the same variable) between baseline survey and resurvey were calculated for both exposure variables and covariates. For both UKB and CKB, resurvey data from

the latest available resurvey were used (i.e., where a participant was included in both resurveys, data from the second resurvey were used). For UKB, the first resurvey was conducted ~4 years after the baseline survey and the second resurvey (imaging visit) ~7 years (Chapter 3.1). For CKB, the first resurvey was conducted ~3 years after the baseline survey and the second resurvey ~8 years (Chapter 3.2). Self-correlations were estimated for resurvey participants (excluding participants with prior major cardiovascular disease at baseline) and for non-hypertensive resurvey participants (additionally excluding participants diagnosed with hypertension at baseline and participants taking blood pressure lowering medications at baseline). Self-correlations were used to assess the magnitude of measurement error in both study populations.

5.3 Results

Associations of height with cardiovascular diseases

There was a total of 14 851 incident major coronary events, 40 814 incident ischaemic strokes, 10 663 incident haemorrhagic strokes, and 15 584 prevalent or incident cases of atrial fibrillation included in both studies. Standing height was approximately log-linearly related to risk of coronary artery disease, stroke and atrial fibrillation in UKB and CKB (Supplementary Figure 5.1). After basic adjustments, taller standing height was inversely associated with incident major coronary events in both UKB (HR per 1 SD taller standing height: 0.84; 95% CI: 0.82, 0.86) and CKB (HR: 0.96; 95% CI: 0.94, 0.98 in CKB), however the strength of these inverse associations differed somewhat (Figure 5.1). When UKB and CKB were combined, the HR of major coronary events per 1 SD taller standing height was: 0.93; 95% CI: 0.92, 0.95) (Figure 5.2). Likewise, taller standing height

was inversely associated with ischaemic stroke in both UKB (HR per 1 SD taller standing height: 0.89; 95% CI: 0.86, 0.93) and CKB (HR: 0.97; 95% CI: 0.96, 0.98 in CKB) (Figure 5.1). When UKB and CKB were combined, the HR of ischaemic stroke per 1 SD taller standing height was: 0.96; 95% CI: 0.95, 0.97) (Figure 5.2). Taller standing height was inversely associated with haemorrhagic stroke in CKB (HR per 1 SD taller standing height: 0.89; 95% CI: 0.87, 0.91; 9789 cases), but was not statistically significantly associated with haemorrhagic stroke in UKB (HR: 0.96; 95% CI: 0.90, 1.02; 869 cases) (Figure 5.1). When UKB and CKB were combined, the HR of haemorrhagic stroke per 1 SD taller standing height was: 0.89; 95% CI: 0.87, 0.90) (Figure 5.2).

Adjustment for major cardiovascular disease risk factors and socio-economic risk factors attenuated the strength of the inverse association of taller standing height with major coronary events in UKB (HR per 1 SD taller standing height: 0.93; 95% CI: 0.90, 0.95), but strengthened the inverse association in CKB slightly (HR: 0.94; 95% CI: 0.92, 0.96) (Figure 5.1). Likewise, the adjustment substantially attenuated the inverse association of taller standing height with ischaemic stroke in UKB (HR per 1 SD taller standing height: 0.97; 95% CI: 0.93, 1.00), but strengthened the inverse association in CKB slightly (HR: 0.96; 95% CI: 0.95, 0.96). The adjustment attenuated the association of taller standing height with haemorrhagic stroke closer to the null value in UKB (HR per 1 SD taller standing height: 0.97; 95% CI: 0.93, 1.00), but strengthened the inverse association in CKB (HR: 0.88; 95% CI: 0.86, 0.90). Overall, adjustment for major cardiovascular disease risk factors and socio-economic risk factors had a greater effect in UKB, and the associations in UKB and CKB were more similar in magnitude after the

adjustment (Figure 5.1). The associations of height with major coronary events, ischaemic stroke and haemorrhagic stroke were largely unaltered after additional adjustment for other risk factors (Supplementary Figure 5.2).

After adjustment for major cardiovascular disease risk factors and socio-economic risk factors, taller sitting height was inversely associated with cardiovascular diseases in UKB and CKB (Figure 5.1, showing HRs per 1 SD sitting height). The pattern of associations was similar to the associations with standing height. However, the associations per 1 SD taller leg height tended to be marginally less extreme (Figure 5.1).

For major coronary events, there were no compelling subgroup differences (taking into the multiple testing involved in subgroup analyses), except a possible tendency to a stronger effect at low BMI, seen in both populations (Figure 5.3). In UKB, the inverse association with major coronary events was much stronger for those with BMI less than 25 kg/m^2 (HR per 1 SD taller standing height for participants with BMI range 20-25 kg/m^2: 0.86; 95% CI: 0.81, 0.91) (Figure 5.3). For ischaemic stroke, there was no suggestion of a parallel trend with BMI group to that seen for major coronary event but, in CKB, the inverse associations with ischaemic stroke were more extreme after participants were stratified by (sex normalised) lean body mass (HRs per 1 SD taller standing height ≤0.93 in all quartiles [Figure 5.4], compared to 0.96 in the adjusted analysis in all participants [Figure 5.1]). In addition, in UKB, the inverse associations with ischaemic stroke were more extreme in participants without atrial fibrillation (HR per 1 SD taller standing height: 0.91; 95% CI: 0.87, 0.95 [Figure 5.4], compared to 0.97 in the

adjusted analysis in all participants [Figure 5.1]). For major haemorrhagic stroke, there were no compelling subgroup differences (Figure 5.5).

Associations of height with atrial fibrillation

After basic adjustments, taller standing height was positively associated with atrial fibrillation in both UKB and CKB (Figure 5.6). Adjustment for major cardiovascular disease risk factors and socio-economic risk factors strengthened the positive association of taller standing height with atrial fibrillation in UKB (OR per 1 SD taller standing height: 1.33; 95% CI: 1.31, 1.36) and attenuated the association in CKB (OR: 1.34; 95% CI: 1.24, 1.45). When UKB and CKB were combined, the OR of atrial fibrillation per 1 SD taller standing height was: 1.33; 95% CI: 1.31, 1.36) (Table 5.1). The positive associations were weaker for sitting height and leg height across both studies (Figure 5.6).

Additional adiposity adjustments identified lean body mass and body weight as having the largest attenuation effects on the association of standing height with atrial fibrillation. The combined OR per 1 SD taller standing height attenuated very substantially from 1.33 (95% CI: 1.31, 1.36) to 1.16 (95% CI: 1.14, 1.19) when adjusted for lean body mass, and the combined OR attenuated to 1.23 (95% CI: 1.21, 1.25) when adjusted for weight (Table 5.1).

In UKB, after basic adjustments and additional adjustment for cardiovascular disease risk factors, socio-economic risk factors, and lean body mass, the positive associations of standing height with atrial fibrillation were similar in men and women (OR per 1 SD taller standing height in men: 1.16; 95% CI: 1.12, 1.19;

OR in women: 1.16; 95% CI: 1.12, 1.20) (Figure 5.7). There was no association of height with atrial fibrillation among participants with diabetes (OR per 1 SD taller standing height: 0.98; 95% CI: 0.91, 1.05), but an association persisted in participants without diabetes (OR: 1.18; 95% CI: 1.15, 1.34). The positive association of height with atrial fibrillation appeared to be more extreme in participants with BMI less than 30 kg/m^2 and in participants in lower quartiles of lean body mass.

Self-correlations between baseline survey and resurvey of exposure variables and covariates

The self-correlations of standing height between baseline survey and resurvey (for resurvey participants) were 0.99 in UKB and 0.98 in CKB, and greater than 0.85 for sitting height and leg height across both studies (Supplementary Table 5.1). The self-correlations of systolic blood pressure (0.63 in UKB and 0.57 in CKB), diastolic blood pressure (0.59 in UKB and 0.54 in CKB) and diagnosed hypertension (0.64 in UKB and 0.46 in CKB) between baseline and resurvey were relatively low. The self-correlations of waist-to-hip ratio and use of blood pressure lowering medications were substantially different in the two studies (waist-to-hip ratio: 0.80 in UKB versus 0.60 in CKB; and taking blood pressure lowering medications: 0.72 in UKB versus 0.30 in CKB). There was little change in the self-correlations after participants with diagnosed hypertension (reported during the baseline survey) or taking blood pressure lowering medications (reported during the baseline survey) were excluded.

5.4 Discussion

The overall number of ischaemic stroke and haemorrhagic stroke cases were 7-fold and 3-fold greater, respectively, than those included in the ERFC meta-analysis (the largest previously reported meta-analysis of studies in European populations which included 5720 ischaemic strokes and 3484 haemorrhagic strokes [including subarachnoid haemorrhage]).[1] The number of ischaemic stroke and haemorrhagic stroke cases in CKB were 19-fold and 6-fold greater, respectively, than those included in the largest previously reported study in Asians (the Lee et al. [2009] meta-analysis which included 1983 ischaemic strokes and 1534 haemorrhagic strokes).[2] The overall number of atrial fibrillation cases was 12-fold greater than those included in the largest study of the association of height with atrial fibrillation to date (Rosenberg et al. [2012] which included 1252 cases).[3] Most of the atrial fibrillation cases were in UKB (14 912 out of 15 584 cases).

The inverse association of taller standing height with ischaemic stroke (HR per 1 SD taller standing height: 0.96; 95% CI: 0.95, 0.97; combined across UKB and CKB after basic adjustments and adjustment for major cardiovascular disease risk factors and socio-economic risk factors) was consistent with the associations reported by the ERFC (HR per 1 SD taller standing height: 0.94; 95% CI: 0.90, 0.97; adjusted for age, sex, year of birth, and smoking status) and by Lee (HR per 1 SD taller standing height [combined across men and women]: 0.96; 95% CI: 0.92, 1.00; adjusted for age, study, and year of birth).[1,2] The associations of taller standing height with ischaemic stroke were more extreme after participants were stratified by lean body mass, which suggests that the component of height

independent of lean body mass may be inversely associated with ischaemic stroke whereas a lean body mass linked height factor may be positively linked to ischaemic stroke.

Taller standing height was strongly positively associated with atrial fibrillation (OR per 1 SD taller height: 1.33; 95% CI: 1.31, 1.36; combined across UKB and CKB after basic adjustments and adjustment for major cardiovascular disease risk factors and socio-economic risk factors) (Table 5.1), and the association was consistent with, and much more precise than, the positive association reported by Rosenberg et al. (1252 atrial fibrillation cases), HR per 6.5 cm taller height (scaled to 6.5 cm and combined across men and women): 1.18; 95% CI: 1.11, 1.25 (adjusted for age, BMI, clinic site, race, and education).[3]

Lean body mass was the most important adiposity risk factor for atrial fibrillation (lean body mass had the largest attenuation effects on the association of standing height with atrial fibrillation), more important than height or body mass index (Table 5.1), confirming a previous report.[79] After adjustment for lean body mass, the combined association of taller standing height with atrial fibrillation attenuated to an OR of 1.16 per 1 SD taller height (95% CI: 1.14, 1.19) (Table 5.1). Lean body mass had a high correlation with standing height (0.55 in UKB, and 0.68 in CKB) (Table 4.4), and thus the association of height with atrial fibrillation may be secondary to an association of lean body mass with atrial fibrillation.

The positive association of taller standing height with atrial fibrillation was stronger in UKB participants with BMI less than 30 kg/m^2 relative to participants

with BMI greater than or equal to 30 kg/m^2. Likewise, the correlations between lean body mass and standing height were also stronger in participants with BMI less than 30 kg/m^2 (correlation of 0.80 for participants with BMI between 25 and 30 kg/m^2) relative to participants with BMI greater than or equal to 30 kg/m^2 (correlation of 0.69). However, the correlations between lean body mass and standing height were the same for participants with and without diabetes (correlations of 0.55 and 0.56, respectively), even though there was no statistically significant association of taller standing height with atrial fibrillation among UKB participants with diabetes.

When restricted to participants without atrial fibrillation in UKB, the association of taller standing height with ischaemic stroke strengthened (became more inverse) (Figure 5.3). In addition, atrial fibrillation is an important risk factor for cardioembolic ischaemic stroke.[80] This suggests that the association with height may vary by ischaemic stroke subtype (cardioembolic stroke, large-vessel stroke, and small-vessel stroke).

The inverse association of taller standing height with haemorrhagic stroke (HR per 1 SD taller standing height: 0.89; 95% CI: 0.87, 0.90; combined across UKB and CKB after basic adjustments and adjustment for major cardiovascular disease risk factors and socio-economic risk factors) was consistent with the associations reported by the ERFC (HR per 1 SD taller standing height [combined for haemorrhagic stroke and subarachnoid haemorrhage]: 0.90; 95% CI: 0.86, 0.94; adjusted for age, sex, year of birth, and smoking status) and by Lee et al.

(HR per 1 SD taller standing height [combined across men and women]: 0.90; 95% CI: 0.85, 0.95; adjusted for age, study, and year of birth).[1,2]

The inverse association of taller standing height with major coronary events (HR per 1 SD taller standing height: 0.93; 95% CI: 0.92, 0.95; combined across UKB and CKB after basic adjustments and adjustment for major cardiovascular disease risk factors and socio-economic risk factors) was consistent with the associations of taller standing height with coronary heart disease reported by the ERFC (HR per 1 SD taller standing height: 0.93; 95% CI: 0.91, 0.94; adjusted for age, sex, year of birth, and smoking status) and by Lee et al. (HR per 1 SD taller standing height [combined across men and women]: 0.93; 95% CI: 0.90, 0.96; adjusted for age, study, and year of birth).[1,2]

Adjustment for major cardiovascular disease risk factors and socio-economic risk factors attenuated the associations of taller standing height with cardiovascular diseases in UKB and strengthened the associations in CKB (i.e. confounding factors acted in different directions in the two populations) (Figure 5.1). The adjustment had a greater effect on the associations in UKB, and the associations in UKB and CKB were more similar in magnitude after adjustment. This suggested that analyses without appropriate adjustment for covariates might be misleading.

In UKB and CKB, the reproducibility of standing height between baseline survey and resurvey was high, with self-correlations greater than 0.98 (Supplementary Table 5.1). Sitting height and leg height also had high reproducibility, with self-

correlations greater than 0.85. Therefore, height was not corrected for measurement error.

However, measurement error in the covariates can also impact the associations of the exposure with disease. Blood pressure was an important risk factor for cardiovascular diseases in both UKB and CKB, but the cross-sectional associations of height with blood pressure were opposite in the two populations (Figure 4.4). In addition, the self-correlations of blood pressure covariates (systolic blood pressure, diastolic blood pressure, and diagnosed hypertension) between baseline survey and resurvey were relatively low, 0.46-0.64 (Supplementary Table 5.1). (Self-correlations represent not only measurement error, but also changes in treatment, such as blood pressure, cholesterol, or diabetes treatment.) The models of the association of height with cardiovascular diseases included three blood pressure related terms (including diagnosed hypertension, a measure of disease). The validity of simple approaches used for correction in this context is uncertain and so this was not undertaken.[81] It is likely that adjustment for usual values, were it possible, would make a somewhat bigger shift in the direction of the adjustment (further attenuating the associations of taller standing height with cardiovascular diseases in UKB, and strengthening the associations in CKB) (Figure 5.1). For example, adjustment for blood pressure terms (systolic blood pressure, diastolic blood pressure, and diagnosed hypertension) attenuated the HRs for cardiovascular diseases in UKB by ~0.02, and strengthened the HRs in CKB by a similar amount.

The associations in UKB and CKB of the components of height (sitting and leg height) were generally consistent with the associations for standing height (Figure 5.1). The subgroup analyses did not identify any appreciable sex differences between the associations of standing height with cardiovascular diseases (Figures 5.2-5.4) and atrial fibrillation (Figure 5.7).

A limitation of any prospective study is the possibility of residual confounding. In the present study, adjustment of potential confounding factors was examined but confounding (due to measured or unmeasured factors) may persist. Mendelian randomisation is a powerful technique which can avoid confounding by using the property that alleles are allocated randomly when passed from parents to offspring. Genetic instruments for height and their application in Mendelian randomisation analyses will be discussed in the next two chapters.

5.5 Tables

Table 5.1 Associations of taller standing height with prevalent or incident atrial fibrillation, shown for different adiposity adjustments.

	Odds ratio (95% CI) per 1 SD taller standing height*		
Adjustments	**UKB (496228 participants and 14912 cases of atrial fibrillation)**	**CKB (513214 participants and 672 cases of atrial fibrillation)**	**Combined (1009442 participants and 15584 cases of atrial fibrillation)**
Basic adjustments*	1.28 (1.26, 1.30)	1.37 (1.27, 1.47)	1.28 (1.26, 1.30)
+ adjustment for major CVD and socio-economic risk factors†	1.33 (1.31, 1.36)	1.34 (1.24, 1.45)	1.33 (1.31, 1.36)
Additionally individually adjusted for:			
+ weight	1.23 (1.20, 1.25)	1.27 (1.17, 1.39)	1.23 (1.21, 1.25)
+ waist circumference	1.30 (1.28, 1.32)	1.32 (1.22, 1.43)	1.30 (1.28, 1.32)
+ waist-to-hip ratio	1.33 (1.31, 1.36)	1.34 (1.24, 1.45)	1.33 (1.31, 1.36)
+ BMI	1.35 (1.32, 1.37)	1.34 (1.24, 1.45)	1.35 (1.32, 1.37)
+ lean body mass	1.16 (1.13, 1.18)	1.22 (1.10, 1.35)	1.16 (1.14, 1.19)
+ fat body mass	1.29 (1.27, 1.31)	1.32 (1.23, 1.43)	1.29 (1.27, 1.31)

*Basic adjustments were age, sex, year of birth, and region in CKB.
†Major cardiovascular disease (CVD) and socio-economic risk factors were smoker status, cigarette equivalents (cigarettes/day), systolic blood pressure, diastolic blood pressure, diagnosed hypertension, diagnosed diabetes, self-rated walking pace (UKB only), household income, and education level.

5.6 Figures

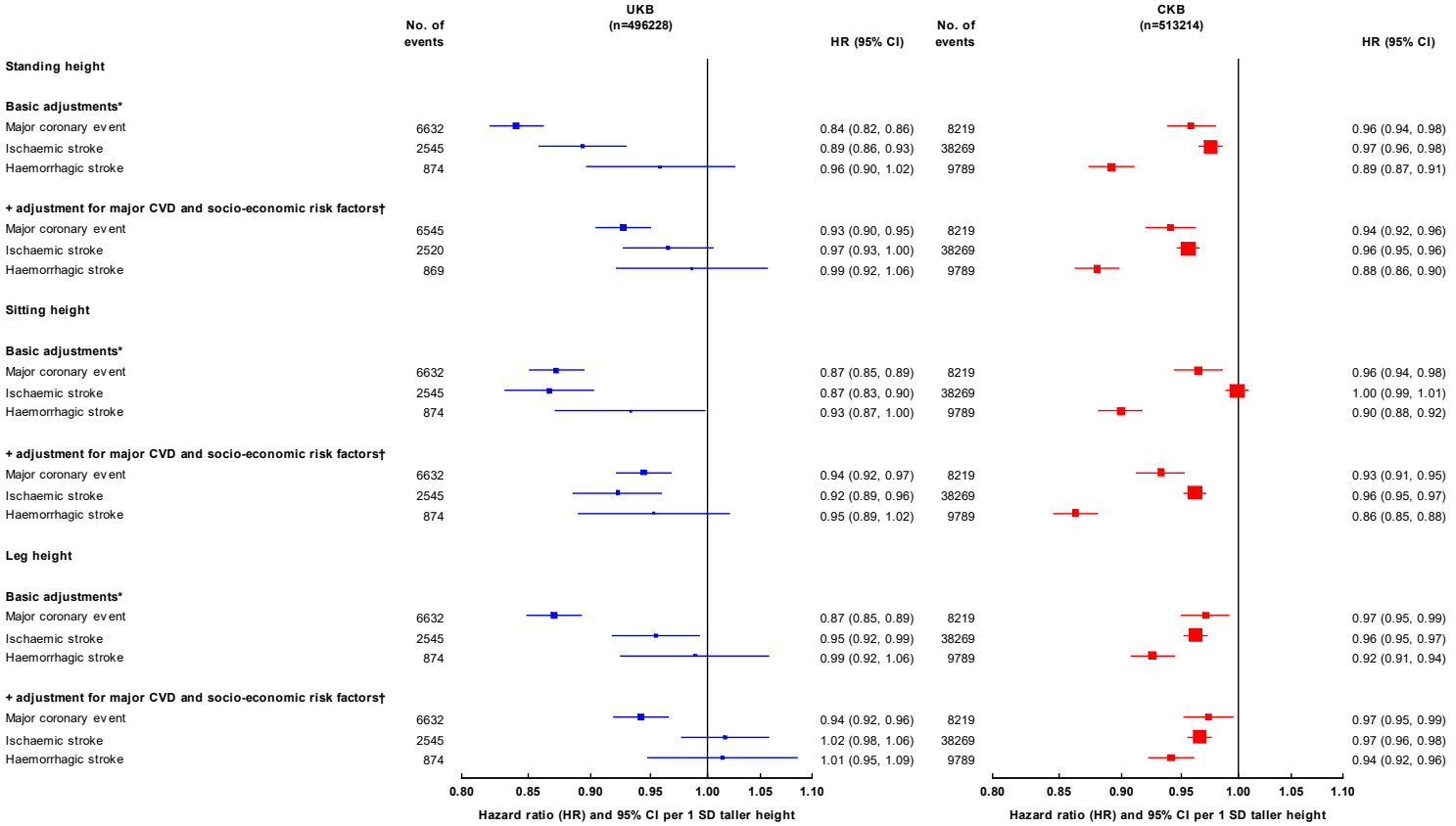


Figure 5.1 Associations of taller standing, sitting, and leg height with incident major coronary events and stroke types, shown for increasing levels of adjustment.

*Basic adjustments were age, sex, year of birth, and region in CKB.
†Major cardiovascular disease (CVD) and socio-economic risk factors were smoker status, cigarette equivalents (cigarettes/day), systolic blood pressure, diastolic blood pressure, diagnosed hypertension, diagnosed diabetes, self-rated walking pace (UKB only), household income, and education level.

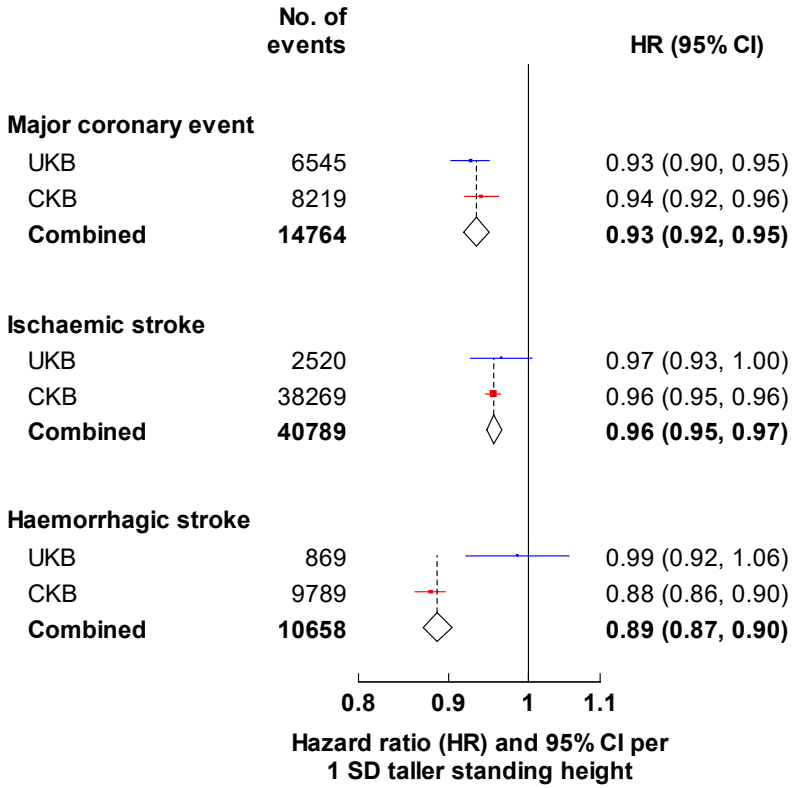


Figure 5.2 Combined associations of taller standing height with incident major coronary events and stroke types in UKB (n=496228) and CKB (n=513214), after basic adjustments and adjustment for major CVD and socio-economic risk factors.

Basic adjustments were age, sex, year of birth, and region in CKB. Major cardiovascular disease (CVD) and socio-economic risk factors were smoker status, cigarette equivalents (cigarettes/day), systolic blood pressure, diastolic blood pressure, diagnosed hypertension, diagnosed diabetes, self-rated walking pace (UKB only), household income, and education level.

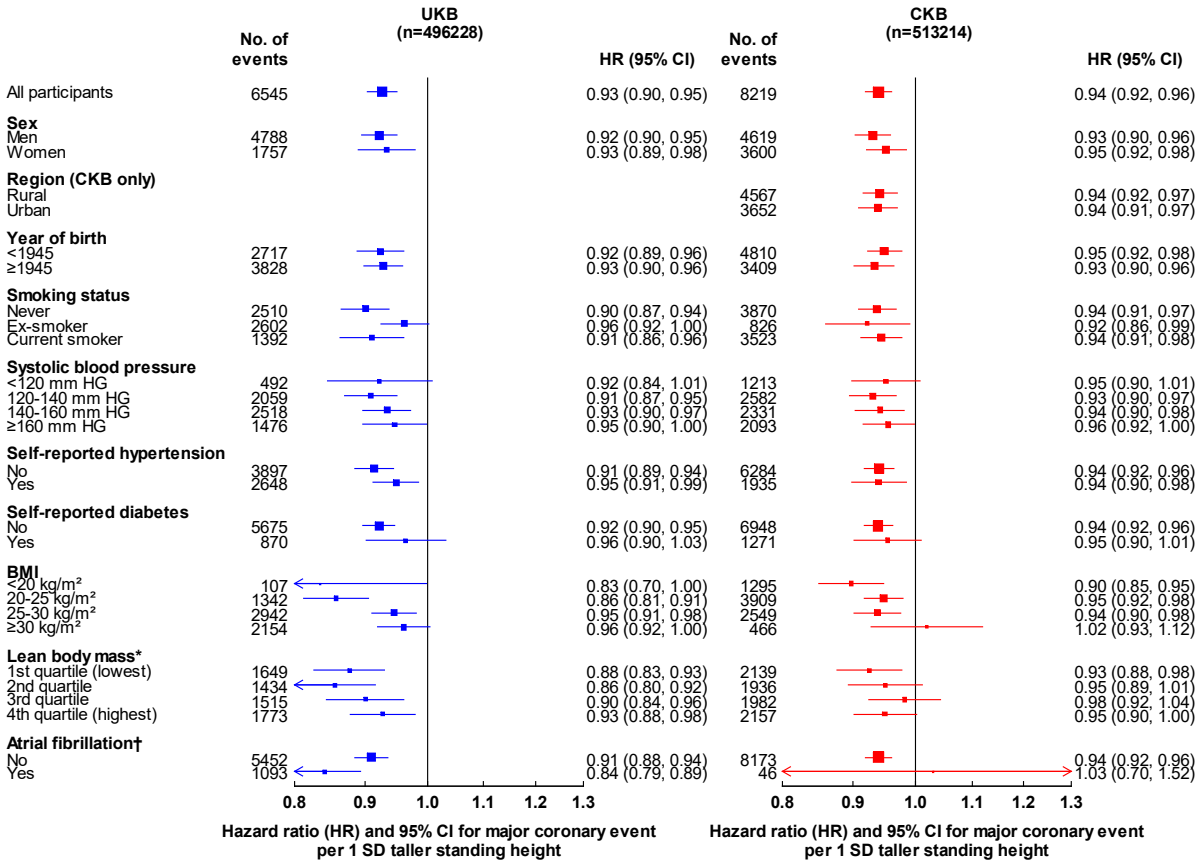


Figure 5.3 Associations of taller standing height with incident major coronary events, by various baseline characteristics, after basic adjustments and adjustment for major CVD and socio-economic risk factors.

Basic adjustments were age, sex, year of birth, and region in CKB. Major cardiovascular disease (CVD) and socio-economic risk factors were smoker status, cigarette equivalents (cigarettes/day), systolic blood pressure, diastolic blood pressure, diagnosed hypertension, diagnosed diabetes, self-rated walking pace (UKB only), household income, and education level.

*Quartiles were based on sex normalised lean body mass, prepared using the same methodology used for the height phenotype.

†Prevalent at baseline or incident atrial fibrillation.

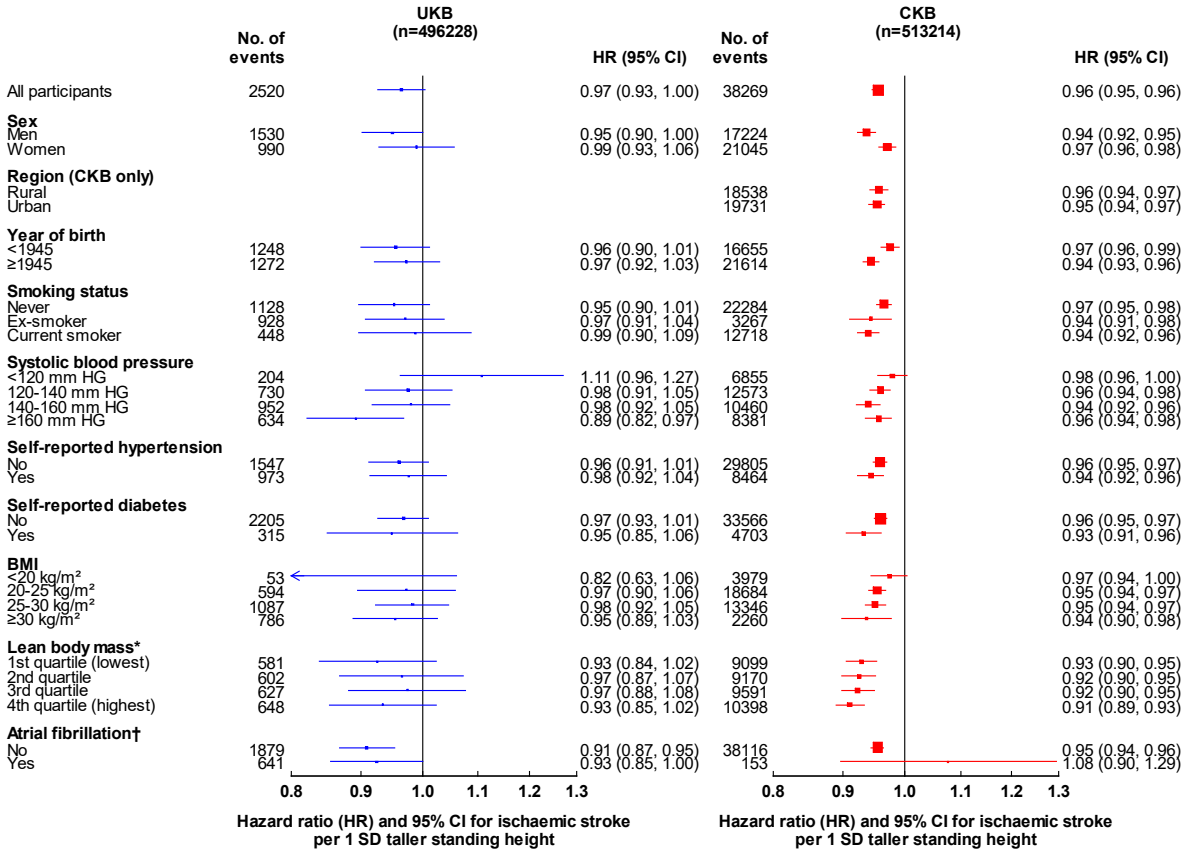


Figure 5.4 Associations of taller standing height with incident ischaemic stroke, by various baseline characteristics, after basic adjustments and adjustment for major CVD and socio-economic risk factors.

Basic adjustments were age, sex, year of birth, and region in CKB. Major cardiovascular disease (CVD) and socio-economic risk factors were smoker status, cigarette equivalents (cigarettes/day), systolic blood pressure, diastolic blood pressure, diagnosed hypertension, diagnosed diabetes, self-rated walking pace (UKB only), household income, and education level.

*Quartiles were based on sex normalised lean body mass, prepared using the same methodology used for the height phenotype.

†Prevalent at baseline or incident atrial fibrillation.

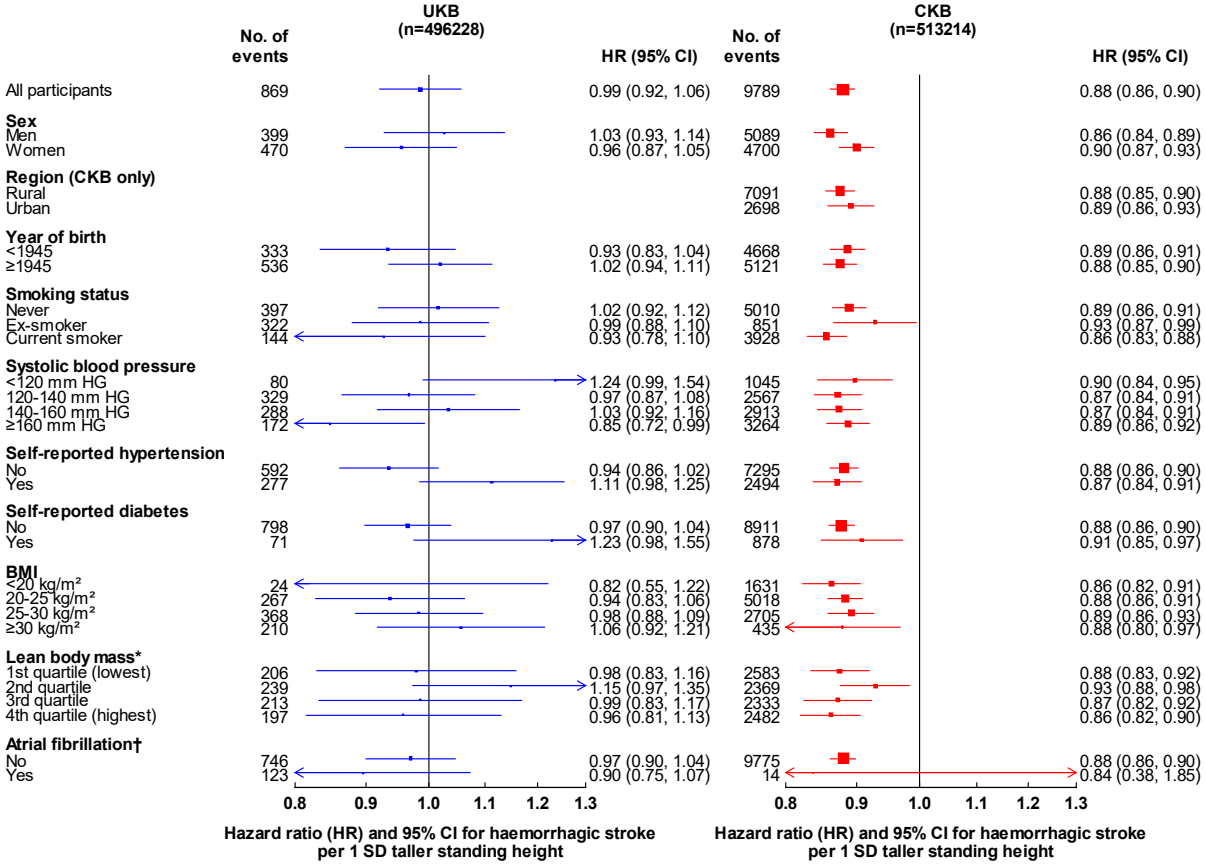


Figure 5.5 Associations of taller standing height with incident haemorrhagic stroke, by various baseline characteristics, after basic adjustments and adjustment for major CVD and socio-economic risk factors.

Basic adjustments were age, sex, year of birth, and region in CKB. Major cardiovascular disease (CVD) and socio-economic risk factors were smoker status, cigarette equivalents (cigarettes/day), systolic blood pressure, diastolic blood pressure, diagnosed hypertension, diagnosed diabetes, self-rated walking pace (UKB only), household income, and education level.

*Quartiles were based on sex normalised lean body mass, prepared using the same methodology used for the height phenotype.

†Prevalent at baseline or incident atrial fibrillation.

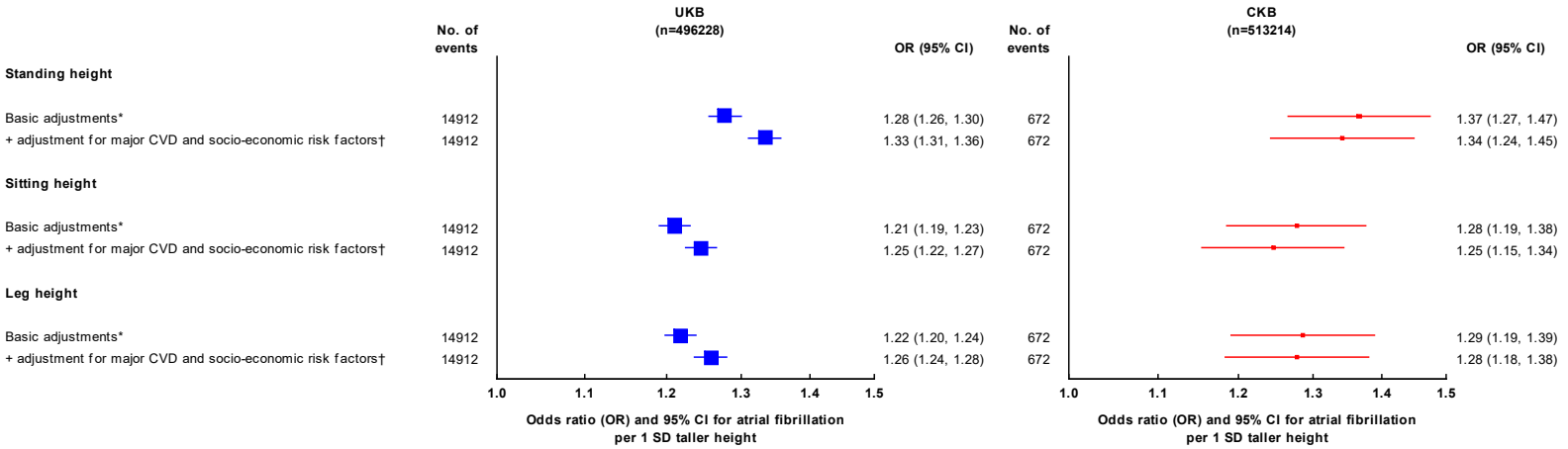


Figure 5.6 Associations of taller standing, sitting, and leg height with prevalent or incident atrial fibrillation, shown for increasing levels of adjustment.

