Applications of genetic data to identify cardiovascular disease mechanisms and therapeutic opportunities

Dipender Gill

Thesis submitted for the degree of Doctor of Philosophy at Imperial College London

Department of Epidemiology and Biostatistics

School of Public Health

Imperial College London

December 2019

Copyright declaration

"Chess holds its master in its own bonds, shackling the mind and brain so that the inner freedom of the very strongest must suffer."

- Albert Einstein (1879 – 1955)

Abstract

Recent years have offered a wealth of genetic association data, with a concurrent explosion in the availability of methods for exploring causal effects through randomly allocated genetic variants that serve as proxies for traits of interest. This thesis investigates the state of this field within the remit of cardiovascular disease. Following an introduction into cardiovascular disease and Mendelian randomization (MR), the research focuses on dietary, social and pharmacological exposures as demonstrative examples for highlighting the breadth of techniques that can be harnessed towards understanding underlying mechanisms and therapeutic opportunities. Both two-sample and one-sample MR analyses are performed, using genetic summary data from large-scale consortia and the UK Biobank. Sample sizes for individual analyses typically exceed tens of thousands of participants. A diverse array of MR methods are employed, appropriate to the setting and objective of each analysis. Considering systemic iron status as a diet-related trait, genetic instruments are identified with consequent MR analyses supporting a protective effect on risk of cardiovascular outcomes related to atherosclerosis but a detrimental effect on outcomes related to thrombosis arising from stasis of blood. Phenome-wide association study further highlights effects of systemic iron status outside the remit of cardiovascular disease. In the investigation of social factors, MR mediation analysis techniques are applied to identify the pathways by which education affects cardiovascular disease risk, with multivariable MR further used to disentangle the direct effects of education and intelligence respectively. In the investigation of pharmacological exposures, genetic instruments for antihypertensive drugs are identified and validated by comparing against corresponding estimates from clinical trials. Phenome-wide association study is used to identify possible side-effects and repurposing opportunities, with a potential detrimental effect of calcium channel blockers identified on risk of diverticulosis. The final section provides an overview of the current state of applied MR, as well as future perspectives.

Declaration of originality

I, Dipender Gill, confirm that this thesis represents my own work. Where any component of it has been obtained from other sources, this is duly indicated as such.

Acknowledgements

My first acknowledgement is that I have been very lucky. I have been offered tremendous opportunities, both in science and life more generally. Any success that I have achieved is very much a function of my situation, which to some large degree, has been out of my control.

I would like to offer my heartfelt thanks and appreciation to my PhD supervisors, Dr Abbas Dehghan and Dr Joanna Tzoulaki. They have shown great patience and kindness in supervising me, taking the time to get to know me, understand me, and put effort into facilitating my unique operating style. Through the freedom and encouragement that they have offered, I have been able to follow my curiosities. I am indebted to them for teaching me how to practically conduct high quality science in the modern era. We still have much exciting research to do together, and I look forward to continuing working with them as collaborators and friends.

I am grateful to the Clinical Academic Training Office and the Wellcome 4i Clinical PhD Programme at Imperial College London for supporting my academic training. In particular, Prof. Matthew Pickering and Prof. Jeremy Levy have been mentors to me.

My wife, Mrs Rubinder Kaur Gill, is my life partner and has shared my PhD experience with me. Whatever work and sacrifice I have put into this effort, hers is still greater. Any reward and recognition should be similarly awarded. In many ways, she has become as knowledgeable in this field as I have!

My parents, Mr Satbir Singh Gill and Mrs Punam Gill, have done everything within their power to optimise my education, from as far back as I can remember. Their whole lives seem to have focused on this goal. It should be appropriate that scientific curiosity is my greatest pleasure, and also my greatest vice. Through my PhD, my parents supported me despite very difficult personal circumstances. They are the strongest characters I know, and are also the inspiration for any strength that I have learnt.

My younger brother, Mr Harvinder Gill, is mischievous.

An integral component of my PhD has been my work with collaborators, who have been a source of education, data and fun. I would like to thank each and every one of you – you will know who you are!

Funding

This work was supported by the Wellcome 4i Clinical PhD Programme at Imperial College London.

Conflict of interest

I attended a cardiovascular research workshop from 27 – 30 June 2019 in Boston, USA, which was organised and paid for by Bayer AG. On 28 August 2019, I visited Wuppertal, Germany to deliver a lecture and attend a series of meetings at the Biomedical Data Sciences Division of Bayer AG. These activities did not in any way influence the contents of this PhD or its related publications. I have no other conflicts of interest to declare.

Related academic work

Publications

1. **Gill D**, Georgakis MK, Koskeridis F, Jiang L, Feng Q, Wei WQ, Theodoratou E, Elliott P, Denny JC, Malik R, Evangelou E, Dehghan A, Dichgans M and Tzoulaki I. Use of Genetic Variants Related to Antihypertensive Drugs to Inform on Efficacy and Side Effects. Circulation. 2019;140:270-279.
2. **Gill D**, Benyamin B, Moore LSP, Monori G, Zhou A, Koskeridis F, Evangelou E, Laffan M, Walker AP, Tsilidis KK, Dehghan A, Elliott P, Hypponen E and Tzoulaki I. Associations of genetically determined iron status across the phenome: A mendelian randomization study. PLoS Med. 2019;16:e1002833.
3. Carter AR, **Gill D**, Davies NM, Taylor AE, Tillmann T, Vaucher J, Wootton RE, Munafo MR, Hemani G, Malik R, Seshadri S, Woo D, Burgess S, Davey Smith G, Holmes MV, Tzoulaki I, Howe LD and Dehghan A. Understanding the consequences of education inequality on cardiovascular disease: mendelian randomisation study. BMJ. 2019;365:l1855.
4. **Gill D**, Efstathiadou A, Cawood K, Tzoulaki I and Dehghan A. Education protects against coronary heart disease and stroke independently of cognitive 5 function: evidence from Mendelian randomization. Int J Epidemiol. 2019. doi: 10.1093/ije/dyz200 [Epub ahead of print].
5. **Gill D**, Brewer CF, Monori G, Tregouet DA, Franceschini N, Giambartolomei C, Consortium I, Tzoulaki I and Dehghan A. Effects of Genetically Determined Iron Status on Risk of Venous Thromboembolism and Carotid Atherosclerotic Disease: A Mendelian Randomization Study. J Am Heart Assoc. 2019;8:e012994.
6. **Gill D**, Monori G, Tzoulaki I and Dehghan A. Iron Status and Risk of Stroke: A Mendelian Randomization Study. Stroke. 2018;49:2815-2821.
7. **Gill D**, Del Greco M F, Walker AP, Srai SKS, Laffan MA and Minelli C. The effect of iron status on risk of coronary artery disease: a mendelian randomization study. Arterioscler Thromb Vasc Biol. 2017;37:1788-1792.

Publications separate to the PhD thesis

8. **Gill D**, James NE, Monori G, Lorentzen E, Fernandez-Cadenas I, Lemmens R, Thijs V, Rost NS, Scott R, Hankey GJ, Lindgren A, Jern C, Maguire JM, International Stroke Genetics Consortium and the GISCOME Network. Genetically Determined Risk of Depression and Functional Outcome After Ischemic Stroke. Stroke. 2019;50:2219-2222.

9. **Gill D**, Georgakis MK, Laffan M, Sabater-Lleal M, Malik R, Tzoulaki I, Veltkamp R and Dehghan A. Genetically Determined FXI (Factor XI) Levels and Risk of Stroke. Stroke. 2018;49:2761-2763.
10. **Gill D**, Monori G, Georgakis MK, Tzoulaki I and Laffan M. Genetically Determined Platelet Count and Risk of Cardiovascular Disease: Mendelian Randomization Study. Arterioscler Thromb Vasc Biol. 2018;38:2862-2869.
11. **Gill D**, Brewer CF, Del Greco MF, Sivakumaran P, Bowden J, Sheehan NA and Minelli C. Age at menarche and adult body mass index: a Mendelian randomization study. Int J Obes (Lond). 2018;42:1574-1581.
12. Georgakis MK, **Gill D**, Rannikmae K, Traylor M, Anderson CD, Lee JM, Kamatani Y, Hopewell JC, Worrall BB, Bernhagen J, Sudlow CLM, Malik R and Dichgans M. Genetically Determined Levels of Circulating Cytokines and Risk of Stroke. Circulation. 2019;139:256-268.
13. Efstathiadou A, **Gill D**, McGrane F, Quinn T and Dawson J. Genetically Determined Uric Acid and the Risk of Cardiovascular and Neurovascular Diseases: A Mendelian Randomization Study of Outcomes Investigated in Randomized Trials. J Am Heart Assoc. 2019;8:e012738.
14. Mahbubani K, Georgiades F, Goh EL, Chidambaram S, Sivakumaran P, Rawson T, Ray S, Hudovsky A and **Gill D**. Clinician-directed improvement in the accuracy of hospital clinical coding. Future Healthc J. 2018;5:47-51.
15. Rawson TM, Sivakumaran P, Lobo R, Mahir G, Rossiter A, Levy J, McGregor AH, Lupton M, Easton G and **Gill D**. Development of a web-based tool for undergraduate engagement in medical research; the ProjectPal experience. BMC Med Educ. 2018;18:166.
16. Dimou NL, Papadimitriou N, **Gill D**, Christakoudi S, Murphy N, Gunter MJ, Travis RC, Key TJ, Fortner RT, Haycock PC, Lewis SJ, Muir K, Martin RM and Tsilidis KK. Sex hormone binding globulin and risk of breast cancer: a Mendelian randomization study. Int J Epidemiol. 2019;48(3):807-816.

Prizes

I was awarded a British Heart Foundation Bursary to attend the International Stroke Genetics Consortium meeting in Cambridge, UK, on 11 – 12 April 2019.

Invited lectures

- "What mediates the effect of education on cardiovascular disease risk? Evidence from Mendelian randomization." Delivered at the Cohorts for Heart and Aging Research in Genomic Epidemiology Meeting. Baltimore, United States of America. 12 October 2018.
- "Mendelian randomization in the current era." Delivered at the Biomedical Data Sciences Division of Bayer. Wuppertal, Germany. 28 August 2019.
- "Applications of Mendelian randomization to inform on disease mechanisms and therapeutic opportunities". To be delivered at the Research Symposium on Mendelian Randomization Methodology. Cambridge, United Kingdom. 28 November 2019 (Scheduled).