*Basic adjustments were age, sex, year of birth, and region in CKB.
†Major cardiovascular disease (CVD) and socio-economic risk factors were smoker status, cigarette equivalents (cigarettes/day), systolic blood pressure, diastolic blood pressure, diagnosed hypertension, diagnosed diabetes, self-rated walking pace (UKB only), household income, and education level.

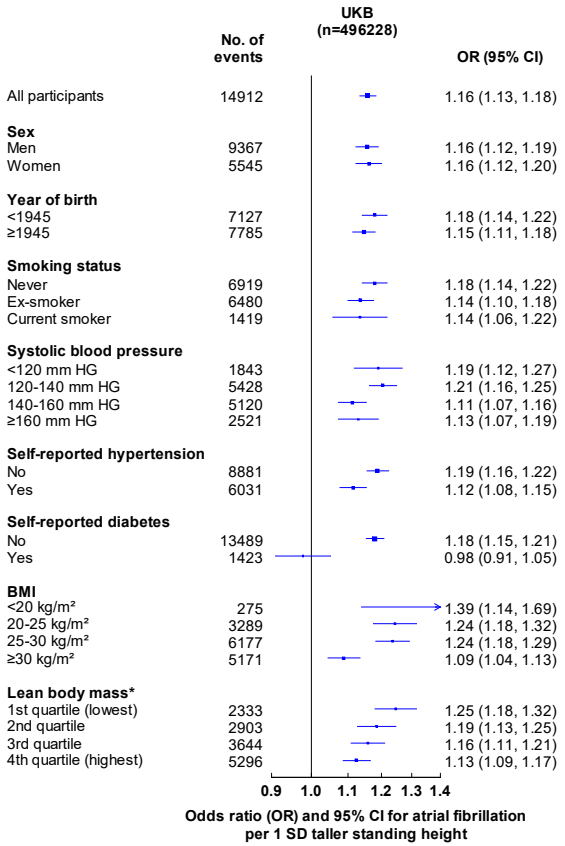


Figure 5.7 Associations of taller standing height with prevalent or incident atrial fibrillation, by various baseline characteristics, after basic adjustments, adjustment for major CVD and socio-economic risk factors, and adjustment for lean body mass.

Basic adjustments were age, sex, and year of birth. Major cardiovascular disease (CVD) and socio-economic risk factors were smoker status, cigarette equivalents (cigarettes/day), systolic blood pressure, diastolic blood pressure, diagnosed hypertension, diagnosed diabetes, self-rated walking pace (UKB only), household income, and education level.
*Quartiles were based on sex normalised lean body mass, prepared using the same methodology used for the height phenotype.

6. Genetic instruments for height in UKB and CKB

6.1 Introduction

Heritability of height

Heritability of height, that is the proportion of variation in height attributable to genetic variation, is substantial and has been estimated in twin studies to be 70-90%.[26–29] In contrast, the heritability of BMI has been estimated to be only ~60%.[29] SNP-based heritability is defined as the proportion of variation explained by a set of SNPs (e.g. a set of SNPs with a genome-wide significant association with a specific trait, or all SNPs imputed from a reference panel or from whole-genome sequencing).[82] SNP-based heritability is typically estimated using a genomic relatedness matrix (GRM) restricted maximum likelihood (GREML) approach, implemented in software such as GCTA.[20,82] By comparing pairs of individuals in a study population, this approach examines the effects of a set of SNPs simultaneously as random effects and estimates the proportion of phenotypic variance explained by the additive effects of all the SNPs regardless of the significance of their association with the phenotype (or trait).[20,83]

The proportion of phenotypic variance explained by a specific set of SNPs may also be reported using R^2 (squared correlation coefficient between genetic predictors and the phenotype).[43] The R^2 statistic is more easily estimated and does not require the use of genetic software. The proportion of phenotypic variance explained (R^2) by a set of SNPs should be estimated in a completely independent population to the SNP discovery population to avoid overestimating the R^2 estimate.[83] Additionally, when the SNP effect sizes on the phenotype are

used to estimate the R^2, for example, when the SNPs are weighted by their effect sizes and combined into a genetic score, the SNP effect sizes should be estimated in a population completely independent from the population used to estimate the R^2 (and should not be re-estimated in the population used to estimate the R^2).[83] However, the R^2 statistic underestimates the true proportion of phenotypic variance explained and will only equal the SNP-based heritability estimate (restricted to a specific set of SNPs) when the SNP effect sizes are estimated without error.[83]

For height, a major genome-wide association study (GWAS) was performed on 0.25M participants of European ancestry by the Genetic Investigation of Anthropometric Traits (GIANT) consortium, an international collaboration set up to discover the genetic loci that influence human body size and shape (Wood et al., 2014).[20] The SNP-based heritability of height in GIANT was estimated to be 50% (estimated using GREML and based on >1M SNPs). Wood et al. identified 697 SNPs associated with height which explained ~16% of the variance of height (estimated using GREML) and had an R^2 of ~14% (estimated using genetic predictors in 5 validation populations) (Table 6.1).[20] When this set of SNPs was expanded to include all those with P-values <1.0E-3 (~9500 SNPs), the GREML percentage of variance estimate increased to ~29%.

Marouli et al. (2017) extended the discovery of SNPs associated with height to include rare (minor allele frequency [MAF] <1%) and low frequency (1%≤ MAF ≤5%) SNPs in a study of 0.46M participants mostly of European ancestry (including 120K UKB participants).[42] Marouli et al. identified 91 rare and low

frequency SNPs associated with height which explained an additional ~1.7% of the variance of height (estimated using GRM) (Table 6.1).

In a pre-publication report (March 2018)[43] and a later published report (August 2018)[30], Yengo et al. used the GIANT dataset together with 0.45M UKB participants to increase the number of SNPs associated with height to 3290. These SNPs explained 24.6% of the variance of height (estimated using GREML) and had an R^2 of 19.7% (estimated using a genetic predictor in the Health and Retirement Study) (Table 6.1).[30]

In the present study of ~0.1M CKB participants, SNP-based heritability of height has been estimated to be 49% (estimated using GREML [BOLT-REML software]).

Ancestry coverage in genome-wide association studies

GWAS consortia have used large datasets, typically involving >0.1M participants, to discover SNPs associated with height and other physical traits and diseases.[20,84–90] Consortia data have chiefly involved individuals of European ancestry. For example, the International Consortium for Blood Pressure was based on 0.2M participants who all had European ancestry.[34,54,90] However, studies that have included individuals of other ancestries have concluded that the effects were similar in different populations, which was the case for the trans-ethnic GWAS of blood pressure, that was based on 0.1M participants, of which approximately two-thirds had Asian ancestry.[84] In the large GWAS of height, the proportion of other ethnicities has been relatively low and may not have been

adequate to capture ethnic differences in the heritability of height (Table 6.2). Previous small GWAS analyses for height have been performed in East Asian populations and included studies by He et al. (2015) (36K participants), Hao et al. (2013) (7K participants), and Okada et al. (2010) (20K participants) (Table 6.1 and Chapter 2).[59–61] Together these studies identified 21 genome-wide significant SNPs which were not directly identified by Wood et al.[20] but may still be in linkage disequilibrium with SNPs identified by Wood et al. (Table 6.1).

Genetically instrumented height

Height-associated SNPs can be used to construct genetic predictors of height (also referred to as genetically determined height or genetically instrumented height). If individual participant data are available, this can be achieved by compiling genetic risk scores for height. In an unweighted genetic risk score, the number of height-increasing alleles a participant possesses is summed. In a weighted genetic risk score, the number of height-associated alleles a participant possesses is first weighted (typically by their SNP effect size on height derived from independent sample to avoid bias [discussed below]) and then summed. The SNP effect sizes on height are typically obtained from large GWAS of height. However, combining SNPs and effect sizes from various GWAS can be problematic, due to possible non-independence of the height-associated SNPs and possible different units of effect sizes on height (due to different height phenotype preparations).

If individual participant data are not available, it is possible to use GWAS summary data (typically available from consortia) for a meta-analysis (usually by

inverse-variance weighting) of the SNP associations with height (and other traits/disease outcomes) to assess the impact of genetically instrumented height on other traits or disease outcomes.[4]

Mendelian randomisation analysis

Mendelian randomisation is a powerful tool to investigate the causal relevance of associations of risk factors with disease or other outcomes.[25] In Mendelian randomisation, genetic variants associated with a risk factor are used to form an instrumental variable (such as a genetic risk score) or variables to assess the associations of the genetically instrumented risk factor with the outcome, thereby avoiding the problems of residual confounding and reverse causation which are often present in observational associations. Mendelian randomisation studies can also help to identify shared biological pathways and discover potentially novel therapeutic targets for disease prevention.

The validity of a Mendelian randomisation study depends on the instrumental variable assumptions being satisfied for all genetic variants.[91] An instrumental variable is assumed to be:[91,92]

(1) Associated with the risk factor,

(2) Independent of any confounders of the association of the risk factor with the outcome, and

(3) Independent of the outcome conditional on the risk factor and the confounders of the association of the risk factor with the outcome.

Mendelian randomisation analyses using genetic risk scores are valid even if all the trait-associated variants are not independent.[93] However, Mendelian randomisation using the meta-analysis of GWAS summary data (by inverse-variance weighting) may be invalid if the trait-associated variants are not independent.[94] Sensitivity analyses, such as the MR-Egger method, can be used to provide a valid test of association even if all the genetic variants are invalid instrumental variables.[95]

Bias in Mendelian randomisation studies

Two-sample Mendelian randomisation studies involve analyses of genetic data on the risk factors and outcomes in separate datasets (typically involving summary data in large genetic consortia). The estimates from instrumental variable analysis are asymptotically unbiased but can suffer from substantial bias in finite samples, known as weak instrument bias.[92,96] The magnitude of this bias depends on the strength of association between the genetic instrument and the risk factor (the F-statistic).[92,96] An F-statistic less than 10 may indicate that an instrument is weak and that causal inferences could be problematic,[91,94] although the exact value of this threshold is debatable.[96] In two-sample Mendelian randomisation, weak instrument bias is a bias towards the null.[92,97] However, when the two-sample Mendelian randomisation datasets are not independent the direction and extent of the bias is uncertain.[92]

One-sample Mendelian randomisation studies involve analyses of risk factors and outcomes in the same dataset. Weak instrument bias in one-sample Mendelian randomisation studies is biased in the direction of the observational

associations of the risk factor with the outcome.[92,97] Weak instrument bias is exacerbated if the same datasets are used for both SNP discovery and effect size estimation (i.e. the SNPs used to compile the instrument and estimation of their effect sizes were both obtained from the same population). This is due to the *winner's curse* phenomenon, whereby SNP effect sizes in the discovery dataset are often overestimated relative to their effect size in replication studies through random error.[98] Cross-validation, based on individual participant data, can be used to mitigate weak instrument bias,[92,99,100] but some bias may persist.[83]

Potential genetic instruments (for height) for use in UKB and CKB, and the properties of these instruments, are examined in the rest of the chapter.

6.2 Methods

Study populations

All UKB participants and ~0.1M CKB participants (as part of a nested case-control study) were genotyped and were available for the present analyses. The CKB nested case-control study included a random population sample of 77K participants (which included the 25K participants who attended the second resurvey ~8 years after the baseline survey) and an additional ~20K cases. The following cases were selected into the case-control study: fatal coronary artery disease (~1K), myocardial infarction (~1K), ischaemic stroke (~5K), intracerebral and subarachnoid haemorrhage (~5.5K), and chronic obstructive pulmonary disease (~5K). In the present chapter, all participants were included in the analyses, regardless of whether they were selected into the case-control study as a case or as part of the random population sample. In the subsequent chapter

(Chapter 7), for a specific disease outcome, the random population sample together with participants selected into the case-control study as a case of the specific disease outcome were retained for the Mendelian randomisation analyses. In addition to UKB and CKB, Mendelian randomisation analyses (Chapter 7) will also be performed on data from two genetic consortia, C+C4D[86] (61K cases of coronary artery disease and 124K controls) and METASTROKE[101] (12K cases of ischaemic stroke and 62K controls).

Genotyping available in UKB

In UKB, all participants were genotyped using an 0.8M SNP panel in 106 batches (0.05M participants were genotyped using Affymetrix UK BiLEVE Axiom array and 0.44M were genotyped using an Affymetrix UK Biobank Axiom array). The UKB study was imputed using two reference panels, namely, the Haplotype Reference Consortium reference panel and a merged version of UK10K and 1000 Genomes Phase 3 reference panels. Where SNPs were present in both panels, the Haplotype Reference Consortium reference panel was preferentially used. After imputation, 96M SNPs were made available. Genomic principal components were derived, by the UKB study, at a national level and the first 10 principal components were used in the analyses to avoid confounding by population stratification.[102]

Participant samples failing quality control (QC), performed by the UKB study, were excluded. Possible reasons for failing QC included duplicates, poor quality genetic information (missing rates >2% [i.e. genotyping call rate ≤98%] or unusually high heterozygosity after adjustment for population structure),

mismatch between genetic and reported sex, laboratory errors, or withdrawn consent. The UKB study identified and excluded poor quality SNPs at the batch-level using statistical tests for batch effects, plate effects, departures from Hardy-Weinberg equilibrium, sex effects, array effects, and discordance across control replicates (the average number of SNPs that failed QC per batch was 7704 [SD 721]).[37]

The UKB study identified related individuals by estimating kinship coefficients for all pairs of individuals in the study population.[37] A total of 107 162 pairs (147 731 participants) were related (3rd degree or closer).[37] The present analyses were restricted to 336 433 UKB participants of white British ancestry (78 674 participants were not of white British ancestry) who passed QC and were identified by the UKB study as being unrelated (73 270 participants failed QC or were related).

Genotyping available in CKB

In CKB, 77K participants from a random population sample and an additional ~20K cases were genotyped using a 0.8M SNP panel in batches of ~50 plates (Affymetrix custom-designed biobank array for CKB). The CKB study was imputed with 1000 Genome Phase 3 East Asian population (EAS) as the reference panel, using SNPs with MAF >0 in EAS. After imputation, 24M SNPs were made available. Genomic principal components were derived, by the CKB genetics team, at a national level and the first 8 principal components were used in the analyses to avoid confounding by population stratification.[102]

Participant samples failing QC, performed by the CKB genetics team, were excluded. Possible reasons for failing QC included duplicates, poor quality genetic information (missing rates >5% [i.e. genotyping call rate ≤95%] or unusually high/low heterozygosity), mismatch between genetic and reported sex, or population outliers (in each ancestry group). The CKB genetics team identified and excluded poor quality SNPs at the batch-level using statistical tests for batch effects, plate effects, and departures from Hardy-Weinberg equilibrium. Where appropriate, these tests were conducted in a single ancestry subset or relatedness-pruned subset of participants (i.e. with no pairs of second cousins or closer). In addition, the CKB genetics team excluded SNPs if they were poorly-clustered, had no call in ≥10%, had an overall genotyping call rate ≤98% (in those batches where they pass initial QC), or had a MAF deviating significantly from the reference datasets (MAF difference of >~0.2).

The CKB genetics team identified related individuals by estimating kinship coefficients for all pairs of individuals in the study population. Among the 100 706 participants genotyped, 49 731 pairs were found to have a kinship coefficient larger than 0.05 (as a comparator, first cousins have a kinship coefficient of 0.0625). A set of 72 473 participants (who passed QC) were identified by the CKB genetics team as being unrelated to each other (this set was identified such that the largest set of unrelated participants was retained). The present analyses were restricted to 95 914 CKB participants who passed QC (4792 participants failed QC). Sensitivity analyses were performed in the 72 473 unrelated CKB participants.

Genetic instruments for height constructed from published GWAS

GIANT (Wood et al., 2014) and Marouli et al. each performed a GWAS of height in European populations to identify SNPs associated with height at the genome-wide significance level of P<5.0E-8.[20,42] Based on 0.25M participants (data from the GIANT consortium [2014]), Wood et al. used GCTA conditional and joint (COJO) analysis (which accounts for linkage disequilibrium between SNPs) and identified 697 SNPs associated with height at the genome-wide significance level.[20] (Of these SNPs, 623 also reached genome-wide significance for univariate associations.) Based on 0.46M participants (data from the GIANT consortium [2017] and 120K UKB participants), Marouli et al. identified 91 rare and low frequency SNPs associated with height using RAREMETALWORKER software.[42] (3 of these SNPs had been previously identified in Wood et al.) A total of 692 Wood SNPs and 87 Marouli SNPs were available for analyses in UKB participants, and 677 Wood SNPs and 23 Marouli SNPs were available for analyses in CKB participants (Table 6.3).

Smaller GWAS of height in East Asian populations have also identified SNPs associated with height at the genome-wide significance level of P<5.0E-8 (Table 2.5). He et al. (2015)[59] (based on 36K East Asian participants) reported 17 novel SNPs, Hao et al. (2013)[60] (based on 6.5K Chinese participants) reported 3 novel SNPs, and Okada et al. (2010)[61] (based on 20K Japanese participants) reported 1 novel SNP. These SNPs were reported as novel at the time of their publications but may not be independent of SNPs identified in subsequent large GWAS in European populations. All 21 of the East Asian[59–61] height-associated SNPs were available for analyses in CKB participants (Table 6.3).

In a pre-publication report, GIANT (Yengo et al., March 2018) reported 3290 independent genome-wide significant loci based on the GIANT consortium (2018) and UKB datasets.[43] At this point the authors did not identify the names of the 3290 SNPs (although these have since been made available on the GIANT consortium's website[103]), but did make available the full GWAS summary data on the associations between 2.3M SNPs and height. The height-associated SNPs were identified using GCTA COJO analysis and a revised genome-wide significance threshold of P<1.0E-8. Out of the 3290 height-associated SNPs, 2388 reached genome-wide significance for univariate associations. These SNPs were present in 712 loci (defined as 1 Mb windows) and most of the SNPs were *nearly independent*, but some SNPs within the same locus may be in partial linkage disequilibrium.

In the present thesis, as the SNPs identified by GIANT (Yengo et al., 2018)[30] were not available, a clumping procedure was applied to the GWAS summary associations (using PLINK software) to identify independent associations, using a significance threshold of P<1.0E-8, linkage disequilibrium threshold of r^2>0.01, distance from the index SNP of <250 kb, and 1000 Genome Phase 3 European population (CEU) as the reference population for linkage disequilibrium calculations. At this significance threshold, 117 766 SNPs were associated with height and the clumping procedure identified 3211 independent associations. These were used to compile a GIANT+UKB genetic risk score. Of these SNPs, 3210 and 3039 were available in UKB and CKB, respectively (Table 6.4).

Selection of appropriate genetic instruments

To avoid bias (through over-estimating the proportion of the variance of height explained), the SNP effect sizes on height were obtained from published GWAS whose discovery samples were independent of the study in which the genetic risk score was applied (i.e. UKB or CKB). Cross-validation, based on individual participant data, can be used to mitigate weak instrument bias,[92,99,100] but some bias may persist.[83] Since the effect sizes on height for the GIANT+UKB[30] and Marouli[42] genetic risk scores were estimated, in part, using UKB data, the UKB analyses were based primarily on the Wood[20] genetic risk score and the CKB analyses were based primarily on the clumped GIANT+UKB[30] genetic risk score. However, as a comparison, genetic risk scores based on Marouli[42], East Asian[59–61] SNPs (in CKB) and GIANT+UKB[30] SNPs (in UKB) were also provided.

Genetic risk scores were constructed at an individual participant level by weighting the number of height-associated effect alleles by their univariate effect size on height as determined by published GWAS of height (Tables 6.3 and 6.4). For each genetic risk score, the height-associated effect alleles were defined as the effect alleles reported in the GWAS. Each of the height-associated SNPs in the dosage data was checked to ensure that the same pair of alleles was reported and, when necessary, the dosage was reversed to reflect the effect allele used in the GWAS (genetic data for the biobanks were provided for the forward strand). The effect allele frequencies of the height-associated SNPs in the biobanks were compared with the published effect allele frequencies in the literature.

Height phenotypes

The height phenotypes were prepared following the methodology proposed by the Global Lipids Genetics Consortium and GIANT consortium to account for height differences by sex and age (Global Lipids Genetics/GIANT Consortium HRC and 1000 Genome Phase 3 Imputation and Analysis plan [unpublished analysis plan circulated by Doctor Robin Walters, April 2017]). For each biobank, the entire cohort was stratified by sex (and also by region in CKB). Separately for each stratum, height was adjusted for age and age squared, and the residuals were inverse-normally transformed, giving a height phenotype in SD units (Chapter 3.3). The height phenotypes were not exactly the same as the phenotypes prepared in other studies. In GIANT (Wood et al., 2014), the height phenotype was described as sex standardised height.[20] In Marouli et al., height was residualised (for age, genomic principal components, and study-specific covariates [e.g. recruitment centre]) separately for sex (except for family-based studies where sex was included as a covariate in the model) and case-control status (if relevant), and inverse-normally transformed.[42] In GIANT (Yengo et al., 2018), height was residualised (for age, sex inferred from SNP data, 10 genomic principal components, recruitment centre, and genotyping batch) separately for males and females, and inverse-normally transformed.[30] All of these studies adjusted or stratified by sex and yielded a phenotype in SD units. Differences in the height phenotypes between the present study and the previous studies were ignored.

Statistical methods

The height phenotype was regressed separately using linear regression on each literature SNP, after adjustment for genomic principal components (10 in UKB and 8 in CKB) and genotyping array type. The resulting SNP effect sizes are referred to as univariate effect sizes. Heterogeneity between the estimated univariate effect sizes and the published effect sizes was assessed by performing a T-test for each SNP at the Bonferroni corrected significance level for the number of SNPs being tested (to account for multiple testing).

When using the GIANT+UKB genetic instrument in the UKB dataset, cross-validation was used to reduce the bias arising from using the instrument in the same dataset from which the effect sizes were in part estimated. The genotyped dataset was randomly divided into 100 equal groups (100-fold cross-validation) and, for each group, the univariate effect sizes were estimated from the remaining 99 groups.

SNPs with good and weak information

The size of the CKB genetic study (~0.1M) was much smaller than the size for large consortia in European populations (0.33M for GIANT [Wood et al., 2014] and 0.7M for GIANT+UKB [Yengo et al., 2018]).[20,30] In addition, ancestry differences between CKB and consortia data may result in differences in MAF for the height-associated SNPs of interest. Both study size and MAF determine the power of a study to detect an association of an individual SNP with height. Therefore, to compare the consistency of CKB estimates with previously published estimates, SNPs were categorised into those with good and weak

information, where SNPs with good information have MAF and effect in the literature large enough to expect to detect an association of a SNP with height in CKB with allowance for multiple testing: P-values of SNP effect sizes were tested using a conservative significance level of 1.5E-5, which is smaller than 0.05/3039 (the Bonferroni corrected significance levels for the 3039 GIANT+UKB SNPs in CKB). For a given true per allele effect of a SNP on a trait, the Z-value of the association is proportional to:

$$\sqrt{MAF(1-MAF).n}$$

where *n* is the size of the study. For a SNP to have a significant association at P<0.05 in CKB after Bonferroni correction for multiple tests, it would require a P<1.5E-5 ($Z_{1.5E\text{-}5}$ being the corresponding Z-value). In order to fulfil these criteria in CKB:

$$\left|\frac{\mathrm{Z}_{1.5\mathrm{E}-5}}{\mathrm{Z}_{\mathrm{Lit}}}\right| = \sqrt{\frac{MAF_{CKB}(1-MAF_{CKB}).n_{CKB}}{MAF_{Lit}(1-MAF_{Lit}).n_{Lit}}}$$

where *Lit* refers to literature values. Combined analyses of GIANT+UKB were based on 702 081 individuals, and there were 96 046 individuals in CKB. Therefore, applying these numbers, we would expect to be able to detect a Bonferroni corrected P<0.05 effect in CKB if:

$$|\mathrm{Z}_{\mathrm{Lit}}| > \frac{|\mathrm{Z}_{1.5\mathrm{E}-5}|}{\sqrt{\frac{MAF_{CKB}(1-MAF_{CKB}).n_{CKB}}{MAF_{Lit}(1-MAF_{Lit}).n_{Lit}}}}$$

The SNPs fulfilling this criterion were categorised as having good information, and those not meeting the criterion were categorised as having weak information.

Percentage of variance explained

The height phenotype was regressed separately on each genetic risk score (using linear regression), adjusted for genomic principal components and genotyping array type. The percentage of the variance of height explained by the genetic risk score was estimated by the R^2 statistic (squared correlation coefficient) between the genetic risk score and height. The R^2 statistic for the genetic risk score was calculated as a percentage of residual variance after the adjustments. In addition, the genetic risk scores were partitioned by the study specific MAFs of SNPs underlying the scores, and the R^2 statistics were calculated for 5% MAF ranges.

The percentage of the variance of height explained (R^2) at a SNP level was estimated by:

$$R^2 = 2\beta^2 MAF(1 - MAF)$$

where β is the SNP effect size on height and MAF is the minor allele frequency of the SNP. The drivers of the differences in R^2 statistics reported in the literature and estimated in CKB and UKB were explored by examining the impact of MAF and SNP effect size differences, as well as the impact of missing and heterogeneous SNPs.

Power calculations

The power of Mendelian randomisation genetic instruments to detect an odds ratio per 1 SD genetically elevated height was calculated for a range of odds ratios and instrument R^2, assuming 6.5K cases (approximate number of haemorrhagic stroke cases in the CKB genotyped dataset), a sample of 95K

(approximate size of the CKB genotyped dataset), and a Type I error rate of 5%.[104]

Additionally, for the approximate number of ischaemic stroke cases in UKB (3K cases), CKB (12K cases), and METASTROKE (12K cases), the power of Mendelian randomisation genetic instruments to detect an odds ratio per 1 SD genetically elevated height was calculated for a range of odds ratios, assuming an appropriate R^2 for the instrument available in the study, an appropriate study size, and a Type I error rate of 5%.[104]

Timelines

The genetic component of the present thesis builds on central processes of CKB genetics team. The CKB genotyped dataset of ~100K participants was first received in December 2017. As mentioned above, the identities of the SNPs discovered by GIANT (Yengo et al., 2018) were not included in the pre-publication report, and therefore a set of SNPs (identified by clumping) was used for the GIANT+UKB genetic risk score.

6.3 Results

Minor allele frequencies of missing SNPs

The number of available and missing height-associated SNPs (Wood[20], Marouli[42], and GIANT+UKB[30]) for UKB and CKB, subdivided by MAF ranges, are presented in Table 6.5. The SNPs with a previously reported low MAF tended not to be available in CKB (e.g. 81 out of 172 GIANT+UKB clumped SNPs missing from the CKB dataset had MAF≤5%). SNPs missing from the CKB dataset were

either monoallelic or not available in the imputation panel (1000 Genome Phase 3 East Asian population reference panel).

Effect allele frequency comparisons

The effect allele frequencies estimated from the UKB data (692 SNPs) were largely concordant with those reported by Wood[20] (correlation coefficient: r=1.00), but the effect allele frequencies estimated from the CKB data (3039 SNPs from clumping) were much less concordant with those reported by GIANT+UKB[30] (r=0.78) (Figure 6.1).

The distributions of MAFs of Wood[20] SNPs in UKB and the literature were similar (average MAF of 28.5% in UKB versus 28.7% in the literature) while the MAFs in CKB were on average lower (average MAF of 22.9% in CKB versus 29.2% in the literature) (Figure 6.2). (The distributions of literature MAFs for Wood SNPs in UKB and CKB differed slightly because of the different SNPs available in each genotyped dataset.) The percentage of the variance of height explained by genetic risk scores (R^2) shown for 5% MAF ranges, increased with increasing MAF, but was lower in CKB than UKB (maximum R^2 per 5% MAF range of 1.1% and 2.9%, respectively) (Figure 6.2). Likewise, for the GIANT+UKB[30] SNPs, the distributions of MAFs and patterns of R^2 for 5% MAF ranges were similar to the Wood[20] SNPs (Figure 6.3).

Consistent with the results of previous studies, SNPs with lower MAFs tended to have lower imputation quality (information statistics[105]), particularly in CKB (Figure 6.4). However, the majority of SNPs had imputation quality ≥0.95 (98%

of Wood[20] SNPs in UKB, and 77% of GIANT+UKB[30] SNPs in CKB). In CKB, a total of 89% of GIANT+UKB SNPs had an imputation quality ≥0.90.

Comparison of effect sizes on height

The univariate effect sizes estimated in UKB (692 SNPs) and CKB (677 SNPs) were compared with those reported by Wood et al.[20] The effect sizes estimated in UKB were concordant with those reported by Wood et al.[20] (r=0.94, with only 4 SNPs exhibiting significant heterogeneity) (Figure 6.5). The regression coefficient of the UKB effect sizes on the literature effect sizes increased from 0.86 to 0.94 when these heterogeneous SNPs were excluded. The effect sizes estimated in CKB were less concordant with those reported by Wood et al.[20] (r=0.68, with 39 SNPs exhibiting significant heterogeneity) (Figure 6.5). The regression coefficient of the CKB effect sizes increased from 0.61 to 0.66 when these heterogeneous SNPs were excluded.

The univariate effect sizes estimated in CKB (3039 SNPs from clumping) were compared with those reported by GIANT+UKB[30] (Figure 6.6). To account for differences in study size (96K CKB participants versus 0.7M GIANT+UKB[30] participants) and genetic ancestry, SNPs were categorised into those with good and weak information, where SNPs with good information have MAF and effect in the GIANT+UKB large enough to expect to detect an association of a SNP with height in CKB. The concordance of the effect sizes differed by SNPs with good and weak information (Figure 6.7). The effect sizes of the SNPs with good information (508 SNPs) had a correlation of 0.85 with 45 SNPs exhibiting significant heterogeneity. In contrast, the effect sizes of the SNPs with weak

information (2531 SNPs) had a correlation of 0.41 with 31 SNPs exhibiting heterogeneity (these SNPs tended to have lower effect sizes than GIANT+UKB). For SNPs with good information, the regression coefficient of effect sizes increased from 0.67 to 0.78 when the heterogeneous SNPs were excluded (Figure 6.7). Likewise, but to a lesser extent, for SNPs with weak information the regression coefficient of effect sizes increased from 0.64 to 0.66 when the heterogeneous SNPs were excluded.

In CKB, the 442 GIANT+UKB[30] SNPs with good information and imputation quality ≥0.95 had a higher regression coefficient of univariate effect sizes of 0.79 (estimated after excluding the 39 heterogeneous SNPs) compared with the 0.67 regression coefficient of the 66 SNPs (estimated after excluding the 6 heterogeneous SNPs) with good information and imputation quality <0.95 (Figure 6.8).

Percentage of the variance of height explained by genetic risk scores

The percentage of the variance of height explained by the genetic risk scores (R^2), constructed using Wood[20] genome-wide significant SNPs and effect sizes, was much lower in CKB than in UKB (6.6% versus 16.9%) (Table 6.6). The R^2 in CKB increased to 7.0% when the heterogeneous SNPs were excluded. The addition of Marouli[42] SNPs to the genetic risk score for UKB increased the percentage of the variance of height explained to 17.7%. The addition of Marouli[42] and East Asian[59–61] SNPs to the genetic risk score for CKB increased the percentage of the variance of height explained to 7.5% (with the heterogeneous SNPs excluded) (Table 6.6).

Similarly, the percentage of the variance of height explained by the genetic risk scores, constructed using clumped GIANT+UKB[30] SNPs and effect sizes, was much lower in CKB than in UKB (9.6% versus 26.2%) (Table 6.6). The R^2 in CKB increased to 10.5% when the heterogeneous SNPs were excluded, and the R^2 in UKB decreased to 25.3% after 100-fold cross-validation (Table 6.6).

In an independent European population, Yengo et al.[30] reported that the GIANT (Yengo et al., 2018)[30] SNPs explained 19.7% of the variance of height. The percentages of variance of height attributable to factors driving the lower percentage of the variance explained in CKB, relative to European populations, were estimated and presented in Table 6.7. A genetic risk score constructed from GIANT+UKB[30] SNPs (3210 SNPs from clumping) and GIANT (Wood et al., 2014)[20] effect sizes was applied in UKB, and explained 24.4% of the variance of height (Table 6.7). The percentages of variance of height in an independent European population (Yengo et al., 2018)[30] were most relevant, but the attribution of the variance could only be estimated in UKB. Therefore, the percentage of variance of height attributable to a specific factor in an independent European population was estimated by scaling the total percentage of variance of height explained reported by Yengo et al.[30] (19.7%) by the ratio of the percentage variance of height attributable to the factor in UKB to the total percentage of variance of height explained in UKB (24.4%). Out of the 3210 SNPs, those that were missing in CKB (171 SNPs) or were heterogeneous in CKB (76 SNPs) were removed from the genetic risk score. This reduced the percentage of variance of height explained in UKB to 24.0% and the estimate reported by Yengo et al.[30] to

~19.4% (Table 6.7). The impact due to differences in the MAFs of the GIANT+UKB[30] SNPs in CKB relative to a European population (GIANT [Wood et al., 2014][20]) was estimated to be ~3.5% (expressed in terms of the percentage of variance of height estimate reported by Yengo et al.[30]) (Table 6.7). Likewise, the impact due to differences in effect sizes of the GIANT+UKB[30] SNPs in CKB relative to a European population (GIANT [Wood et al., 2014][20]) was estimated to be ~5.9% (also expressed in terms of the percentage of variance of height estimate reported by Yengo et al.[30]) (Table 6.7).

Power calculations

The Wood[20] genetic instrument, which explained 6.6% of the variance of height in CKB, provides 71% power to detect an odds ratio for haemorrhagic stroke (assuming 6.5K cases) in CKB of at least as extreme as 0.875 (which was approximately equivalent to the odds ratio reported in the observational analyses) per 1 SD taller genetically instrumented height (Table 6.8). The power to detect an effect of 0.90 per 1 SD taller genetically instrumented height was 52%.

The GIANT+UKB[30] genetic instrument (from clumping) provides 89% power to detect an odds ratio for haemorrhagic stroke in CKB of at least as extreme as 0.875 per 1 SD taller genetically instrumented height (and 72% power to detect an effect of 0.90) (Table 6.8). Likewise, for ischaemic stroke, the instrument provides 39% power in CKB and 62% power in METASTROKE[101] to detect an odds ratio of at least as extreme as 0.95 (approximate effect in the observational analyses) per 1 SD taller genetically instrumented height (Table 6.9).

Associations of the genetic risk scores with height in UKB and CKB

1 unit of the respective genetic risk scores were associated with 0.82 SD of measured height in UKB and 0.59 SD in CKB (Table 6.10). The F-statistics for these instruments were 68 600 and 9303 for UKB and CKB, respectively (Table 6.10), providing strong evidence that the causal estimates using these instruments should be unaffected by weak instrument bias.[91,96]

6.4 Discussion

The genetic risk score planned for use in UKB was constructed from Wood[20] SNPs and effect sizes, and the genetic risk score planned for use in CKB was constructed from clumped GIANT+UKB[30] SNPs and effect sizes, excluding the SNPs showing heterogeneity. These genetic risk scores explained 16.9% of the variance of height in UKB and 10.5% in CKB (Table 6.6). The GIANT+UKB[30] genetic instrument provides 89% power to detect an odds ratio for haemorrhagic stroke in CKB at least as extreme as the observational association reported in Chapter 5.

Height was estimated to have similar heritability in Europeans and East Asians (SNP-based heritability of 50% in 0.25M Europeans[20] and 49% in CKB [from the CKB genetics team]). However, the percentage of the variance of height explained (by the proposed genetic risk score for CKB) was much lower in CKB than in European populations (10.5% in CKB versus 19.7% reported in an independent population[30]) (Tables 6.1 and 6.7).

In CKB, a total of 171 GIANT+UKB[30] SNPs were missing and 76 SNPs showed heterogeneity of effect sizes (Table 6.7). In our attempted attribution (Table 6.7), these SNPs did not account for much loss of percentage of height variance explained (when estimated, it reduced the percentage of height variance explained from 19.7% to 19.4%). Although the removal of the heterogeneous SNPs from the genetic risk score in CKB increased the percentage of the variance explained by a worthwhile amount (from 9.6% to 10.5%) (Table 6.6). In CKB, 5.4% of the GIANT+UKB[30] SNPs were missing while only 2.9% of the Wood[20] SNPs were missing, possibly because when the CKB panel was designed it deliberately covered many SNPs with known GWAS associations.

The discovery of Wood and GIANT+UKB[30] SNPs was based on European populations and therefore the UKB MAFs follow the Wood[20] and GIANT+UKB MAFs closely (Figures 6.2 and 6.3). For a given effect size, height-associated SNPs with higher MAFs have greater power of being detected and so will be over-represented among discovered SNPs in the discovery population. The MAFs of Wood SNPs in CKB were lower than UKB due to linkage differences and the lack of representation of Chinese participants in the discovery populations. The lower MAFs of the SNPs in the CKB population, relative to a European population (GIANT [Wood et al., 2014][20]), were estimated to account for ~3.5% of the difference in variance explained (Table 6.7). In a detailed investigation of blood lipids (based on a genome-wide discovery in 48K East Asians and a meta-analysis with 300K Europeans), Lu et al. (2017)[31] reported that most rare and low frequency SNPs associated with blood lipids were population specific, and that SNPs identified in European populations tended to have lower MAFs in East

Asian populations. Although this research was conducted on blood lipids, their insights may well be applicable to other traits. It is possible that height-associated SNPs with high MAFs in East Asian populations and very low MAF in Europeans are yet to be discovered.

Lower average effect sizes in the CKB population, relative to a European population (GIANT [Wood et al., 2014][20]), were estimated to account for ~5.9% of the difference in variance explained (Table 6.7). Lu et al. (2017)[31] reported that within the same genes, lead SNPs often differed between Europeans and East Asians. Identifying different lead SNPs in East Asians could increase the strength of the genetic instrument. Since the majority (77%) of SNPs in CKB had information statistics (imputation quality) ≥0.95 (Figure 6.4), poor imputation quality did not provide an explanation for lower percentage of the variance explained.