Oral presentations

I have delivered oral presentations related to the contents of this PhD at the following conferences and meetings:

- "Testing for violations of the InSIDE assumption in Mendelian randomization". Mendelian Randomization Conference. Bristol, United Kingdom. 11-12 July 2017.
- "Blood pressure traits differentially affect risk of different ischaemic stroke subtypes". 23rd Workshop of the International Stroke Genetics Consortium. Kyoto, Japan. 12-13 April 2018.
- "Blood pressure traits differentially affect risk of different ischaemic stroke subtypes". The 4th European Stroke Organisation Conference. Gothenburg, Sweden. 16-18 May 2018.
- "The causal effect of systemic iron status in cardiovascular disease subtypes". 25th Workshop of the International Stroke Genetics Consortium. Cambridge, United Kingdom. 10-12 April 2019.
- "Strategies for selecting instruments in Mendelian randomization analyses". Mendelian randomization Conference. Bristol, United Kingdom. 17-19 July 2019.

Podcast

I was interviewed for the "Circulation on the Run" podcast, which was released with the 23 July 2019 issue of the journal Circulation. The full audio recording and transcript is available at https://circulation.libsyn.com/circulation-july-23-2019-issue.

Press releases

Imperial College London issued the following press releases related to research undertaken by me as part of my PhD study:

- “Higher iron levels associated with increased risk of certain types of stroke”. 27 October 2018. https://www.imperial.ac.uk/news/188777/higher-iron-/
- “More years spent in education associated with lower weight and blood pressure”. 22 May 2019. https://www.imperial.ac.uk/news/191308/more-years-spent-education-associated-with/
- “Blood pressure drug linked with increased risk of bowel condition”. 3 July 2019. https://www.imperial.ac.uk/news/191872/blood-pressure-drug-linked-with-increased/
- “Higher iron levels may boost heart health – but also increase risk of stroke”. 17 July 2019. https://www.imperial.ac.uk/news/192055/higher-iron-levels-boost-heart-health/

Table of contents

Abbreviations

ACEI: angiotensin-converting enzyme inhibitor

AF: atrial fibrillation

ARB: angiotensin receptor blocker

BB: beta-blocker

BioVU: Vanderbilt University Biobank

BMI: body mass index

CAD: coronary artery disease

CARDIoGRAMplusC4D: Coronary Artery Disease Genome-wide Replication and Meta-analysis plus The Coronary Artery Disease

CCB: calcium channel blocker

CES: cardioembolic stroke

CI: confidence interval

cIMT: carotid intima-media thickness

COGENT: Cognitive Genomics Consortium

CVD: cardiovascular disease

DBP: diastolic blood pressure

FDR: false discovery rate

GIANT: Genetic Investigation of Anthropometric Traits

GIS: Genetics of Iron Status

GRS: genetic risk score

GWAS: genome-wide association study

HES: hospital episode statistics

ICD: International Classification of Diseases

ICH: intracerebral haemorrhage

InSIDE: instrument strength independent of direct effect

INVENT: International Network against Thrombosis

IS: ischemic stroke

ISCED: International Standard Classification of Education

IVW: inverse-variance weighted

LAS: large artery stroke

LD: linkage disequilibrium

MI: myocardial infarction

MR: Mendelian randomization

MVMR: Multivariable Mendelian randomization

NOS: not otherwise specified

OR: odds ratio

PheWAS: phenome-wide association study

PRESSO: pleiotropy residual sum and outlier

RCT: randomised controlled trial

RR: relative risk

SBP: systolic blood pressure

SD: standard deviation

SE: standard error

SNP: single-nucleotide polymorphism

SSGAC: Social Sciences Genetic Association Consortium

SVS: small vessel stroke

TD: thiazide diuretic

VTE: venous thromboembolism

Structure of the thesis

This thesis explores the application of Mendelian randomization (MR) techniques for identifying cardiovascular disease (CVD) mechanisms and therapeutic opportunities. It is divided into three parts. The first is made up of Chapter 1, and provides an introductory overview of CVD, MR and application of MR for studying CVD. Having highlighted diet, education and blood pressure as important cardiovascular risk factors that can be studied to demonstrate the breadth of scenarios within which MR can be applied, the second part of the thesis consists of Chapters 2-4, which in turn investigate previously unexplored aspects of these. Chapter 2 looks at the effect of systemic iron status on CVD subtypes and health outcomes more generally, Chapter 3 investigates mediators of the effect of educational attainment on CVD risk and effects not mediated via cognition, and Chapter 4 explores the efficacy, side-effects and repurposing opportunities for common antihypertensive drugs. The final part of the thesis, Chapter 5, looks at the recent advances and available strategies for applying MR to infer causal effects within the field of cardiovascular medicine, finishing to consider future perspective.

Chapter 1: Introduction

All of the work presented in this chapter is my own, unless otherwise indicated in the text

1.1 Cardiovascular disease

Cardiovascular disease (CVD) is the leading cause of death worldwide, with an age-standardized mortality rate of 278 per 100,000 population per year (1). Furthermore, there is relatively little variation in its incidence across countries of different economic status, although case fatalities do vary, with more favourable outcomes in higher income countries (2). The term CVD encompasses a range of conditions, of which coronary artery disease (CAD) and stroke are the most common, with a global prevalence of approximately 150 million and 80 million respectively (3), and together accounting for 85% of CVD deaths (1). CAD describes atherosclerotic disease of the coronary arteries, resulting in disruption of the cardiac blood supply with associated ischaemia and increased risk of thrombosis. In contrast, stroke is defined as a neurological deficit of the central nervous system caused by an acute vascular injury (4). Myocardial infarction (MI) refers to death of heart tissue due to an inadequate blood supply, and represents the most severe consequence of CAD. Although CAD is the most common cause of MI, the specific mechanisms by which this can occur are heterogeneous, and MI can be sub-classified accordingly. Type 1 MI relates to the rupture or erosion of an atherosclerotic plaque in the coronary arteries and consequent occlusion of adequate blood flow to the heart tissue (5). In contrast, type 2 MI relates to a mismatch between myocardial oxygen demand and blood supply that is not related to an acute thrombotic event, such as can occur due to coronary artery atherosclerosis, spasm, dissection, or due to systemic causes such as hypotension of hypoxia (5).

Stroke itself is also a highly heterogeneous disease, with its two major subtypes being ischaemic stroke (IS) and intracerebral haemorrhage (ICH), making up 85% and 15% of the total stroke burden respectively in European-ancestry populations (6). IS relates to insufficient blood supply to the brain, and ICH describes a bleed within the brain parenchyma (4). IS and ICH may be further subdivided in relation to the underlying causative mechanism. The Causative Classification of Stroke system uses an algorithm to categorise IS subtypes based on the most likely underlying mechanism of occlusion to cerebral blood flow: large artery stroke (LAS), cardioembolic stroke (CES), small vessel stroke (SVS), and other uncommon cause or undetermined cause (7). This approach uses examination and investigation results from individual patients (7), and has moderate to strong correlation of allocated phenotypes as compared to the less widely used Trial of Org 10172 in Acute Stroke Treatment system (8), which employs clinician judgement to assign aetiology (9). Identifying the underlying pathophysiologic aetiology of IS can be challenging in scenarios where multiple mechanisms may be at play. For example, although atrial fibrillation (AF) is the major risk factor for CES (10, 11), patients with AF also tend to have other cerebrovascular risk factors and may also be susceptible to IS from LAS or SVS (12, 13). Thus, while aetiological classification systems aim to offer insight into the most likely cause of IS, they cannot be perfectly accurate (14). While symptom based classification systems for IS are also available, they still offer less information on the underlying mechanism (15). ICH subtypes also differ in their pathophysiology – most cases are attributable to hypertension, with cerebral amyloid angiopathy accounting for approximately 10% of cases as the next most common cause (16). Similar to IS, classification systems for ICH are either based on an anatomical or mechanistic subdivision, and have excellent inter- and intra-rater reliability within specialist centres (17). Anatomical ICH subtypes are defined by the location of haemorrhage; lobar, deep, cerebellar or brainstem, with the latter three categories combined as 'non-lobar' in some systems (17). Mechanistic ICH classification is typically related in some form to the 'SMASH-U' system that subtypes based on most likely cause of ICH: structural lesion, medication, amyloid angiopathy, systemic disease, hypertension, or undetermined aetiology (18). In general terms, lobar ICH is more commonly associated with cerebral amyloid angiopathy than non-lobar ICH (19), while non-lobar ICH is more likely to be related to hypertension than lobar ICH (20).

The Emerging Risk Factors Collaboration and other groups have performed meta-analyses of observational studies to offer insight into the association of individual risk factors with CVD, identifying blood pressure (21), dyslipidaemia (22), blood glucose levels (23), cigarette smoking (24), and obesity to all show significant positive associations (25). There is also mediation between risk factors, with changes in body mass index (BMI) affecting blood pressure, for example (26). Risk factor and susceptibility profiles also vary for different CVD subtypes and across distinct population groups (27-30). For example, observational research has found smoking to be more strongly associated with LAS than other IS subtypes, and hypertension similarly more commonly seen in SVS (27, 30). For ICH, alcohol use and hypertension have been identified as prominent risk factors (31), with obesity possibly being protective (32).

The most comprehensive observational studies to investigate the association of modifiable traits with CVD risk have been performed by the Global Burden of Disease Study (33, 34). These efforts have consistently identified metabolic, social and dietary factors to explain the majority of cardiovascular risk (34). However, potential limitations to these studies include the amalgamation of data from diverse populations, with distinct methods used in different geographical locations to obtain, record and analyse data (34). To complement such efforts and overcome these limitations, a recent multinational, prospective cohort analysis consisting of 155,722 individuals in the Prospective Urban Rural Epidemiology Study explored the association between modifiable risk factors and incident CVD events (35). This effort provided evidence to support that 70% of CVD events can be attributed to variation in modifiable risk factors (35). In particular, metabolic traits represented over 40% of risk, with hypertension making up over 20% of this (35). Of the considered socioeconomic and psychosocial factors, low educational attainment was the greatest risk factor for CVD, contributing over 10% of total risk (35). Finally, of the behavioural risk factors studied, diet explained approximately 5% of the CVD risk (35). Indeed, these findings are consistent with previous Global Burden of Disease Study efforts in identifying the large contributing effects of these traits on CVD risk worldwide.

The Comparative Risk Assessment Study has also made progress towards improving comparability across analyses of cardiovascular risk factors (36, 37), further supporting that traditional CVD risk factors such as hypertension are among the leading causes of global mortality and morbidity (36-38). Consistent with the Prospective Urban Rural Epidemiology Study, the Comparative Risk Assessment Study found that 80% of CAD death and 70% of stroke deaths were attributable to the joint effect of a select number of risk factors, including hypertension, raised serum cholesterol, cigarette smoking, obesity, alcohol consumption, low fruit and vegetable intake and physical inactivity (39). Of relevance, when considering the various risk factors separately, rather than in a joint model, the sum of their effects came to 226% for CAD and 165% for stroke (rather than 80% and 70% respectively as in the joint model), highlighting that many of these risk factors interact, overlap and cluster within the same individuals. Following on from this work, research priorities have been put forward towards improving understanding of the determinants of CVD risk, and in particular highlighted that methods less vulnerable to the confounding, reverse causation and measurement error biases encountered in traditional epidemiological research are required (40). The next sections will discuss how the emergence of genetic data may facilitate such an objective.

1.2 Genetic data

With the growing availability of genetic data, it is becoming increasingly feasible to identify the genetic correlates of phenotypic traits. Figure 1.1 shows the number of scientific papers listed on the PubMed database when searching for 'genome-wide association study' as a general search term (accessed 14 May 2019), with just 18 in 2000, to 52 in 2003 when the human genome was first sequenced, and increasing to 3310 in 2018. The first GWAS was entered into the curated GWAS Catalog database in 2005, with the total number of entered research papers exceeding 3600 in 2018 (41).

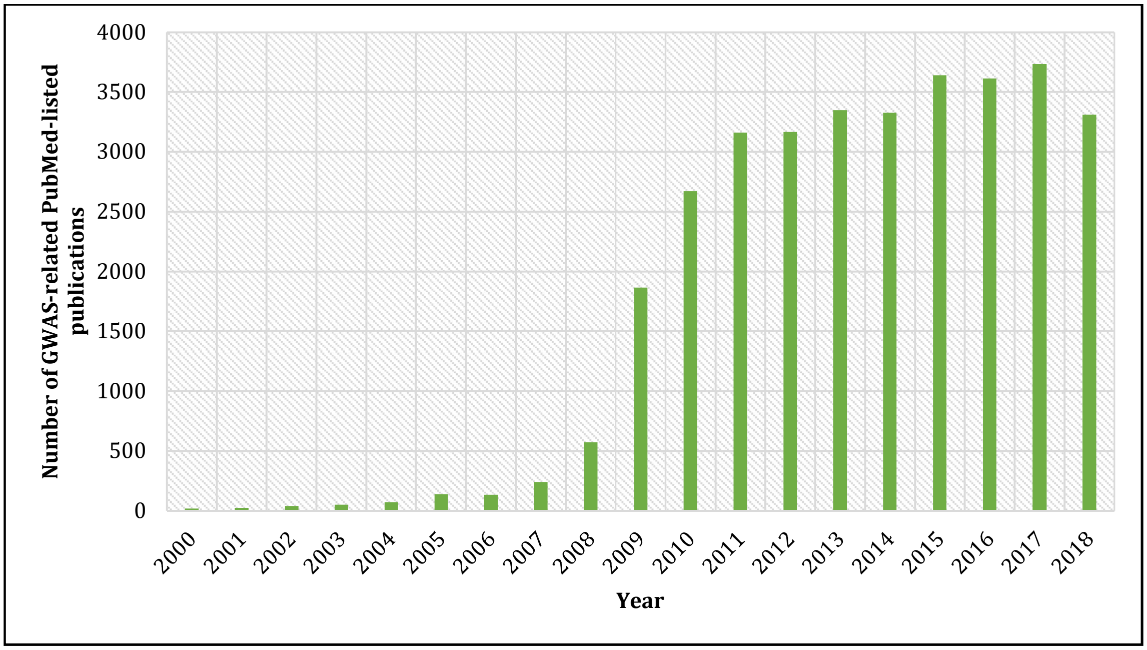


Figure 1.1. The number of scientific papers, listed by year, on the PubMed database when searching for 'genome-wide association study' as a general search term

In turn, this has been associated with opportunities for identifying disease mechanisms and therapeutic targets, as well for individualised risk stratification and personalised medicine (41). Relating to CVD, there have been large-scale GWAS meta-analyses for abdominal aortic aneurysm (42), AF (43), CAD (44), carotid plaque (45), IS and its subtypes (46), ICH and its subtypes (29), and peripheral arterial disease (47). Details relating to these studies are provided in Table 1.1. While there have been a number of different GWASs investigating each of these traits, a recent study for each outcome is listed for demonstrative purposes. While such work has clearly highlighted the role of genetic factors in predisposing to CVD risk, the findings have also been used to study the interplay between genetic and environmental determinants, with these each shown to cumulatively affect outcomes (48). Genetic studies investigating CVD subtypes have also supported aetiological overlap. For example, there is a strong genetic association between AF and CES (11), which is consistent with AF being a major predisposing factor for CES. Similarly, there is substantial genetic overlap between ICH and SVS (29), which both relate to pathology of the cerebral small vessels, or LAS and CAD (49), which both relate to atherosclerosis.

Table 1.1. Published genome-wide association studies exploring cardiovascular disease outcomes.

Trait	Subtype	Ethnicity	Cases	Controls	Summary data availability	Reference
Abdominal aortic aneurysm	-	European-ancestry	10,204	107,766	Available on request	(42)
Atrial fibrillation	-	Mixed (88% European-ancestry)	65,446	522,744	Publicly available	(43)
Carotid plaque	-	European-ancestry	21,540	26,894	Available on request	(45)
Coronary artery disease	-	Mixed (76% European-ancestry)	60,801	123,504	Publicly available	(44)
Ischaemic stroke	Any	Mixed (86% European-ancestry)	67,162	454,450	Publicly available	(46)
	Cardioembolic		9,006			
	Large artery		6,688			
	Small vessel		11,710			
Intracerebral haemorrhage	Any	Mixed (73% European-ancestry)	3,226	3,742	Publicly available	(29)
	Lobar		1,148			
	Non-lobar		2,075			
Peripheral artery disease	-	Mixed (91% European-ancestry)	36,424	601,044	Available on request	(47)

Table 1.1 highlights that the majority of GWASs exploring CVD outcomes have been conducted in populations either entirely or predominantly of European ancestry, thus in turn having implications regarding the generalisability of these findings to other ethnic groups. This is more reflective of the distribution and allocation of resources available for such study, rather than the representation of this European-ancestry group within the global population. Also of interest is the marked variation in the sizes of the respective studies, ranging from the 3,226 cases included in the analysis of ICH (29), up to 67,162 cases included in the study of IS (46), more than a 20-fold difference. Neither can this discrepancy be attributed to differences in the relative incidence or prevalence of these outcomes, as ICH and IS are estimated to make up approximately 15% and 85% respectively of all stroke cases in European populations (6). Rather, it is again a function of differences in resource allocation and prioritisation of research objectives. Importantly for work that requires summary genetic association estimates related to these studies, their data have been made publicly available in the majority of cases, and are

otherwise obtainable on request from the relevant study authors or consortia in the remaining cases (Table 1.1).

1.3 Mendelian randomization

While observational research is useful for identifying associations between environmental or physiological traits and disease outcomes, such study is intrinsically limited in its ability to infer causal effects because of the possibility that any observed associations arise due to confounding or reverse causation (50). Instrumental variable analysis can overcome such limitations (51), and more specifically, where randomly allocated genetic variants are used as instruments for an environmental or physiological exposure (52), this Mendelian randomization (MR) principle can be used to estimate the effect of an intervention (53). It is specifically because the genetic variants used as instruments are randomly allocated at conception that they are independent of environmental factors and also precede the outcome of interest, thus allowing the MR framework to overcome confounding and reverse causation biases respectively. Furthermore, the same approach can be used to investigate the direction of causality (54), and can also accommodate negative and positive control analyses (55). With the availability of genetic association studies that investigate particular CVD subtypes, it is also possible to perform MR analysis that explore distinctions in underlying CVD mechanisms.

MR requires that the instrument is associated with the exposure of interest (relevance assumption), but not directly with the outcome of interest independently of the exposure (exclusion-restriction), nor any confounders of the relationship between the exposure and the outcome (independence assumption) (52) (Figure 1.2). With the widespread availability of genetic association data, MR studies have shown an exponential and persistent increase in number over recent years. Figure 1.3 shows the number of MR-related scientific papers identified on PubMed by searching ‘Mendelian randomization’ as a general search term (accessed 14 May 2019), growing to 350 in 2018 from zero in 2002.

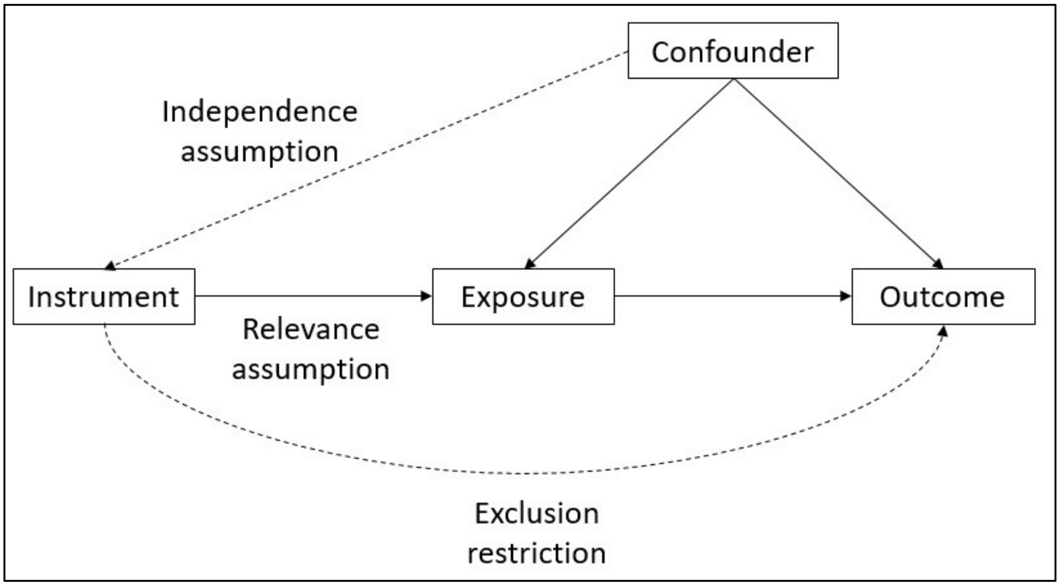


Figure 1.2. Instrumental variable assumptions of Mendelian randomization depicted in a directed acyclic graph. The solid lines represent causal associations, whereas the dashed lines represent violations of the underlying assumptions.