The height-associated SNPs included in the genetic risk scores were mostly independent. For the genetic risk score constructed from Wood SNPs, the SNPs, according to Wood, were mostly uncorrelated but some SNPs within the same locus (defined as 1 Mb windows) may be in partial linkage disequilibrium.[20] The SNPs included in the GIANT+UKB[30] genetic risk score had linkage disequilibrium of $r^2 \leq 0.01$ (from the clumping procedure used) or were more than 250 kb apart (a clumping window of 500 kb was used). Even if these SNPs were not independent, their use in a genetic risk score for Mendelian randomisation would be valid.[93] Although Mendelian randomisation using the meta-analysis of consortia GWAS summary data (by inverse-variance weighting) may be invalid if the SNPs are not

independent.[94] Since the Wood[20] and GIANT+UKB[30] SNPs were mostly independent, it was also appropriate to use these sets of SNPs in consortia data (C+C4D[86] and METASTROKE[101]).

The SNPs included in the GIANT+UKB[30] genetic risk score were identified by clumping. This approach was taken in part because the SNPs identified by Yengo et al. (2018)[30] were not available until too late for the analyses in this thesis (the SNPs have subsequently been uploaded on the GIANT consortium's website[103]). Yengo et al.[30] used a conditional and joint analysis (GCTA COJO analysis) to identify the genome-wide height-associated SNPs. These SNPs might be expected to provide a somewhat stronger instrument in a very similar independent population, but the strength of the instrument may be sensitive to differences in linkage disequilibrium structure. Hence, these SNPs may not be of benefit in the Chinese population. Had the information about the Yengo et al.[30] SNPs been available earlier, sensitivity analyses comparing the instrument strength could have been undertaken, but would not necessarily have been expected to yield a great improvement over the approach taken. In any case, the GIANT+UKB[30] genetic instrument, which explained 10.5% of the variance of height in CKB, was a large improvement on the Wood[20] genetic instrument (which explained 6.6% of the variance of height in CKB).

A genome-wide discovery in the CKB population of 100K might hold some potential to improve the instrument for CKB. For example, the genome-wide discovery could identify novel East Asian height-associated SNPs and could help to improve estimates of effect sizes at known loci. This was not attempted for the

present thesis because the timing of the availability of the GWAS data and access to computer resources was too late, and it was beyond the scope of the thesis to add in this task. Additionally, a genome-wide discovery would require an independent East Asian population for replication, which was not available. The GWAS by Lango et al. (2010)[62] identified 180 genome-wide significant SNPs explaining 10.5% of the variance of height, based on 134K European participants in the discovery dataset. Performing a GWAS in CKB of 100K participants could be expected to yield a similar number of discovery results. However, the loci identified would likely largely overlap those discovered in Europeans and the extent to which this could improve upon the GIANT+UKB[30] score is uncertain. The future potential of this is discussed further in Chapter 8.

In summary, for UKB and CKB respectively, strong genetic instruments based on 692 Wood[20] SNPs and 2963 clumped GIANT+UKB[30] SNPs and explaining 16.8% and 10.5% of variance have been defined and offer excellent power for the planned Mendelian randomisation analyses.

6.5 Tables

Table 6.1 Size and ancestry of genome-wide association studies (GWAS) for height.

Author, year	Individuals in GWAS discovery	Individuals in total (GWAS discovery + replication)	No. of genome-wide significant SNPs (P<5E-8)	European ancestry in discovery	East Asian ancestry in discovery	Percentage of the variance of height explained (estimated by GRM/GRML†)	Percentage of the variance of height explained (R^2)
Yengo, 2018[30]	693529	702081	3290*	693529	0	24.6%	19.7%
Marouli, 2017[42]	458927	711428	91	381625	8767	1.7%	-
He, 2015[59]	36227	93926	98	0	36227	-	8.9%
Wood, 2014[20]	253288	333355	697	253288	0	~16%	~14%
Hao, 2013[60]	6534	8415	5	0	6534	-	0.9%
Lango, 2010[62]	133653	183727	180	133653	0	-	10.5%
Okada, 2010[61]	19633	19633	8	0	19633	-	1.8%

*Study used a revised genome-wide significance level of P<1.0E-8.
†GRM=genomic relatedness matrix; GREML=genomic relatedness matrix restricted maximum likelihood.

Table 6.2 Size and ancestry of large genome-wide association studies (GWAS) consortia for various traits.

			N				
Author, year	**Consortium***	**Trait**	**Individuals in GWAS discovery**	**European ancestry**	**East Asian ancestry**	**South Asian ancestry**	**Asian ancestry (incl. South and East)**
Kato, 2015[84]	BIOS, CARDIoGRAMplusC4D, LifeLines Cohort Study, InterAct Consortium	Blood pressure	99994	35352	31516	33126	64642
Locke, 2015[85]	GIANT	BMI	339224	251026	-	-	57668
Yengo, 2018[30]	GIANT+UKB	BMI	681275	681275	0	0	0
Nikpay, 2015[86]	CARDIoGRAMplusC4D	CAD†	184305	141915	11058	23960	35018
Mahajan, 2014[87]	DIAGRAM, AGEN-T2D, SAT2D, MAT2D, T2D-GENES	Type 2 diabetes	110452	69033	18817	20019	38836
Wood, 2014[20]	GIANT	Height	253288	253288	0	0	0
Yengo, 2018[30]	GIANT+UKB	Height	693529	693529	0	0	0
Willer, 2013[88]	GLGC	Blood lipids	196475	188577	0	0	0
Morris, 2012[89]	DIAGRAM	Type 2 diabetes	149821	149821	0	0	0
Evangelou, 2018[34]	ICBP+UKB	Blood pressure	757601	757601	0	0	0

*BIOS=Biobank-based Integrative Omics Studies; CARDIoGRAMplusC4D=Coronary ARtery DIsease Genome-wide Replication and Meta-analysis (CARDIoGRAM) plus The Coronary Artery Disease (C4D) Genetics; DIAGRAM=DIAbetes Genetics Replication And Meta-analysis; AGEN-T2D=Asian Genetic Epidemiology Network Type 2 Diabetes; SAT2D=South Asian Type 2 Diabetes; MAT2D=Mexican American Type 2 Diabetes; T2D-GENES=Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples; GIANT=Genetic Investigation of Anthropometric Traits; GLGC=Global Lipids Genetics Consortium; ICBP=International Consortium for Blood Pressure.
†CAD=coronary artery disease.

Table 6.3 SNPs associated with height that reached genome-wide significance (identified in the literature) available in UKB and CKB.

Author, year	SNPs genome-wide significant (P<5E-8)	SNPs genome-wide significant and potentially novel (relative to Wood[20])	SNPs available in UKB	SNPs available in CKB
Wood, 2014[20]	697	-	692	677
Marouli, 2017[42]	91	88	87	23
He, 2015[59]	98	17	-	17
Hao, 2013[60]	5	3	-	3
Lango, 2010[62]	180	0	-	-
Okada, 2010[61]	8	1	-	1
Total	1079	109	779	721

Table 6.4 SNPs associated with height that reached genome-wide significance (identified in GIANT+UKB[30] and by clumping) available in UKB and CKB.

	N			
Author, year, or clumping procedure	**SNPs genome-wide significant but not independent associations***	**SNPs genome-wide significant and independent associations***	**SNPs available in UKB**	**SNPs available in CKB**
Yengo, 2018[30]	117767	3290†	-	-
Clumping procedure	117767	3211	3210	3039

*Study used a revised genome-wide significance level of P<1.0E-8.
†List of 3290 independent SNPs not reported in the pre-publication report.[43]

Table 6.5 Number of available and missing height-associated SNPs in UKB and CKB, subdivided by minor allele frequency (MAF) ranges.

Description of height-associated SNPs	No. of SNPs	Average MAF in literature	No. of SNPs in the following literature MAF ranges:										
			<1%	1%-5%	5%-10%	10%-15%	15%-20%	20%-25%	25%-30%	30%-35%	35%-40%	40%-45%	>45%
UKB													
Wood[20] SNPs available in UKB	692/697	0.29	0	23	47	50	62	87	90	87	83	81	82
Wood[20] SNPs missing in UKB	5/697	0.26	0	0	1	1	0	0	0	2	0	1	0
Marouli[42] SNPs available in UKB	87/88	0.02	30	56	1	0	0	0	0	0	0	0	0
Marouli[42] SNPs missing in UKB	1/88	0.01	1	0	0	0	0	0	0	0	0	0	0
Clumped GIANT+UKB[30] SNPs available in UKB	3201/3202	0.26	0	204	326	315	326	354	326	347	320	342	341
Clumped GIANT+UKB[30] SNPs missing in UKB	1/3202	0.07	0	0	1	0	0	0	0	0	0	0	0
CKB													
Wood[20] SNPs available in CKB	677/697	0.29	0	17	42	49	58	86	89	89	83	82	82
Wood[20] SNPs missing in CKB	20/697	0.10	0	6	6	2	4	1	1	0	0	0	0
Marouli[42] SNPs available in CKB	23/88	0.03	0	22	1	0	0	0	0	0	0	0	0
Marouli[42] SNPs missing in CKB	65/88	0.01	31	34	0	0	0	0	0	0	0	0	0
East Asian[59–61] SNPs available in CKB	21/21	0.34	0	0	0	1	2	2	1	4	4	3	4
East Asian[59–61] SNPs missing in CKB	0/21	-	0	0	0	0	0	0	0	0	0	0	0
Clumped GIANT+UKB[30] SNPs available in CKB	3039/3211	0.27	0	123	280	289	320	348	326	347	323	342	341
Clumped GIANT+UKB[30] SNPs missing in CKB	172/3211	0.07	0	81	47	27	8	7	1	1	0	0	0

Table 6.6 Percentage of the variance of height explained by genetic risk scores constructed using literature SNPs and effect sizes in UKB and CKB.

	UKB (n=336433)		CKB (n=96046)	
Genetic risk score constructed from literature SNPs and effect sizes	**No. of SNPs**	**R^{2*}**	**No. of SNPs**	**R^{2*}**
Wood[20]	692/697	16.94%	677/697	6.55%
Wood[20] excluding heterogeneous SNPs	-	-	638/697	7.01%
Marouli[42]	87/88	1.42%	23/88	0.19%
Wood[20] and Marouli[42]	779/785	17.73%	700/785	6.50%
East Asian[59–61]	-	-	21/21	0.52%
Wood[20], Marouli[42] and East Asian[59–61]	-	-	721/806	6.61%
Wood[20], Marouli[42] and East Asian[59–61] excluding heterogeneous SNPs	-	-	677/806	7.47%
Clumped GIANT+UKB[30] SNPs	3210/3211	26.22%	3039/3211	9.56%
Clumped GIANT+UKB[30] SNPs excluding heterogeneous SNPs	-	-	2963/3211	10.51%
Clumped GIANT+UKB[30] SNPs data derived effect sizes (100-fold cross-validation)	3210/3211	25.34%	-	-

*R^2 adjusted for genomic principal components (first 10 principal components in UKB and first 8 in CKB) and genotyping array type.

Table 6.7 Estimated attribution of the lower R^2 of GIANT+UKB[30] SNP based genetic instrument in CKB compared to in European ancestry populations.

SNP set, source of effect sizes, and population the instrument was applied in	No. of SNPs	R^2	
		No scaling	Scaling*
European ancestry populations			
1. Yengo[30] SNPs and effect sizes applied in an independent population, as cited in Yengo et al. (2018)[30]†	3290	19.7%	-
2. GIANT+UKB[30] SNPs and Yengo[30] effect sizes applied in CKB¶	2963	10.5%	-
Possible ways of using the clumped GIANT+UKB genetics variants in a genetic risk score in UKB			
3. GIANT+UKB[30] SNPs and Yengo[30] effect sizes applied in UKB	3210	25.8%	-
4. GIANT+UKB[30] SNPs and UKB estimated effect sizes applied in UKB	3210	26.2%	-
5. GIANT+UKB[30] SNPs and 100-fold cross-validated UKB estimated effect sizes applied in UKB	3210	25.3%	-
6. GIANT+UKB[30] SNPs and effect sizes from GIANT (Wood et al., 2014)[20] applied in UKB	3210	24.4%	-
Components not available in CKB			
7. GIANT+UKB[30] SNPs for missing SNPs with effect sizes from GIANT (Wood et al., 2014)[20] applied in UKB	171	1.2%	1.0%
8. GIANT+UKB[30] SNPs for heterogeneous SNPs with effect sizes from GIANT (Wood et al., 2014)[20] applied in UKB	76	1.8%	1.5%
9. GIANT+UKB[30] SNPs for the remaining SNPs with effect sizes from GIANT (Wood et al., 2014)[20] applied in UKB	2963	24.0%	19.4%
Among SNPs available in CKB			
10. GIANT+UKB[30] SNPs and GIANT (Wood et al., 2014)[20] effect sizes (β) to calculate the loss of R^2 due to MAF differences in CKB relative to GIANT (Wood et al., 2014)[20]‡	2963	4.3%	3.5%
11. GIANT+UKB[30] SNPs and GIANT (Wood et al., 2014)[20] effect sizes to calculate the loss of R^2 due to lower average effect sizes in CKB relative to UKB§	2963	7.3%	5.9%

*R^2 scaled down by the ratio of R^2 in (1) to (6), where the R^2 in (6) was presumed to be the most independent estimate.
†R^2 from Yengo et al. (2018)[30] (Table 1 in the article, prediction accuracy (R^2) in Health and Retirement Study from genome-wide significant SNPs at a revised genome-wide significance threshold of P<1.0E-8). Note that the ~24.6% of the variance of height quoted in the abstract of the article refers to SNP-based heritability, not R^2.

¶The set of GIANT+UKB[30] SNPs was derived by clumping the GWAS summary associations reported in Yengo et al. (2018)[30]. When applied in UKB and CKB, except in (10), the GIANT+UKB[30] SNPs were constructed, at an individual participant level, into genetic risk scores by weighting the number of height-associated effect alleles by their effect size on height (as defined in each row of the table).
‡The loss of R^2 due to MAF differences in CKB relative to GIANT (Wood et al., 2014)[20] was calculated as (one minus the ratio of $\sum 2\beta^2 MAF[1-MAF]$ with CKB MAFs to GIANT MAFs) multiplied by (9).
§The loss of R^2 due to lower average effect sizes in CKB relative to UKB was calculated from the ratio of the regression coefficients of height regressed on the genetic risk score (constructed from GIANT+UKB[30] SNPs and GIANT [Wood et al., 2014][20] effect sizes) in CKB (0.56) to UKB (0.68), giving a ratio of 0.82, using the formula (one minus the square of the ratio of the regression coefficients) multiplied by the percentage in (9).

Table 6.8 Power of genetic instruments to detect odds ratios per 1 SD genetically elevated height, assuming 6.5K cases (approximate number of haemorrhagic stroke cases in the CKB genotyped dataset), a sample of 95K, and a Type I error rate of 5%.

Genetic instrument constructed from:	No. of SNPs	Genetic instrument applied in:	R^2	Power (%) of genetic instrument to detect an odds ratio of:				
				0.85	0.875	0.90	0.925	0.95
Wood[20] SNPs and effect sizes	697	European population	~14%*	99%	96%	83%	59%	31%
Wood[20] SNPs and effect sizes	677	CKB	6.6%	85%	71%	52%	32%	17%
Clumped GIANT+UKB[30] SNPs and effect sizes	3211	European population	19.7%†	100%	99%	94%	75%	42%
Clumped GIANT+UKB[30] SNPs and effect sizes¶	2963	CKB	10.5%	97%	89%	72%	48%	24%

*R^2 from Wood et al. (2014)[20] (Figure 2d, page 1175, prediction R^2 weighted average over genome-wide significant SNPs [P<5.0E-8]).
†R^2 from Yengo et al. (2018)[30] (Table 1 in the article, prediction accuracy [R^2] in Health and Retirement Study from genome-wide significant SNPs at a revised genome-wide significance threshold of P<1.0E-8). Note that the ~24.6% of the variance of height quoted in the abstract of the article refers to SNP-based heritability, not R^2.
¶Excluding SNPs showing heterogeneity in CKB.

Table 6.9 Power of genetic instruments to detect odds ratios for ischaemic stroke per 1 SD genetically elevated height, in METASTROKE, UKB and CKB, assuming a Type I error rate of 5%.

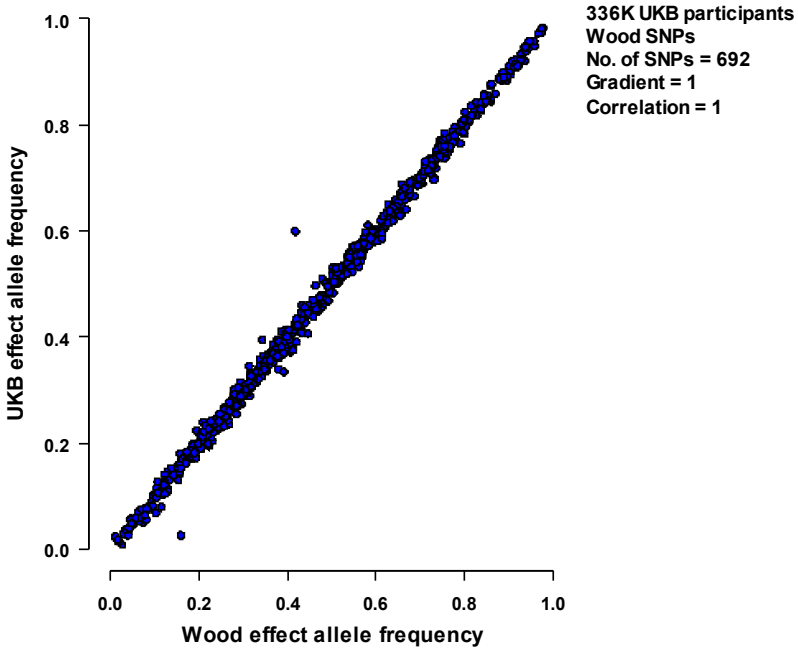
Study	Approx. no. of ischaemic stroke cases	No. of participants (rounded to nearest 5K)	Genetic instrument constructed from:	No. of SNPs	R^2	Power (%) of genetic instrument to detect an odds ratio of:				
						0.875	0.90	0.925	0.95	0.975
UKB	3K	335K	Wood[20] SNPs and effect sizes	692	16.9%	80%	61%	39%	20%	9%
CKB	12K	95K	Clumped GIANT+UKB[30] SNPs and effect sizes¶	2963	10.5%	99%	92%	71%	39%	13%
METASTROKE[101]	12K	75K	Clumped GIANT+UKB[30] SNPs and effect sizes	3150*	19.7%†	100%	100%	93%	62%	20%

*3150 out of 3211 SNPs identified through clumping were available in METASTROKE[101].
†R^2 from Yengo et al. (2018)[30] (Table 1 in the article, prediction accuracy [R^2] in Health and Retirement Study from genome-wide significant SNPs at a revised genome-wide significance threshold of P<1.0E-8). Note that the ~24.6% of the variance of height quoted in the abstract of the article refers to SNP-based heritability, not R^2.
¶Excluding SNPs showing heterogeneity in CKB.

Table 6.10 Associations of the genetic risk scores with height in UKB and CKB.

Estimate	UKB genetic risk score constructed from:		CKB genetic risk score constructed from:
	Wood[20] SNPs and effect sizes	Clumped GIANT+UKB[30] SNPs and 100-fold cross-validated UKB estimated effect sizes	Clumped GIANT+UKB[30] SNPs and effect sizes
Regression coefficient of the height phenotype regressed on the genetic risk score	0.82	0.59	0.59
F-statistic of the height phenotype regressed on the genetic risk score	68600	114174	9303
1 / regression coefficient of the height phenotype regressed on the genetic risk score	1.22	1.69	1.69

6.6 Figures



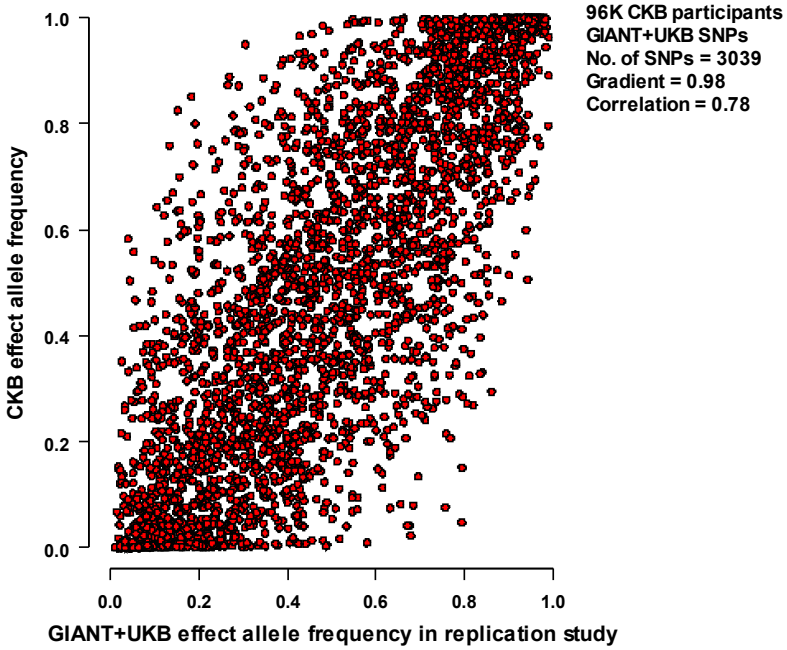


Figure 6.1 Effect allele frequency in UKB (336433 participants) versus Wood et al.[20] and CKB (96046 participants) versus GIANT+UKB[30].

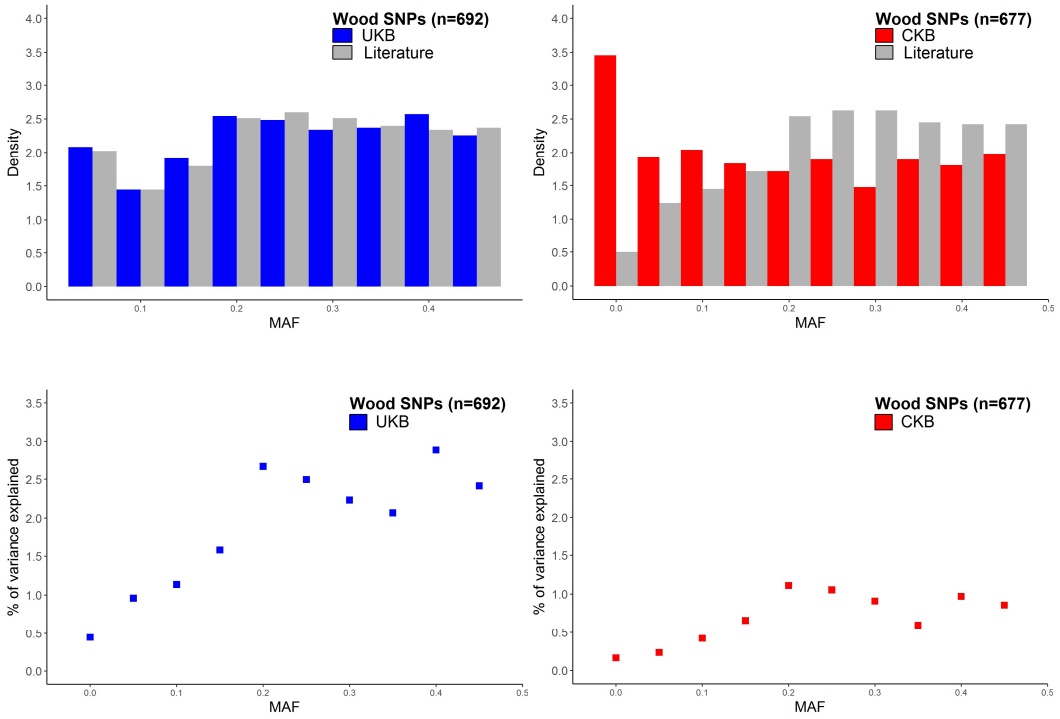


Figure 6.2 The distributions of minor allele frequencies (MAFs) of Wood[20] SNPs in UKB, CKB, and the literature, and the corresponding percentages of the variance of height explained by genetic risk scores (constructed using Wood[20] effect sizes), shown for 5% MAF ranges.

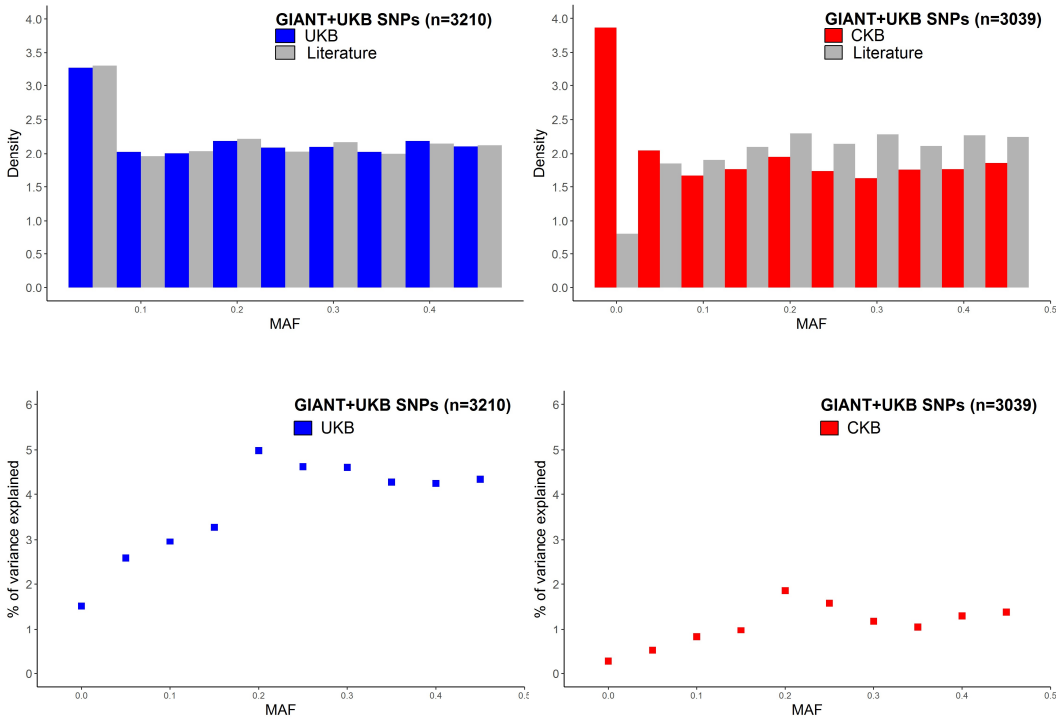
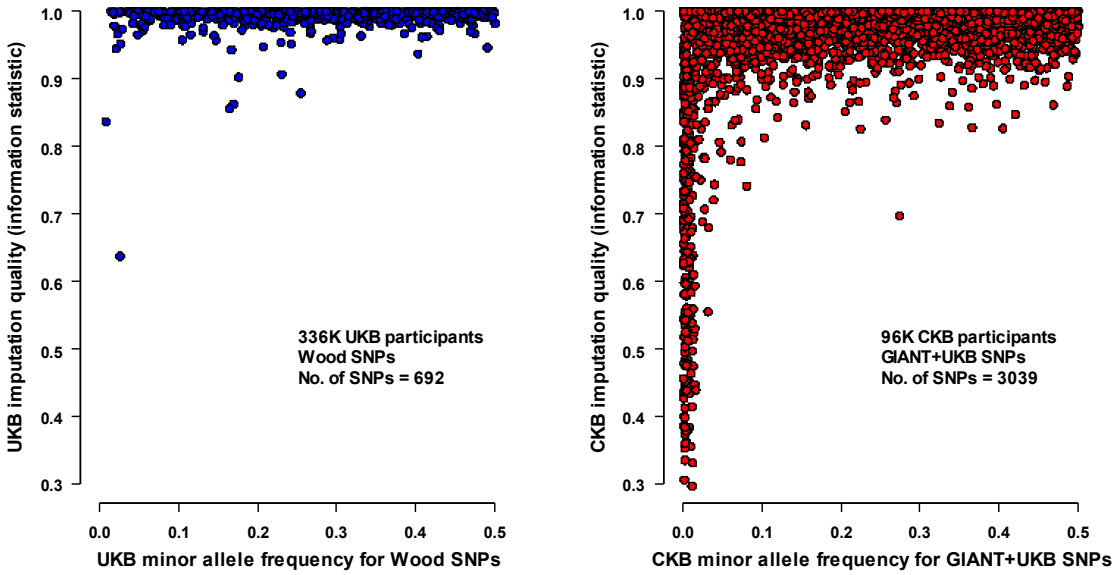


Figure 6.3 The distributions of minor allele frequencies (MAFs) of clumped GIANT+UKB[30] SNPs in UKB, CKB, and the literature, and the corresponding percentages of the variance of height explained by genetic risk scores (constructed using GIANT+UKB[30] effect sizes), shown for 5% MAF ranges.

The GIANT+UKB MAF was reported in the replication study of 8552 unrelated participants of the Health and Retirement Study;[30] and the R^2 estimates for UKB do not represent unbiased estimates because the genetic risk score was constructed from effect sizes derived, in part, form the UKB dataset.



Proportion of SNPs shown by imputation quality.

	Information statistic (imputation quality)			
Study	**<0.8**	**0.8-0.9**	**0.9-0.95**	**≥0.95**
UKB	0.00	0.01	0.01	0.98
CKB	0.06	0.05	0.12	0.77

Figure 6.4 Imputation quality and minor allele frequency for Wood[20] SNPs in UKB (336433 participants) and clumped GIANT+UKB[30] SNPs in CKB (96046 participants).

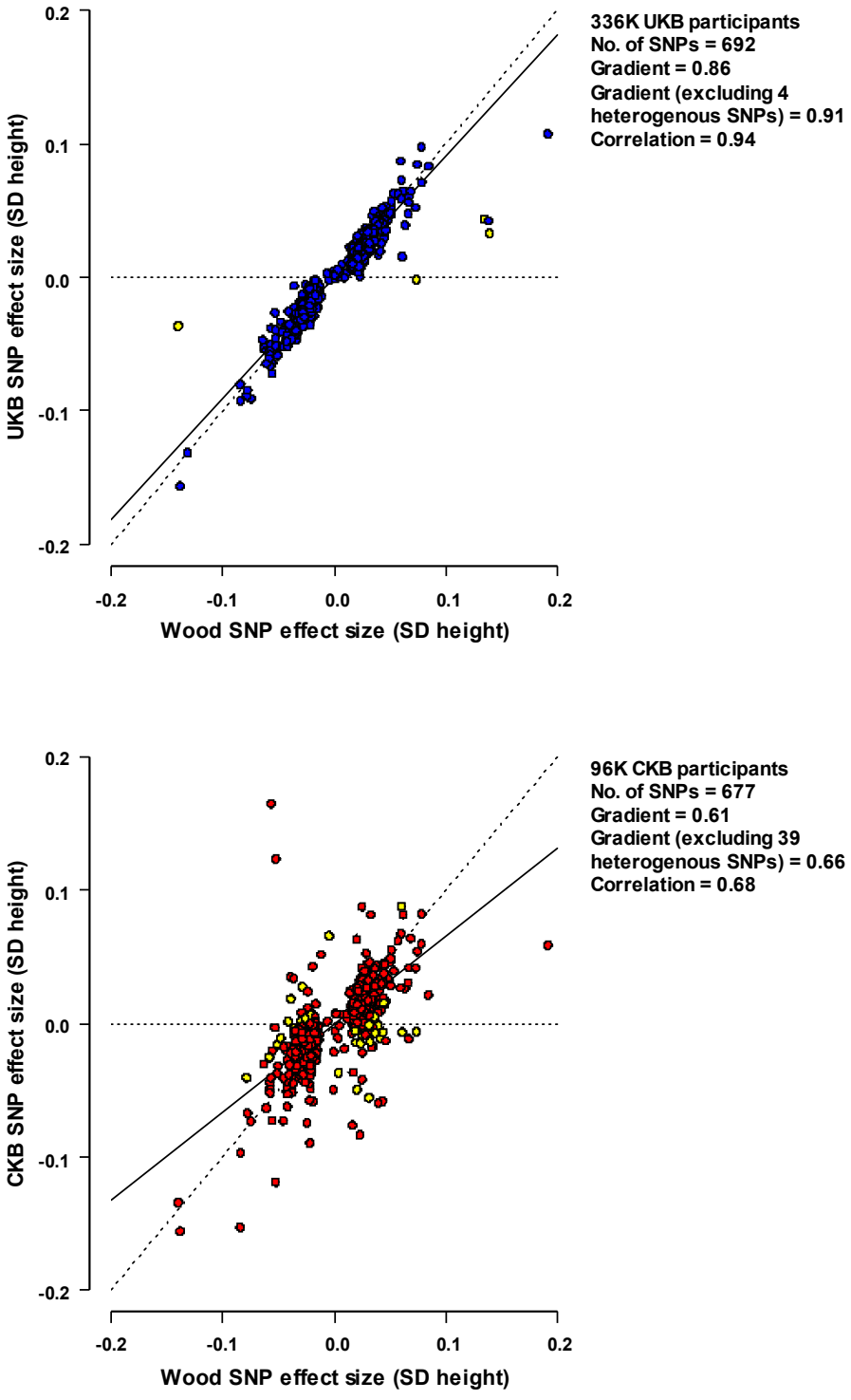


Figure 6.5 Univariate effect sizes of available Wood[20] SNPs on standardised standing height (336433 UKB participants and 96046 CKB participants).

SNPs in yellow indicate heterogeneity. Plotted gradients (solid black lines) excluded heterogeneous SNPs.

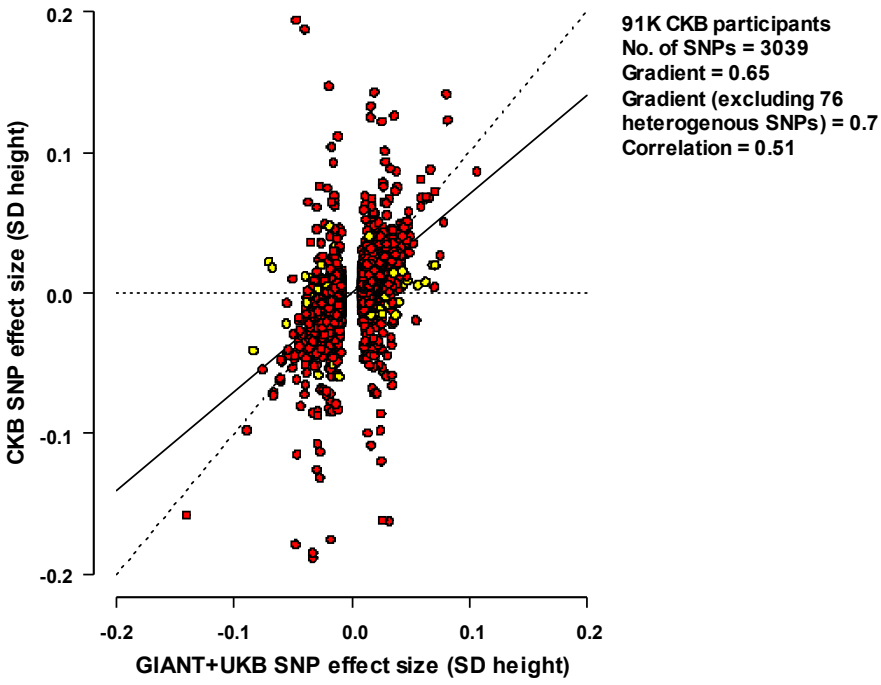


Figure 6.6 Univariate effect sizes of clumped GIANT+UKB[30] SNPs on standardised standing height (96046 CKB participants).

SNPs in yellow indicate heterogeneity. Plotted gradients (solid black lines) excluded heterogeneous SNPs.

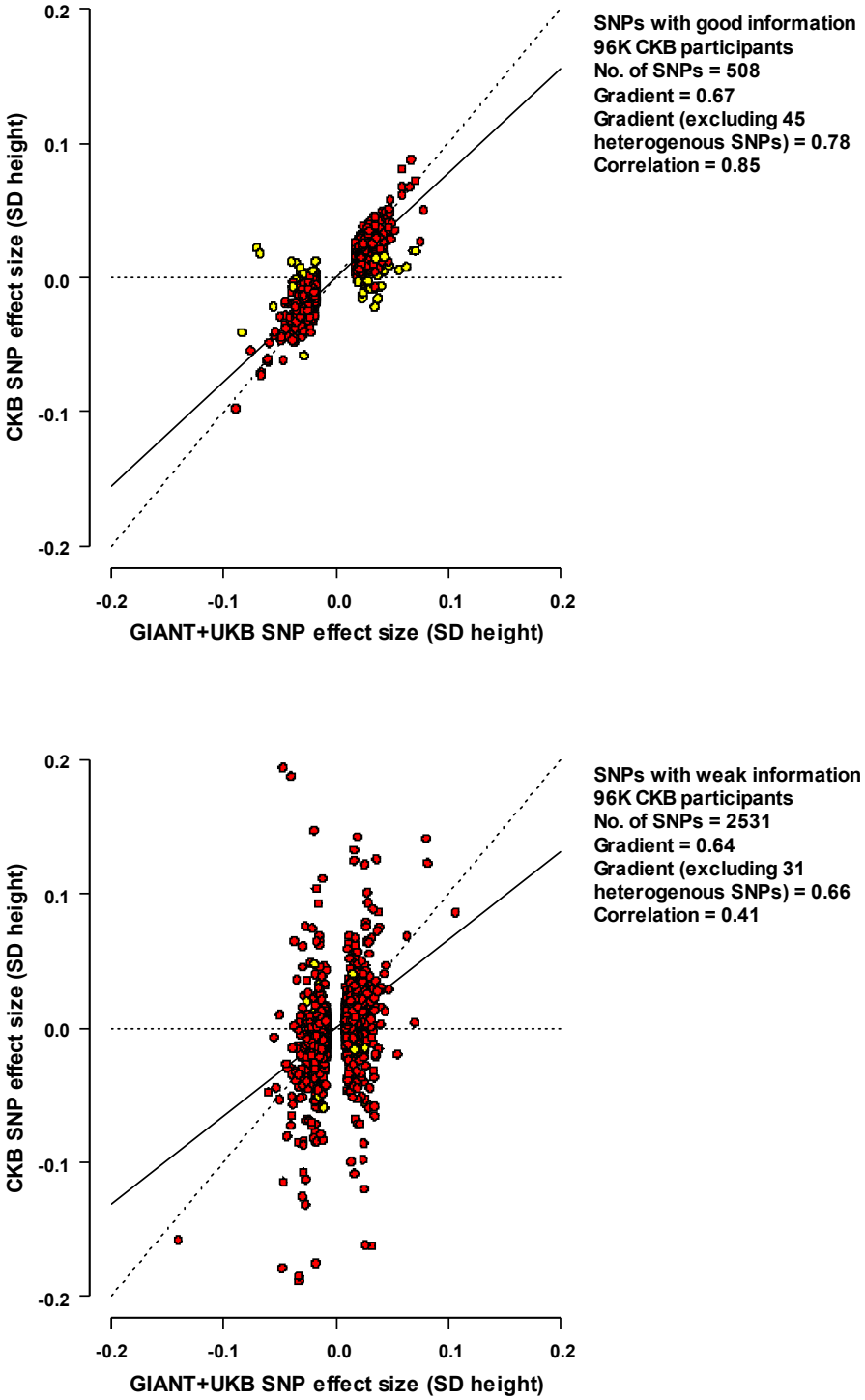


Figure 6.7 Univariate effect sizes of clumped GIANT+UKB[30] SNPs on standardised standing height (96046 CKB participants), classified by SNPs with good and weak information.