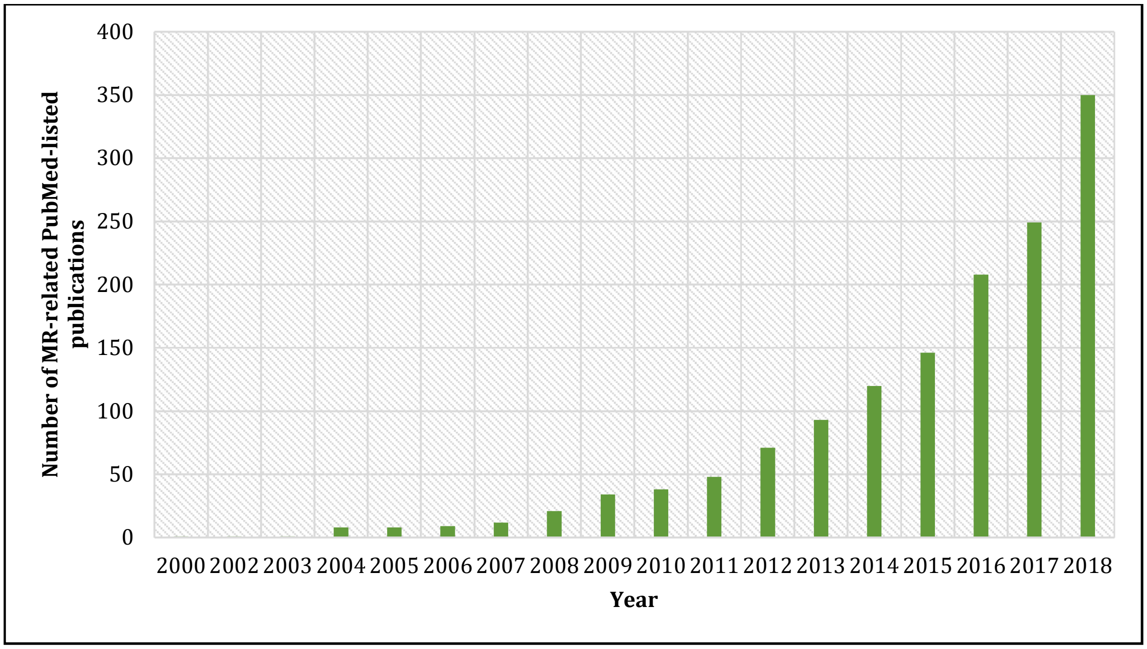


Figure 1.3. The number of Mendelian randomization scientific papers, listed by year, identified on PubMed by searching 'Mendelian randomization' as a general search term.

1.4 Mendelian randomization approaches

Where genetic association estimates for both the exposure and the outcome are obtained from the same population, the effect of the exposure on the outcome can be estimated using the ratio method, by regressing the outcome on the instrument and dividing the consequent coefficient from that obtained by regressing the exposure on the instrument (56). Incidentally, the same estimate can also be obtained using the two-stage least squares method, where the exposure is first regressed on the instrument, and the outcome is then regressed on the fitted values from this first-stage regression (56). These methods can be applied to consider multiple genetic instruments separately, or also after creating a genetic risk score, which has the advantage of minimising bias from use of many weak instruments (57).

The availability of summary genetic data has further made it possible to perform MR analysis using genetic association estimates for the exposure and outcome of interest arriving from separate studies, thus also increasing the maximum available sample size and consequent statistical power. However, to avoid the introduction of bias when using such two-sample approaches, it is important that consistency is maintained in the populations used to obtain genetic association estimates for the exposure and the outcome, including demographics, co-morbidities, lifestyle factors and ethnicity, as these approaches assume homogenous effects and genetic associations with the exposure and outcome in both populations (58, 59). A range of MR methods are now available for both the one-sample and two-sample summary data settings, each with distinct properties and characteristics within different scenarios (58, 60, 61).

As detailed in Figure 1.2, central to MR analysis is that the instruments are associated with the outcome only through the exposure, and not by some pleiotropic pathway. Violation of this exclusion-restriction assumption can potentially result in biased MR estimates (62). MR exposures for consideration can be broadly categorized as either ‘distal’ or ‘proximal’ in respect to their relation with the genetic variant. Distal exposures arise from the culmination of many inter-related processes that are mechanistically far downstream of the genetic variant used as an instrument, with examples including educational attainment (63, 64), BMI index or age at menarche (65, 66). In contrast, ‘proximal’ exposures are closely related to the genetic variant in terms of mechanism, and may for example relate to effects on the structure or function of proteins. Such clinically relevant examples include HMGCR inhibition (statin treatment) (67), or beta-blockade (beta-blocker treatment) for example. Instruments for distal exposures can be selected as related genetic variants from throughout the genome (63, 65). Such instruments typically also associate to the outcome of interest in MR through pleiotropic mechanisms, thus generating heterogeneity in their consequent MR estimates (68). In contrast, instruments for proximal exposures can be selected as variants at the corresponding gene locus (*cis*-acting)

(69). Such mechanistic proximity to the relevant gene increases assurance in the validity of the instrument, although the potential for bias from pleiotropic effects remains.

A further distinction is between MR analyses considering a change in exposure through a particular pathway, and those considering a change in exposure through any mechanism. An example highlighting this in low-density liproprotein cholesterol (LDL-C) lowering through HMGCR inhibition (a particular pathway), in contrast to LDL-C lowering by any means. For the former, genetic variants specifically related to the HMGCR locus and LDL-C levels (thus proxying the effect of HMGCR inhibition) would be preferred, while in the latter scenario, variants from throughout the genome that are associated with LDL-C levels may be employed as instruments.

The proportion of exposure variance explained by the genetic instruments is a key factor in determining the consequent statistical power of MR analysis (70, 71). A commonly employed strategy to thus increase statistical power is the pooling of MR estimates generated from different genetic instruments. Such pooling of MR estimates can be achieved using a number of methods (72), with inverse-variance weighted (IVW) meta-analysis commonly implemented in applied studies where the ratio method is used to calculate MR estimates (73). With the ratio method, the MR estimate is derived using the Wald ratio (74), with standard errors typically estimated using either the first order or second order expansion of the Taylor series (75), although modified second order weights have also been suggested (75). However, in the context of selecting such instruments from a specific gene locus, the proximity of the genetic variants with each other makes it likely that they will suffer some degree of linkage disequilibrium (LD) and will thus be correlated. Pooling of MR estimates from correlated instruments without accounting for this will thus cause shrinkage of the resultant MR estimate standard errors and an inflated risk of false positive results. This issue may be overcome using alternative summary data methods, such as by combining estimates from multiple instruments using a generalized linear regression approach weighted for the correlation between instruments (57), or through use of principal component analysis to construct instruments (76).

Where plausible genetic variants that explain a sufficiently large proportion of the variance in the exposure of interest are available for use as instruments, MR analyses can have sufficient statistical power with fewer instruments that have known biological functions (69). In contrasting scenarios where the exposure of interest is a trait distal to the genetic variant that represents the culmination of various genetic and environmental factors, many genetic variants with less clearly defined biological functions may be incorporated as instruments. Examples of such MR analyses include an investigation into the causal effect of age at menarche on lung function (77), or the effect of time spent in education on coronary heart disease risk (64). Indeed, for such distal traits, individual genetic variants may only explain a small proportion of the variation on the exposure of interest (78, 79), and thus using many genetic variants as instruments serves to increase the statistical power of the MR study (70, 71).

Other considerations for identifying instruments include the *P*-value thresholds for association with the exposure. No consensus criteria have yet been established for this, with individual authors also varying in their approaches for different circumstances (80-82). Generally speaking, arbitrary criteria that attempt to minimise bias while optimising statistical power may be selected, with appropriate sensitivity analyses performed to investigate the effects of altering these thresholds. As further discussed below, the F-statistic can also be used as a measure of instrument strength, and is related but preferable to *P*-value thresholds for this purpose (83).

Additionally, information on secondary traits can be incorporated to improve the validity and robustness of genetic instruments for MR, thus also reducing bias in consequent analyses. This relates to the exclusion-restriction assumption of MR, where instruments do not relate to confounders of the exposure-outcome relationship, and the relevance assumption of MR, where the instruments must be related to the exposure of interest. A practical example demonstrating this could be the identification of genetic variants for the diuretic antihypertensive drug furosemide, for which genetic variants at the corresponding *SLC12A1* gene locus would also be expected to relate to lower SBP and higher urinary sodium, given the known mechanism and clinical effects of this agent.

1.5 Pleiotropy in Mendelian randomization

The availability of multiple instruments in MR analysis allows for statistical investigation of possible pleiotropy (84), where instruments may be associated with the outcome through some pathway at least partly independent of the exposure (85). Where the instruments are all valid, little heterogeneity would be expected in their individual MR estimates. The presence and magnitude of any heterogeneity may thus be used to estimate the presence and magnitude of instrument pleiotropy that may be biasing the MR estimate (68). In cases where there is significant heterogeneity in the individual MR estimates generated by different instruments but no MR evidence of directional pleiotropy, random-effects IVW meta-analysis of MR estimates is preferred for generating an overall MR estimate (72). This is in contrast to the fixed-effects model that would be favoured if there was no convincing evidence of heterogeneity (which is used as a proxy for pleiotropy) in the MR estimates of different instruments (72).

Numerous pleiotropy robust MR methods have now been described, with the number of available methods continuously growing (61). One popular method is MR-Egger, which is an adaptation of the Egger regression used to assess for publication bias following meta-analysis

(86). Under the assumption that instrument strength is independent of any direct effect on the outcome (InSIDE), the MR-Egger technique regresses each instrument-outcome association by the corresponding instrument-exposure association, weighted for the precision of the instrument-outcome association (86). A non-zero intercept provides evidence of any directional pleiotropy that may be biasing the MR estimate, and thus allows MR-Egger to serve as a test for this. However, this approach is particularly sensitive to violations of the InSIDE assumption, and can produce misleading results in scenarios where it does not hold (87). A similar regression-based approach is applied in the MR-pleiotropy residual sum and outlier (PRESSO) method, which regresses the variant-outcome estimates on the variant-exposure estimates, with the gradient of the regression line representing the MR estimate (88). Furthermore, MR-PRESSO is able to identify outlier variants based on their observed distance from the regression line, as compared to their expected distance based on the assumption of no horizontal pleiotropy, and can thus generate outlier-corrected MR estimates that exclude such variants (88).

Where there are known biological pathways through which an instrument may be affecting an outcome independently of the exposure of interest, a technique called multivariable MR (MVMR) may be used to generate MR estimates that are independent of these pleiotropic pathways (89), with a regression based MVMR method applicable for the two-sample summary data setting (86). This approach may be used for example, to generate MR estimates for the causal effect of age at menarche on adult BMI, independent of the genetic association of the instruments with childhood BMI (65). Furthermore, MVMR may also be used to dissect causal pathways, such as in the investigation of the role of BMI in mediating the effect of age at menarche on cancer risk (90). An extension of this approach now also incorporates elements of the MR-Egger technique to investigate if there is any residual directional pleiotropy after adjusting for possible mediators or pleiotropic pathways (91).

The IVW meta-analysis MR, MR-Egger, MR-PRESSO and MVMR techniques all rely on the InSIDE assumption and there is no widely applicable method to test whether this is being violated (92). Fortunately, other statistical approaches are available that make orthogonal assumptions on the inclusion of any pleiotropic instruments (84). The median estimator calculates the mid-point of the distribution of MR estimates, and can be weighted for their precision, allowing it to offer consistent estimates when more than half of the estimates for the analysis come from valid instruments (93). Another MR sensitivity analysis is the mode-based estimator, which centres its overall MR estimate on the greatest number of individual instruments giving similar MR estimates, thus providing robust results when these instruments are valid (94). The mode-based estimator can be similarly weighted for the precision of the MR estimates (94). Table 1.2 summarises the most commonly applied MR methods within the two-sample summary data setting. Other statistical approaches to deal with pleiotropy are also available (61), as are Bayesian models that average the MR estimates from various analyses with differing underlying assumptions on the nature of pleiotropy, to reduce bias at the cost of precision (95).

Recent developments have enabled MR to also be used for performing mediation analysis, both for the purpose of disentangling causal mechanisms, as well as for estimating mediation effects. The most commonly applied methods for this are MVMR and network MR (90, 96). In MVMR, the variant-outcome association estimates are regressed against the variant-exposure estimates adjusted for the variant-mediator estimates. Attenuation of the direct estimates obtained in MVMR as compared to the total exposure-outcome effect estimate generated in conventional (univariable) MR methods would provide an indication of the degree of mediation through the considered mediator (90). For network MR, the exposure-mediator MR estimate is multiplied by the mediator-outcome estimate adjusted for the exposure using MVMR, to estimate the effect of the exposure on the outcome that is mediated through the mediator (96). Standard errors can be estimated using bootstrapping or the propagation of error method. In all MVMR, superior power is generally achieved when instruments for both the exposure and any considered mediators are included in the same model.

Table 1.2. Commonly applied two-sample Mendelian randomization approaches.

Method	Pleiotropy assumption	Strengths	Weaknesses	Reference
Inverse-weighted	All variants are valid instruments	Tends to offer precise estimates	Least robust to pleiotropy	(73)
Egger	Instrument strength is independent of any direct effect on the outcome	Can still be reliable when all variants are invalid instruments, offers a test for bias related to pleiotropy	Sensitive to outliers, often imprecise	(86)
PRESSO	Only outlier variants are pleiotropic	Tends to offer precise estimates, can identify outlier variants	High false positive rate with a large proportion of invalid instruments	(88)
Multivariable	Genetic association estimates for all pleiotropic pathways are available	Can estimate both direct and indirect effects	Susceptible to bias related to measurement error, can be unstable with highly correlated traits	(89)
Median	Majority of the variants are valid instruments	Robust to pleiotropic outliers	Sensitive to removal of genetic variants	(93)
Mode	Variants generating the most frequently observed estimate are valid	Robust to pleiotropic outliers	Generally conservative, sensitive to pre-defined modelling parameters	(94)

1.6 Other sources of bias in Mendelian randomization

When translating the results of MR analyses to clinical practice, it is also important to appreciate the various other implicit assumptions made by the technique. The MR effect estimate represents the cumulative lifetime effect of genetic variants serving as an instrument for the exposure (52). Bias may therefore be introduced where developmental compensation buffers the effect of genetic variants in a process called canalization, thus attenuating MR causal effect estimates towards the null (52). Under the assumption of monotonicity, MR calculates an average causal effect estimate for the population considered (i.e. the local-average treatment effect), and there may be subgroups that deviate from this (97). This assumption may be notably violated in MR analyses that consider a binary exposure (such as the presence of a disease phenotype), as in such scenarios it is unusual for any instrument to perfectly predict the presence of the trait (98). Furthermore, MR assumes that the effect of the exposure on the outcome is linear and homogenous across participants (homogeneity assumption) (96). The linearity assumption may be particularly susceptible to violation in the scenarios considering metabolites or other small molecules as exposure in MR (99), and the homogeneity assumption may be compromised when there are relevant distinctions in the populations used to obtain genetic association estimates for the exposure and outcome, for example.

As discussed above, the MR technique is a form of instrumental variable analysis and is therefore subject to bias from the inclusion of weak instruments, where the genetic variants do not relate to the exposure under consideration with sufficient strength (100). The F-statistic can be used to quantify the level of such bias for each instrument (100), with a value of 10 previously shown to correspond to an approximate 10% relative bias (83). Approximations for the F-statistic are also available to facilitate the practical measurement of instrument strength in applied MR analyses (66, 101). Any weak instrument bias of the MR estimates will be towards the observational association between the exposure and the outcome for MR analyses that derive all genetic association estimates from a single population (i.e. one-sample MR) (102). For such one-sample MR, weak instrument bias can be reduced by combining multiple genetic instruments into a single genetic risk score that has greater overall strength as an instrument (60). In the case of two-sample ratio method MR analysis, where the genetic association estimates for the exposure and outcome are derived from separate populations, any bias from the inclusion of weak instrument would push the overall MR estimate towards the null hypothesis (102). However, for two-sample MR analyses where there is overlap in the populations used to derive instrument-exposure and instrument-outcome genetic association estimates, any bias would be towards the observational estimate, and is a linear function of the degree of overlap between the samples for continuous outcomes, whereas for binary outcomes such bias only arises if there is sample overlap with the exposure population for outcome cases (102).

As the genetic association estimates used to fit models in two-sample MR are associated with measurement error, the method used for weighting the contribution of individual instruments in pooled analyses can have implications for the overall accuracy (75). To this end, modified second order (Taylor series) weighting has been shown to be preferable for estimating heterogeneity and causal effects (75). For the MR-Egger technique in particular, measurement error can result in regression dilution bias (103), and the I^2 heterogeneity statistic for instrument-exposure estimates can be used to explore the implications of this interpreting MR-Egger estimates, with caution recommended for I^2 estimates less than 90% (103).

The use of instrument-exposure genetic association estimates from discovery analyses can introduce Winner's curse bias to inflate estimates and bias two-sample MR analysis towards the null hypothesis (104). In scenarios where the instruments for an MR analysis are valid, but their genetic association estimates with the exposure are biased or unknown, an unweighted allele score of instrument-outcome genetic association estimates may be used to investigate for any causal effect of an exposure on an outcome, but not measure its magnitude (60). Such an unweighted allele score can therefore be used as a sensitivity analysis in MR analyses where instrument-exposure estimates may be biased because of Winner's curse (66).

In the field of statistics and causal inference, colliders are variables that are themselves affected by two or more variables (105). In the context of MR, any conditioning on such colliders can also result in bias (Figure 1.4) (105, 106). This includes any adjustment or stratification undertaken in obtaining genetic association estimates for MR analysis or when selecting populations in which to perform MR analysis (Figure 1.4).

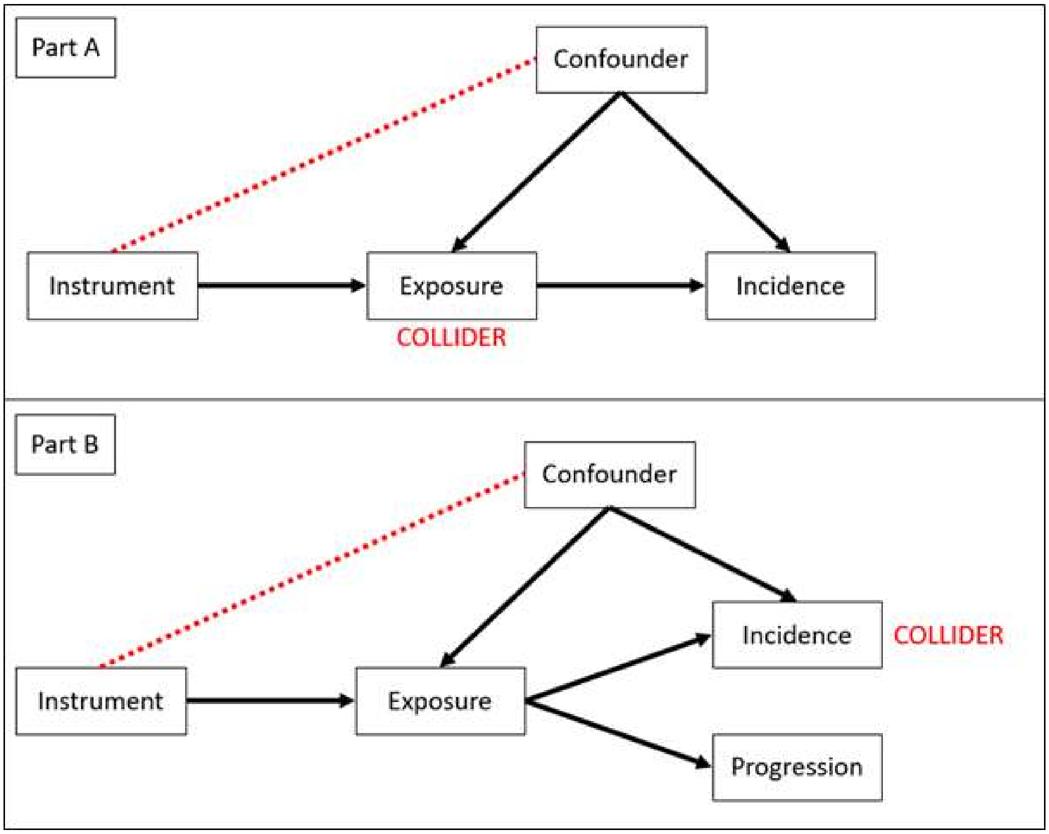


Figure 1.4. Examples of collider bias. Part A: Conditioning on the exposure (a collider), such as by restricting the Mendelian randomization analysis to a subpopulation with a certain level of the exposure, will result in collider bias by introducing an association between the instrument and confounders of the exposure and outcome. Part B: Conditioning on disease incidence (a collider) when performing Mendelian randomization to study disease progression will introduce collider bias if the exposure under consideration also affects disease incidence.

Finally, where the exposure under investigation in MR analyses affects survival, and particularly survival long enough to develop the outcome of study, survivor bias can also be a relevant consideration (107, 108). Where the exposure has the same direction of effect on both the outcome and survival, bias is towards the null hypothesis (no effect of the exposure on the outcome), with greater bias seen in studies that consider older populations (108). Simulation analyses have suggested that such bias might be in the order of approximately 5% when considering the effect of BMI on Parkinson's disease risk, for example (109).

1.7 Mendelian randomization and cardiovascular disease

Given the burden of CVD discussed above, and the applicability and advantages of its investigation using MR, it follows that this approach has been used extensively to study underlying disease mechanisms and therapeutic opportunities. In order to formally explore the application of MR in the context of CVD, PubMed was searched up to 19 May 2019 using the following search terms: (((Mendelian randomization) OR (Mendelian randomisation))) AND ((Cardiovascular disease) OR (Coronary heart disease) OR (Coronary artery disease) OR (Stroke)). No restrictions were applied for language or article type. Rather than a formal systematic review of the literature, this semi-systematic search was merely performed to provide a broad overview of the application of MR to the field of CVD. A total of 695 search results were produced, for which all titles and abstracts were read. Of these, 260 were Comments, Commentaries, Editorials, Letters or Review articles, 40 were Methodological papers, 18 were not MR studies, 2 were exactly duplicated entries and 2 were Corrections (Figure 1.5). Of the 373 remaining original research articles, 258 were considering cardiovascular and thrombotic diseases as the outcome of interest.