SNPs in yellow indicate heterogeneity. Plotted gradients (solid black lines) excluded heterogeneous SNPs. A SNP with good information was able to detect an association with height (given the minor allele frequency and sample size differences between CKB and the GWAS study, GIANT+UKB [Yengo et al., 2018][30]).

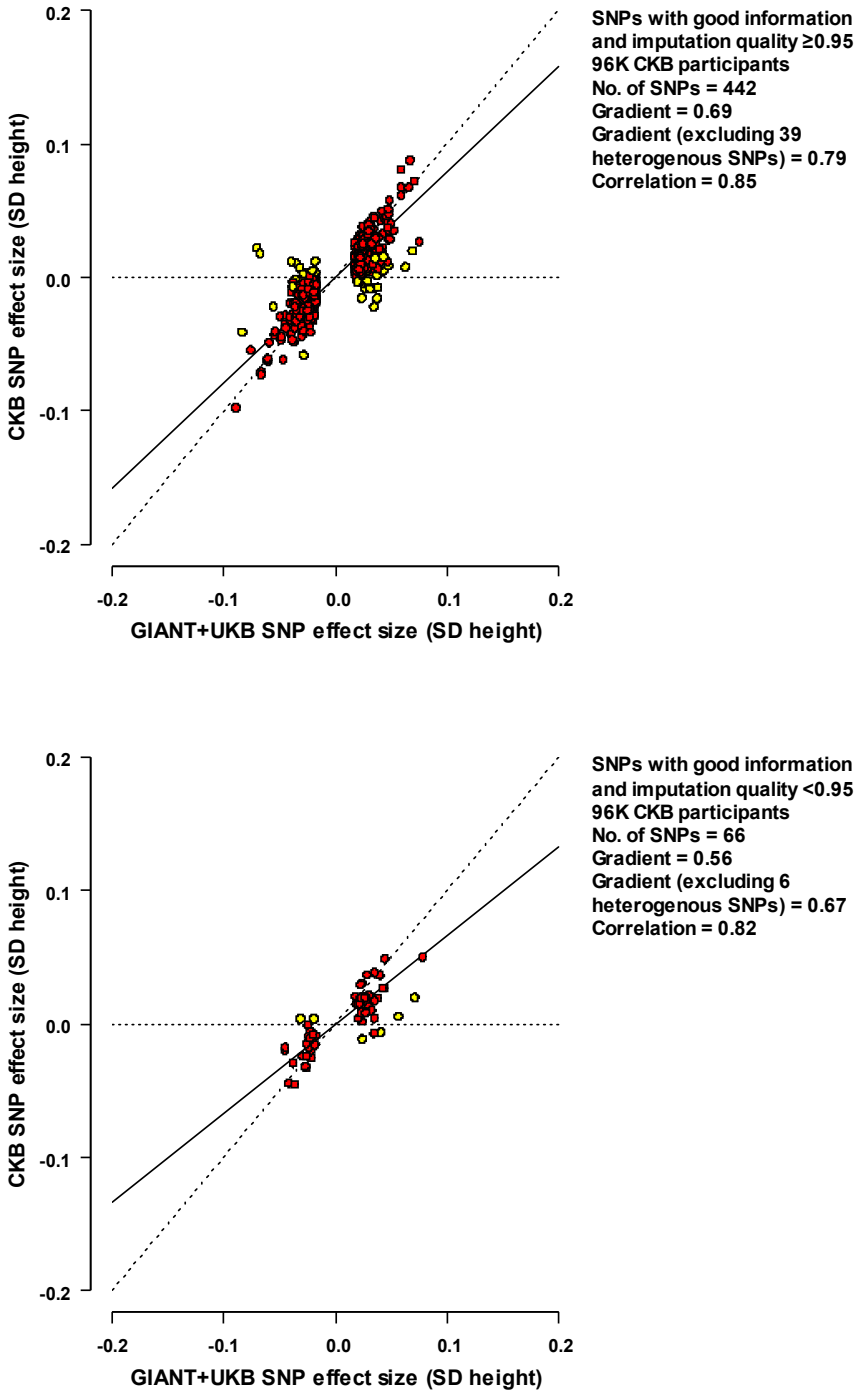


Figure 6.8 Univariate effect sizes of clumped GIANT+UKB[30] SNPs on standardised standing height (96046 CKB participants), shown for SNPs with good information and classified by imputation quality ≥0.95 and <0.95.

SNPs in yellow indicate heterogeneity. Plotted gradients (solid black lines) excluded heterogeneous SNPs. A SNP with good information was able to detect an association with height (given the minor allele frequency and sample size differences between CKB and the GWAS study, GIANT+UKB [Yengo et al., 2018][30]).

7. Associations of genetically instrumented height with cardiovascular diseases

7.1 Introduction

Observational analyses conducted in the present thesis (Chapter 5) have demonstrated inverse associations of taller standing height with major coronary events in UKB and CKB (combined HR per 1 SD taller standing height: 0.93; 95% CI: 0.92, 0.95), consistent with the results of associations reported by the ERFC meta-analysis (HR for coronary heart disease per 1 SD [6.5 cm] taller standing height: 0.93; 95% CI: 0.91, 0.94).[1] The causal relevance of these associations is supported by Mendelian randomisation studies (see systematic review in Chapter 2). A Mendelian randomisation analysis of height and major coronary events conducted by Nelson et al. (2015), involving 65 066 cases in C+C4D and 180 height-associated SNPs, demonstrated that one unit (6.5 cm) taller genetically instrumented height was associated with a 12% lower risk of coronary artery disease (OR: 0.88; 95% CI: 0.82, 0.95).[4] Thus, the strength of association estimated from genetically instrumented height was almost twice as strong as those reported in traditional observational analyses both in UKB and CKB, and in the ERFC meta-analysis.[1,4] To identify potential mechanisms, Nelson et al. investigated the extent to which the associations of genetically elevated height were mediated by established cardiovascular risk factors using summary data on the associations of each SNP in a genome-wide panel with individual traits, such as blood pressure, cholesterol, and BMI.[4] Nelson et al. concluded that about 30% of the inverse associations of genetically elevated height with coronary artery

disease were explained by LDL cholesterol and triglycerides (quantified using previously reported predictions of lipid-related markers and cardiovascular disease[33]), but that height was unrelated with blood pressure.[4] Nüesch et al. (2016) investigated the associations of genetically elevated height with cardiovascular risk factors and also reported inverse associations with LDL cholesterol and triglycerides and a null association with blood pressure.[4] In a recent large GWAS consortium meta-analysis of blood pressure, the International Consortium of Blood Pressure reported that some blood pressure SNPs were strongly positively associated and some strongly negatively associated with height.[34]

Observational analyses conducted in the present thesis (Chapter 5) have demonstrated inverse associations of taller standing height with ischaemic stroke and haemorrhagic stroke in UKB and CKB (combined HR per 1 SD taller standing height: 0.96; 95% CI: 0.95, 0.97 for ischaemic stroke, and HR: 0.89; 95% CI: 0.87, 0.90 for haemorrhagic stroke), consistent with the associations reported by the ERFC meta-analysis.[1] In a Mendelian randomisation study based on 4848 cases of stroke from the C+C4D Consortium and 69 height-associated SNPs, Nüesch et al. reported no statistically significant association of genetically instrumented height with stroke (relative risk per unit [6.5 cm] genetically elevated height: 0.97; 95% CI: 0.79, 1.19).[32] However, the associations of genetically elevated height with stroke types have not been previously reported and may differ between ischaemic and haemorrhagic stroke types (see systematic review in Chapter 2).

Observational analyses conducted in the present thesis (Chapter 5) have also demonstrated a positive association of taller standing height with atrial fibrillation in both UKB and CKB (combined OR per 1 SD taller standing height: 1.33; 95% CI: 1.31, 1.36), which was stronger than the observational association reported by Rosenberg et al. (2012) based on 1252 atrial fibrillation cases from the Cardiovascular Health Study (HR per 6.5 cm taller height [scaled to 6.5 cm and combined across men and women]: 1.18; 95% CI: 1.11, 1.25).[3] Lean body mass is one of the most important anthropometric risk factors for atrial fibrillation.[79] Analyses of these associations of height and atrial fibrillation in UKB and CKB in the present thesis demonstrated that the adjustment for lean body mass greatly attenuated the positive observational association of taller standing height with atrial fibrillation (Chapter 5). The causal relevance of height with atrial fibrillation has been confirmed in a Mendelian randomisation study based on 209 height-associated SNPs and 879 atrial fibrillation cases from the Cardiovascular Health Study, where Rosenberg et al. (2014) reported an OR per unit (scaled to 6.5 cm) genetically elevated height of 1.19 (95% CI: 1.02, 1.38).[10] However, this role of lean body mass, as a potential mediator of the association of height with atrial fibrillation, has not previously been examined in Mendelian randomisation studies.

Atrial fibrillation is an important risk factor for cardioembolic ischaemic stroke.[80] The opposing associations of height with coronary artery disease and atrial fibrillation suggests the possibility that the association of height may vary by ischaemic stroke subtype, which previous studies have not investigated.

The UKB and CKB genotyped datasets in combination have much larger numbers of cases of ischaemic stroke, haemorrhagic stroke, and atrial fibrillation (>16K prevalent and incident cases of ischaemic stroke, >8K cases of haemorrhagic stroke, and >13K cases of atrial fibrillation) than have previously been investigated in Mendelian randomisation studies of height and these diseases (see systematic review in Chapter 2). Since the publication of the most recent Mendelian randomisation studies of height and cardiovascular diseases,[4,32,51] the number of known genome-wide significant height-associated SNPs has increased substantially (from ~200 to ~3000 SNPs in March 2018), which has increased the power of the genetic instruments for height. In UKB, a genetic risk score for height based on 692 Wood[20] SNPs and effect sizes explained 16.9% of the variance in height and, in CKB, a genetic risk score based on 2963 clumped GIANT+UKB[30] SNPs and effect sizes explained 10.5% of the variance in height (Chapter 6). Data from METASTROKE, a large genetic consortium of 12K ischaemic stroke cases, will also be included in the present study. Moreover, UKB and CKB also have individual participant data on a wide range of anthropometric and cardiovascular risk factors and, therefore, permit reliable analyses of the extent to which the associations of height with different types of cardiovascular diseases are mediated by established cardiovascular risk factors.

The aims of the present chapter are to: (i) examine the association of genetically instrumented height with coronary artery diseases, ischaemic stroke, haemorrhagic stroke, and atrial fibrillation; (ii) examine the associations of genetically instrumented height with anthropometric and cardiovascular risk

factors; and (iii) assess the extent to which quantitative associations of genetically instrumented height with cardiovascular diseases could be mediated by established cardiovascular risk factors.

7.2 Methods

Study populations for UKB and CKB

336 433 genotyped UKB participants (after QC and relatedness exclusions) and 96 046 genotyped CKB participants (after QC exclusions) (77 419 participants from a random population sample and 18 627 additional cases, which were included only in disease outcome analyses) were available for the present Mendelian randomisation analyses. In addition, genomic principal components (first 10 principal components in UKB and first 8 in CKB) and genotyping array type were used in the Mendelian randomisation analyses, as detailed in Chapter 6.2.

The height phenotype was normalised within strata of sex (and region in CKB), as described in Chapter 3.3. Anthropometric traits were also normalised within strata of sex (and region in CKB). In addition, when a trait was included as a covariate in a regression model it was also normalised within strata of sex (and region in CKB). Further investigations and sensitivity analyses (MR-Egger, MR-PRESSO, funnel plots, and plots of effect sizes) were carried out on selected traits showing strong evidence of an association with genetically elevated height. Each selected trait was normalised within strata of sex (and region in CKB), and the investigations were carried out using the summary associations of the trait with each SNP. For genetic analyses, systolic blood pressure was increased by

15 mm Hg and diastolic blood pressure increased by 10 mm Hg for participants taking blood pressure lowering medications (which was consistent with the methodology used by the International Consortium for Blood Pressure[54,90]).

Study design for CARDIoGRAMplusC4D and METASTROKE

C+C4D collected data from 48 studies, totalling 60 801 coronary artery disease cases and 123 504 controls.[86] Summary GWAS data (i.e. the associations of 9.5M SNPs with coronary artery disease) were made available online. The C+C4D summary data contained 687 (out of 697) Wood[20] SNPs, and 3204 (out of 3211) clumped GIANT+UKB[30] SNPs.

METASTROKE collected data from 15 ischaemic stroke cohorts, yielding a total of 12 389 ischaemic stroke cases and 62 004 controls.[101] Some of the ischaemic stroke cases were classified into ischaemic stroke subtypes (2365 cardioembolic strokes, 2167 large-vessel strokes, and 1894 small-vessel strokes). Summary GWAS data (i.e. the associations of 2.4M SNPs with ischaemic stroke and its subtypes) were made available online. The METASTROKE summary data contained 692 Wood[20] SNPs, and 3150 clumped GIANT+UKB[30] SNPs.

Statistical methods for UKB and CKB

The genetic risk scores used for Mendelian randomisation analyses were constructed at an individual participant level by weighting the number of height-associated effect alleles (for each of the 692 Wood[20] SNPs in UKB, and 2963 GIANT+UKB[30] SNPs [from clumping] in CKB) by their corresponding literature effect sizes (expressed in SD units of height as defined by Wood et al.[20] and

Yengo et al.[30], for UKB and CKB, respectively) (Chapter 6.2). SNPs showing heterogeneity (between the estimated univariate effect sizes and the published effect sizes) were excluded from the genetic risk score in CKB, as discussed in Chapter 6.2.

Logistic regression was used to quantify the association of each genetic risk score with each disease outcome, adjusted for genomic principal components (first 10 principal components in UKB and first 8 in CKB) and genotyping array type. Disease outcomes included coronary artery disease (myocardial infarction, fatal ischaemic heart disease, and coronary revascularisation procedures [percutaneous transluminal coronary angioplasty or coronary artery bypass grafting]), stroke types (ischaemic and haemorrhagic stroke), and atrial fibrillation. These associations were estimated separately for the combination of prevalent cases (at baseline) and incident cases. For some outcomes (such as ischaemic and haemorrhagic stroke in CKB), prevalent cases at baseline data were not available, so analyses were restricted to incident cases. As a sensitivity analysis for CKB, the associations of genetically elevated height with the disease outcomes were re-estimated excluding related participants (as described in Chapter 6.2).

As part of the instrumental variable process (the ratio method),[100,106] measured height was regressed on the genetic risk score for height in both studies (Table 6.10). One unit of the respective genetic risk scores was associated with 0.82 SD of measured height in UKB and 0.59 SD in CKB (Table 6.10). When estimating the causal effects from the instrumental variable analyses, the regression

coefficients of the genetic risk scores were scaled by the inverse of these values, so that 1 unit of the genetic risk score represented 1 SD of measured height (Table 3.1). The second order variance term that is formally in an instrumental variable estimate was ignored because the contribution from this term would be negligible given the strength (large F-statistics) of the instruments (Table 6.10).[100,107,108]

The shape of the associations of the height genetic risk score with cardiovascular diseases were investigated by estimating the ORs (and 95% CIs) for each quartile of the genetic risk score. Floating absolute risk estimates were calculated relative to the lowest height quartile. The associations with cardiovascular diseases were presented as ORs (and 95% CIs) per unit genetically elevated height. The ORs from the UKB and CKB studies were combined using fixed-effects meta-analysis.

Since prior literature and observational results from Chapter 5 suggested opposite effects of height on coronary artery disease and atrial fibrillation (and an association of lean body mass with atrial fibrillation), the associations of genetically instrumented height with cardiovascular diseases were re-estimated after excluding participants with atrial fibrillation (prevalent and incident cases). In addition, the associations of genetically instrumented height with atrial fibrillation were estimated with different adjustments for adiposity and blood pressure.

Linear regression was used to quantify the association of each genetic risk score with each continuous trait (anthropometric measures and cardiovascular risk

factors), adjusted for genomic principal components and genotyping array type. For variables that were not standardised, within each study, the ORs were estimated separately for men and women and combined using fixed-effects meta-analysis. (For standardised variables, ORs were not estimated separately for men and women). The associations were presented as coefficients (and 95% CIs) per unit genetically elevated height.

Further investigations were carried out on selected traits showing strong evidence of an association with genetically elevated height (which were therefore potential mediators of the effect of height on disease.) This was done in UKB because it was more informative than CKB. Univariate effect sizes of the height-associated SNPs on the selected traits (using standardised variables) were calculated for all the SNPs as well as the subset of SNPs with the greatest (top decile) statistical information, estimated by:

$$2\beta^2 MAF(1 - MAF)$$

where β is the SNP effect size on the trait, and MAF is the minor allele frequency of the SNP.

Heterogeneity between the estimated univariate effect size of the height-associated SNPs on the trait of interest and the published effect size on height scaled by the common effect, was assessed by performing a T-test for each SNP at a Bonferroni corrected significance level. (The common effect was defined as the ratio of the Z-values for the association of the genetic risk score with the trait of interest and with height.)

Statistical methods for CARDIoGRAMplusC4D and METASTROKE

Following the approach used by Nelson et al. for summary data,[4] for each SNP the log odds ratio (β_3) of height and the disease outcome was derived as follows:

$$\beta_3 = \frac{\beta_2}{\beta_1} \qquad SE(\beta_3) = \sqrt{\frac{SE(\beta_2)^2}{\beta_1{}^2}}$$

where β_1 is the log odds ratio of the SNP and the disease outcome (from the consortia data), and β_2 is the literature effect size on height of the SNP in SD units (from Wood et al.[20] and GIANT+UKB[30]). β_3 of each SNP was combined using an inverse-variance weighted fixed-effects meta-analysis and the results presented as ORs (and 95% CIs) for coronary artery disease (C+C4D), ischaemic stroke (METASTROKE), and ischaemic stroke subtypes (METASTROKE) per 1 SD taller genetically instrumented height. These associations were meta-analysed with the UKB and CKB associations for the corresponding diseases.

Mendelian randomisation sensitivity analyses

As a check of the instrumental variable assumptions (Chapter 6.1), the height phenotype was included as a covariate in the logistic regression models together with the genetic instrument.[91] MR-Egger and MR-PRESSO methods were applied to traits that were most strongly associated with the genetic instrument for height.[95,109] Since the MR-Egger and MR-PRESSO methods make use of per SNP summary data, to facilitate comparisons, the associations of the selected traits with height were re-estimated using inverse-variance weighted meta-analysis.

The MR-Egger method provides a test of directional pleiotropy (the MR-Egger intercept test) and an estimate of the causal effect.[95] If the MR-Egger intercept is

significantly different from zero then at least one genetic variant is invalid and the standard inverse-variance weighted (meta-analysed) estimate of the causal effect is biased. Under the assumption that the association of each genetic variant with the exposure is independent of the pleiotropic effect of the variant (the instrument strength independent of direct effect [InSIDE] assumption), the MR-Egger method provides a valid test of association and a consistent causal effect estimate even if all the genetic variants are invalid instrumental variables.[95] The MR-Egger method uses a random-effects model as pleiotropy leads to heterogeneity between the causal estimates of the genetic variants. (In this scenario a fixed-effects model is not valid.)

The Mendelian randomisation pleiotropy residual sum and outlier (MR-PRESSO) test is used to identify horizontal pleiotropic outliers in summary-level Mendelian randomisation testing.[109] The MR-PRESSO method provides a causal estimate before and after correction for horizontal pleiotropy via outlier removal.

Both the MR-Egger and MR-PRESSO methods were implemented in R version 3.3.3, using the MendelianRandomization and MRPRESSO packages, respectively. In addition, plots of effect sizes and funnel plots were created for traits that were most strongly associated with the genetic instrument for height. For the funnel plots, instrumental variable strength (absolute value of the effect size on height divided by the standard error of the effect size on the trait) (also known as instrument precision) was plotted against instrumental variable estimates (the effect size on the trait divided by the effect size on height).[110]

7.3 Results

Associations of genetically elevated height with cardiovascular diseases in consortia data

Genetically elevated height, based on 3208 SNPs, was inversely associated with risk of coronary artery disease in the C+C4D dataset (60 801 cases; OR per unit genetically elevated height: 0.86; 95% CI: 0.84, 0.88) (Table 7.1). This estimate was similar to the association reported by Nelson et al.,[4] based on 180 height-associated SNPs in C+C4D (OR per unit genetically elevated height: 0.88; 95% CI: 0.82, 0.95).

In contrast, genetically elevated height, based on 3150 SNPs, was not statistically significantly associated with ischaemic stroke in METASTROKE (12 389 cases; OR per unit genetically elevated height: 0.98; 95% CI: 0.94, 1.02) (Table 7.1). The estimate was similar when based on 692 SNPs (OR per unit genetically elevated height: 1.00; 95% CI: 0.95, 1.05).

Associations of genetically elevated height with cardiovascular diseases in UKB and CKB

In a combined analysis of the UKB and CKB genotyped datasets, there were 18 159 coronary artery disease events (much less than the 60 801 events in C+C4D), 14 537 ischaemic strokes (slightly more than the 12 389 events in METASTROKE), and 7819 haemorrhagic strokes (Figures 7.1 and 7.2). These were prevalent cases (at baseline) or incident cases.

The genetic risk score for height was approximately log-linearly related to risk of coronary artery disease, stroke types and atrial fibrillation in UKB and CKB (Supplementary Figure 7.1). The causal effects from instrumental variable analyses were presented per unit genetically elevated height, where the unit was 1 SD of measured height (Table 3.1). The association of genetically elevated height with coronary artery disease (prevalent and incident) in UKB was similar to the association in C+C4D (Figure 7.1). In CKB the association with coronary artery disease was null but was based on a smaller number of cases. When combined (78 960 cases), genetically elevated height was inversely associated with coronary artery disease (OR per unit genetically elevated height: 0.87; 95% CI: 0.84, 0.89). The association of genetically elevated height with coronary disease was stronger and less precise than the observational association (HR per 1 SD taller measured height: 0.93; 95% CI: 0.92, 0.95; combined across UKB and CKB after basic adjustments and adjustment for cardiovascular disease risk factors and socio-economic risk factors) (Figure 7.2).

The association of genetically elevated height with ischaemic stroke was similar in UKB, CKB and METASTROKE. When combined (26 926 cases), genetically elevated height was not statistically significantly associated with ischaemic stroke (OR per unit genetically elevated height: 0.97; 95% CI: 0.94, 1.00). The association of genetically elevated height with ischaemic stroke was slightly weaker and less precise than the observational association (HR per 1 SD taller measured height: 0.96; 95% CI: 0.95, 0.97; combined across UKB and CKB after basic adjustments and adjustment for cardiovascular disease risk factors and socio-economic risk factors) (Figure 7.2).

When UKB (1447 cases) and CKB (6372 cases) were combined, genetically elevated height was inversely associated with haemorrhagic stroke (OR per unit genetically elevated height: 0.91; 95% CI: 0.85, 0.98) (Figure 7.1). The association of genetically elevated height with haemorrhagic stroke was slightly weaker and much less precise than the observational association (HR per 1 SD taller measured height: 0.89; 95% CI: 0.87, 0.90; combined across UKB and CKB after basic adjustments and adjustment for cardiovascular disease risk factors and socio-economic risk factors) (Figure 7.2).

Associations of genetically elevated height with atrial fibrillation in UKB and CKB

There were 13 225 and 140 prevalent (at baseline) or incident cases of atrial fibrillation in UKB and CKB, respectively (Figure 7.3). Genetically elevated height was positively associated with atrial fibrillation in UKB and CKB, and the association was similar in both populations (combined OR per unit genetically elevated height: 1.32; 95% CI: 1.27, 1.38) (Figure 7.3). The association of genetically elevated height with atrial fibrillation was similar to the observational association (OR per 1 SD taller measured height: 1.33; 95% CI: 1.31, 1.36; combined across UKB and CKB after basic adjustments and additional adjustment for major cardiovascular disease risk factors and socio-economic risk factors) (Table 5.1).

Associations of genetically elevated height with ischaemic stroke subtypes in METASTROKE and with related outcomes in UKB

In contrast to the null association of genetically elevated height with ischaemic stroke overall in METASTROKE (12 389 cases), genetically elevated height, based on 3150 SNPs, was positively associated with cardioembolic stroke (2365 cases; OR per unit genetically elevated height: 1.13; 95% CI: 1.05, 1.22), but inversely associated with large-vessel stroke (2167 cases; OR: 0.91; 95% CI: 0.84, 0.99) and small-vessel stroke (1894 cases; OR: 0.89; 95% CI: 0.81, 0.97) (Figure 7.4). In UKB, when the 13 225 participants with atrial fibrillation were excluded, all associations of genetically elevated height with ischaemic stroke, coronary artery disease and haemorrhagic stroke became more extreme (Figure 7.4).

Associations of genetically elevated height with continuous traits in UKB and CKB

The effects of genetically elevated height with anthropometric measures and cardiovascular risk factors were directionally concordant across UKB and CKB (Table 7.2). Both study size and genetic instrument strength were greater in UKB than CKB: the Z-value for measured height was 262 in UKB and 97 in CKB, and the magnitude of the effects of genetically elevated height tended to be greater in UKB. Z-values within each study provide an indication of the relative strength of the associations with different traits.

In both studies, genetically elevated height was positively associated with most directly measured anthropometric traits (weight, waist circumference, lean body mass, fat body mass, hand grip strength, FEV_1, and FVC), but was inversely

associated with waist-to-hip ratio and BMI (Table 7.2). Of these traits, lean body mass had the strongest association (Z-value of 133 in UKB, and Z-value of 62 in CKB). The observational associations (adjusted for age, sex, year of birth, and region in CKB) of taller measured height with lean body mass and fat body mass were similar in UKB and CKB; whereas the corresponding associations with waist-to-hip ratio and BMI were in the same direction as the genetic associations in UKB but were in the opposite direction in CKB.

Genetically elevated height was strongly inversely associated with systolic and diastolic blood pressure in both UKB and CKB (-1.39 [SE: 0.17] and -0.27 [SE: 0.24] mm Hg systolic blood pressure per unit genetically elevated height, in the two populations respectively) (Table 7.2) The observational associations of taller measured height with systolic blood pressure were in the same direction as the genetic associations in UKB (-1.46 [SE: 0.03] mm Hg systolic blood pressure per 1 SD taller measured height) but were in the opposite direction in CKB (0.41 [SE: 0.03] mm Hg).

The inverse association of genetically elevated height with systolic and diastolic blood pressure in UKB attenuated but remained statistically significant when the blood pressure variables were pre-adjusted for BMI, age, and age squared, as often done in GWAS of blood pressure[54,90] (-1.11 [SE: 0.16] mm Hg systolic blood pressure per unit genetically elevated height) (Supplementary Table 7.1). In CKB, the associations were no longer statistically significant (-0.08 [SE: 0.22] mm Hg systolic blood pressure). Additional adjustment for socio-economic risk factors (income and education) slightly further attenuated the association in UKB (-1.07

[SE: 0.16] mm Hg systolic blood pressure per unit genetically elevated height), and, in CKB, the effect of adjustment could not be reliably inferred (Supplementary Table 7.1).

Genetically elevated height was inversely associated with heart rate in both studies (-0.25 [SE: 0.09] bpm per unit genetically elevated height in UKB; and -0.46 [SE: 0.13] bpm in CKB) (Table 7.2). The observational association of taller measured height with heart rate was approximately twice as large as the genetic association in UKB (-0.58 [SE: 0.02] bpm per 1 SD taller measured height), but similar to the genetic association in CKB (-0.38 [SE: 0.02] bpm).

Genetically elevated height was not statistically significantly associated with cIMT or carotid plaque score (available in CKB only); however these traits were only available in a subset of the genotyped dataset (1836 UKB participants and 21 299 CKB participants) (Table 7.2). Measured height was positively associated with cIMT in both studies, and not statistically significantly associated with or carotid plaque score.

Blood lipids were available in a subset of CKB (6062 participants) and not available in UKB. Genetically elevated height was inversely associated with LDL cholesterol (-0.09 [SE: 0.025] mmol/L per unit genetically elevated height) and not statistically significantly associated with triglycerides in CKB (Table 7.2). In contrast, measured height was positively associated with LDL cholesterol (0.01 [SE: 0.005] mmol/L per 1 SD taller measured height) and triglycerides (0.04 [SE: 0.011] mmol/L).

Associations of genetically elevated height with atrial fibrillation in UKB and CKB, adjusted for adiposity and blood pressure

Most of the association of genetically elevated height with atrial fibrillation was accounted for by adjustment for lean body mass (the combined OR per unit genetically elevated height attenuated from 1.32 [95% CI: 1.27, 1.38] to 1.11 [95% CI: 1.06, 1.16]) (Figure 7.3) and was similar to the observational association adjusted for lean body mass (combined OR per 1 SD taller measured height: 1.16; 95% CI: 1.14, 1.19; after basic adjustments and additional adjustment for major cardiovascular disease risk factors and socio-economic risk factors) (Table 5.1). Further adiposity and blood pressure adjustments identified lean body mass and weight as having the largest attenuation effects on the association of genetically instrumented height with atrial fibrillation, while adjustment for blood pressure had little effect (Table 7.3).

Mendelian randomisation sensitivity analyses for cardiovascular outcomes

Restriction of the CKB analyses to unrelated participants had a small impact on the associations of genetically elevated height with cardiovascular diseases (Supplementary Table 7.2). For haemorrhagic stroke, exclusion of related participants strengthened the association per unit genetically elevated height from 0.94 (95% CI: 0.86, 1.02; n=6372) to 0.90 (95% CI: 0.81, 0.99; n=5035).

In UKB, the inclusion of the height phenotype as a covariate in the regression model attenuated the association of genetically elevated height with haemorrhagic stroke and atrial fibrillation, but made the association with coronary

artery disease and ischaemic stroke positive (Supplementary Table 7.3). In CKB, the associations of the genetic risk score with coronary artery disease, stroke types, and atrial fibrillation were not significant before or after the inclusion of the height phenotype as a covariate in the regression model (Supplementary Table 7.3).

Mendelian randomisation sensitivity analyses for lean body mass

Mendelian randomisation sensitivity analyses (heterogeneity investigations of effect sizes, MR-Egger, MR-PRESSO, and funnel plots) were performed for lean body mass and systolic blood pressure in UKB (because it had more data). The lean body mass effect sizes had a correlation of 0.86 with the height effect sizes of the SNPs included in the genetic instrument for height in UKB (Figure 7.5). Of these SNPs, 66 showed heterogeneity with the common effect. For lean body mass in UKB, the genetic risk score estimate per unit genetically elevated height in the main analyses was 0.551 (SE 0.008) SD units (Tables 7.1 and 7.4). The fixed-effects and random-effects meta-analyses of SNP effect sizes, produced consistent estimates per unit genetically elevated height (0.627 [SE 0.005] SD and 0.627 [SE 0.013] SD, respectively) (Table 7.4). The MR-Egger estimate for lean body mass was slightly higher (0.672 [SE 0.032] SD), and the MR-Egger intercept estimate was small (-0.001 [SE 0.001] SD) and not significant (P-value of 0.12) indicating no evidence to invalidate the original genetic risk score estimates. The funnel plot of instrumental variable strength against instrumental variable estimates for lean body mass did not show any obvious asymmetry (Figure 7.7).

An MR-PRESSO analysis identified 67 outliers, and the outlier adjusted lean body mass estimate was 0.594 (SE 0.010) SD units. SNPs identified by MR-PRESSO as outliers or showing heterogeneity with the common effect were listed in Supplementary Table 7.4.

Mendelian randomisation sensitivity analyses for systolic blood pressure

The systolic blood pressure effect sizes had a correlation of -0.27 with the height effect sizes of the SNPs included in the genetic instrument for height in UKB (Figure 7.6). Of these SNPs, 61 showed heterogeneity with the common effect. Five Wood[20] SNPs had previously been identified as being associated with blood pressure (in GWASs of blood pressure).[54,111] For systolic blood pressure in UKB, the genetic risk score estimate per unit genetically elevated height was -0.076 (SE 0.008) SD units (Table 7.4). The fixed-effects and random-effects meta-analyses of SNP effect sizes, produced consistent systolic blood pressure estimates per unit genetically elevated height (-0.085 [SE 0.005] SD and -0.085 [SE 0.010] SD, respectively) (Table 7.4). The MR-Egger estimate for systolic blood pressure was slightly weaker (-0.065 [SE 0.024] SD), and the MR-Egger intercept estimate was small (-0.001 [SE 0.001] SD) and not significant (P-value of 0.35) indicating no evidence to invalidate the original genetic risk score estimates. An MR-PRESSO analysis identified 56 outliers, and the outlier adjusted systolic blood pressure estimate was -0.072 (SE 0.008) SD units. SNPs identified by MR-PRESSO as outliers or showing heterogeneity with the common effect were listed in Supplementary Table 7.5. The funnel plot of instrumental variable strength against instrumental variable estimates for systolic blood pressure did not show any obvious asymmetry (Figure 7.7).

7.4 Discussion

There were distinct effect sizes of height on coronary artery disease (13% reduction in risk per unit genetically elevated height), ischaemic stroke (3% reduction in risk), and atrial fibrillation (32% increase in risk) (OR for coronary artery disease [78 960 cases] per unit genetically elevated height: 0.87; 95% CI: 0.84, 0.89; OR for ischaemic stroke [26 926 cases]: 0.97; 95% CI: 0.94, 1.00; and OR for atrial fibrillation [13 363 cases]: 1.32; 95% CI: 1.27, 1.38) (Figures 7.1 and 7.3). The distinct effect sizes indicate the involvement, or lack of involvement, of different mechanisms for the associations of height with these diseases. The association of height with haemorrhagic stroke was less definite (OR per unit genetically elevated height [7819 cases]: 0.91; 95% CI: 0.85, 0.98) (Figure 7.1).

Atrial fibrillation is a major risk factor for cardioembolic ischaemic stroke.[80] The opposing associations of height with coronary artery disease and atrial fibrillation suggested, a priori, that the association may vary by ischaemic stroke subtype. This was confirmed in METASTROKE where genetically elevated height was positively associated with cardioembolic stroke, but inversely associated with large-vessel stroke and small-vessel stroke (Figure 7.4). In addition, in UKB, all associations of genetically elevated height with ischaemic stroke, coronary artery disease and haemorrhagic stroke became more extreme when participants with atrial fibrillation were excluded (Figure 7.4). The rates of atrial fibrillation may be lower in East Asia than Western countries,[112] but it is unlikely to be as low as suggested by the number of cases in CKB relative to UKB. Therefore, it is likely that atrial fibrillation is not adequately recorded in CKB.

Mechanisms of the associations of height with cardiovascular diseases were explored in Table 7.2. Of the anthropometric traits, lean body mass was the most strongly associated with genetically instrumented height. The associations of genetically elevated height with lower waist-to-hip ratio and BMI were consistent with the traditional observational associations in UKB but opposite in CKB. The observational associations in CKB were misleading and did not reflect the causal effects from the Mendelian randomisation analyses, perhaps due to inadequate adjustment.

The association of obesity with atrial fibrillation is well-known,[113] but studies have suggested that obesity may not be the anthropometric mechanism underlying this association.[79] After adjustment for measured lean body mass, the OR for atrial fibrillation per unit genetically elevated height reduced to 1.11 (95% CI: 1.06, 1.16) (Figure 7.3). This provided suggestive support for a causal role of higher lean body mass associated with greater atrial fibrillation. Lean body mass was confirmed as the most important anthropometric risk factor for atrial fibrillation (more important than height or body mass index) (Table 7.3), which was consistent with a recent observational study by Fenger-Grøn et al.[79]

The association of genetically elevated height with lean body mass was robust and remained largely unchanged when sensitivity analyses (MR-Egger and MR-PRESSO methods) were applied (Table 7.4). The MR-Egger intercept was small and not statistically significant, which indicated that there was no directional pleiotropy (Table 7.4). In addition, the funnel plot did not show non-directional pleiotropy (Figure 7.7).

The genetic instrument for height may be acting as a proxy for the effects of lean body mass on atrial fibrillation (the correlation between height and lean body mass was 0.55 in UKB and 0.68 in CKB [Table 4.4]). However, height-associated SNPs are better suited for instrumental variable analysis as they explain a much greater percentage of variance than those SNPs currently identified as specifically associated with lean body mass (5 SNPs explaining <1% of its variance).[114]

Nelson et al.[4], based on ~95K European participants obtained from publicly available data, reported inverse associations of genetically elevated height with LDL cholesterol (-0.06 SD units per unit genetically elevated height; 95% CI: -0.09, -0.04) and triglycerides (-0.05 SD units; 95% CI: -0.08, -0.03) (Supplementary Table 2.1). Based on the same dataset, Bulik-Sullivan et al.[21] reported compatible associations, as did Nüesch et al.[32], based on ~42K European participants (Supplementary Table 2.1). The present study had much less power than these studies to detect the association with blood lipids as data were only available in ~6K CKB participants, and not available in UKB. However, the estimates in CKB were consistent with the same magnitude of effect as previously reported: there was an inverse association of genetically elevated height with LDL cholesterol (-0.09 [SE: 0.025] mmol/L per unit genetically elevated height) and an inverse, but not statistically significant, association with triglycerides (-0.05 [SE: 0.055] mmol/L) (Table 7.2). The observational associations of taller measured height with LDL cholesterol and triglycerides were positive in CKB. This highlights the value of genetic instrumental analysis to

understand the underlying causal effects, whereas the observational effects can be misleading.