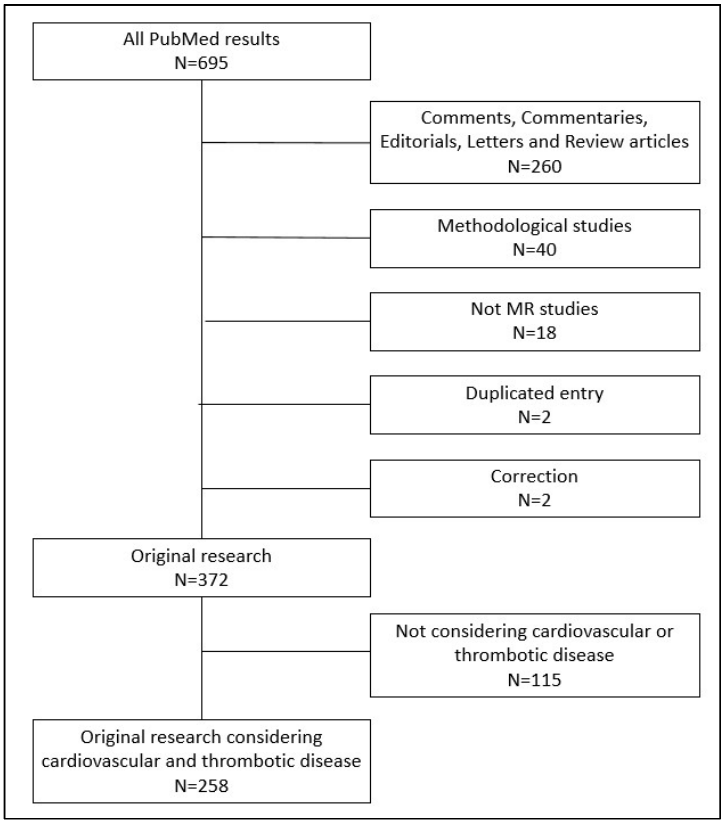

All PubMed results
N=695
Comments, Commentaries, Editorials, Letters and Review articles
N=260
Methodological studies
N=40
Not MR studies
N=18
Duplicated entry
N=2
Correction
N=2
Original research
N=372
Not considering cardiovascular or thrombotic disease
N=115
Original research considering cardiovascular and thrombotic disease
N=258

Figure 1.5. A flow chart depicting the results of the PubMed search of Mendelian randomization studies investigating cardiovascular disease. The boxes to the right side relate to those studies that were excluded.

The 258 articles broadly fell into five categories (presented in tabular format in Appendix 1):

1. Circulating factors (156 articles, Appendix Table 1),
2. Physiological traits and diseases (56 articles, Appendix Table 2),
3. Social and behavioural traits (19 articles, Appendix Table 3),
4. Cellular characteristics (14 articles, Appendix Table 4), and
5. Existing drugs (13 articles, Appendix Table 5).

The results of the literature review highlight the wide variety of exposures that have been studied using MR within the context of CVD. While the majority of studies investigate circulating factors, such as metabolites, chemical messengers, enzymes and hormones (156/258; 60%), fewer investigated social and behavioural traits (19/258; 7%) or existing drugs (13/258; 5%). The number of MR studies in each of the 5 categories published over time are depicted in Figure 1.6.

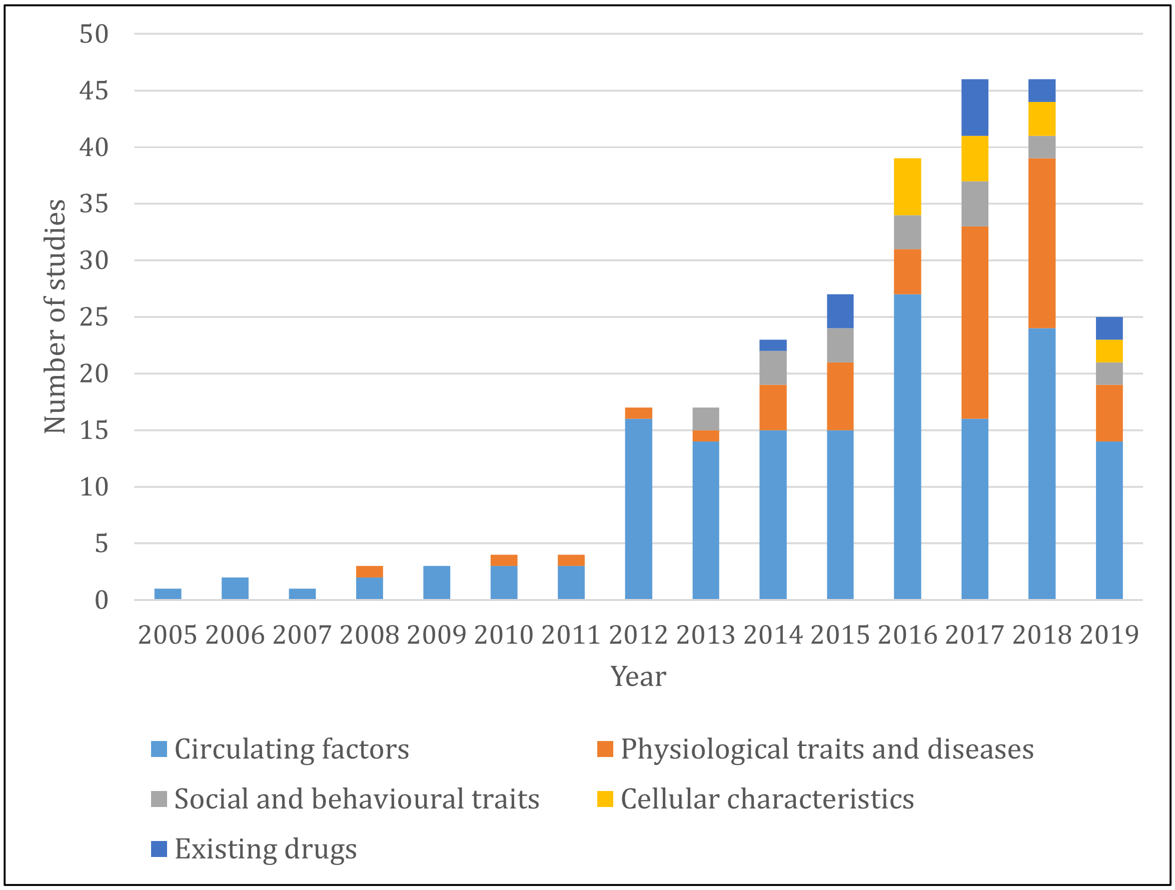


Figure 1.6. The number of Mendelian randomization studies falling within different exposure categories published by year.

While MR studies considering circulating factors and physiological traits and diseases have been published since 2005 and 2008 respectively, those considering social and behavioural traits, existing drug targets and cellular characteristics have only become prominent more recently (Figure 1.6). While this is partly reflective of the growing availability of genetic association data and instruments for these different exposure categories with time, it may also relate to how interest in various different types of exposure trait has shifted. To highlight the transition, while 94% (16/17) of the published MR studies considering CVD outcomes in 2012 related to circulating factors as exposures, this proportion fell to 52% (24/46) in 2018.

There are also exposures that are relatively over-represented in the MR literature. One example of this is serum urate, a product of purine metabolism, for which the literature review identified 9 MR studies where this was considered as the exposure of interest (110-118), thus representing 6% (9/156) of all MR studies within the category of circulating factors, and 3% (9/258) of all included studies. However, these studies did vary in the populations and specific disease outcomes studied, thus offering complementary insight. At the other end of the spectrum, there was also a relative sparsity of MR investigation into some of the modifiable risk factors that have previously been highlighted to have most influence on CVD risk (34, 35). While dietary factors have been shown to be an important determinants for CVD (34, 35), with MR now extensively used to study the effect of variations in circulating levels of nutrients (Appendix Table 1), no MR studies had previously investigated the effect of systemic iron status on cardiovascular outcomes. Of relevance, iron status is highly variable, having a coefficient of variation greater than 30% in both men and women (119). Furthermore, iron deficiency affects approximately a fifth of the world's population, representing a significant health burden in its own right (120). Most importantly, iron status can be effectively modified through both dietary and pharmacological interventions (121).

Of the behavioural and social risk factors, educational attainment has been shown to have the greatest effect on CVD risk (35), and indeed MR evidence supports this finding (64). However, the mediators of this relationship are not known and can now also be explored using mediation analysis methods within the MR framework (90, 96). Similarly, it is not known whether education itself is affecting cardiovascular risk or whether it is the closely related measure of intelligence that is causing any benefit, something that can now be disentangled using the MVMR framework. Distinguishing the mediators and distinct effects of education and intelligence would have important public health implications, as this would identify clear targets for public health and educational policy, as well as resource allocation. In addition, where educational attainment is not amenable to modification but the mediators of education's protective effect on CVD are known, these can be targeted to minimise any societal consequence of educational inequality.

Finally, hypertension has been identified as the risk factor having the single greatest effect on CVD risk (35), and furthermore there are numerous pharmacological treatments that have been shown to be efficacious for reducing blood pressure and consequently CVD risk (122). However, less is known about possible side-effects and repurposing potential for these medications, particularly as trials are often limited by their time and resource constraints to focus on demonstrating efficacy in high risk populations. To this end, the MR approach can now be applied to instrument drug effects (123), as has already previously been done for lipid-lowering drugs (Appendix Table 5) (124). However, such an approach had not previously been adopted to antihypertensive medications.

1.8 Scope, overall aims and objectives of the thesis

The literature review above highlighted relative gaps in the application of MR for specific purposes when studying mechanisms and therapeutic targets for CVD. Specifically, within the domains of dietary, social and pharmacological exposures, there was a relative sparsity in the use of MR to explore the implications of variation in systemic iron status on cardiovascular health, mediators of the effect of education, and effects and repurposing potential of antihypertensive drugs, respectively. The aim of this thesis was therefore to integrate developments in MR methodology with the availability of large-scale genetic data to explore these three areas further:

1. Cardiovascular and general health consequences of variation in iron status
2. Factors mediating the effect of educational attainment on CVD subtypes
3. Antihypertensive drug efficacy, side-effects and repurposing potential

The diversity of these exposures importantly offers opportunity for understanding and applying the MR framework in distinct contexts, while specifically considering effects on CVD as the outcome (Figure 1.7). This was expected to highlight the full potential of the MR approach, as well as how its application and overall strategy may need to be adapted in different settings.

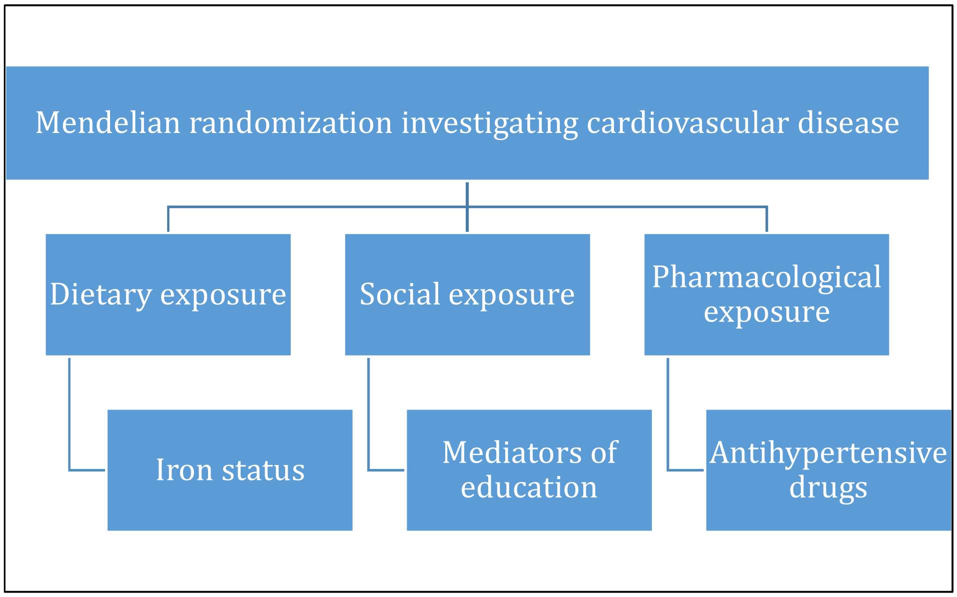


Figure 1.7. Structure of the thesis, in terms of the exposures explored in order to appreciate the full breadth of contexts to which MR can be applied. These make up Chapters 2-4 respectively, while Chapter 5 offers an overview of the areas covered and possible future directions.

The data sources used to investigate such a range of exposures and indeed CVD outcomes were similarly broad, and are summarised below in Table 1.3. Data analysis for this work was performed using R statistical software (version 3.4.3, The R Foundation for Statistical Computing). For two-sample MR analyses using the IVW, MR-Egger and weighted median approaches, the 'TwoSampleMR' package was applied (125). For MR-PRESSO, the 'MRPRESSO' package was used (88). For analysis of associations with multiple phenotypes in cohort data (also called 'phenome-wide association analysis'), the 'phecode' package was used (126). All other analyses were performed using raw code specifically generated for that purpose. Harmonization of genetic data from distinct sources were performed by aligning effect alleles, with no exclusions performed for palindromic variants. UK Biobank data were accessed through application 236. Any GWAS summary data originating from the UK Biobank that was used in this work were retrieved from analyses performed in previous studies, and have been cited appropriately. Individual participant data from the UK Biobank was used for phenome-wide association analysis. As detailed in the individual chapters, I performed this myself for the analysis of iron status (Chapter 2). For the phenome-wide association analysis of antihypertensive drug targets, this was performed by Fotios Koskeridis (University of Ioannina, Greece) for data relating to the UK Biobank cohort, with that in the Vanderbilt University Biobank performed by Lan Jiang, Qiping Feng, Wei-Qi Wei and Joshua C. Denny (Vanderbilt University Medical Center, United States of America).

1.9 Ethical approval

Ethical approval and participant consent for use of all data in this work had been previously obtained in their respective primary studies, and therefore was not required to be sought again here. The primary studies from which data were obtained have been cited in the Methods sections for the Chapters where they are introduced.

Table 1.3. Data sources used to obtain genetic association estimates for performing the various Mendelian randomization analyses in the thesis.

Chapter	Trait	Population ethnicity	Sample size	Reference
2	Serum iron biomarkers	European	48,972	(127)
2	Coronary artery disease	Multi-ethnic	60,801 cases and 123,504 controls	(44)
2	Stroke	Multi-ethnic	67,162 cases and 454,450 controls	(46)
2	Venous thromboembolism	European	of 7,507 cases and 52,632 controls	(128)
2	Carotid intima media thickness	European	71,128	(45)
2	Carotid plaque	European	21,540 cases and 26,894 controls	(45)
2	Multiple	White British	424,439	(129)
3	Education	European	1,131,881	(130)
3	Cognition	European	257,841	(130)
3	Coronary artery disease	Multi-ethnic	60,801 cases and 123,504 controls	(44)
3	Stroke	Multi-ethnic	67,162 cases and 454,450 controls	(46)
3	Body mass index	European	681,275	(131)
3	Systolic blood pressure	White British	318,417	(63)
3	Smoking	White British	462,690	(132)
4	Systolic blood pressure	European	757,601	(133)
4	Coronary artery disease	Multi-ethnic	60,801 cases and 123,504 controls	(44)
4	Stroke	Multi-ethnic	67,162 cases and 454,450 controls	(46)
4	Multiple – UK Biobank	White British	424,439	(129)
4	Multiple – Vanderbilt University Biobank	European	45,517	(126)

1.10 References

1. GBD. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017;390(10100):1151-210.

2. Dagenais GR, Leong DP, Rangarajan S, Lanas F, Lopez-Jaramillo P, Gupta R, et al. Variations in common diseases, hospital admissions, and deaths in middle-aged adults in 21 countries from five continents (PURE): a prospective cohort study. Lancet. 2019:S0140-6736(19)32007-0 [Epub ahead of print].

3. GBD. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017;390(10100):1211-59.

4. Sacco RL, Kasner SE, Broderick JP, Caplan LR, Connors JJ, Culebras A, et al. An updated definition of stroke for the 21st century: a statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2013;44(7):2064-89.

5. DeFilippis AP, Chapman AR, Mills NL, de Lemos JA, Arbab-Zadeh A, Newby LK, et al. Assessment and Treatment of Patients with Type 2 Myocardial Infarction and Acute Non-Ischemic Myocardial Injury. Circulation. 2019:10.1161/CIRCULATIONAHA.119.040631. [Epub ahead of print].

6. Sudlow CL, Warlow CP. Comparable studies of the incidence of stroke and its pathological types: results from an international collaboration. International Stroke Incidence Collaboration. Stroke. 1997;28(3):491-9.

7. Ay H, Benner T, Arsava EM, Furie KL, Singhal AB, Jensen MB, et al. A computerized algorithm for etiologic classification of ischemic stroke: the Causative Classification of Stroke System. Stroke. 2007;38(11):2979-84.

8. McArdle PF, Kittner SJ, Ay H, Brown RD, Jr., Meschia JF, Rundek T, et al. Agreement between TOAST and CCS ischemic stroke classification: the NINDS SiGN study. Neurology. 2014;83(18):1653-60.

9. Adams HP, Jr., Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. Stroke. 1993;24(1):35-41.

10. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke - the Framingham study. Stroke. 1991;22(8):983-8.

11. Lubitz SA, Yin X, Lin HJ, Kolek M, Smith JG, Trompet S, et al. Genetic Risk Prediction of Atrial Fibrillation. Circulation. 2017;135(14):1311-20.

12. Chang YJ, Ryu SJ, Lin SK. Carotid artery stenosis in ischemic stroke patients with nonvalvular atrial fibrillation. Cerebrovasc Dis. 2002;13(1):16-20.

13. Park YS, Chung PW, Kim YB, Moon HS, Suh BC, Yoon WT, et al. Small deep infarction in patients with atrial fibrillation: evidence of lacunar pathogenesis. Cerebrovasc Dis. 2013;36(3):205-10.

14. Moncayo J, Devuyst G, Van Melle G, Bogousslavsky J. Coexisting causes of ischemic stroke. Arch Neurol. 2000;57(8):1139-44.

15. Bamford J, Sandercock P, Dennis M, Burn J, Warlow C. Classification and natural history of clinically identifiable subtypes of cerebral infarction. Lancet. 1991;337(8756):1521-6.

16. Yeh SJ, Tang SC, Tsai LK, Jeng JS. Pathogenetical subtypes of recurrent intracerebral hemorrhage: designations by SMASH-U classification system. Stroke. 2014;45(9):2636-42.

17. Rannikmae K, Woodfield R, Anderson CS, Charidimou A, Chiewvit P, Greenberg SM, et al. Reliability of intracerebral hemorrhage classification systems: A systematic review. Int J Stroke. 2016;11(6):626-36.

18. Meretoja A, Strbian D, Putaala J, Curtze S, Haapaniemi E, Mustanoja S, et al. SMASH-U: a proposal for etiologic classification of intracerebral hemorrhage. Stroke. 2012;43(10):2592-7.

19. Vinters HV. Cerebral amyloid angiopathy. A critical review. Stroke. 1987;18(2):311-24.

20. Fisher CM. Pathological observations in hypertensive cerebral hemorrhage. J Neuropathol Exp Neurol. 1971;30(3):536-50.

21. Lewington S, Clarke R, Qizilbash N, Peto R, Collins R, Prospective Studies C. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. Lancet. 2002;360(9349):1903-13.

22. Emerging Risk Factors C, Di Angelantonio E, Sarwar N, Perry P, Kaptoge S, Ray KK, et al. Major lipids, apolipoproteins, and risk of vascular disease. JAMA. 2009;302(18):1993-2000.

23. Emerging Risk Factors C, Sarwar N, Gao P, Seshasai SR, Gobin R, Kaptoge S, et al. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. Lancet. 2010;375(9733):2215-22.

24. Herrington W, Lacey B, Sherliker P, Armitage J, Lewington S. Epidemiology of Atherosclerosis and the Potential to Reduce the Global Burden of Atherothrombotic Disease. Circ Res. 2016;118(4):535-46.

25. Prospective Studies C, Whitlock G, Lewington S, Sherliker P, Clarke R, Emberson J, et al. Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies. Lancet. 2009;373(9669):1083-96.

26. Droyvold WB, Midthjell K, Nilsen TI, Holmen J. Change in body mass index and its impact on blood pressure: a prospective population study. Int J Obes (Lond). 2005;29(6):650-5.

27. Schulz UGR, Rothwell PM. Differences in vascular risk factors between etiological subtypes of ischemic stroke - Importance of population-based studies. Stroke. 2003;34(8):2050-9.

28. NINDS-Stroke Genetics Network, International Stroke Genetics Consortium. Loci associated with ischaemic stroke and its subtypes: a genome-wide association study. Lancet Neurol. 2016;15(2):174-84.

29. Woo D, Falcone GJ, Devan WJ, Brown WM, Biffi A, Howard TD, et al. Meta-analysis of genome-wide association studies identifies 1q22 as a susceptibility locus for intracerebral hemorrhage. Am J Hum Genet. 2014;94(4):511-21.

30. Grau AJ, Weimar C, Buggle F, Heinrich A, Goertler M, Neumaier S, et al. Risk factors, outcome, and treatment in subtypes of ischemic stroke: the German stroke data bank. Stroke. 2001;32(11):2559-66.