Nelson et al.[4], based on the assumption that the associations in the ERFC reports on blood lipids and cardiovascular diseases[33,115] are causal, estimated that the associations of 1 SD genetically determined height with LDL cholesterol and triglycerides would be expected to be associated with a 2.2% and 1.5% reduction in coronary artery disease risk, respectively.[4] Nelson et al. estimated that ~30% of the association of genetically elevated height with coronary artery disease could be accounted for by LDL cholesterol and triglycerides.[4] Using the same methodology reported by Nelson et al.[4], the associations of 1 unit genetically elevated height with LDL cholesterol and triglycerides in the present study of CKB would be expected to be associated with a 3.0% and 0.8% reduction in coronary artery disease risk, respectively.

Nelson et al.[4], based on ~70K European participants from the International Consortium for Blood Pressure, reported a positive but not statistically significant association of genetically elevated height with systolic blood pressure adjusted for BMI (0.34 mm Hg per unit genetically elevated height; 95% CI: -0.31, 1.00) (Supplementary Table 2.1). Nüesch et al.[32], based on ~52K European participants, reported an inverse but not statistically significant association of genetically elevated height with systolic blood pressure (-0.93 mm Hg per unit genetically elevated height; 95% CI: -1.86, 0.00, converted from SD units assuming 1 SD systolic blood pressure was 18.6 mm Hg as was the case in UKB, a similar European population [Table 1.1]). There was a strong inverse

association of genetically elevated height with systolic blood pressure in UKB (-1.39 [SE: 0.17] mm Hg systolic blood pressure per unit genetically elevated height) and a weaker, but also inverse, association in CKB (-0.27 [SE: 0.24] mm Hg) (Table 7.2). The observational association of taller measured height with systolic blood pressure was consistent with the genetic effects in UKB but opposite in CKB. Attenuated inverse associations of genetically elevated height with systolic and diastolic blood pressure in UKB persisted when the blood pressure variables were first adjusted for BMI, age, and age squared, and additionally adjusted for socio-economic status (Supplementary Table 7.1). The majority of evidence appears to support an inverse association, but estimates vary (from -0.3 to -1.4 mm Hg per unit genetically elevated height). The association of genetically elevated height with systolic blood pressure was robust and remained largely unchanged when sensitivity analyses (MR-Egger and MR-PRESSO methods) were applied (Table 7.4). In a recent large GWAS consortium meta-analysis of blood pressure,[34] the International Consortium of Blood Pressure reported that some blood pressure SNPs were strongly positively associated and some strongly negatively associated with height, which might explain some differences by population.

In large-scale observational analyses by the Prospective Studies Collaboration (of ~1M mostly European participants), each 20 mm Hg higher usual systolic blood pressure was associated with an approximate doubling of the risk of fatal stroke, fatal ischaemic heart disease, or other vascular mortality in middle age (ages 40-69 years).[116] The observational impact of blood pressure in East Asians may differ due to different prevalence of diseases, levels of treatment, and

confounding factors. Mendelian randomisation studies of blood pressure based on East Asians may avoid these issues. In such a study, involving ~80K participants in CKB, it was reported that each 20 mm Hg higher genetically instrumented systolic blood pressure was associated with relative risks of 1.85, 2.48, and 2.85, for major coronary event, ischaemic stroke, and haemorrhagic stroke, respectively.[117] Assuming the effects in European and East Asian populations reflect causal associations, then this implies that a reduction of usual systolic blood pressure by 0.3 to 1.4 mm Hg would be expected to be associated with a 1.0-4.7% risk reduction for cardiovascular diseases in UKB, and a 0.9-4.2%, 1.4-6.2%, and 1.6-7.1% risk reduction in CKB for major coronary event, ischaemic stroke and haemorrhagic stroke, respectively. This equated to systolic blood pressure potentially accounting for 10-40% of the association of genetically elevated height with coronary artery disease in UKB (there was a null association of genetically elevated height with coronary artery disease in CKB), and between 25-100% of the association of genetically elevated height with ischaemic stroke and haemorrhagic stroke in CKB. The null association of genetically elevated height with coronary artery disease in CKB was possibly due to lower power (driven by a weaker genetic instrument in CKB) or poorer phenotyping (of myocardial infarction relative to stroke in CKB [Chapter 3.4]).

There was a strong positive association of genetically elevated height with lung function (FVC) in UKB (0.41 [SE: 0.008] SD units per unit genetically elevated height) and a weaker, but also positive, association in CKB (0.35 [SE: 0.011] SD units) (Table 7.2). Observational meta-analyses for the association of lung function with cardiovascular diseases have reported that higher lung function is

associated with reduced risk of cardiovascular diseases (relative risk of cardiovascular mortality for participants in the highest lung function group relative to participants in the lowest group: 0.50; 95% CI: 0.44, 0.58; 80K participants).[118] These estimates were presented for lung function categories which are not easily converted to per SD estimates.[118] However, a Mendelian randomisation study based on data from the C+C4D consortium, reported an inverse but not statically significant association of lung function with coronary artery disease (OR per 1 SD greater FVC: 0.82; 95% CI: 0.64, 1.06).[119] If this this estimate reflects a causal association then this would imply that higher lung function (FVC) of 0.41 and 0.35 SD units (the associations per unit genetically elevated height in UKB and CKB, respectively) would be expected to be associated with a 7-8% risk reduction for coronary artery disease in both UKB and CKB. This equated to lung function potentially accounting for 55-65% of the association of height with coronary artery disease in UKB. Hence, this is a potentially relevant component of mediation that requires further evidence.

There was an inverse association of genetically elevated height with heart rate in UKB (-0.25 [SE: 0.09] bpm per unit genetically elevated height) and CKB (0.46 [SE: 0.13] bpm) (Table 7.2). Heart rate is highly related to blood pressure[120] and it is likely that this association of genetically elevated height with heart rate is in conjunction with systolic blood pressure.

Tyrrell et al., in a Mendelian randomisation study of ~120K UKB participants, reported that genetically elevated height was associated with greater income.[17] This finding was in European populations, but was replicated in the observational

associations in CKB (Table 4.2). Income could be a mediating factor of height and some diseases. For example, higher income may lead to unhealthier lifestyles in developing countries (where people with higher income may be more likely to smoke, drink and eat more) and healthier lifestyles in developed countries (where people with higher income may be more health conscious). If this is the case, then income would lead to opposite contributions to the effects of height on cardiovascular diseases in UKB and CKB, which could account for some weaker genetic associations in CKB.

A potential source of bias in Mendelian randomisation studies is assortative mating, which occurs when individuals do not select their partners at random but rather based on particular traits.[121] For example, people might select on several desirable traits such as, taller height, healthier, higher income, healthier lifestyle so that these factors would then become linked genetically in offspring, as well as being able to contribute to nurture through genes not inherited by the offspring. Height is a trait subject to assortative mating,[122,123] which could possibly explain a general tendency to better risk factors in taller individuals. However, it seems less likely that assortative mating could explain the associations of height with atrial fibrillation and cancer – and very unlikely that assortative mating could explain all the diverse effects seen. Kong et al. (2018)[124], using family data, reported that indirect genetic effects of non-transmitted alleles (referred to as *genetic nurture*) could explain ~30% of the genetic effect on educational attainment, and ~6% of the genetic effect on height. Most of the indirect genetic effects on height were reported to be due to confounding by assortative mating.[124]

In both studies, the associations of genetically instrumented height with blood lipids and blood pressure were in directions consistent with the inverse associations of height with coronary artery disease and ischaemic stroke (Figure 7.1). Lung function could potentially also mediate the associations of height with coronary artery disease and ischaemic stroke, but the evidence for this is less substantial. However, associations of genetically instrumented height with blood lipids, blood pressure, and lung function were not compatible as an explanation of the positive associations of height with atrial fibrillation, which appears to be related to the strong association of height with lean body mass. A diagram of the associations of genetically elevated height with possible mediating factors and with the cardiovascular diseases considered in this thesis is presented in Figure 7.8. In the diagram, the height-associated genetic variants affect shared biological pathways which in turn affect the risk of cardiovascular diseases, either through height and direct mediating factors, or through indirect mediating factors (represented as the solid black arrows in Figure 7.8). It is possible that assortative mating may be affecting this process (represented as the solid black arrows in Figure 7.8). It is also possible that the height-associated variants affect the risk of cardiovascular diseases independently of the mediating factors or through different mechanisms (represented as the dashed blue arrows in Figure 7.8), or through lifestyle or behavioural factors (represented as the dashed black arrow).

Hindy et al.[125], in a Mendelian randomisation study of 16 851 cases of ischaemic stroke, reported that LDL cholesterol impacts large-vessel stroke (2410 cases), more than small-vessel stroke (3186 cases) or cardioembolic stroke (3427 cases). Therefore, the association of genetically instrumented height with blood

lipids is more likely to be a mediating factor of the inverse association of height with large-vessel stroke than other ischaemic stroke subtypes. The association of genetically instrumented height with atrial fibrillation, a major risk factor for cardioembolic stroke,[80] is likely to be a mediating factor for the positive association of genetically elevated height with cardioembolic stroke (represented as the solid orange arrow in Figure 7.8).

The association of genetically elevated height with haemorrhagic stroke was a combination of an inverse association in UKB (1447 cases) and no statistically significant association in CKB (6372 cases). Even though CKB had 4-fold more cases than UKB, the lower power of the genetic instrument in CKB (which explained 10.5% of the variance of height compared to 16.9% explained by the UKB instrument) may have resulted in the null association (if the true genetically instrumented height effect on haemorrhagic stroke was 0.95 then the CKB instrument only had ~24% power to detect this [Table 6.8]). Blood pressure is a major risk factor for haemorrhagic stroke,[126] but LDL cholesterol levels have been inversely associated with risk of haemorrhagic stroke.[126–128] Therefore, the mediating factors contributions from blood pressure and LDL cholesterol had opposing effects on the association of height with haemorrhagic stroke. More than 90% of incident haemorrhagic stroke cases in CKB have had brain imaging, and stroke phenotyping in CKB had a positive predictive value of >90%. Even though haemorrhagic stroke has a high fatality rate,[126] most haemorrhagic stroke cases in UKB were prevalent cases at baseline and hence non-fatal, which suggests that the haemorrhagic stroke captured from hospital episode statistics and self-reports in UKB may be rather different from those in CKB.

For coronary disease, the traditional observational results did not provide an accurate picture of the causal effects from Mendelian randomisation (the observational associations were weaker) (Figure 7.2). This may be due to inadequate adjustment of the observational associations, which were adjusted for basic adjustments (age, sex, year of birth, and region in CKB), major cardiovascular disease risk factors and socio-economic risk factors, including blood pressure but not LDL cholesterol. This may also be due to biases in the Mendelian randomisation, such as due to assortative mating. For ischaemic and haemorrhagic stroke, the traditional observational results were similar to the causal effects from Mendelian randomisation (however the observational estimates were more accurate) (Figure 7.2).

The validity of the Mendelian randomisation analyses depend on the instrumental variable assumptions being satisfied for all genetic variants (Chapter 6.1).[91] In the present thesis, the selection of genetic variants for height has been motivated by statistical rather than biological criteria. The large number of genetic variants makes it difficult to justify the instrumental variable assumptions from a biological perspective, especially since height is a highly polygenic trait and is expected to have shared pathways with other traits. Such studies have been referred to as *joint association studies*[129] to distinguish them from Mendelian randomisation studies based on a small number of variants (often selected using biological criteria). However, non-null findings in joint association studies still provide suggestive evidence of a causal effect or of shared causal mechanisms.[129]

7.5 Tables

Table 7.1 Associations of height-associated SNPs with ischaemic stroke and coronary artery disease using data from the CARDIoGRAMplusC4D (C+C4D) and METASTROKE consortia.

Consortium (source of odds ratio)	Source of height-associated SNPs	No. of SNPs	No. of events	Odds ratio (95% CI) per 1 SD taller genetically instrumented height*
Coronary artery disease				
C+C4D (reported in Nelson[4])	Lango[62]	180	60801	0.88 (0.82, 0.95)
C+C4D (calculated)	Wood[20]	687	60801	0.87 (0.85, 0.90)
C+C4D (calculated)	GIANT+UKB[30]	3208	60801	0.86 (0.84, 0.88)†
Ischaemic stroke				
METASTROKE (calculated)	Wood[20]	692	12389	1.00 (0.95, 1.05)
METASTROKE (calculated)	GIANT+UKB[30]	3150	12389	0.98 (0.94, 1.02)†

*The odds ratios were calculated as inverse-variance weighted fixed-effects meta-analyses of associations of height-associated SNPs with ischaemic stroke and coronary artery disease. The odds ratio reported in Nelson was calculated as a random effects meta-analysis.
†Used in the meta-analysis of associations in UKB and CKB (Figures 7.1 and 7.3).

Table 7.2 Associations of genetically instrumented height with continuous traits in UKB and CKB.

	UKB (n=336433)				CKB (n=77419)			
Continuous trait*	**SD (men/women)**	**Beta (SE) per 1 SD taller genetically instrumented height†**	**Z-value for genetic associ-ation**	**Beta (SE) per 1 SD taller measured height¶**	**SD (men/women)**	**Beta (SE) per 1 SD taller genetically instrumented height†**	**Z-value for genetic associ-ation**	**Beta (SE) per 1 SD taller measured height¶**
Height (SD units)	6.8/6.3 cm	1.00 (0.008)	261.9	1.00 (0.000)	6.5/6.0 cm	0.99 (0.010)	96.5	1.00 (0.000)
Weight (SD units)	14/14 kg	0.38 (0.008)	89.6	0.36 (0.001)	11/9 kg	0.39 (0.011)	36.1	0.47 (0.001)
Waist circumference (SD units)	11/13 cm	0.13 (0.008)	31.1	0.10 (0.001)	10/9 cm	0.13 (0.011)	11.9	0.21 (0.001)
Waist-to-hip ratio (SD units)	0.06/0.07 ratio	-0.03 (0.008)	-7.3	-0.07 (0.001)	0.06/0.07 ratio	-0.04 (0.011)	-3.8	0.02 (0.001)
BMI (SD units)	4/5 kg/m^2	-0.08 (0.008)	-18.0	-0.09 (0.001)	3/3 kg/m^2	-0.07 (0.011)	-6.0	0.02 (0.001)
Lean body mass (SD units)	8/5 kg	0.55 (0.008)	133.3	0.56 (0.001)	6/4 kg	0.66 (0.011)	61.8	0.70 (0.001)
Fat body mass (SD units)	8/10 kg	0.18 (0.008)	41.2	0.15 (0.001)	6/7 kg	0.10 (0.011)	9.2	0.19 (0.001)
Hand grip strength (SD units)‡	9/6 kg	0.24 (0.008)	56.6	0.29 (0.001)	9/6 kg	0.30 (0.021)	14.5	0.33 (0.006)
FEV_1 (SD units)	0.8/0.5 L	0.34 (0.008)	78.9	0.42 (0.001)	0.7/0.5 L	0.31 (0.011)	28.9	0.37 (0.001)
FVC (SD units)	1.0/0.7 L	0.41 (0.008)	97.2	0.47 (0.001)	0.8/0.5 L	0.35 (0.011)	32.2	0.39 (0.001)
Systolic blood pressure (mm Hg)§	19/21 mm Hg	-1.39 (0.17)	-15.9	-1.14 (0.02)	20/23 mm Hg	-0.27 (0.24)	-1.1	0.33 (0.03)
Diastolic blood pressure (mm Hg)§	11/11 mm Hg	-0.61 (0.09)	-12.8	-0.36 (0.01)	12/11mm Hg	-0.11 (0.13)	-0.9	0.43 (0.02)
Heart rate (bpm)	12/11 bpm	-0.25 (0.09)	-5.2	-0.58 (0.02)	12/12 bpm	-0.46 (0.13)	-3.6	-0.38 (0.02)
cIMT (mm)‡**	0.15/0.11 mm	0.001 (0.015)	0.1	0.009 (0.002)	0.17/0.15 mm	0.003 (0.003)	1.0	0.006 (0.001)
Carotid plaque score (mm)‡	-	-	-	-	1.16/0.96 mm	0.002 (0.021)	0.1	-0.006 (0.006)
LDL cholesterol (mmol/L)††	-	-	-	-	0.7/0.7 mmol/L	-0.09 (0.025)	-3.5	0.01 (0.005)
Triglycerides (mmol/L)††	-	-	-	-	1.7/1.4 mmol/L	-0.05 (0.055)	-1.0	0.04 (0.011)

*Traits represented in SD units were normalised within strata of sex and region in CKB (as described in Chapter 3.3).
†The UKB genetic risk score was constructed from Wood[20] SNPs and effect sizes, and the CKB genetic risk score was constructed from clumped GIANT+UKB[30] SNPs and effect sizes. All traits had missing data <0.1%, except for, in UKB: weight (0.1%), BMI (0.1%), lean body mass (1.6%), fat body mass (1.7%), hand grip strength (0.5%), FEV1 (8.6%), FCV (8.6%), and pulse rate (0.2%); and in CKB: hang grip strength (1.4%), cIMT (0.4%), and carotid plaque score (0.4%). Associations with traits represented in natural

units were meta-analysed across sex. The causal effects from instrumental variable analyses were presented per unit genetically instrumented height, where the unit was 1 SD of measured height (Table 3.1). The associations were adjusted for genomic principal components and genotyping array type.
¶Associations based on the full UKB (n=481667) and CKB (n=490067) datasets after exclusions for prior cardiovascular diseases. Associations with traits represented in SD units were adjusted for age, sex, year of birth, and region in CKB. Associations with traits represented in natural units were meta-analysed across sex and adjusted for age, year of birth, and region in CKB.
‡Associations based on CKB second resurvey (measurements were available in 21299 participants).
§For associations with genetically instrumented height, systolic blood pressure was increased by 15 mm Hg and diastolic blood pressure increased by 10 mm Hg for participants taking blood pressure lowering medications.
**Associations based on UKB imaging visit data (measurements were available in 1836 participants). cIMT=carotid intima media thickness.
††Associations based on blood lipids data available in 6062 CKB participants.

Table 7.3 Associations of genetically instrumented height with atrial fibrillation in UKB and CKB, shown for different adiposity and blood pressure adjustments.

	Odds ratio (95% CI) per 1 SD taller genetically instrumented height*		
Adjustments¶	**UKB (336433 participants and 13225 cases of atrial fibrillation†)**	**CKB (77559 participants and 140 cases of atrial fibrillation†)**	**Combined (413992 participants and 13365 cases of atrial fibrillation†)**
Basic adjustments‡	1.32 (1.27, 1.38)	1.36 (0.82, 2.26)	1.32 (1.27, 1.38)
Basic adjustments + weight	1.18 (1.13, 1.23)	0.94 (0.56, 1.56)	1.18 (1.13, 1.23)
Basic adjustments + waist circumference	1.27 (1.22, 1.33)	1.02 (0.62, 1.70)	1.28 (1.22, 1.33)
Basic adjustments + waist-to-hip ratio	1.33 (1.27, 1.39)	1.06 (0.64, 1.76)	1.33 (1.27, 1.39)
Basic adjustments + BMI	1.35 (1.30, 1.41)	1.07 (0.65, 1.78)	1.35 (1.30, 1.41)
Basic adjustments + lean body mass	1.11 (1.06, 1.16)	0.85 (0.51, 1.44)	1.11 (1.06, 1.16)
Basic adjustments + fat body mass	1.25 (1.19, 1.31)	1.03 (0.62, 1.71)	1.25 (1.20, 1.31)
Basic adjustments + systolic blood pressure	1.31 (1.25, 1.37)	1.06 (0.64, 1.75)	1.31 (1.26, 1.37)
Basic adjustments + diastolic blood pressure	1.32 (1.26, 1.38)	1.06 (0.64, 1.75)	1.32 (1.26, 1.38)
Basic adjustments + all the above adjustments	1.15 (1.06, 1.26)	0.80 (0.47, 1.37)	1.15 (1.06, 1.25)

*The UKB genetic risk score was constructed from Wood[20] SNPs and effect sizes, and the CKB genetic risk score was constructed from clumped GIANT+UKB[30] SNPs and effect sizes. The causal effects from instrumental variable analyses were presented per unit genetically instrumented height, where the unit was 1 SD of measured height (Table 3.1).
†Atrial fibrillation cases included prevalent and incident cases. For CKB, only incident cases were available.
¶The adiposity and blood pressure variables were normalised within strata of sex and region in CKB (as described in Chapter 3.3).
‡Adjustment for genomic principal components and genotyping array type.

Table 7.4 Sensitivity analyses of standardised lean body mass and standardised systolic blood pressure for UKB (336433 participants).

	Beta (SE) per 1 SD taller genetically instrumented height†							
Trait*	**Genetic risk score**	**Inverse-variance weighted fixed-effects meta-analysis**	**Inverse-variance weighted random-effects meta-analysis**	**MR-Egger¶**	**MR-Egger intercept**	**MR-Egger intercept P-value**	**MR-PRESSO unadjusted**	**MR-PRESSO outlier adjusted**
Lean body mass (SD units)	0.551 (0.008)	0.627 (0.005)	0.627 (0.013)	0.672 (0.032)	-0.001 (0.001)	0.12	0.626 (0.013)	0.594 (0.010)
Systolic blood pressure (SD units)	-0.076 (0.008)	-0.085 (0.005)	-0.085 (0.010)	-0.065 (0.024)	-0.001 (0.001)	0.35	-0.086 (0.010)	-0.072 (0.008)

*Systolic blood pressure was increased by 15 mm Hg and diastolic blood pressure increased by 10 mm Hg for participants taking blood pressure lowering medications. Systolic blood pressure and lean body mass were normalised within strata of sex and region in CKB (as described in Chapter 3.3).
†The genetic risk score was constructed from Wood[20] SNPs and effect sizes. The meta-analyses, MR-Egger, and MR-PRESSO models used individual Wood[20] SNP univariate effect sizes (estimated from UKB) for systolic blood pressure and lean body mass, and Wood[20] effect sizes for height. The causal effects from instrumental variable analyses were presented per unit genetically instrumented height, where the unit was 1 SD of measured height (Table 3.1).
¶MR-Egger is a random-effects model.

7.6 Figures

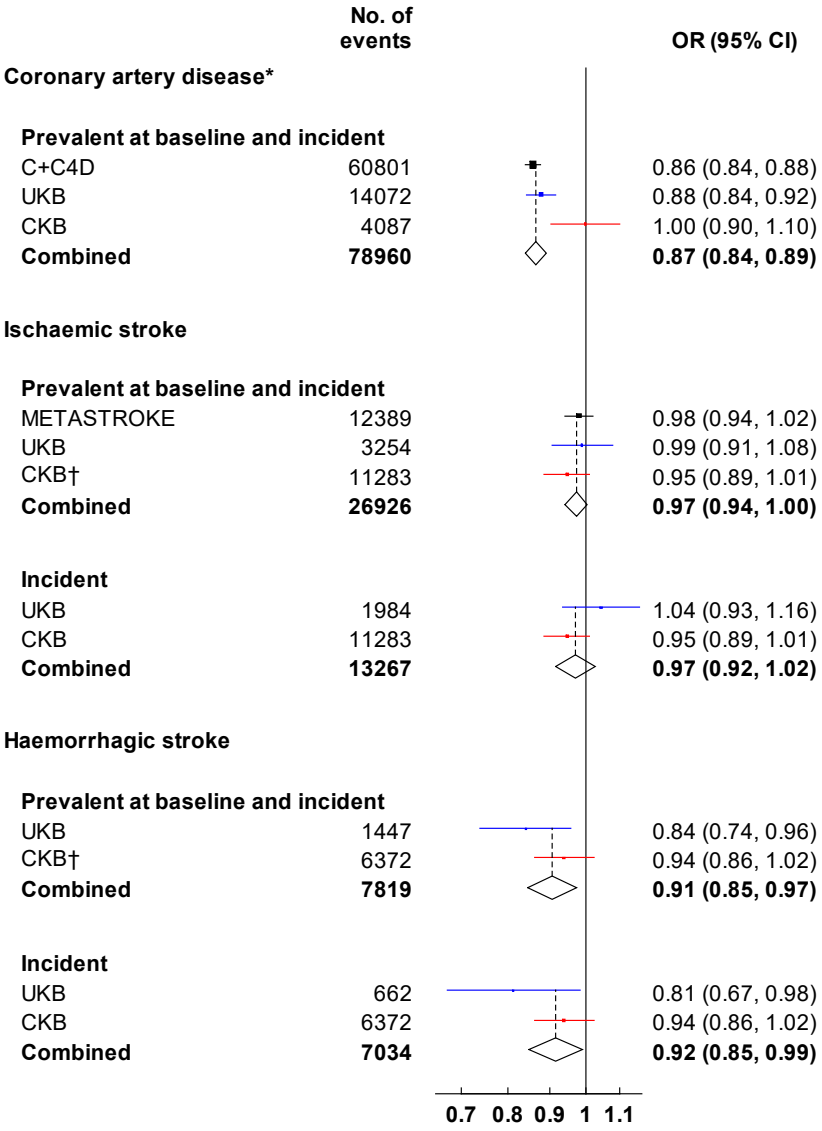


Figure 7.1 Associations of genetically instrumented height with coronary artery disease and stroke types in UKB (336433 participants) and CKB (77419 controls).

The UKB genetic risk score was constructed from Wood[20] SNPs and effect sizes, and the CKB genetic risk score was constructed from clumped GIANT+UKB[30] SNPs and effect sizes. The causal effects from instrumental variable analyses were presented per unit genetically instrumented height, where the unit was 1 SD of measured height (Table 3.1). Consortia genetically instrumented height was constructed from clumped GIANT+UKB[30] SNPs and effect sizes. Associations were adjusted for genomic principal components and genotyping array type. C+C4D=CARDIoGRAMplusC4D.

*Coronary artery disease included incident or prevalent (UKB only) myocardial infarction, fatal ischemic heart disease cases, and incident coronary revascularisation cases.

†Prevalent ischaemic and haemorrhagic stroke cases were not available in CKB (only incident cases).

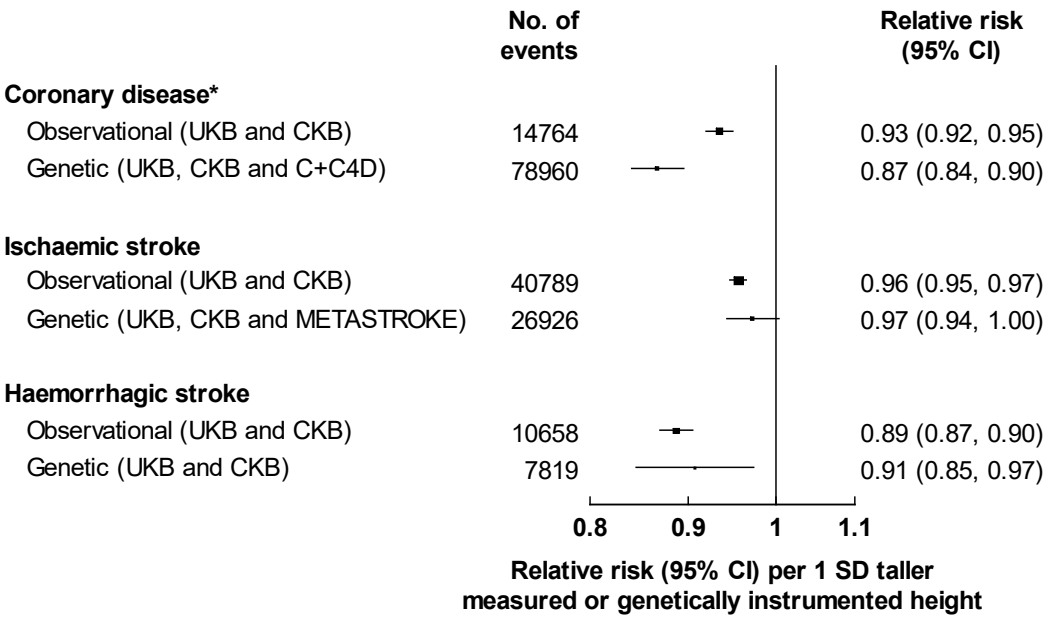


Figure 7.2 Comparisons of observation and genetic associations of taller height with coronary disease and stroke types.

Observational associations were based on incident outcomes and presented as hazard ratios (95% CIs) per 1 SD taller measured height adjusted for basic adjustments (age, sex, year of birth, and region in CKB), major cardiovascular disease risk factors and socio-economic risk factors (smoker status, cigarette equivalents [cigarettes/day], systolic blood pressure, diastolic blood pressure, diagnosed hypertension, diagnosed diabetes, self-rated walking pace [UKB only], household income, and education level). Genetic associations were based on prevalent at baseline or incident outcomes and were presented at odds ratios (95% CIs) per genetically instrumented height (where the unit was 1 SD of measured height [Table 3.1]) adjusted for genomic principal components and genotyping array type. The genetic risk score was constructed from clumped GIANT+UKB[30] SNPs and effect sizes. C+C4D=CARDIoGRAMplusC4D.

*Coronary disease was defined as major coronary events for observational analyses, which included incident myocardial infarction and fatal ischemic heart disease. For genetic analyses, coronary disease was defined as coronary artery disease, which in addition to major coronary events also included incident coronary revascularisation cases.

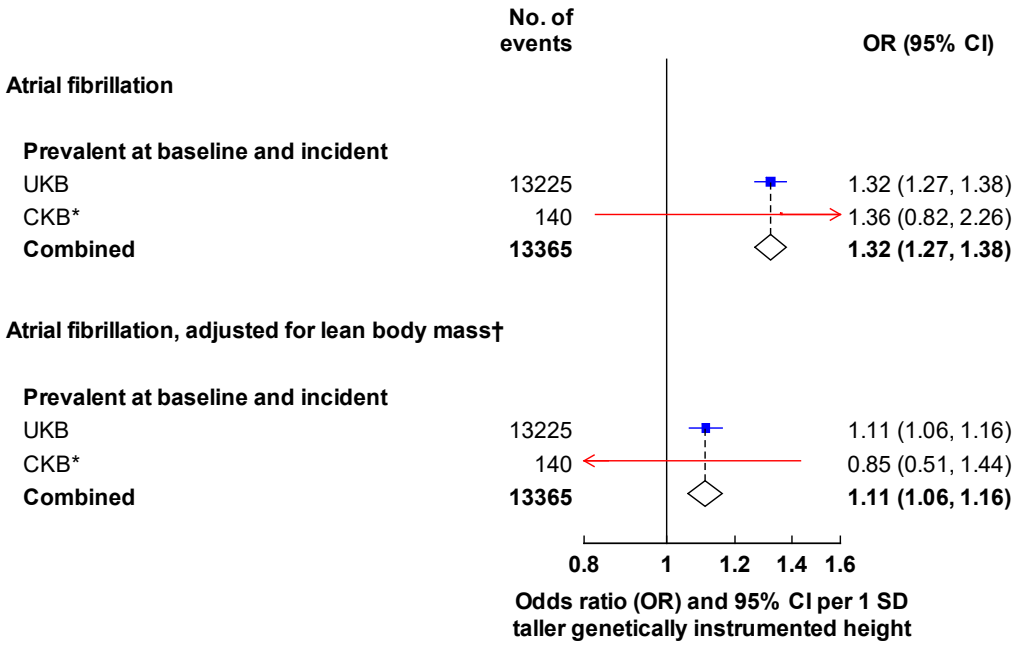


Figure 7.3 Associations of genetically instrumented height with atrial fibrillation in UKB (336433 participants) and CKB (77419 controls).

The UKB genetic risk score was constructed from Wood[20] SNPs and effect sizes, and the CKB genetic risk score was constructed from clumped GIANT+UKB[30] SNPs and effect sizes. The causal effects from instrumental variable analyses were presented per unit genetically instrumented height, where the unit was 1 SD of measured height (Table 3.1). The associations were adjusted for genomic principal components and genotyping array type.
*Prevalent atrial fibrillation cases were not available in CKB (only incident cases).
†Lean body mass was normalised within strata of sex and region in CKB (as described in Chapter 3.3).

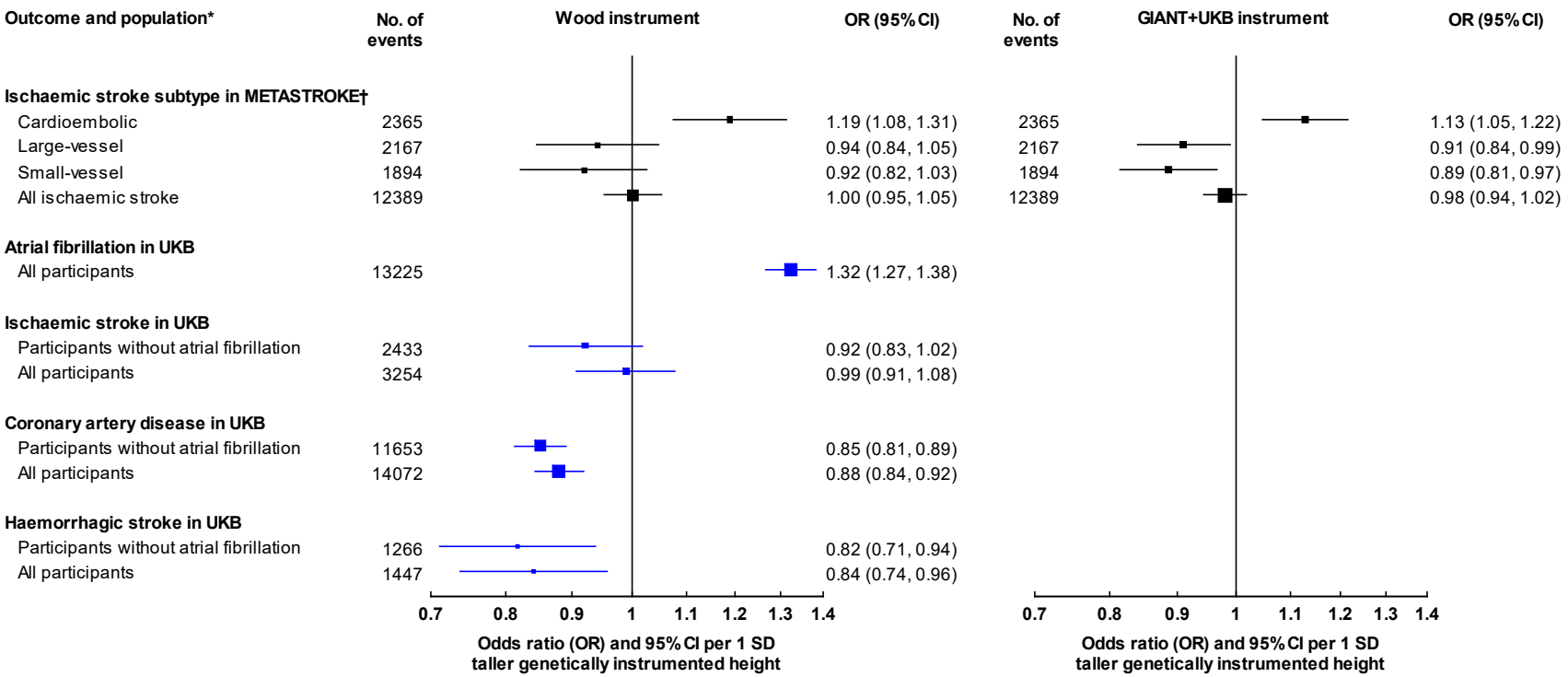


Figure 7.4 Associations of genetically instrumented height with ischaemic stroke subtypes in METASTROKE, and with related outcomes in UKB (336433 participants).

*UKB outcomes were prevalent at baseline or incident. Coronary artery disease included incident or prevalent myocardial infarction, fatal ischemic heart disease cases, and incident coronary revascularisation cases.

†For METASTROKE, the Wood[20] instrument was based on 692 Wood[20] SNPs and effect sizes, and the GIANT+UKB[30] instrument was based on 3150 GIANT+UKB[30] SNPs and effect sizes. For UKB, the Wood[20] genetic risk score was based on 692 Wood[20] SNPs effect sizes, and the causal effects in UKB from instrumental variable analyses were presented per unit genetically instrumented height (where the unit was 1 SD of measured height [Table 3.1]), adjusted for genomic principal components and genotyping array type.

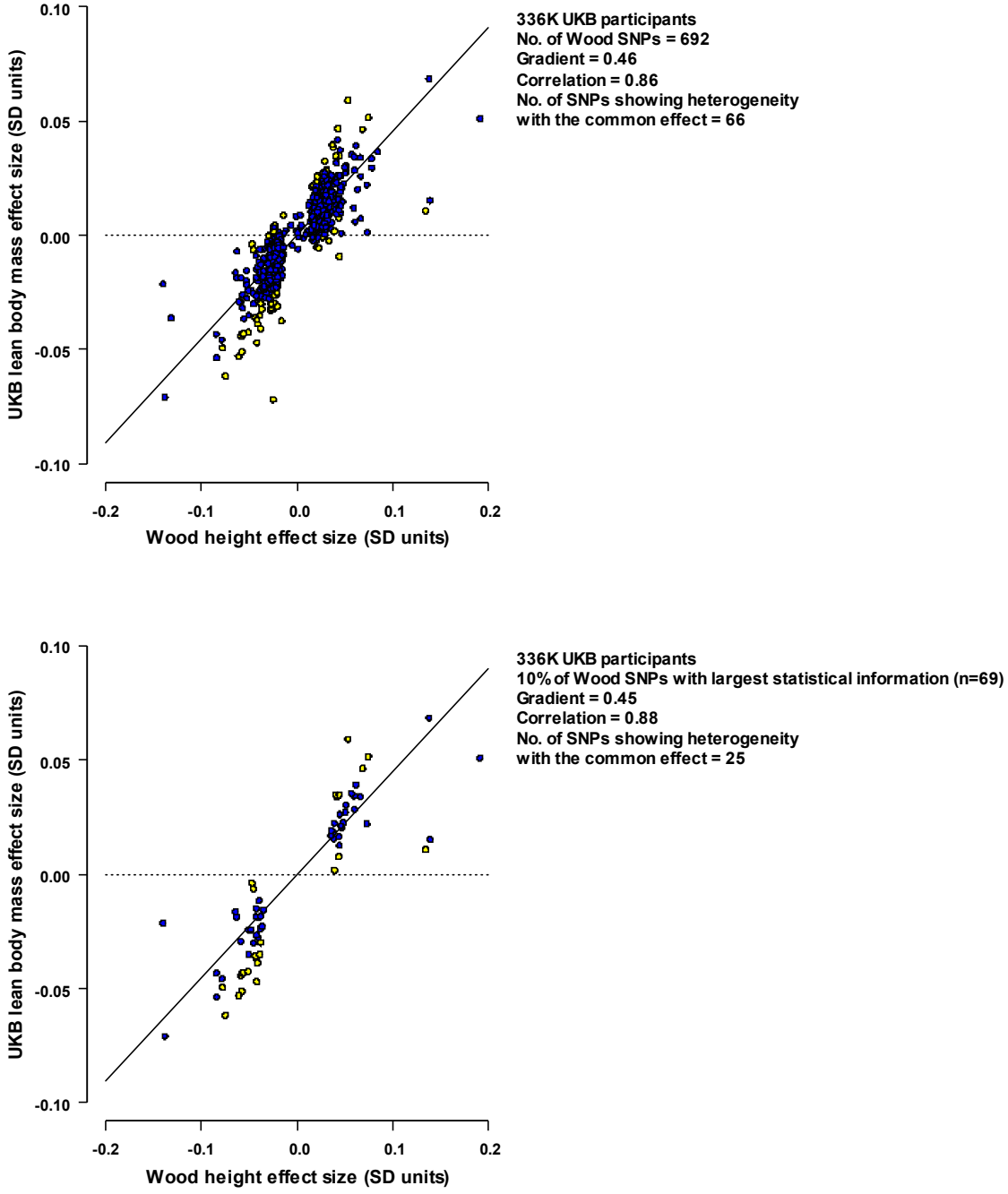


Figure 7.5 Univariate effect sizes of Wood[20] height-associated SNPs on standardised lean body mass and standardised standing height in 336433 UKB participants.

Lean body mass was normalised within strata of sex and region in CKB (as described in Chapter 3.3). SNPs in yellow showed heterogeneity with the common effect. Statistical information per SNP was approximated by $2\beta^2 MAF(1-MAF)$ where β is the SNP effect size on lean body mass, and MAF is the minor allele frequency of the SNP.

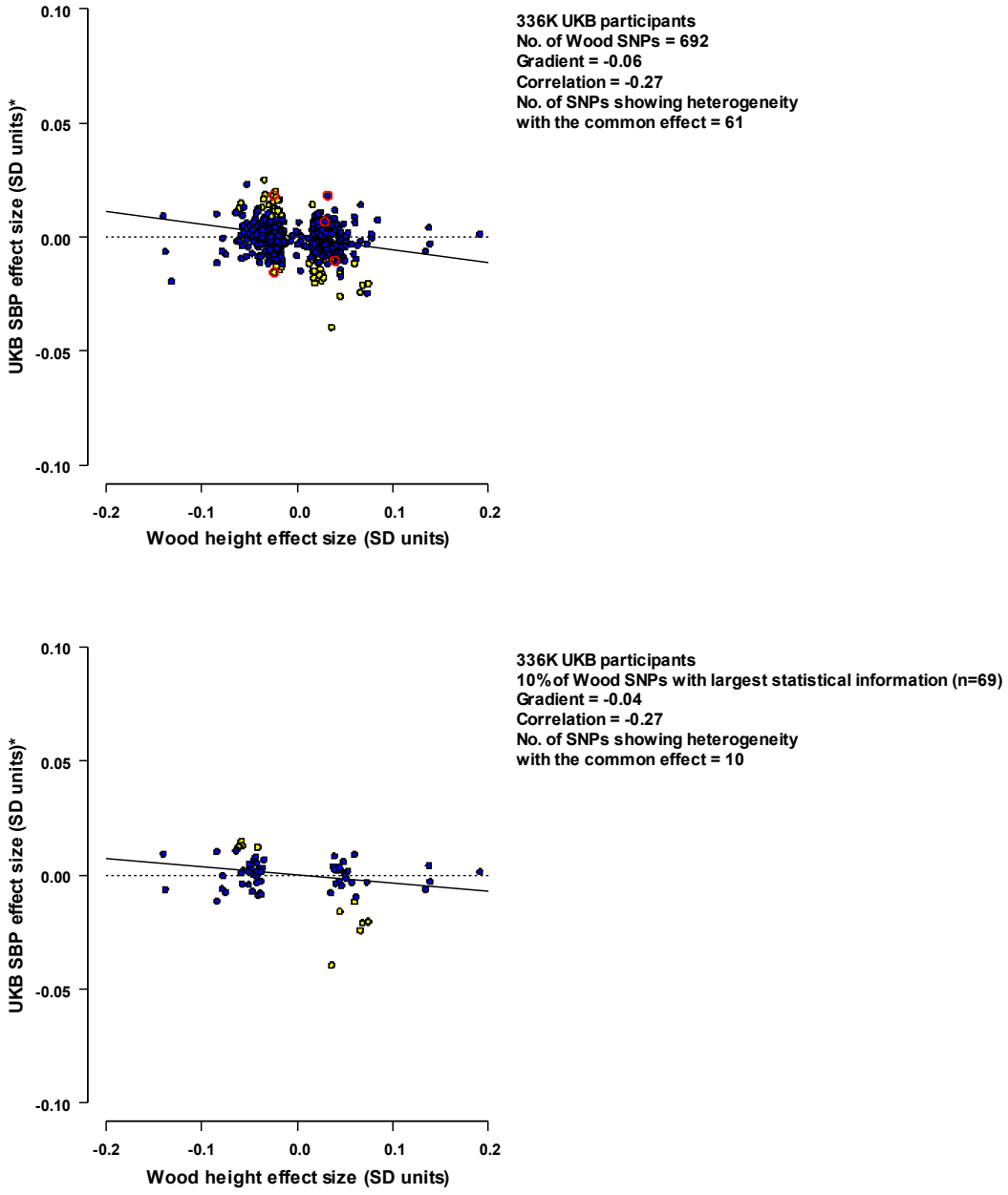


Figure 7.6 Univariate effect sizes of Wood[20] height-associated SNPs on standardised systolic blood pressure (SBP) and standardised standing height in 336433 UKB participants.

Systolic blood pressure was increased by 15 mm Hg and diastolic blood pressure increased by 10 mm Hg for participants taking blood pressure lowering medications. Systolic blood pressure was normalised within strata of sex and region in CKB (as described in Chapter 3.3). SNPs in yellow showed heterogeneity with the common effect. SNPs circled in red were SNPs known to be associated with blood pressure (identified in GWAS of blood pressure). Statistical information per SNP was approximated by $2\beta^2MAF(1\text{-}MAF)$ where β is the SNP effect size on systolic blood pressure, and MAF is the minor allele frequency of the SNP.

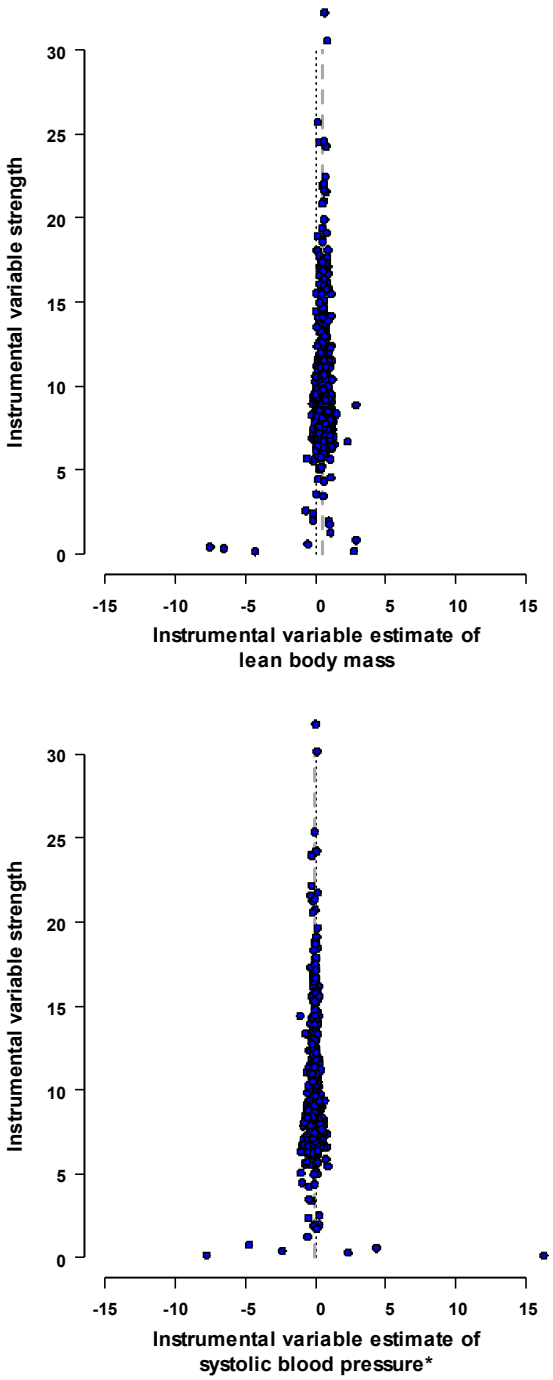


Figure 7.7 Funnel plots of instrumental variable strength against instrumental variable estimates for standardised lean body mass and standardised systolic blood pressure for Wood[20] height-associated SNPs in 336433 UKB participants.

Systolic blood pressure was increased by 15 mm Hg and diastolic blood pressure increased by 10 mm Hg for participants taking blood pressure lowering medications. Systolic blood pressure and lean body mass were normalised within strata of sex and region in CKB (as described in Chapter 3.3). Dashed grey line indicates the meta-analysed instrumental variable estimate.

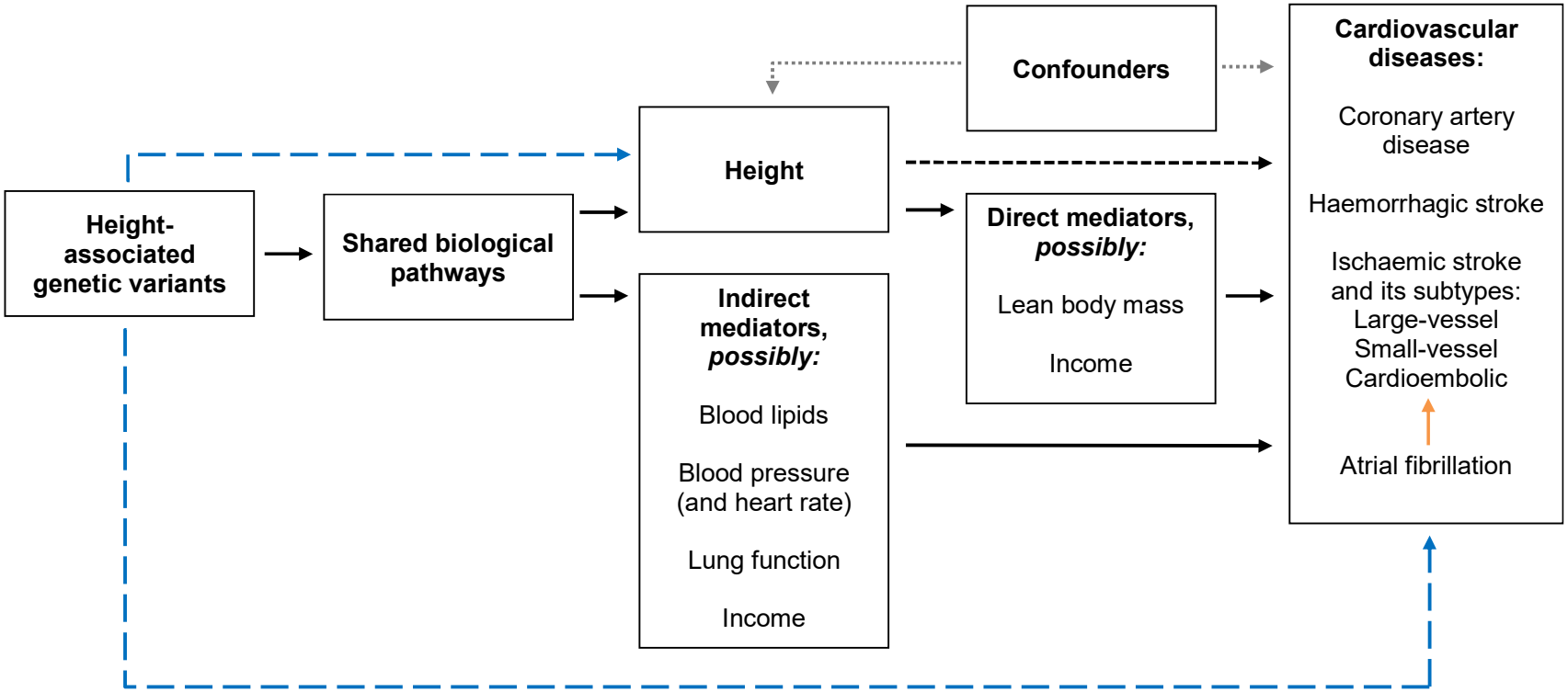


Figure 7.8 Diagrammatic representation of the associations of genetically instrumented height with coronary artery disease, ischaemic stroke subtypes, haemorrhagic stroke, and atrial fibrillation.

The lines with arrows represent associations between the different factors.

8. Conclusions

8.1 Summary of findings

The inverse association of genetically elevated height with coronary artery disease reported by C+C4D (OR per unit genetically elevated height: 0.88; 95% CI: 0.82, 0.95)[4] was replicated in the present thesis (OR: 0.87; 95% CI: 0.84, 0.89; 78 960 cases in a combined analysis of UKB, CKB, and C+C4D) (Figure 7.2). Likewise, the positive association of genetically elevated height with atrial fibrillation (OR per unit [scaled to 6.5 cm] genetically elevated height: 1.19; 95% CI: 1.02, 1.38; 879 cases)[10] was also replicated in the present analysis of 13 365 cases (OR per unit genetically elevated height: 1.32; 95% CI: 1.27, 1.38; combined across UKB and CKB) (Figure 7.3). While the association of genetically elevated height with all cases of ischaemic stroke was null, analyses of consortia (METASTROKE) data demonstrated different associations with ischaemic stroke subtypes. Large-vessel stroke, for which LDL cholesterol is an important risk factor (relative to other ischaemic stroke subtypes),[125] was inversely associated with genetically elevated height (OR per unit genetically elevated height: 0.91; 95% CI: 0.84, 0.99; 2167 cases), but cardioembolic stroke, for which atrial fibrillation is an important risk factor,[80] was positively associated with genetically elevated height (OR: 1.13; 95% CI: 1.05, 1.22; 2365 cases) (Figure 7.4).

The association of genetically elevated height with blood pressure and blood lipids suggested that the association with coronary artery disease could be partially mediated by these factors (blood pressure explained potentially a third of the association [based on UKB estimates], and blood lipids explained a quarter

to a third of the association [based on the estimate by Nelson et al.[4]]). The association of taller height with higher income[17] and the possible substantial impact of assortative mating[122–124] could explain a general tendency to better risk factors in taller individuals.

The association of height with atrial fibrillation was strongly related to its association with lean body mass (as the association of height with atrial fibrillation attenuated very substantially after adjustment for lean body mass in both observational and genetic analyses [Table 5.1 and Figure 7.3]).

The present research has identified several potential mediating factors, but does not fully explain the mechanisms underlying these pathways. Nelson et al. reported that height-associated SNPs were associated with growth factors and factors promoting cardiogenesis,[4] and Evangelou et al. reported that some SNPs associated with blood pressure were also strongly associated (both positively and inversely) with height.[34] Further investigations into the underlying mechanisms are needed.

The number of haemorrhagic stroke cases in CKB was large (>6K cases in genotyped participants compared to ~2K cases in the ERFC observational meta-analysis[1]) and there was a borderline statistically significant inverse associations with height in both observational and genetic analyses. However, the power of the genetic instrument was low (10.5% in CKB versus 25.3% estimated in UKB and 19.7% reported in the replication study by GIANT [Yengo et al., 2018][30]), even though height is estimated to have the same SNP-based heritability in

Europeans and in CKB (both ~50%). Most of the difference in variance explained by the instruments was due to lower average effect sizes and lower MAF of the SNPs in the CKB population, relative to a European population (GIANT [Wood et al., 2014][20]) (Table 6.7). This is likely because different lead SNPs for height may exist in European and East Asian populations (as has previously been reported for blood lipids[31]).

In UKB and CKB, the observational associations of sitting and leg height were generally consistent with the associations for standing height (Figure 5.1). Genetic associations of sitting and leg height were not investigated because large GWAS of these traits were not available.

8.2 Strength and limitations of the study

A major strength of the present thesis is the comparison of two large prospective studies of ethnically diverse populations, involving the China Kadoorie Biobank and UK Biobank studies, supplemented with additional data from other large consortia (C+C4D and METASTROKE). The present study is the largest study to date of the genetic associations of height with haemorrhagic stroke and atrial fibrillation, and the largest such study to date to be conducted in East Asians.

A limitation of the present thesis is that the genetic instrument for height in East Asians was weak relative to Europeans. Height has been estimated to have a similar heritability in Europeans and East Asians, but a smaller number of East Asians have been included in the available GWAS discovery studies. With greater contributions from East Asians or specifically Chinese in GWAS discovery

studies, it should be possible to increase the percentage of the variance of height explained in a Chinese population such as CKB, and thereby increasing the power of the instruments to detect associations with such diseases. A GWAS discovery in CKB would need to be integrated with the current European-discovered SNPs as well as a replication study in a suitable East Asian population, which was beyond the scope of the present thesis but is planned for the future.

An additional limitation of the present thesis was that ischaemic stroke subtype information was only available in METASTROKE, and was not available in UKB or CKB. There was also an inconsistency between the genotyped datasets used for Mendelian randomisation analyses in UKB and CKB. The UKB genotyped dataset excluded related participants (3rd degree or closer) while the CKB genotyped dataset did not. Additionally, the present thesis makes use of a set of SNPs (identified by clumping) for the GIANT+UKB[30] genetic risk score instead of the actual SNPs identified by Yengo et al. (as these SNPs were not included in their pre-publication report[43], but have subsequently been made available on the GIANT consortium's website[103]).

8.3 Further research

Future work might include analysing ischaemic stroke subtype data in CKB, which should be available shortly. The set of disease outcomes considered in the Mendelian randomisation analyses could be extended to the wider set of disease outcomes associated with height in the traditional observational analyses (including cancer types, respiratory diseases, diabetes, and pulmonary

embolism).[1] Pulmonary embolism has been linked to atrial fibrillation,[130] so this would be an interesting outcome to investigate. The UKB and CKB genotyped datasets have accumulated a substantial number of these diseases (Table 8.1), however some of these associations in UKB have very recently been published by Lai et al (2018).[39] This study confirmed similar results to those in the present thesis for coronary artery disease and atrial fibrillation in UKB.[39] Additionally, Lai et al. (2018) reported an inverse association of genetically elevated height with hypertension, a positive association with venous thromboembolism, and no significant associations with respiratory diseases.[39]

Further insight into the potential mediating factors might be gained by exploring whether the associations of genetically elevated height with cardiovascular diseases vary by subsets of height-associated SNPs (if rational distinctions of sets of SNPs can be made, e.g. by subdividing the height-associated SNPs into those that are strongly and weakly associated with a particular mediating factor, such as LDL cholesterol or blood pressure). However, if assortative mating is a driving force behind the multiple associated traits, then this may not be possible.

Currently identified SNPs specifically associated with lean body mass only explain <1% of its variance (5 SNPs identified in a discovery sample of ~40K Europeans).[114] Obtaining a genetic instrument for the phenotype of lean body mass adjusted for height should be particularly informative to further investigate the association of height with atrial fibrillation. This would initially require a discovery GWAS of lean body mass adjusted for height, which could be performed in UKB and CKB. In addition, UKB has an extensive biochemistry

panel (biomarkers for cardiovascular disease, cancer, diabetes, liver function, renal function, and bones and joint) which will soon be available for further mediation analyses of the effects of biochemical mediators of height on disease.

It is not clear whether East Asians are a sufficiently homogenous to be considered as a single ethnic group for genetic studies. For example, CKB has much genetic diversity across the 10 geographical regions of the study. In any case, improving the genetic instrument for height in East Asians, and specifically Chinese, would increase the power of the instrument to detect associations with such diseases. Application of transethnic methods in UKB and CKB, such as transethnic meta-analyses of GWAS results (which have previously been conducted for kidney function and type 2 diabetes[131]), might improve the genetic instrument for height in East Asian populations. The total numbers of SNPs available in UKB and CKB (96M and 24M, respectively) were much greater than the number of SNPs used by the GIANT consortium (2018)[30] in their meta-analysis of GWAS for height (~17M SNPs). This presents an additional opportunity for potential instrument improvement.

8.4 Tables

Table 8.1 Number of selected events for phenome-wide analyses of height in UKB and CKB participants with genotyping.

	No. of events			
	UKB (n=336432)		CKB (n=77419)	
Phenome-wide outcome*	Prevalent at baseline	Incident	Prevalent at baseline	Incident
Vascular death	-	3181	-	2117
Non-vascular death	-	6498	-	4186
Cancer	25386	26256	360	4425
Chronic obstructive pulmonary disease	-	6673	5738	2521
Asthma	38591	16308	408	429
Diabetes	15249	13034	4572	4887
Pulmonary embolism	2688	2133	-	-
Pulmonary heart disease†	2688	2628	-	385
Fracture	32081	10362	5019	2884

*Non-fatal outcomes were prevalent at baseline or incident.
†Pulmonary embolism or other pulmonary heart diseases (ICD-10: I26-I27).

Appendix

Supplementary tables for chapter 2

Supplementary Table 2.1 Associations of genetically elevated height and cardiovascular risk factors, extracted from results of systematic search 3 (published studies of the causal relationship between the genetics of height and major diseases).

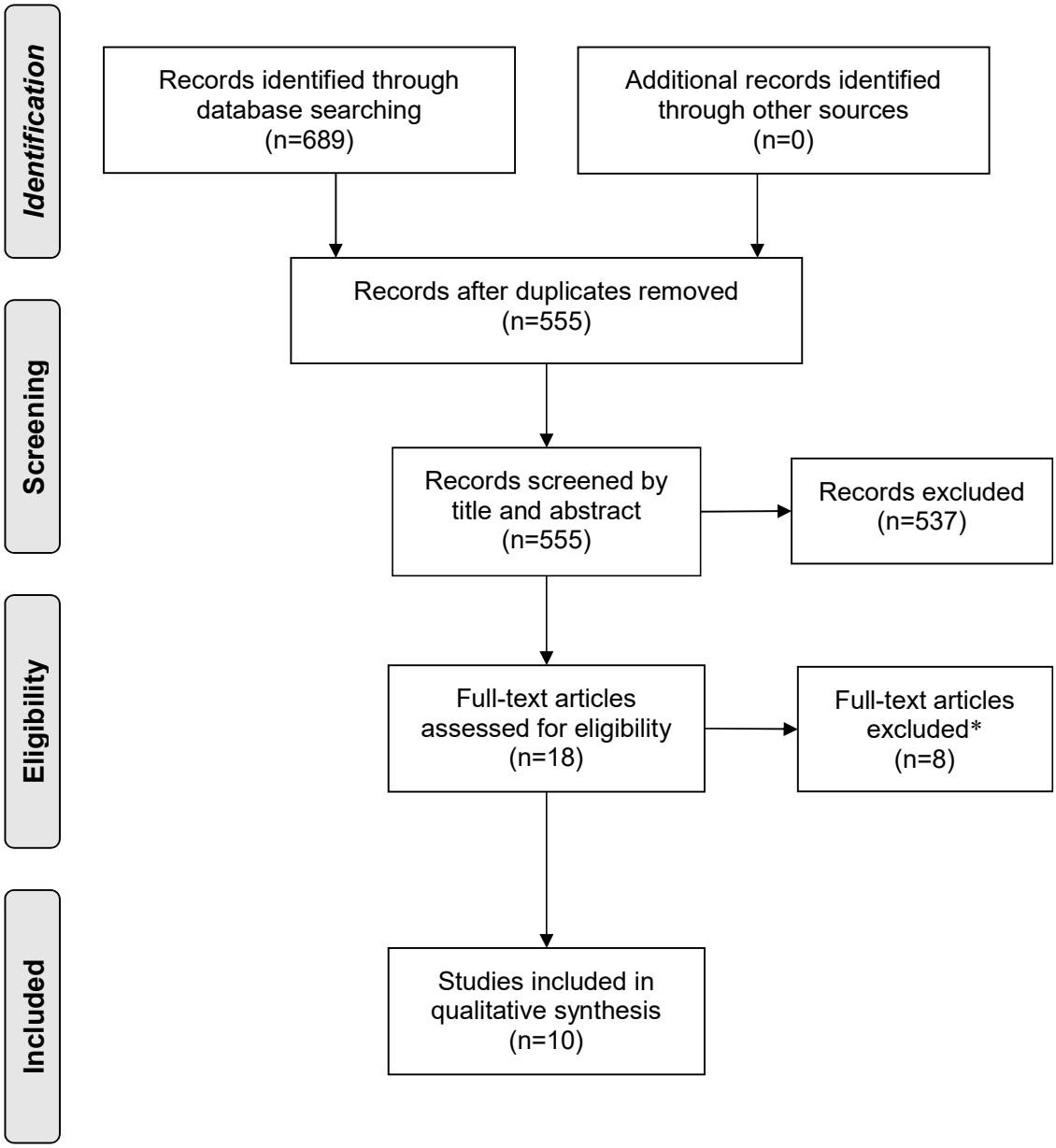
Author, year	Study description*	No. of height-associated SNPs	Cardiovascular disease risk factor or trait	No. of participants	Estimate (95% CI) per unit (6.5 cm) genetically elevated height	Units	Data source*	Reference for source of cardiovascular disease risk factor or trait data (author, year)
Mendelian randomisation studies of cardiometabolic diseases								
Nelson, 2015[4]	Mendelian randomization analysis of CARDIoGRAMplusC4D consortium data	180	BMI	249796	0.01 (-0.02, 0.03)	SD units	GIANT consortium	Speliotes, 2010[132]
			Systolic blood pressure‡	69899	0.34 (-0.31, 1.00)	mm Hg	ICBP	Ehret, 2011[54]
			Diastolic blood pressure‡	69909	0.14 (-0.27, 0.56)	mm Hg	ICBP	Ehret, 2011[54]
			Mean arterial pressure‡	29182	0.20 (-0.19, 0.60)	mm Hg	ICBP	Wain, 2011[133]
			Pulse pressure‡	74079	0.23 (-0.06, 0.52)	mm Hg	ICBP	Wain, 2011[133]
			LDL cholesterol	95454	-0.06 (-0.09, -0.04)	SD units	46 participating studies	Teslovich, 2010[53]
			HDL cholesterol	99900	-0.02 (-0.05, 0.02)	SD units	46 participating studies	Teslovich, 2010[53]
			Triglycerides	96598	-0.05 (-0.08, -0.03)	SD units	46 participating studies	Teslovich, 2010[53]
			Fasting glucose	96496	0.01 (-0.01, 0.02)	mmol/L	MAGIC	Manning, 2012[134]
			Fasting insulin	85573	0.01 (-0.01, 0.02)	log pmol/L	MAGIC	Manning, 2012[134]
			Smoking quantity	41150	0.04 (-0.01, 0.09)	Score: 0-3§	20 participating studies	Liu, 2010[135]

Nüesch, 2016[32]	Mendelian randomization analysis of CARDIoGRAMplusC4D consortium data†	69	BMI	54099	-0.10 (-0.15, -0.05)	All estimates are in SD units	21 participating studies	
			FEV_1	11131	0.26 (0.15, 0.36)			
			FVC	11129	0.30 (0.20, 0.41)			
			Systolic blood pressure	52345	-0.05 (-0.10, 0.00)			
			Non-HDL cholesterol	41477	-0.12 (-0.17, -0.06)			
			HDL cholesterol	43030	0.02 (-0.03, 0.08)			
			Triglycerides	42117	-0.10 (-0.16, -0.05)			
			C-reactive protein	35538	-0.07 (-0.13, -0.00)			
			Fasting glucose	40451	-0.04 (-0.10, 0.01)			
Mendelian randomisation studies of cancer and respiratory diseases								
Zhang, 2015[55]	Mendelian randomization analysis of BCAC and DRIVE consortia data	168	Weight	51634	0.046 (0.040, 0.051)¶	Kg	BCAC and DRIVE consortia	
			BMI	47221	-0.005 (-0.006, -0.004)	Kg/m^2		
			Ever (vs never) smoker	39562	1.000 (0.999, 1.001)	Odds ratio		
Davies, 2015[52]	Mendelian randomization analysis of PRACTICAL consortium data	179	Passive smoker	752	1.00 (0.86, 1.16)	All estimates are odds ratios	PRACTICAL consortium	
			Ever smoker	780	1.08 (0.93, 1.25)			
			Current smoker	552	1.21 (0.97, 1.51)			
			≥3 drinks in the last week	820	1.05 (0.91, 1.22)			
Other relevant studies:								
Bulik-Sullivan, 2015[21]	Genetic correlations across human diseases and traits, calculated from consortia data	697	Waist circumference	253288	0.22 (0.18, 0.27)	All estimates are correlations	GIANT consortium	-
			Hip circumference	253288	0.39 (0.35, 0.44)		GIANT consortium	-
			Waist-to-hip ratio	253288	-0.08 (-0.13, -0.03)		GIANT consortium	-
			Obesity class 1	236231	-0.10 (-0.14, -0.05)		GIANT consortium	Locke, 2015[85]
			Obesity class 2	236231	-0.09 (-0.15, -0.03)		GIANT consortium	Locke, 2015[85]
			Obesity class 3	236231	-0.15 (-0.22, -0.07)		GIANT consortium	Locke, 2015[85]
			Overweight	236231	-0.09 (-0.14, -0.04)		GIANT consortium	Locke, 2015[85]
			LDL cholesterol	95454	-0.09 (-0.16, -0.02)		46 participating studies	Teslovich, 2010[53]
			Total cholesterol	95454	-0.09 (-0.15, -0.04)		46 participating studies	Teslovich, 2010[53]
			Triglycerides	96598	-0.08 (-0.13, -0.03)		46 participating studies	Teslovich, 2010[53]
			Years of education	101069	0.13 (0.08, 0.19)		42 participating studies	Rietveld, 2013[136]

*CARDIoGRAMplusC4D=Coronary ARtery DIsease Genome-wide Replication and Meta-analysis (CARDIoGRAM) plus The Coronary Artery Disease (C4D) Genetics; GIANT=Genetic Investigation of Anthropometric Traits; ICBP=International Consortium for Blood Pressure; MAGIC=Meta-Analyses of Glucose and Insulin-related traits

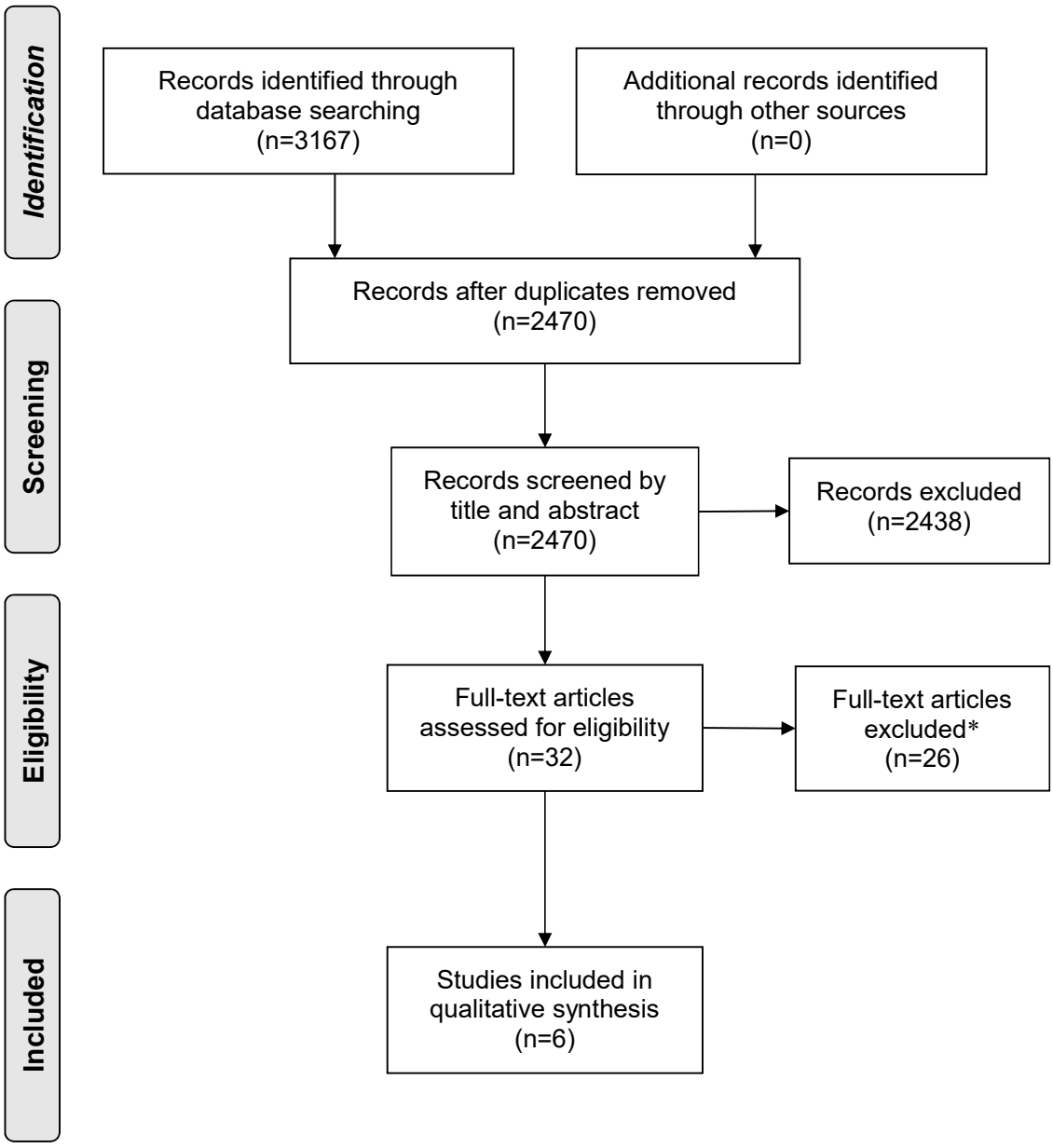
Consortium; BCAC=Breast Cancer Association Consortium; DRIVE=Discovery, Biology, and Risk of Inherited Variants in Breast Cancer; PRACTICAL=Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome; GLGC=Global Lipids Genetics Consortium.
†Study greatly overlaps Nelson using CARDIoGRAM or C4D consortia cases.
‡Blood pressure measure corrected for use of blood pressure lowering medication and adjusted for age, age2, sex, and BMI.
§Smoking quantity score: 0 (1-10 cigarettes/day), 1 (11-20 cigarettes), 2 (21-30 cigarettes), and 3 (>30 cigarettes).
¶The log relative risks were scaled pro rata to 6.5 cm genetically elevated height.

Supplementary figures for chapter 2



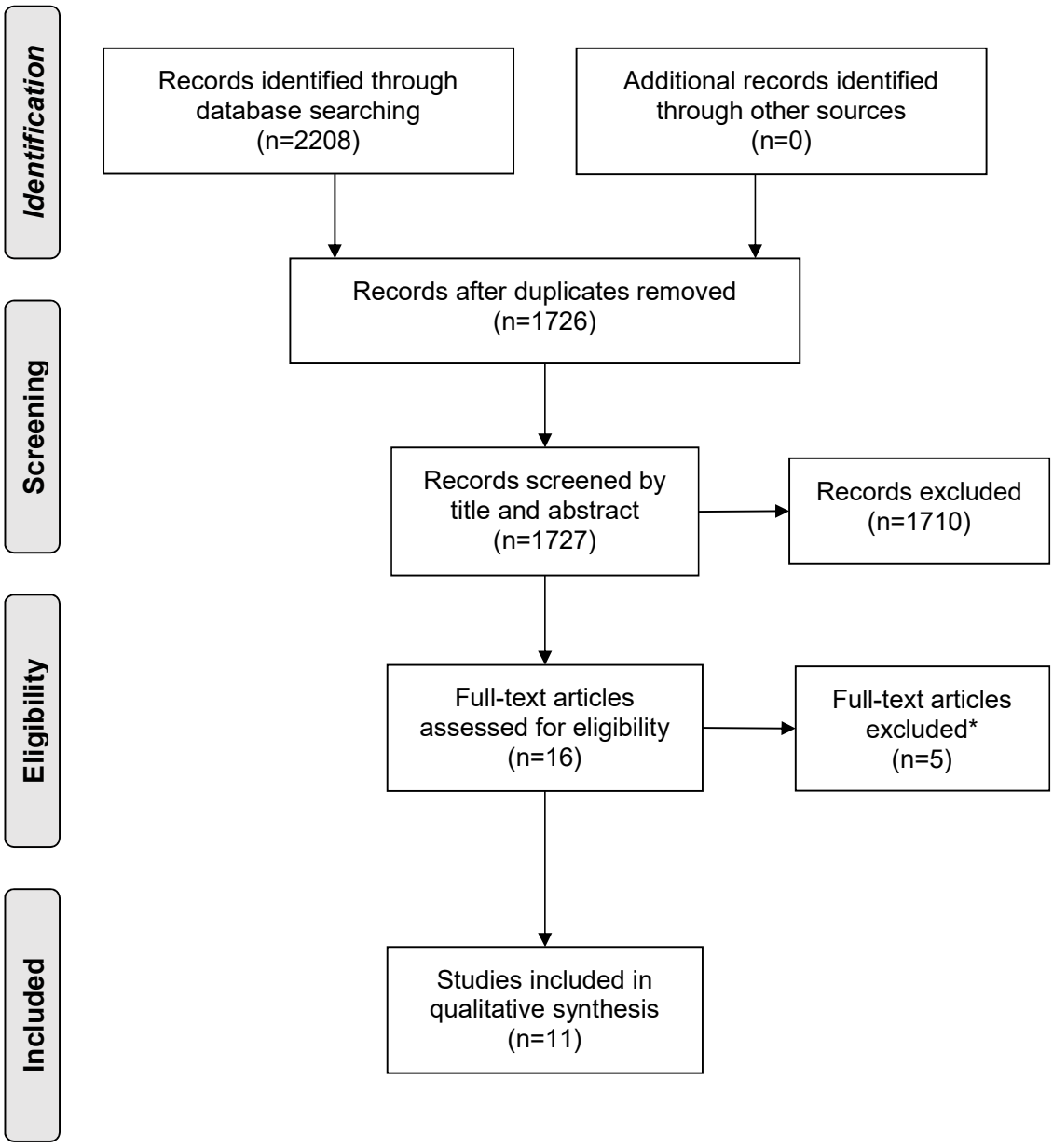
Supplementary Figure 2.1 Flow diagram illustrating search 1: published studies of the associations of the components of adult height (standing, sitting, and leg height) with major diseases.

*Full-text articles were excluded if the study: did not meet the pre-defined minimum number of events (n=5); did not provide number of cases or relative risks for sitting or leg height (n=2); was included in a systematic review already included in the search 1 results (n=1).



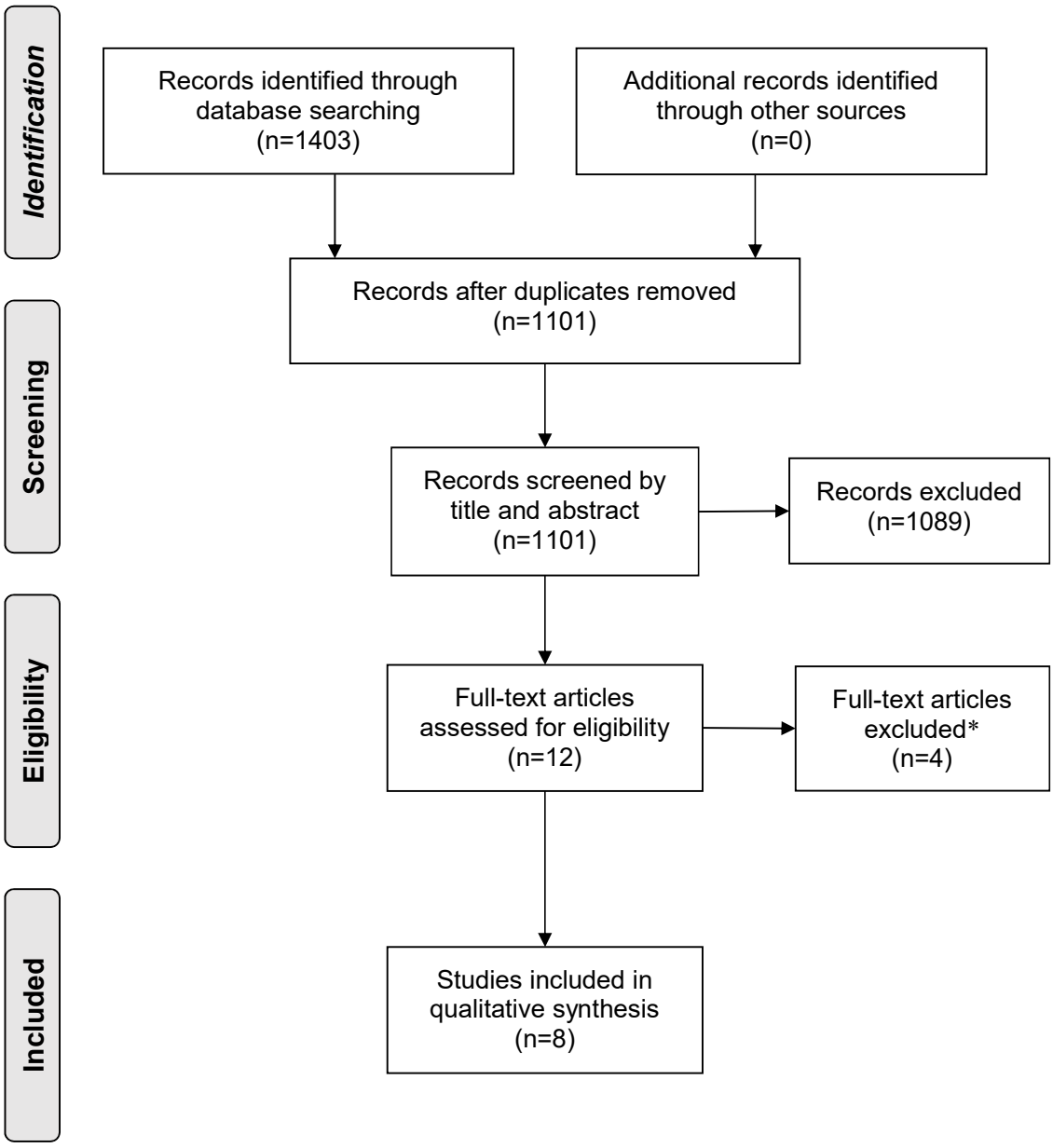
Supplementary Figure 2.2 Flow diagram illustrating search 2: published studies of the associations of adult height with major diseases in East Asian populations.

*Full-text articles were excluded if the study: did not meet the pre-defined minimum number of events (n=17); was a review article of studies already in the search results (n=4); was a genetic study covered in systematic search 3 (n=3); examined peripheral arterial disease in diabetics (n=1); overlapped with a study already included in the results (n=1).



Supplementary Figure 2.3 Flow diagram illustrating search 3: published studies of the causal relevance of the associations of adult height with major diseases.

*Full-text articles were excluded if the study: was a review article of studies already in the search results (n=2); did not meet the pre-defined minimum number of events (n=1); was a review article of pathways between height genes and cancer (n=1); did not have a sufficient number of discordant twin pairs, with regard to outcome, for analysis (n=1).



Supplementary Figure 2.4 Flow diagram illustrating search 4: published studies of the genetics of height in East Asian populations.

*Full-text articles were excluded if the study: was a replication study (n=2); identified variants in a population that was not East Asian (n=1); did not identify East Asian specific variants (n=1).

Supplementary tables for chapter 3

Supplementary Table 3.1 Additional information about UKB outcome variables.

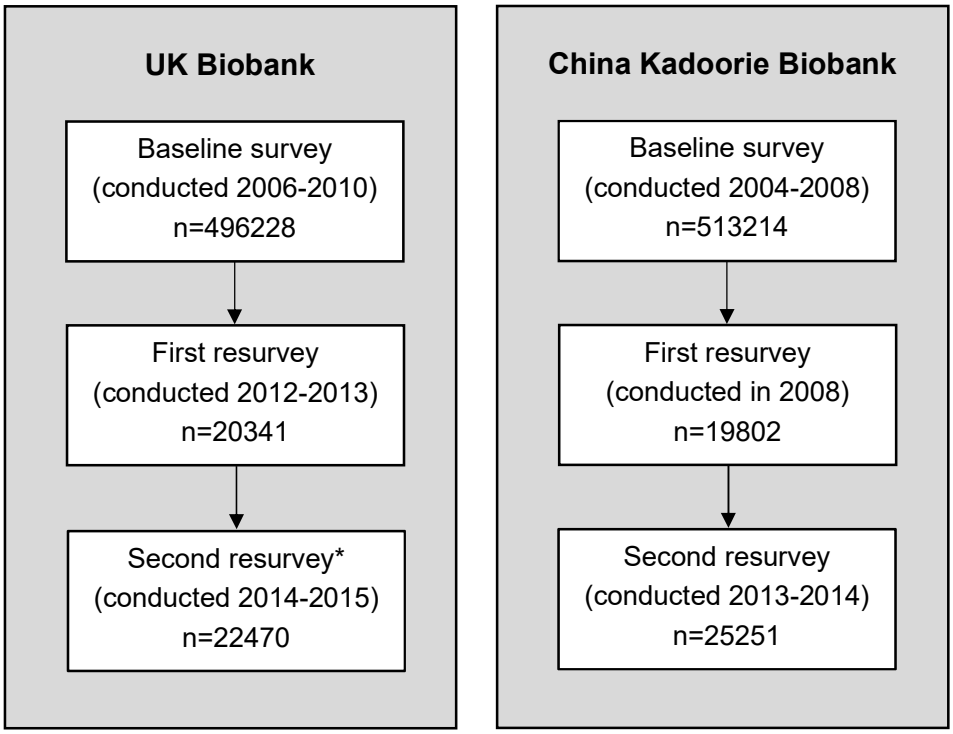
Outcome (ICD-10 or OPCS-4 code)	Link to UKB data showcase website[137]	Data sources
Algorithmically defined myocardial infarction (I21-I22)	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=42000	Self-reported, hospital episode statistics (diagnoses and procedures), and death registries
Coronary heart disease death (I20-I25 [fatal])	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=40001	Death registries
Major coronary event (I21-I22, I20-I25 [fatal])	Union of algorithmically defined myocardial infarction and coronary heart disease death	Self-reported, hospital episode statistics (diagnoses and procedures), and death registries
Coronary artery bypass grafting (K40-K46)	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=41200	Hospital episode statistics (procedures)
Percutaneous transluminal coronary angioplasty (K49, K50.1, K75)	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=41200	Hospital episode statistics (procedures)
Coronary artery disease (I21-I22, I20-I25 [fatal], K40-K46, K49, K50.1, K75)	Union of major coronary event, coronary artery bypass grafting and percutaneous transluminal coronary angioplasty	Self-reported, hospital episode statistics (diagnoses and procedures), and death registries
Algorithmically defined ischaemic stroke (I63)	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=42008	Self-reported, hospital episode statistics (diagnoses and procedures), and death registries
Algorithmically defined haemorrhagic stroke (I60-I61)	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=42010	Self-reported, hospital episode statistics (diagnoses and procedures), and death registries
Atrial fibrillation (I48)	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=41202	Self-reported, hospital episode statistics (diagnoses and procedures), and death registries

Supplementary Table 3.2 Additional information about the CKB outcome variables.

Outcome (ICD-10 code or health insurance procedures curated by the CKB data team)	CKB data request form follow-up code[138]	Data sources
Acute myocardial infarction* (I21)	EP0004	Death registries, disease registries, and the national health insurance claim database
Subsequent myocardial infarction* (I22)	Not available in data request form	Death registries, disease registries, and the national health insurance claim database
Coronary heart disease death (I20-I25 [fatal])	EP0003 (fatal)	Death registries
Major coronary event (I21-I22, I20-I25 [fatal])	Union of myocardial infarction and coronary heart disease death	Death registries, disease registries, and the national health insurance claim database
Coronary artery bypass grafting (health insurance procedures curated by the CKB data team)	Not available in data request form	National health insurance claim database
Percutaneous transluminal coronary angioplasty (health insurance procedures curated by the CKB data team)	Not available in data request form	National health insurance claim database
Coronary artery disease (I21-I22, I20-I25 [fatal], health insurance procedures curated by the CKB data team)	Union of major coronary event, coronary artery bypass grafting and percutaneous transluminal coronary angioplasty	Death registries, disease registries, the national health insurance claim database, and data curated by the CKB data team
Ischaemic stroke (I63)	EP0009	Death registries, disease registries, and the national health insurance claim database
Haemorrhagic stroke (I60-I61)	EP0007 and EP0008	Death registries, disease registries, and the national health insurance claim database
Atrial fibrillation (I48)	Not available in data request form	Death registries and the national health insurance claim database

*Baseline self-reported myocardial infarction data were also available in CKB.

Supplementary figures for chapter 3



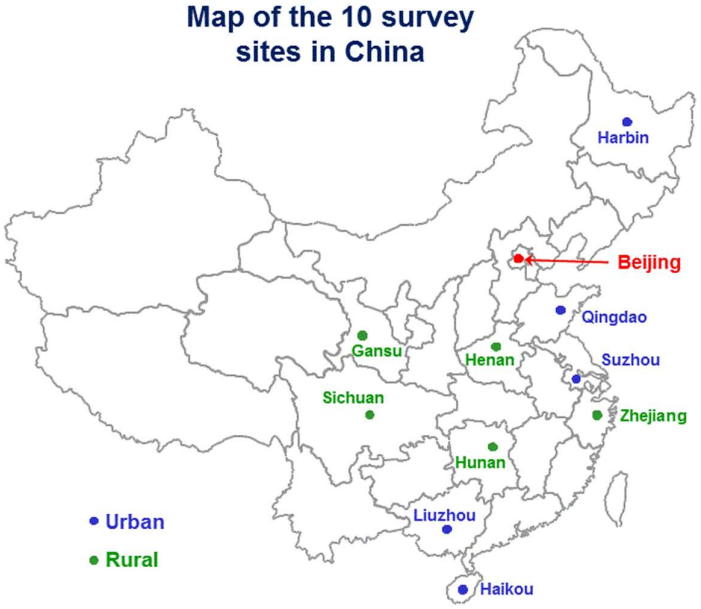
Supplementary Figure 3.1 Data available in the UKB and CKB baseline surveys and resurveys.

*Imaging visit.

Supplementary tables for chapter 4

Supplementary Table 4.1 Mean standing height across 10 regions in China, shown separately for men and women, after adjusting for age and year of birth.

	Men		Women	
Regions in China	No. of participants	Standing height mean, cm	No. of participants	Standing height mean, cm
Individual regions				
Qingdao (urban)	14909	169.6	18449	158.3
Harbin (urban)	20207	169.4	29488	157.5
Haikou (urban)	10434	165.2	18514	153.6
Liuzhou (urban)	17789	165.1	28621	154.2
Suzhou (urban)	21737	164.7	30185	154.1
Henan (rural)	26033	166.1	33841	154.9
Gansu (rural)	18775	165.5	29922	154.0
Zhejiang (rural)	23505	165.3	33223	154.8
Hunan (rural)	25096	162.7	31937	152.4
Sichuan (rural)	20985	160.6	34031	149.8
Aggregated regions				
All urban regions	85511	166.8	126086	155.5
All rural regions	114853	164.1	163617	153.1
All regions	200364	165.2	289703	154.2



Source of map:
http://www.ckbiobank.org/site/About+the+Study/Study+Design

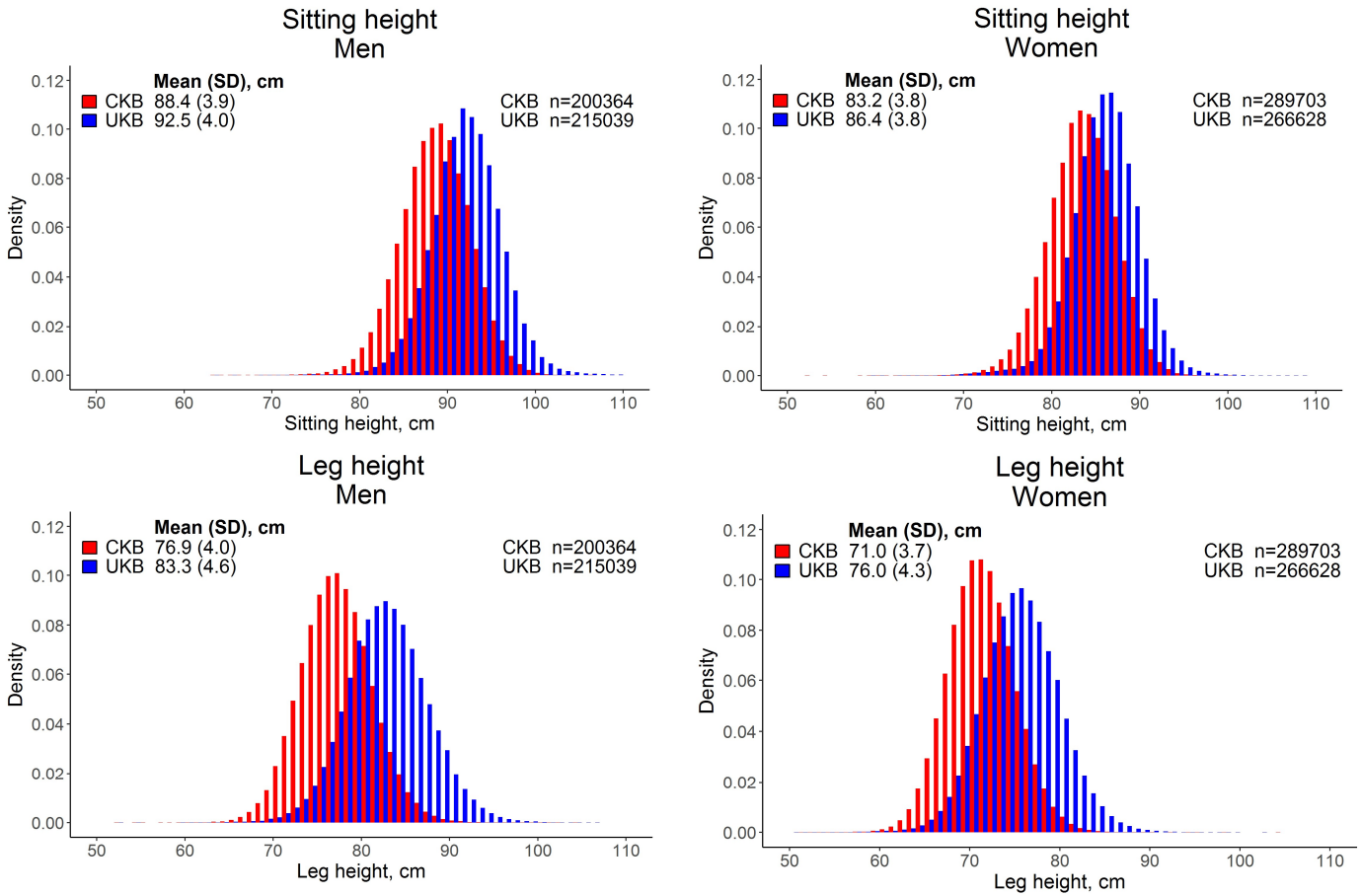
Supplementary Table 4.2 Number of participants by individual year of birth and 5-year age group in UKB and CKB.

Year of birth	5-year age group													
	UKB						CKB							
	<45	45-50	50-55	55-60	60-65	≥65	<40	40-45	45-50	50-55	55-60	60-65	65-70	≥70
≤1920	0	0	0	0	0	0	0	0	0	0	0	0	0	3
1921	0	0	0	0	0	0	0	0	0	0	0	0	0	5
1922	0	0	0	0	0	0	0	0	0	0	0	0	0	8
1923	0	0	0	0	0	0	0	0	0	0	0	0	0	22
1924	0	0	0	0	0	0	0	0	0	0	0	0	0	17
1925	0	0	0	0	0	0	0	0	0	0	0	0	0	16
1926	0	0	0	0	0	0	0	0	0	0	0	0	0	29
1927	0	0	0	0	0	0	0	0	0	0	0	0	0	44
1928	0	0	0	0	0	0	0	0	0	0	0	0	0	106
1929	0	0	0	0	0	0	0	0	0	0	0	0	0	177
1930	0	0	0	0	0	0	0	0	0	0	0	0	0	1548
1931	0	0	0	0	0	0	0	0	0	0	0	0	0	2438
1932	0	0	0	0	0	0	0	0	0	0	0	0	0	3527
1933	0	0	0	0	0	0	0	0	0	0	0	0	0	4771
1934	0	0	0	0	0	1	0	0	0	0	0	0	61	5237
1935	0	0	0	0	0	0	0	0	0	0	0	0	1121	4818
1936	0	0	0	0	0	3	0	0	0	0	0	0	3307	3631
1937	0	0	0	0	0	559	0	0	0	0	0	0	5273	2048
1938	0	0	0	0	0	4887	0	0	0	0	0	0	7130	178
1939	0	0	0	0	0	11085	0	0	0	0	0	103	6820	0
1940	0	0	0	0	0	14960	0	0	0	0	0	1393	6309	0
1941	0	0	0	0	0	15868	0	0	0	0	0	3977	4391	0
1942	0	0	0	0	484	18128	0	0	0	0	0	6189	2168	0
1943	0	0	0	0	5571	14681	0	0	0	0	0	7872	216	0
1944	0	0	0	0	14573	8061	0	0	0	0	103	9241	0	0

1945	0	0	0	0	18660	1073	0	0	0	0	1904	7731	0	0
1946	0	0	0	0	24266	0	0	0	0	0	4992	5523	0	0
1947	0	0	0	570	26192	0	0	0	0	0	8798	3133	0	0
1948	0	0	0	5664	17154	0	0	0	0	0	12074	325	0	0
1949	0	0	0	12483	7699	0	0	0	0	179	14146	0	0	0
1950	0	0	0	17034	1101	0	0	0	0	2885	11126	0	0	0
1951	0	0	0	17191	0	0	0	0	0	7009	8002	0	0	0
1952	0	0	444	16349	0	0	0	0	0	12926	4486	0	0	0
1953	0	0	4460	11828	0	0	0	0	0	16543	449	0	0	0
1954	0	0	10330	5466	0	0	0	0	183	17756	0	0	0	0
1955	0	0	14403	769	0	0	0	0	3081	13463	0	0	0	0
1956	0	0	14794	0	0	0	0	0	7441	8746	0	0	0	0
1957	0	435	14205	0	0	0	0	0	12899	4546	0	0	0	0
1958	0	4023	10047	0	0	0	0	0	14113	539	0	0	0	0
1959	0	8892	4673	0	0	0	0	97	10686	0	0	0	0	0
1960	0	12686	720	0	0	0	0	1710	8744	0	0	0	0	0
1961	0	12628	0	0	0	0	0	3720	4794	0	0	0	0	0
1962	354	12302	0	0	0	0	0	13220	4601	0	0	0	0	0
1963	3315	8829	0	0	0	0	0	21096	737	0	0	0	0	0
1964	7974	4106	0	0	0	0	176	17261	0	0	0	0	0	0
1965	11124	625	0	0	0	0	2884	13419	0	0	0	0	0	0
1966	11304	0	0	0	0	0	6762	9057	0	0	0	0	0	0
1967	9163	0	0	0	0	0	10004	3456	0	0	0	0	0	0
1968	4757	0	0	0	0	0	15840	411	0	0	0	0	0	0
1969	2546	0	0	0	0	0	14179	0	0	0	0	0	0	0
1970	166	0	0	0	0	0	11651	0	0	0	0	0	0	0
1971	2	0	0	0	0	0	7571	0	0	0	0	0	0	0
1972	0	0	0	0	0	0	4702	0	0	0	0	0	0	0
1973	0	0	0	0	0	0	2062	0	0	0	0	0	0	0
1974	0	0	0	0	0	0	925	0	0	0	0	0	0	0
1975	0	0	0	0	0	0	355	0	0	0	0	0	0	0
1976	0	0	0	0	0	0	221	0	0	0	0	0	0	0

1977	0	0	0	0	0	0	132	0	0	0	0	0	0	0
1978	0	0	0	0	0	0	88	0	0	0	0	0	0	0
1979	0	0	0	0	0	0	56	0	0	0	0	0	0	0
1980	0	0	0	0	0	0	27	0	0	0	0	0	0	0
1981	0	0	0	0	0	0	52	0	0	0	0	0	0	0
1982	0	0	0	0	0	0	27	0	0	0	0	0	0	0
1983	0	0	0	0	0	0	26	0	0	0	0	0	0	0
1984	0	0	0	0	0	0	6	0	0	0	0	0	0	0
1985	0	0	0	0	0	0	7	0	0	0	0	0	0	0
1986	0	0	0	0	0	0	6	0	0	0	0	0	0	0
1987	0	0	0	0	0	0	4	0	0	0	0	0	0	0

Supplementary figures for chapter 4



Supplementary Figure 4.1 Distribution of sitting and leg height, shown separately for men and women in UKB and CKB.

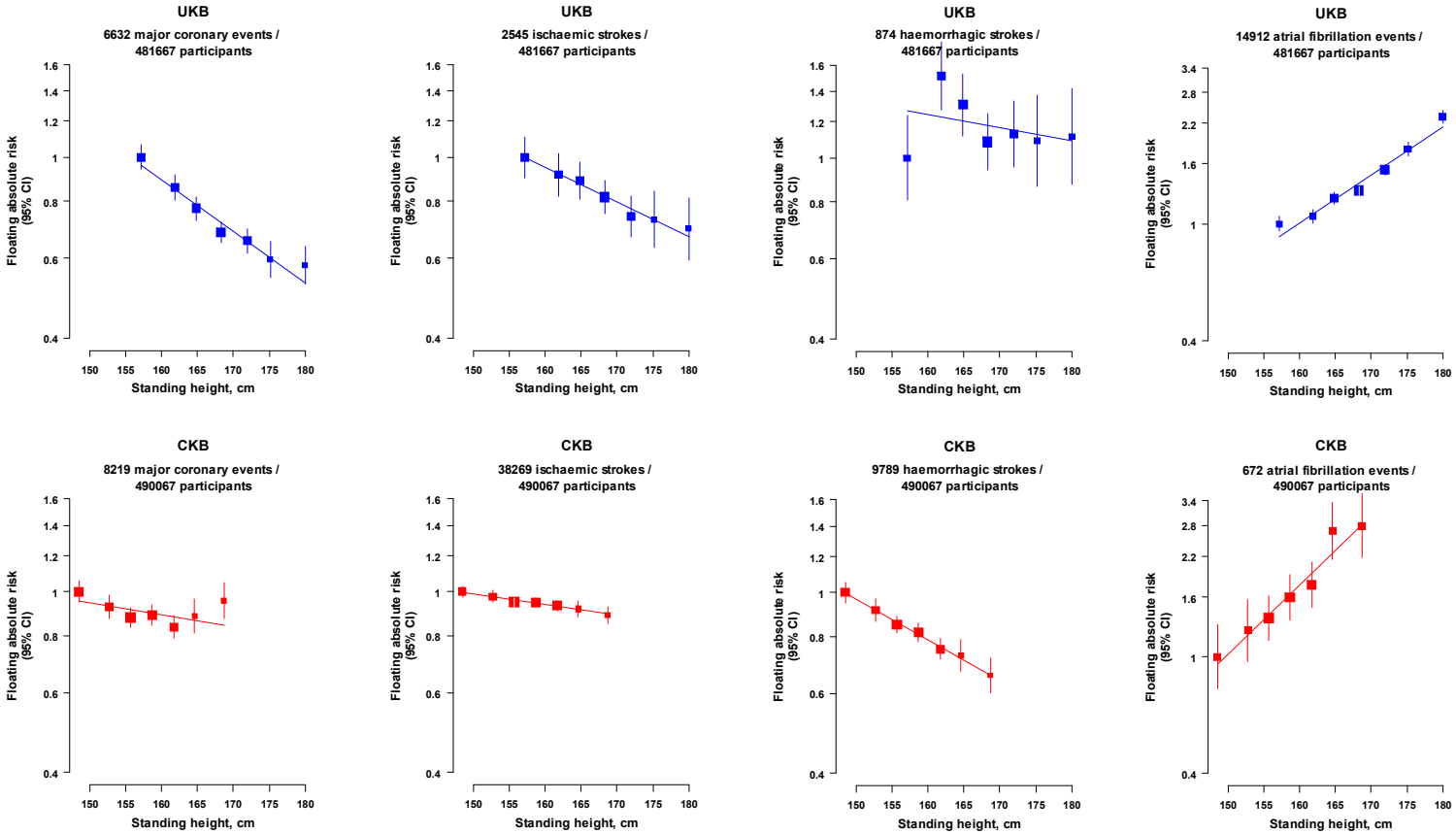
Supplementary tables for chapter 5

Supplementary Table 5.1 Pearson self-correlations between baseline and resurvey* for exposure variables and covariates in UKB and CKB, shown for resurvey participants and for non-hypertensive† participants at baseline.

	Self-correlation between baseline and resurvey* for resurvey participants		Self-correlation between baseline and resurvey* for non-hypertensive¶ resurvey participants	
Exposure variable or covariate	**UKB (n=36681) ~6 years of follow-up**	**CKB (n=29975) ~7 years of follow-up**	**UKB (n=28817) ~6 years of follow-up**	**CKB (n=27207) ~7 years of follow-up**
Standing height	0.99	0.98	0.99	0.98
Sitting height	0.88	0.85	0.88	0.85
Leg height	0.91	0.86	0.91	0.86
Smoking status	0.87	0.80	0.87	0.80
Cigarette equivalents	0.69	0.59	0.68	0.59
Systolic blood pressure	0.63	0.57	0.64	0.55
Diastolic blood pressure	0.59	0.54	0.60	0.54
Diagnosed hypertension	0.64	0.46	-	-
Diagnosed diabetes	0.74	0.59	0.69	0.58
Self-rated walking pace	0.59	-	0.57	-
Household income	0.59	0.55	0.59	0.55
Education	0.89	0.88	0.89	0.88
Waist-to-hip ratio	0.80	0.60	0.79	0.59
Body mass index	0.91	0.88	0.91	0.88
Taking blood pressure lowering medication	0.72	0.30	-	-

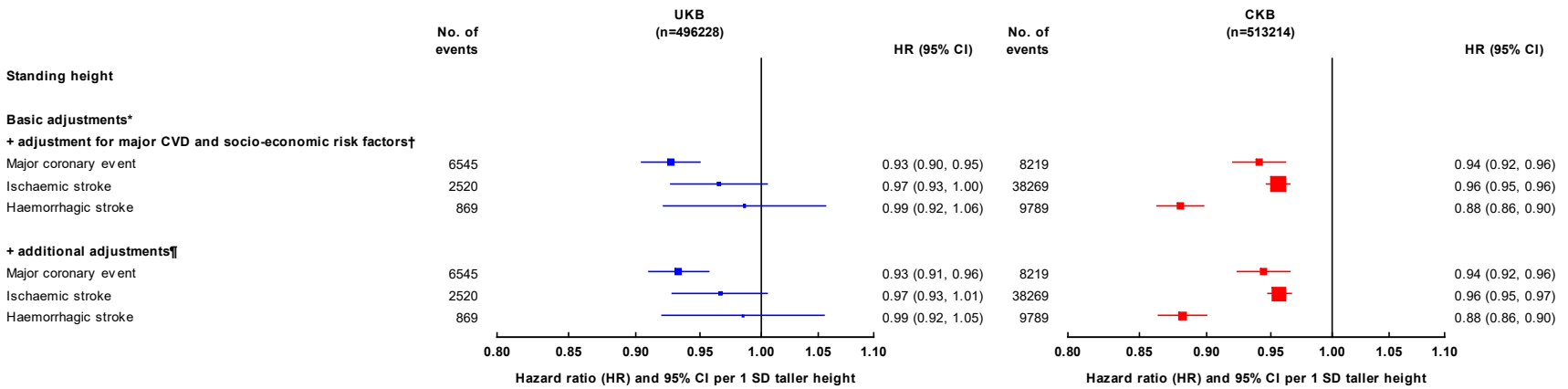
*Resurvey data for a participant was defined as either the first resurvey or the second resurvey. Resurvey participants with prior major cardiovascular disease at baseline were excluded. Where a participant was included in both resurveys, data from the second resurvey was used. Approximately 17% (UKB) and 50% (CKB) of participants in the second resurvey were also in the first resurvey.
†Non-hypertensive participants excluded resurvey participants diagnosed with hypertension at baseline and participants taking blood pressure lowering medications at baseline.

Supplementary figures for chapter 5



Supplementary Figure 5.1 The shape of the associations of height with cardiovascular diseases (major coronary events, stroke types, and atrial fibrillation), after basic adjustments.

Basic adjustments were age, sex, year of birth, and region in CKB.



Supplementary Figure 5.2 Associations of taller standing height with incident major coronary events and stroke types, shown for further levels of adjustment.

*Basic adjustments were age, sex, year of birth, and region in CKB.
†Major cardiovascular disease (CVD) and socio-economic risk factors were smoker status, cigarette equivalents (cigarettes/day), systolic blood pressure, diastolic blood pressure, diagnosed hypertension, diagnosed diabetes, self-rated walking pace (UKB only), household income, and education level.
¶Additional adjustments were self-rated health, physical activity (MET hours/ day), alcohol consumption, and taking blood pressure, cholesterol or diabetes medication.

Supplementary appendix for chapter 5

Supplementary appendix 5.1

This appendix describes the details of the adjustment variables, grouped into major cardiovascular disease risk factors, socio-economic risk factors, and additional risk factors.

Major cardiovascular disease risk factors included smoking status (UKB and CKB: never, ex-smoker, or current smoker), cigarette equivalents (cigarettes/day), systolic blood pressure, diastolic blood pressure, self-reported hypertension diagnosis, self-reported diabetes diagnosis, and self-rated walking pace (UKB only: slow pace, steady average pace, or brisk pace). In exploratory analyses, self-rated walking pace was identified as an important cardiovascular disease risk factor. The latter findings were consistent with a previous study that reported self-rated walking pace as a strong predictor of mortality in UKB.[139]

Socio-economic risk factors included household income (UKB: <£18 000, £18 000-£30 999, £31 000-£51 999, £52 000-£100 000, >£100 000; CKB: <¥2500, ¥2500-¥4999, ¥5000-¥9999, ¥10 000-¥19 999, ¥20 000-¥34 999, ≥¥35 000), and education level (UKB: education to degree level [yes/no]; CKB: no formal school, education to primary school level, middle/high school level, or technical school/college/university level).

Additional risk factors included self-rated health (UKB and CKB: excellent, good, fair, or poor), physical activity (MET hours/day), alcohol consumption (UKB:

alcohol units/week; CKB: never a regular drinker, ex-regular drinker, occasional drinker, monthly drinker, reduced intake drinker, or weekly drinker), and taking blood pressure, cholesterol or diabetes medication (UKB: taking blood pressure medication, cholesterol medication, or diabetes medication; CKB: taking blood pressure medication [calcium antagonists, ace inhibitors, beta blockers, or diuretics], cholesterol medication [statins], or diabetes medication [insulin, or metformin]). All covariates were measured at baseline. In exploratory analyses, the impact of individually adjusting for each of the additional risk factors was assessed to confirm that no single risk factor had a large effect on the associations of height with the cardiovascular diseases of interest.

Supplementary tables for chapter 7

Supplementary Table 7.1 Associations of genetically instrumented height with different blood pressure phenotypes in UKB and CKB.

	UKB (n=336433)				CKB (n=77419)			
Blood pressure (BP) phenotype	**SD (men/ women) mm Hg**	**Beta (SE) per 1 SD taller genetically instrumented height***	**Z-value for genetic associ-ation**	**Beta (SE) per 1 SD taller measured height†**	**SD (men/ women) mm Hg**	**Beta (SE) per 1 SD taller genetically instrumented height***	**Z-value for genetic associ-ation**	**Beta (SE) per 1 SD taller measured height†**
Medication adjusted systolic BP (mm Hg)¶	19/21	-1.39 (0.17)	-15.9	-1.46 (0.03)	20/23	-0.27 (0.24)	-1.1	0.41 (0.03)
Medication adjusted diastolic BP (mm Hg)¶	11/11	-0.61 (0.09)	-12.8	-0.58 (0.02)	12/11	-0.11 (0.13)	-0.9	0.49 (0.02)
Systolic BP (mm Hg)	17/19	-1.13 (0.15)	-14.3	-1.14 (0.02)	20/22	-0.29 (0.23)	-1.3	0.33 (0.03)
Diastolic BP (mm Hg)	10/10	-0.43 (0.08)	-10.2	-0.36 (0.01)	11/11	-0.13 (0.12)	-1.1	0.43 (0.02)
Systolic BP adjusted for BMI, age, and age squared (mm Hg)¶	18/19	-1.11 (0.16)	-14.0	-1.06 (0.03)	19/20	-0.08 (0.22)	-0.4	0.29 (0.03)
Diastolic BP adjusted for BMI, age, and age squared (mm Hg)¶	10/10	-0.35 (0.09)	-7.8	-0.23 (0.01)	11/11	0.03 (0.12)	0.2	0.43 (0.02)
Systolic BP adjusted for socio-economic risk factors, BMI, age, and age squared (mm Hg)¶‡	18/19	-1.07 (0.16)	-13.5	-0.98 (0.03)	19/20	-0.07 (0.22)	-0.3	0.46 (0.03)
Diastolic BP adjusted for socio-economic risk factors, BMI, age, and age squared (mm Hg)¶‡	10/10	-0.34 (0.09)	-7.7	-0.22 (0.01)	11/11	0.03 (0.12)	0.2	0.45 (0.02)

*Meta-analysed across sex. The UKB genetic risk score was constructed from Wood[20] SNPs and effect sizes, and the CKB genetic risk score was constructed from clumped GIANT+UKB[30] SNPs and effect sizes. The causal effects from instrumental variable analyses were presented per unit genetically instrumented height, where the unit was 1 SD of measured height (Table 3.1). The associations were adjusted for genomic principal components and genotyping array type.
†Associations based on the full UKB (n=481667) and CKB (n=490067) datasets after exclusions for prior cardiovascular diseases. Associations were meta-analysed across sex and adjusted for age, year of birth, and region in CKB.

¶Systolic blood pressure was increased by 15 mm Hg and diastolic blood pressure increased by 10 mm Hg for participants taking blood pressure lowering medications.
‡Socio-economic risk factors included household income (UKB: <£18000, £18000-£30999, £31000-£51999, £52000-£100000, >£100000; CKB: <¥2500, ¥2500-¥4999, ¥5000-¥9999, ¥10000-¥19999, ¥20000-¥34999, ≥¥35000), and education level (UKB: education to degree level [yes/no]; CKB: no formal school, education to primary school level, middle/high school level, or technical school/college/university level).

Supplementary Table 7.2 Associations of genetically instrumented height* with coronary artery disease, stroke types and atrial fibrillation in CKB before and after exclusion of related individuals.

	CKB including related individuals (77419 controls)		CKB excluding related individuals (56815 controls)	
Disease outcome†	No. of events	Odds ratio (95% CI) per 1 SD taller genetically instrumented height*	No. of events	Odds ratio (95% CI) per 1 SD taller genetically instrumented height*
Coronary artery disease	4087	1.00 (0.90, 1.10)	3283	0.96 (0.86, 1.07)
Ischaemic stroke	11283	0.95 (0.89, 1.01)	8999	0.97 (0.90, 1.05)
Haemorrhagic stroke	6372	0.94 (0.86, 1.02)	5035	0.90 (0.81, 0.99)
Atrial fibrillation	140	1.36 (0.82, 2.26)	111	0.94 (0.53, 1.68)

*The genetic risk score was constructed from clumped GIANT+UKB[30] SNPs and effect sizes. The causal effects from instrumental variable analyses were presented per unit genetically instrumented height, where the unit was 1 SD of measured height (Table 3.1). The associations were adjusted for principal components and array type.

†Coronary artery disease cases were prevalent at baseline or incident. Ischaemic stroke, haemorrhagic stroke, and atrial fibrillation cases were incident (only incident cases for these outcomes were available in CKB).

Supplementary Table 7.3 Associations of genetically instrumented height* with coronary artery disease, stroke types and atrial fibrillation before and after the inclusion of the height phenotype as a covariate in the model.

	UKB		CKB	
Disease outcome†	Odds ratio (95% CI) per 1 SD taller genetically instrumented height	Odds ratio (95% CI) per 1 SD taller genetically instrumented height, adjusted for the height phenotype	Odds ratio (95% CI) per 1 SD taller genetically instrumented height	Odds ratio (95% CI) per 1 SD taller genetically instrumented height, adjusted for the height phenotype
Coronary artery disease	0.88 (0.84, 0.92)	1.08 (1.03, 1.13)	1.00 (0.90, 1.10)	1.00 (0.90, 1.11)
Ischaemic stroke	0.99 (0.91, 1.08)	1.20 (1.09, 1.32)	0.95 (0.89, 1.01)	0.98 (0.92, 1.05)
Haemorrhagic stroke	0.84 (0.74, 0.96)	0.91 (0.79, 1.05)	0.94 (0.86, 1.02)	1.06 (0.97, 1.16)
Atrial fibrillation	1.32 (1.27, 1.38)	1.11 (1.06, 1.16)	1.36 (0.82, 2.26)	0.77 (0.45, 1.32)

*The UKB genetic risk score was constructed from Wood[20] SNPs and effect sizes, and the CKB genetic risk score was constructed from clumped GIANT+UKB[30] SNPs and effect sizes. The causal effects from instrumental variable analyses were presented per unit genetically instrumented height, where the unit was 1 SD of measured height (Table 3.1).
†Prevalent and incident cases were used, except for ischaemic stroke and haemorrhagic stroke in CKB where only incident cases were available.

Supplementary Table 7.4 List of height-associated SNPs from Wood et al.[20] showing heterogeneity with the common lean body mass effect or identified as outliers by the MR-PRESSO method.

RSID	Chromosome	Position	Nearby gene	Heterogeneous SNP (1=yes; 0=no)	MR-PRESSO outlier (1=yes; 0=no)
rs10779751	1	11206923	FRAP1	1	1
rs6600365	1	41328840	SCMH1	1	1
rs12855	1	51212681	CDKN2C	1	0
rs17391694	1	78396214	GIPC2	1	1
rs1409156	1	119293307	TBX15	1	1
rs7534365	1	148142748	SV2A	1	1
rs3767627	1	148205522	OTUD7B	0	1
rs4656220	1	168915901	PRRX1	1	1
rs1325596	1	175060689	PAPPA2	1	1
rs1935157	1	219383881	HLX	1	1
rs17038954	2	1624680	PXDN	1	1
rs780094	2	27594741	GCKR	1	1
rs11683207	2	97699722	ZAP70	1	1
rs994533	2	217992523	TNS1	1	1
rs2240919	3	52806741	ITIH3	1	1
rs9880211	3	137590239	STAG1	1	1
rs724016	3	142588260	ZBTB38	1	0
rs11714558	3	142645735	ZBTB38	1	1
rs509035	3	173646143	GHSR	1	1
rs7692995	4	17545732	LCORL	1	0
rs9993613	4	73694878	ADAMTS3	0	1
rs2167645	4	82403073	PRKG2	0	1
rs17499117	4	82423115	PRKG2	0	1
rs12639764	4	106435654	TET2	1	1
rs7654571	4	145540456	HHIP	0	1
rs4240326	4	146058714	ANAPC10	1	1
rs17574650	5	42472673	GHR	1	1
rs7727731	5	64710202	ADAMTS6	1	1
rs12519505	5	77541632	AP3B1	1	1
rs1582931	5	122685098	CCDC100	1	1
rs9392918	6	7653630	BMP6	1	0
rs6457374	6	31380240	HLA-C	1	1
rs3800461	6	34724300	C6orf106	1	1
rs16895130	6	42032909	CCND3	1	1
rs479744	6	109126725	FOXO3	1	1
rs6920372	6	109830632	PPIL6	0	1
rs1155939	6	126907826	C6orf173	1	1
rs1415701	6	130387528	L3MBTL3	1	1
rs7740107	6	130416154	L3MBTL3	1	1
rs486359	6	160694431	SLC22A3	1	1
rs42039	7	92082358	CDK6	1	0
rs6955948	7	150139653	TMEM176A	1	0
rs9650315	8	57318152	CHCHD7	1	1
rs4735677	8	78310746	PXMP3	1	1
rs2737220	8	116706860	TRPS1	1	1
rs1036821	8	135719665	ZFAT	1	0

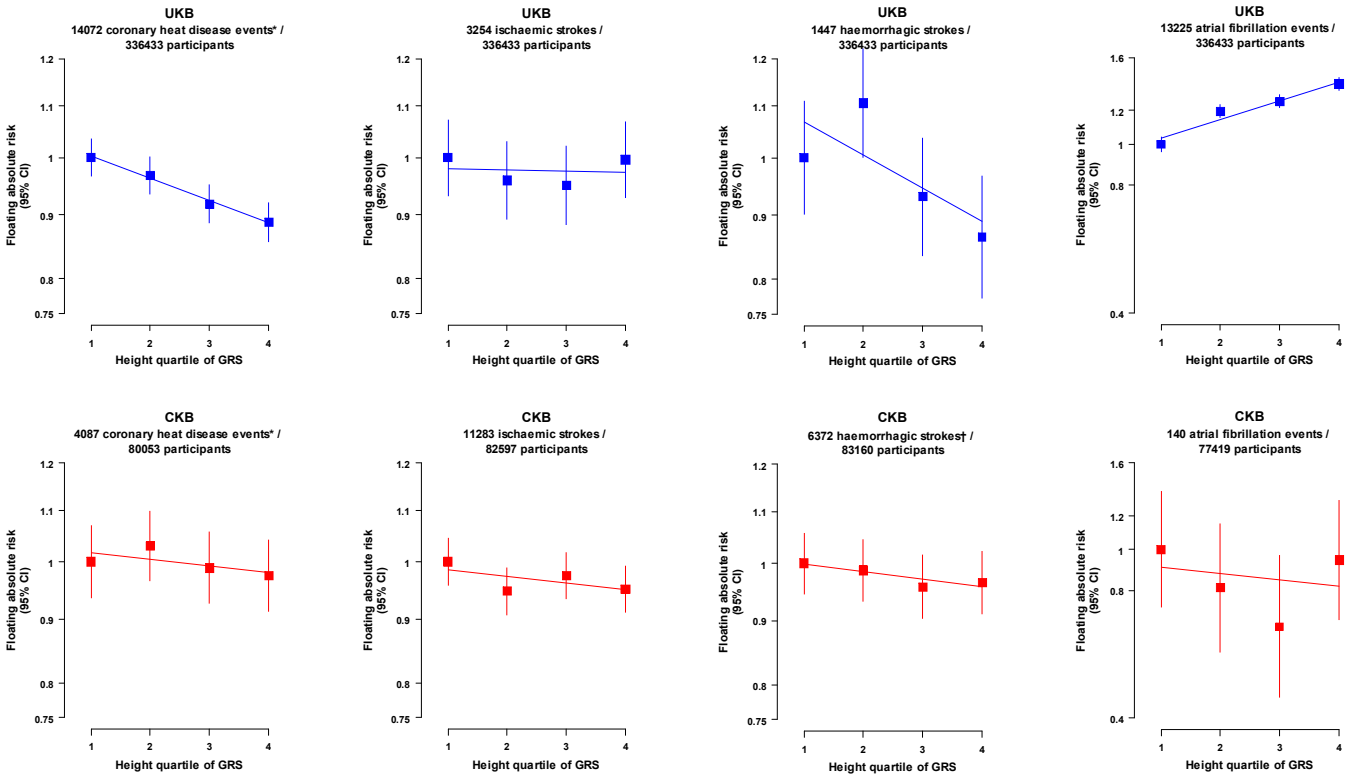
rs7033487	9	118169078	PAPPA	1	1
rs7849585	9	138251691	QSOX2	0	1
rs10883563	10	102674370	FAM178A	1	1
rs757081	11	17308259	NUCB2	1	1
rs2510396	11	68174228	GAL	1	1
rs10880969	12	45113290	SLC38A2	1	1
rs8756	12	64646019	HMGA2	1	1
rs3825199	12	92501085	SOCS2	1	1
rs12323101	13	32041406	PDS5B	1	1
rs3118905	13	50003335	DLEU7	1	1
rs3818416	13	77372469	EDNRB	1	1
rs2093210	14	60027032	C14orf39	1	1
rs2257011	15	82057149	SH3GL3	1	1
rs7162542	15	82305294	ADAMTSL3	1	1
rs3817428	15	87216251	ACAN	1	1
rs2573625	15	98330681	ADAMTS17	0	1
rs2531992	16	3961735	ADCY9	1	1
rs11642612	16	29937696	FLJ25404	1	1
rs4785393	16	48816984	PAPD5	1	1
rs8058684	16	52072619	RBL2	1	1
rs2070776	17	59361230	CD79B	1	1
rs4239020	17	77769930	CCDC57	0	1
rs4369779	18	18989406	CABLES1	1	1
rs11152213	18	56003928	MC4R	1	1
rs8102380	19	10662185	ILF3	1	0
rs7253628	19	35739109	ZNF536	0	1
rs6137287	20	21128259	C20orf19	1	1
rs1074683	20	31768314	PXMP4	1	0
rs143384	20	33489170	GDF5	1	1
rs2425163	20	33896084	PHF20	1	1

Supplementary Table 7.5 List of height-associated SNPs from Wood et al.[20] showing heterogeneity with the common systolic blood pressure effect or identified as outliers by the MR-PRESSO method.

RSID	Chromosome	Position	Nearby gene	Heterogeneous SNP (1=yes; 0=no)	MR-PRESSO outlier (1=yes; 0=no)
rs425277	1	2059032	PRKCZ	1	0
rs4601530	1	24916698	CLIC4	1	1
rs12119525	1	27376249	SLC9A1	0	1
rs16834765	1	32144029	PTP4A2	1	1
rs3014219	1	45797041	AKR1A1	1	0
rs7551732	1	88911629	PKN2	1	1
rs12120956	1	113004094	CAPZA1	1	1
rs6694089	1	170350504	DNM3	0	1
rs2421992	1	170507874	DNM3	0	1
rs1935157	1	219383881	HLX	1	1
rs7561273	2	24101018	LOC388931	1	1
rs6714546	2	33214929	LTBP1	0	1
rs9309101	2	43483116	THADA	1	1
rs3791679	2	55950396	EFEMP1	1	0
rs833152	2	182927346	PDE1A	1	0
rs12693589	2	191540907	STAT1	1	1
rs992157	2	218863025	PNKD/TMBIM1	1	0
rs3915129	3	41218746	CTNNB1	0	1
rs2581830	3	53109138	RFT1	1	0
rs9835332	3	56642722	C3orf63	0	1
rs9825951	3	100752611	COL8A1	1	0
rs4974480	3	135661252	ANAPC13	1	0
rs2302580	4	8659534	CPZ	1	0
rs7692995	4	17545732	LCORL	1	1
rs1996422	4	48382108	FRYL	1	1
rs12639764	4	106435654	TET2	1	0
rs7731703	5	32730699	NPR3	1	1
rs9292468	5	32854830	C5orf23	1	1
rs13183624	5	32856925	C5orf23	1	1
rs6594336	5	108100984	FER	1	0
rs9392918	6	7653630	BMP6	0	1
rs806794	6	26308656	HIST1H2BF	0	1
rs6457374	6	31380240	HLA-C	1	0
rs2857693	6	31696363	BAT2	1	1
rs12214804	6	34296844	HMGA1	0	1
rs479744	6	109126725	FOXO3	1	0
rs1405212	6	117597357	VGLL2	0	1
rs389663	6	117974744	DCBLD1	0	1
rs4895801	6	126258096	NCOA7	0	1
rs3020418	6	152386855	ESR1	0	1
rs2888877	7	92066336	CDK6	1	1
rs42039	7	92082358	CDK6	1	1
rs9650315	8	57318152	CHCHD7	1	0
rs16939034	8	76203138	CRISPLD1	1	0
rs1599473	8	120544539	NOV	1	1
rs894343	8	135681777	ZFAT	1	1

rs1036821	8	135719665	ZFAT	1	0
rs1571892	9	93298657	NFIL3	0	1
rs10766065	11	13234537	ARNTL	1	1
rs2915404	11	14361401	RRAS2	1	0
rs757081	11	17308259	NUCB2	1	1
rs10767838	11	30304503	C11orf46	1	1
rs8756	12	64646019	HMGA2	1	1
rs7980687	12	122388664	SBNO1	1	0
rs1199734	13	20468246	LATS2	1	1
rs3118905	13	50003335	DLEU7	1	0
rs7985356	13	114045564	CDC16	1	1
rs8017130	14	22828996	HOMEZ	1	0
rs2093210	14	60027032	C14orf39	0	1
rs1036477	15	46702218	FBN1	0	1
rs10744956	15	49056921	AP4E1	1	0
rs7162825	15	61226239	LACTB	1	1
rs8042424	15	99580062	CHSY1	1	0
rs2014467	16	2276395	ABCA3	1	1
rs960006	16	4851196	UBN1	1	1
rs11642612	16	29937696	FLJ25404	1	1
rs3790086	16	68445208	WWP2	1	0
rs11640018	16	73885809	CFDP1	0	1
rs3760318	17	26271841	CENTA2	0	1
rs4986172	17	40571807	ACBD4	1	1
rs199515	17	42211804	WNT3	1	1
rs6504389	17	43998363	HOXB3	1	1
rs318095	17	44329733	ATP5G1	1	1
rs2079795	17	56851431	C17orf82	1	1
rs1552173	17	74230437	PSCD1	1	1
rs2123731	19	4880473	UHRF1	1	1
rs1346490	19	7195233	INSR	1	1
rs7273787	20	4046567	SMOX	1	0
rs143384	20	33489170	GDF5	0	1
rs17450430	20	47205671	STAU1	1	1

Supplementary figures for chapter 7



Supplementary Figure 7.1 Associations of increasing quartiles of the height genetic risk score (GRS) with coronary artery disease, stroke types, and atrial fibrillation in UKB (n=336433) and CKB (77419 controls).

The UKB GRS was constructed from Wood[20] SNPs and effect sizes, and the CKB GRS was constructed from clumped GIANT+UKB[30] SNPs and effect sizes. Associations were adjusted for genomic principal components and genotyping array type.

*Coronary artery disease included incident or prevalent (UKB only) myocardial infarction, fatal ischemic heart disease cases, and incident coronary revascularisation cases.

†Prevalent haemorrhagic stroke cases were not available in CKB (only incident cases).

Acknowledgements

The author would like to thank his supervisors, Professor Sarah Parish and Professor Robert Clarke, for their continued support, attention to detail, and guidance during his DPhil journey. This research was supported by the Clarendon Fund, Brasenose College, and the Nuffield Department of Population Health, without which, this DPhil would not have been possible. This research included use of the UK Biobank resource under application 10 061.

References

1. Emerging Risk Factors Collaboration. Adult height and the risk of cause-specific death and vascular morbidity in 1 million people: individual participant meta-analysis. J Epidemiol 2012;41(5):1419–33.

2. Lee CMY, Barzi F, Woodward M, et al. Adult height and the risks of cardiovascular disease and major causes of death in the Asia-Pacific region: 21,000 deaths in 510,000 men and women. Int J Epidemiol 2009;38(4):1060–71.

3. Rosenberg MA, Patton KK, Sotoodehnia N, et al. The impact of height on the risk of atrial fibrillation: the Cardiovascular Health Study. Eur Heart J 2012;33(21):2709–17.

4. Nelson CP, Hamby SE, Saleheen D, et al. Genetically determined height and coronary artery disease. N Engl J Med 2015;372(17):1608–18.

5. Feigin VL, Krishnamurthi R. Stroke Prevention in the Developing World. Stroke 2011;42(12):3655–8.

6. Black R, Victora CG, Walker S, et al. Maternal and child undernutrition and overweight in low-income and middle-income countries. Lancet 2013;382(9890):427–451.

7. Ashton B, Hill K, Piazza A, Zeitz R. Famine in China, 1958-61. Popul Dev Rev 1984;10(4):613–645.

8. Chen Y, Zhou L-A. The long-term health and economic consequences of the 1959–1961 famine in China. J Health Econ 2007;26(4):659–81.

9. Gertler MM, Garn SM, White PD. Young candidates for coronary heart disease. J Am Med Assoc 1951;147(7):621–5.

10. Rosenberg MA, Kaplan RC, Siscovick DS, et al. Genetic variants related to height and risk of atrial fibrillation: the cardiovascular health study. J Epidemiol 2014;180(2):215–22.

11. Marmot MG, Shipley MJ, Rose G. Inequalities in death: Specific explanations of a general pattern? Lancet 1984;1(8384):1003–1006.

12. Batty GD, Shipley MJ, Gunnell D, et al. Height, wealth, and health: an overview with new data from three longitudinal studies. [Review] [160 refs]. Econ Hum Biol 2009;7(2):137–52.

13. Paajanen TA, Oksala NKJ, Kuukasjärvi P, Karhunen PJ, Oksala NKJ. Short stature is associated with coronary heart disease: A systematic review of the literature and a meta-analysis. Eur Heart J 2010;31(14):1802–1809.

14. Green J, Cairns BJ, Casabonne D, et al. Height and cancer incidence in the Million Women Study: prospective cohort, and meta-analysis of prospective studies of height and total cancer risk. Lancet Oncol 2011;12(8):785–94.

15. NCD Risk Factor Collaboration, Bentham J, Cesare MD, et al. A century of trends in adult human height [Internet]. eLife. 2016 [cited 2018 Sep 11];Available from: https://elifesciences.org/articles/13410

16. Case A, Paxson C. Stature and status: Height, ability, and labor market outcomes. J Polit Econ 2008;116(3):499–532.

17. Tyrrell J, Jones SE, Beaumont R, et al. Height, body mass index, and socioeconomic status: mendelian randomisation study in UK Biobank. BMJ 2016;352:i582.

18. Patel PC, Devaraj S. Height-income association in developing countries: Evidence from 14 countries. Am J Hum Biol Off J Hum Biol Counc 2018;30(3):e23093.

19. Pampel FC, Denney JT, Krueger PM. Obesity, SES, and economic development: a test of the reversal hypothesis. Soc Sci Med 1982 2012;74(7):1073–81.

20. Wood AR, Esko T, Yang J, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet 2014;46(11):1173–86.

21. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet 2015;47(11):1236–41.

22. Burgess S, Butterworth A, Malarstig A, Thompson SG. Use of Mendelian randomisation to assess potential benefit of clinical intervention. BMJ 2012;345:e7325.

23. Bennett DA, Holmes MV. Mendelian randomisation in cardiovascular research: an introduction for clinicians. Heart 2017;103(18):1400–7.

24. Jansen H, Samani NJ, Schunkert H. Mendelian randomization studies in coronary artery disease. Eur Heart J 2014;35(29):1917–24.

25. Kathiresan S, Srivastava D. Genetics of Human Cardiovascular Disease. Cell 2012;148(6):1242–57.

26. Silventoinen K, Kaprio J, Koskenvuo M, Lahelma E. The association between body height and coronary heart disease among Finnish twins and singletons. J Epidemiol 2003;32(1):78–82.

27. Silventoinen K, Zdravkovic S, Skytthe A, et al. Association between height and coronary heart disease mortality: a prospective study of 35,000 twin pairs. J Epidemiol 2006;163(7):615–21.

28. Macgregor S, Cornes BK, Martin NG, Visscher PM. Bias, precision and heritability of self-reported and clinically measured height in Australian twins. Hum Genet 2006;120(4):571–80.

29. Schousboe K, Visscher PM, Erbas B, et al. Twin study of genetic and environmental influences on adult body size, shape, and composition. Int J Obes 2004;28(1):39–48.

30. Yengo L, Sidorenko J, Kemper KE, et al. Meta-analysis of genome-wide association studies for height and body mass index in ∼700000 individuals of European ancestry. Hum Mol Genet 2018;

31. Lu X, Peloso GM, Liu DJ, et al. Exome chip meta-analysis identifies novel loci and East Asian–specific coding variants that contribute to lipid levels and coronary artery disease. Nat Genet 2017;49(12):1722–30.

32. Nüesch E, Dale C, Palmer TM, et al. Adult height, coronary heart disease and stroke: a multi-locus Mendelian randomization meta-analysis. Int J Epidemiol 2016;45(6):1927–37.

33. Emerging Risk Factors Collaboration. Lipid-Related Markers and Cardiovascular Disease Prediction. JAMA 2012;307(23):2499–506.

34. Evangelou E, Warren HR, Mosen-Ansorena D, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. Nat Genet 2018;50(10):1412–25.

35. China Kadoorie Biobank [Internet]. [cited 2018 Jul 18];Available from: http://www.ckbiobank.org/site/

36. UK Biobank [Internet]. [cited 2018 Jul 12];Available from: https://www.ukbiobank.ac.uk/

37. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 2018;562(7726):203.

38. Chen Z, Chen J, Collins R, et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. Int J Epidemiol 2011;40(6):1652–1666.

39. Lai FY, Nath M, Hamby SE, Thompson JR, Nelson CP, Samani NJ. Adult height and risk of 50 diseases: a combined epidemiological and genetic analysis. BMC Med [Internet] 2018 [cited 2018 Nov 12];16. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6201543/

40. Suzuki S, Yamashita T, Ohtsuka T, et al. Body size and atrial fibrillation in Japanese outpatients. Circ J Off J Jpn Circ Soc 2010;74(1):66–70.

41. Song Y-M, Sung J. Adult height and the risk of mortality in South Korean women. Am J Epidemiol 2008;168(5):497–505.

42. Marouli E, Graff M, Medina-Gomez C, et al. Rare and low-frequency coding variants alter human adult height. 2016 [cited 2018 May 29];Available from: https://dash.harvard.edu/handle/1/34375309

43. Yengo L, Sidorenko J, Kemper KE, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700,000 individuals of European ancestry. bioRxiv 2018;274654.

44. Asao K, Kao WHL, Baptiste-Roberts K, Bandeen-Roche K, Erlinger TP, Brancati FL. Short Stature and the Risk of Adiposity, Insulin Resistance, and Type 2 Diabetes in Middle Age: The Third National Health and Nutrition Examination Survey (NHANES III), 1988–1994. Diabetes Care 2006;29(7):1632–7.

45. Lawlor DA, Ebrahim S, Davey Smith G. The association between components of adult height and Type II diabetes and insulin resistance: British Women's Heart and Health Study. Diabetologia 2002;45(8):1097–106.

46. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009;339:b2700.

47. Sawada N, Wark PA, Merritt MA, et al. The association between adult attained height and sitting height with mortality in the European Prospective Investigation into Cancer and Nutrition (EPIC). PLOS ONE 2017;12(3):e0173117.

48. Weitzman S, Wang C-H, Pankow JS, Schmidt MI, Brancati FL. Are measures of height and leg length related to incident diabetes mellitus? The ARIC (Atherosclerosis Risk in Communities) study. Acta Diabetol 2010;47(3):237–42.

49. Sung J, Song Y-M, Lawlor DA, Smith GD, Ebrahim S. Height and site-specific cancer risk: A cohort study of a korean adult population. Am J Epidemiol 2009;170(1):53–64.

50. Song Y-M, Smith GD, Sung J. Adult height and cause-specific mortality: a large prospective study of South Korean men. Am J Epidemiol 2003;158(5):479–85.

51. van Iperen EPA, Sivapalaratnam S, Holmes MV, Hovingh GK, Zwinderman AH, Asselbergs FW. Genetic analysis of emerging risk factors in coronary artery disease. Atherosclerosis 2016;254:35–41.

52. Davies NM, Gaunt TR, Lewis SJ, et al. The effects of height and BMI on prostate cancer incidence and mortality: a Mendelian randomization study in 20,848 cases and 20,214 controls from the PRACTICAL consortium. Cancer Causes Control CCC 2015;26(11):1603–16.

53. Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature 2010;466(7307):707–13.

54. International Consortium for Blood Pressure Genome-Wide Association Studies, Ehret GB, Munroe PB, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature 2011;478(7367):103–9.

55. Zhang B, Shu X-O, Delahanty RJ, et al. Height and Breast Cancer Risk: Evidence From Prospective Studies and Mendelian Randomization. J Natl Cancer Inst 2015;107(11).

56. Khankari NK, Shu X-O, Wen W, et al. Association between Adult Height and Risk of Colorectal, Lung, and Prostate Cancer: Results from Meta-analyses of Prospective Studies and Mendelian Randomization Analyses. PLoS Med 2016;13(9):e1002118.

57. Thrift AP, Gong J, Peters U, et al. Mendelian randomization study of height and risk of colorectal cancer. Int J Epidemiol 2015;44(2):662–72.

58. Thrift AP, Risch HA, Onstad L, et al. Risk of esophageal adenocarcinoma decreases with height, based on consortium analysis and confirmed by Mendelian randomization. Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc 2014;12(10):1667-1676.e1.

59. He M, Xu M, Zhang B, et al. Meta-analysis of genome-wide association studies of adult height in East Asians identifies 17 novel loci. Hum Mol Genet 2015;24(6):1791–800.

60. Hao Y, Liu X, Lu X, et al. Genome-wide association study in Han Chinese identifies three novel loci for human height. Hum Genet 2013;132(6):681–9.

61. Okada Y, Kamatani Y, Takahashi A, et al. A genome-wide association study in 19 633 Japanese subjects identified LHX3-QSOX2 and IGF1 as adult height loci. Hum Mol Genet 2010;19(11):2303–12.

62. Lango Allen H, Estrada K, Lettre G, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 2010;467(7317):832–8.

63. Wang N, Zhang X, Xiang Y-B, et al. Associations of adult height and its components with mortality: a report from cohort studies of 135,000 Chinese women and men. J Epidemiol 2011;40(6):1715–26.

64. Fagherazzi G, Vilier A, Boutron-Ruault M-C, Clavel-Chapelon F, Mesrine S. Height, sitting height, and leg length in relation with breast cancer risk in the E3N cohort. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol 2012;21(7):1171–5.

65. Zuccolo L, Harris R, Gunnell D, et al. Height and prostate cancer risk: a large nested case-control study (ProtecT) and meta-analysis. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol 2008;17(9):2325–36.

66. Mellemkjær L, Christensen J, Frederiksen K, et al. Leg length, sitting height and postmenopausal breast cancer risk. Br J Cancer 2012;107(1):165–8.

67. Rinaldi S, Lise M, Clavel-Chapelon F, et al. Body size and risk of differentiated thyroid carcinomas: findings from the EPIC study. Int J Cancer 2012;131(6):E1004-1014.

68. Gunnell D, Okasha M, Smith GD, Oliver SE, Sandhu J, Holly JM. Height, leg length, and cancer risk: a systematic review. Epidemiol Rev 2001;23(2):313–42.

69. Honjo K, Iso H, Inoue M, Tsugane S. Adult height and the risk of cardiovascular disease among middle aged men and women in Japan. J Epidemiol 2011;26(1):13–21.

70. Shen C, Zheng X, Gao J, et al. A rare variant in COL11A1 is strongly associated with adult height in Chinese Han population. J Genet Genomics Yi Chuan Xue Bao 2016;43(9):549–54.

71. Kim J-J, Lee H-I, Park T, et al. Identification of 15 loci influencing height in a Korean population. J Hum Genet 2010;55(1):27–31.

72. Hong X, Tsai H-J, Liu X, et al. A large-scale genome-wide linkage analysis to map loci linked to stature in Chinese population. J Clin Endocrinol Metab 2008;93(11):4511–8.

73. Takeuchi F, Nabika T, Isono M, et al. Evaluation of genetic loci influencing adult height in the Japanese population. J Hum Genet 2009;54(12):749–52.

74. Lei S-F, Yang T-L, Tan L-J, et al. Genome-wide association scan for stature in Chinese: evidence for ethnic specific loci. Hum Genet 2009;125(1):1–9.

75. UK Biobank Resources [Internet]. [cited 2018 Aug 28];Available from: http://www.ukbiobank.ac.uk/resources/

76. Standard Operating Procedures [Internet]. [cited 2018 Aug 28];Available from: http://www.ckbiobank.org/site/Study+Resources/SOPs

77. UKB Resource 460 [Internet]. [cited 2018 Aug 7];Available from: https://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=460

78. Easton DF, Peto J, Babiker AG a. G. Floating absolute risk: An alternative to relative risk in survival and case-control analysis avoiding an arbitrary reference group. Stat Med 1991;10(7):1025–35.

79. Fenger-Grøn M, Overvad K, Tjønneland A, Frost L. Lean Body Mass Is the Predominant Anthropometric Risk Factor for Atrial Fibrillation. J Am Coll Cardiol 2017;69(20):2488–97.

80. Kamel H, Healey JS. Cardioembolic Stroke. Circ Res 2017;120(3):514–26.

81. Frost C, White IR. The effect of measurement error in risk factors that change over time in cohort studies: do simple methods overcorrect for 'regression dilution'? Int J Epidemiol 2005;34(6):1359–68.

82. Yang J, Zeng J, Goddard ME, Wray NR, Visscher PM. Concepts, estimation and interpretation of SNP-based heritability. Nat Genet 2017;49(9):1304.

83. Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM. Pitfalls of predicting complex traits from SNPs. Nat Rev Genet 2013;14(7):507–15.

84. Kato N, Loh M, Takeuchi F, et al. Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation. Nat Genet 2015;47(11):1282–93.

85. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature 2015;518(7538):197–206.

86. Nikpay M, Goel A, Won H-H, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet 2015;47(10):1121–30.

87. DIAbetes Genetics Replication And Meta-analysis Consortium, Asian Genetic Epidemiology Network Type 2 Diabetes Consortium, South Asian Type 2 Diabetes Consortium, et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat Genet 2014;46(3):234.

88. Willer CJ, Schmidt EM, Sengupta S, et al. Discovery and refinement of loci associated with lipid levels. Nat Genet 2013;45(11):1274–83.

89. Morris AP, Voight BF, Teslovich TM, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet 2012;44(9):981–90.

90. Ehret GB, Ferreira T, Chasman DI, et al. The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. Nat Genet [Internet] 2016 [cited 2018 Aug 30];Available from: https://dash.harvard.edu/handle/1/32072211

91. Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. Stat Med 2008;27(8):1133–1163.

92. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. Genet Epidemiol 2016;40(7):597–608.

93. Palmer TM, Lawlor DA, Harbord RM, et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. Stat Methods Med Res 2012;21(3):223–42.

94. Burgess S, Butterworth A, Thompson SG. Mendelian Randomization Analysis With Multiple Genetic Variants Using Summarized Data. Genet Epidemiol 2013;37(7):658–65.

95. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol 2015;44(2):512–25.

96. Burgess S, Thompson SG. Avoiding bias from weak instruments in Mendelian randomization studies. Int J Epidemiol 2011;40(3):755–64.

97. Lawlor DA. Commentary: Two-sample Mendelian randomization: opportunities and challenges. Int J Epidemiol 2016;45(3):908–15.

98. Commentary: Curses—Winner's and Otherwise—in Genetic Epidemiology on JSTOR [Internet]. [cited 2018 Jul 12];Available from: https://ezproxy-prd.bodleian.ox.ac.uk:4228/stable/25662608

99. Burgess S, Thompson SG. Use of allele scores as instrumental variables for Mendelian randomization. Int J Epidemiol 2013;42(4):1134–44.

100. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. Stat Methods Med Res 2017;26(5):2333–55.

101. Traylor M, Farrall M, Holliday EG, et al. Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies. Lancet Neurol 2012;11(11):951–62.

102. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38(8):904–9.

103. GIANT consortium [Internet]. [cited 2018 Sep 24];Available from: https://portals.broadinstitute.org/collaboration/giant/index.php/Main_Page

104. Burgess S. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. Int J Epidemiol 2014;43(3):922–9.

105. Lin P, Hartz SM, Zhang Z, et al. A New Statistic to Evaluate Imputation Reliability. PLOS ONE 2010;5(3):e9697.

106. Palmer TM, Sterne JAC, Harbord RM, et al. Instrumental Variable Estimation of Causal Risk Ratios and Causal Odds Ratios in Mendelian Randomization Analyses. Am J Epidemiol 2011;173(12):1392–403.

107. Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. Stat Med 2016;35(11):1880–906.

108. Teumer A. Common Methods for Performing Mendelian Randomization. Front Cardiovasc Med [Internet] 2018 [cited 2018 Nov 27];5. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5985452/

109. Verbanck M, Chen C-Y, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet 2018;50(5):693–8.

110. Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants. Epidemiol Camb Mass 2017;28(1):30–42.

111. Hoffmann TJ, Ehret GB, Nandakumar P, et al. Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. Nat Genet 2017;49(1):54–64.

112. Chugh SS, Roth GA, Gillum RF, Mensah GA. Global burden of atrial fibrillation in developed and developing nations. Glob Heart 2014;9(1):113–9.

113. Chatterjee Neal A., Giulianini Franco, Geelhoed Bastiaan, et al. Genetic Obesity and the Risk of Atrial Fibrillation. Circulation 2017;135(8):741–54.

114. Zillikens MC, Demissie S, Hsu Y-H, et al. Large meta-analysis of genome-wide association studies identifies five loci for lean body mass. Nat Commun 2017;8(1):80.

115. The Emerging Risk Factors Collaboration. Major Lipids, Apolipoproteins, and Risk of Vascular Disease. JAMA 2009;302(18):1993–2000.

116. Prospective Studies Collaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. The Lancet 2002;360(9349):1903–13.

117. Clarke R, Li L, Bennett D, et al. Abstract 16532: Impact of Systolic Blood Pressure on Cardiovascular Disease in a Chinese Population: A Mendelian Randomization Study. Circulation 2014;130(Suppl 2):A16532–A16532.

118. Sin DD, Wu L, Man SFP. The Relationship Between Reduced Lung Function and Cardiovascular Mortality: A Population-Based Study and a Systematic Review of the Literature. Chest 2005;127(6):1952–9.

119. Au Yeung Shiu Lun, Borges Maria-Carolina, Lawlor Debbie A. Association of Genetic Instrumental Variables for Lung Function on Coronary Artery Disease Risk. Circ Genomic Precis Med 2018;11(4):e001952.

120. Palatini P. Heart rate and the cardiometabolic risk. Curr Hypertens Rep 2013;15(3):253–9.

121. Hartwig FP, Davies NM, Davey Smith G. Bias in Mendelian randomization due to assortative mating. Genet Epidemiol 2018;42(7):608–20.

122. Tenesa A, Rawlik K, Navarro P, Canela-Xandri O. Genetic determination of height-mediated mate choice. Genome Biol [Internet] 2015 [cited 2018 Nov 18];16. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4717574/

123. Keller MC, Garver-Apgar CE, Wright MJ, et al. The Genetic Correlation between Height and IQ: Shared Genes or Assortative Mating? (Genetic Correlation between Height and IQ). 2013;9(4):e1003451.

124. Kong A, Thorleifsson G, Frigge ML, et al. The nature of nurture: Effects of parental genotypes. Science 2018;359(6374):424–8.

125. Hindy G, Engström G, Larsson SC, et al. Role of Blood Lipids in the Development of Ischemic Stroke and its Subtypes. Stroke [Internet] 2018 [cited 2018 Sep 7];Available from: https://www.ahajournals.org/doi/10.1161/STROKEAHA.117.019653?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%3dpubmed

126. An SJ, Kim TJ, Yoon B-W. Epidemiology, Risk Factors, and Clinical Features of Intracerebral Hemorrhage: An Update. J Stroke 2017;19(1):3–10.

127. Ma Y, Li Z, Chen L, Li X. Blood lipid levels, statin therapy and the risk of intracerebral hemorrhage. Lipids Health Dis [Internet] 2016 [cited 2018 Nov 16];15. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4774119/

128. Wang X, Dong Y, Qi X, Huang C, Hou L. Cholesterol levels and risk of hemorrhagic stroke: a systematic review and meta-analysis. Stroke 2013;44(7):1833–9.

129. Burgess S, Butterworth AS, Thompson JR. Beyond Mendelian randomization: how to interpret evidence of shared genetic predictors. J Clin Epidemiol 2016;69:208–16.

130. Bikdeli B, Abou Ziki MD, Lip GYH. Pulmonary Embolism and Atrial Fibrillation: Two Sides of the Same Coin? A Systematic Review. Semin Thromb Hemost 2017;43(8):849–63.

131. Mägi R, Horikoshi M, Sofer T, et al. Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power

for discovery and improves fine-mapping resolution. Hum Mol Genet 2017;26(18):3639–50.

132. Speliotes EK, Willer CJ, Berndt SI, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet 2010;42(11):937–48.

133. Wain LV, Verwoert GC, O'Reilly PF, et al. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. Nat Genet 2011;43(10):1005–11.

134. Manning AK, Hivert M-F, Scott RA, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat Genet 2012;44(6):659–69.

135. Liu Y-P, Richart T, Li Y, Zhan W-W, Staessen JA. Is arterial stiffness related to body height? Hypertension 2010;55(6).

136. Rietveld CA, Medland SE, Derringer J, et al. GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment. Science 2013;340(6139):1467–71.

137. UKB Showcase [Internet]. [cited 2018 Aug 7];Available from: http://biobank.ctsu.ox.ac.uk/crystal/

138. Data Access Procedures [Internet]. [cited 2018 Aug 8];Available from: https://www.ckbiobank.org/site/Data+Access/Access+Procedures

139. Ganna A, Ingelsson E. 5 year mortality predictors in 498 103 UK Biobank participants: a prospective population-based study. The Lancet 2015;386(9993):533–40.