31. Ariesen MJ, Claus SP, Rinkel GJ, Algra A. Risk factors for intracerebral hemorrhage in the general population: a systematic review. Stroke. 2003;34(8):2060-5.

32. Chen Z, Iona A, Parish S, Chen Y, Guo Y, Bragg F, et al. Adiposity and risk of ischaemic and haemorrhagic stroke in 0.5 million Chinese men and women: a prospective cohort study. Lancet Glob Health. 2018;6(6):e630-e40.

33. GBD. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018;392(10159):1736-88.

34. GBD. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018;392(10159):1923-94.

35. Yusuf S, Joseph P, Rangarajan S, Islam S, Mente A, Hystad P, et al. Modifiable risk factors, cardiovascular disease, and mortality in 155 722 individuals from 21 high-income, middle-income, and low-income countries (PURE): a prospective cohort study. Lancet. 2019:S0140-6736(19)32008-2. [Epub ahead of print].

36. Ezzati M, Lopez A, Rodgers A, Murray C. Comparative Quantification of Health Risks. Global and Regional Burden of Disease Attributable to Selected Major Risk Factors. vols 1 and 2. Geneva, Switzerland: World Health Organization; 2004:2248.

37. Ezzati M, Lopez AD, Rodgers A, Vander Hoorn S, Murray CJ, Comparative Risk Assessment Collaborating G. Selected major risk factors and global and regional burden of disease. Lancet. 2002;360(9343):1347-60.

38. Ezzati M, Henley SJ, Thun MJ, Lopez AD. Role of smoking in global and regional cardiovascular mortality. Circulation. 2005;112(4):489-97.

39. Ezzati M, Hoorn SV, Rodgers A, Lopez AD, Mathers CD, Murray CJ, et al. Estimates of global and regional potential health gains from reducing multiple major risk factors. Lancet. 2003;362(9380):271-80.

40. Tzoulaki I, Elliott P, Kontis V, Ezzati M. Worldwide Exposures to Cardiovascular Risk Factors and Associated Health Effects: Current Knowledge and Data Gaps. Circulation. 2016;133(23):2314-33.

41. Mills MC, Rahal C. A scientometric review of genome-wide association studies. Commun Biol. 2019;2:9.

42. Jones GT, Tromp G, Kuivaniemi H, Gretarsdottir S, Baas AF, Giusti B, et al. Meta-Analysis of Genome-Wide Association Studies for Abdominal Aortic Aneurysm Identifies Four New Disease-Specific Risk Loci. Circ Res. 2017;120(2):341-53.

43. Roselli C, Chaffin MD, Weng LC, Aeschbacher S, Ahlberg G, Albert CM, et al. Multi-ethnic genome-wide association study for atrial fibrillation. Nat Genet. 2018;50(9):1225-33.

44. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet. 2015;47(10):1121-30.

45. Franceschini N, Giambartolomei C, de Vries PS, Finan C, Bis JC, Huntley RP, et al. GWAS and colocalization analyses implicate carotid intima-media thickness and carotid plaque loci in cardiovascular outcomes. Nat Commun. 2018;9(1):5141.

46. Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. Nat Genet. 2018;50(4):524-37.

47. Klarin D, Lynch J, Aragam K, Chaffin M, Assimes TL, Huang J, et al. Genome-wide association study of peripheral artery disease in the Million Veteran Program. Nat Med. 2019;25(8):1274-9.

48. Khera AV, Emdin CA, Drake I, Natarajan P, Bick AG, Cook NR, et al. Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. N Engl J Med. 2016;375(24):2349-58.

49. Dichgans M, Malik R, Konig IR, Rosand J, Clarke R, Gretarsdottir S, et al. Shared genetic susceptibility to ischemic stroke and coronary artery disease: a genome-wide analysis of common variants. Stroke. 2014;45(1):24-36.

50. Davey Smith G, Timpson N, Ebrahim S. Strengthening causal inference in cardiovascular epidemiology through Mendelian randomization. Ann Med. 2008;40(7):524-41.

51. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. J Am Stat Assoc. 1996;91(434):444-55.

52. Davey Smith G, Ebrahim S. Mendelian randomization: can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003;32(1):1-22.

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

54. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. PLOS Genet. 2017;13(11):e1007081.

55. Gage SH, Jones HJ, Burgess S, Bowden J, Davey Smith G, Zammit S, et al. Assessing causality in associations between cannabis use and schizophrenia risk: a two-sample Mendelian randomization study. Psychol Med. 2017;47(5):971-80.

56. Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. Stat Methods Med Res. 2007;16(4):309-30.

57. 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.

58. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013;37(7):658-65.

59. Thompson JR, Minelli C, Del Greco MF. Mendelian Randomization using Public Data from Genetic Consortia. Int J Biostat. 2016;12(2).

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

61. Slob EAW, Burgess S. A Comparison Of Robust Mendelian Randomization Methods Using Summary Data. bioRxiv. 2019:577940.

62. Sheehan NA, Didelez V, Burton PR, Tobin MD. Mendelian randomisation and causal inference in observational epidemiology. PLoS Med. 2008;5(8):e177.

63. Carter AR, Gill D, Davies NM, Taylor AE, Tillmann T, Vaucher J, et al. Understanding the consequences of education inequality on cardiovascular disease: mendelian randomisation study. BMJ. 2019;365:l1855.

64. Tillmann T, Vaucher J, Okbay A, Pikhart H, Peasey A, Kubinova R, et al. Education and coronary heart disease: mendelian randomisation study. BMJ. 2017;358:j3542.

65. Gill D, Brewer CF, Del Greco MF, Sivakumaran P, Bowden J, Sheehan NA, et al. Age at menarche and adult body mass index: a Mendelian randomization study. Int J Obes (Lond). 2018;42(9):1574-81.

66. Gill D, Del Greco M F, Rawson TM, Sivakumaran P, Brown A, Sheehan NA, et al. Age at menarche and time spent in education: a Mendelian randomization study. Behav Genet. 2017:47(5):480-5.

67. Swerdlow DI, Preiss D, Kuchenbaecker KB, Holmes MV, Engmann JE, Shah T, et al. HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: evidence from genetic analysis and randomised trials. Lancet. 2015;385(9965):351-61.

68. Del Greco M F, Minelli C, Sheehan NA, Thompson JR. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. Stat Med. 2015;34(21):2926-40.

69. Swerdlow DI, Kuchenbaecker KB, Shah S, Sofat R, Holmes MV, White J, et al. Selecting instruments for Mendelian randomization in the wake of genome-wide association studies. Int J Epidemiol. 2016;45(5):1600-16.

70. 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.

71. Brion MJ, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. Int J Epidemiol. 2013;42(5):1497-501.

72. Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan N, Thompson J. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. Stat Med. 2017:36(11):1783-802.

73. Burgess S, Bowden J. Integrating summarized data from multiple genetic variants in Mendelian randomization: bias and coverage properties of inverse-variance weighted methods. 2015.

74. Wald A. The Fitting of Straight Lines if Both Variables are Subject to Error. Ann Math Stat. 1940;11(3):284-300.

75. Bowden J, Del Greco MF, Minelli C, Zhao Q, Lawlor DA, Sheehan NA, et al. Improving the accuracy of two-sample summary-data Mendelian randomization: moving beyond the NOME assumption. Int J Epidemiol. 2018:48(3):728-42.

76. Burgess S, Zuber V, Valdes-Marquez E, Sun BB, Hopewell JC. Mendelian randomization with fine-mapped genetic data: Choosing from large numbers of correlated instrumental variables. Genet Epidemiol. 2017;41(8):714-25.

77. Gill D, Sheehan NA, Wielscher M, Shrine N, Amaral AFS, Thompson JR, et al. Age at menarche and lung function: a Mendelian randomization study. Eur J Epidemiol. 2017:32(8):701-10.

78. Perry JR, Day F, Elks CE, Sulem P, Thompson DJ, Ferreira T, et al. Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. Nature. 2014;514(7520):92-7.

79. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, et al. Genome-wide association study identifies 74 loci associated with educational attainment. Nature. 2016;533(7604):539-42.

80. Ference BA, Kastelein JJP, Ginsberg HN, Chapman MJ, Nicholls SJ, Ray KK, et al. Association of Genetic Variants Related to CETP Inhibitors and Statins With Lipoprotein Levels and Cardiovascular Risk. JAMA. 2017;318(10):947-56.

81. Ference BA, Majeed F, Penumetcha R, Flack JM, Brook RD. Effect of naturally random allocation to lower low-density lipoprotein cholesterol on the risk of coronary heart disease mediated by polymorphisms in NPC1L1, HMGCR, or both: a 2 x 2 factorial Mendelian randomization study. J Am Coll Cardiol. 2015;65(15):1552-61.

82. Ference BA, Robinson JG, Brook RD, Catapano AL, Chapman MJ, Neff DR, et al. Variation in PCSK9 and HMGCR and Risk of Cardiovascular Disease and Diabetes. N Engl J Med. 2016;375(22):2144-53.

83. Palmer TM, Lawlor DA, Harbord RM, Sheehan NA, Tobias JH, Timpson NJ, et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. Stat Methods Med Res. 2012;21(3):223-42.

84. Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity analyses for bobust causal inference from Mendelian randomization analyses with multiple genetic variants. Epidemiology. 2017;28(1):30-42.

85. Paaby AB, Rockman MV. The many faces of pleiotropy. Trends Genet. 2013;29(2):66-73.

86. 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.

87. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. Eur J Epidemiol. 2017;32(5):377-89.

88. Verbanck M, Chen CY, 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.

89. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. Am J Epidemiol. 2015;181(4):251-60.

90. Burgess S, Thompson DJ, Rees JMB, Day FR, Perry JR, Ong KK. Dissecting Causal Pathways Using Mendelian Randomization with Summarized Genetic Data: Application to Age at Menarche and Risk of Breast Cancer. Genetics. 2017;207(2):481-7.

91. Rees JMB, Wood AM, Burgess S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. Stat Med. 2017;36(29):4705-18.

92. Bowden J, Burgess S, Davey Smith G. Difficulties in Testing the Instrument Strength Independent of Direct Effect Assumption in Mendelian Randomization. JAMA Cardiol. 2017;2(8):929-30.

93. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol. 2016;40(4):304-14.

94. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. Int J Epidemiol. 2017;46(6):1985-98.

95. Thompson JR, Minelli C, Bowden J, Del Greco FM, Gill D, Jones EM, et al. Mendelian randomization incorporating uncertainty about pleiotropy. Stat Med. 2017;36(29):4627-45.

96. Burgess S, Daniel RM, Butterworth AS, Thompson SG, Consortium E-IA. Network Mendelian randomization: using genetic variants as instrumental variables to investigate mediation in causal pathways. Int J Epidemiol. 2015;44(2):484-95.

97. Swanson SA, Hernan MA. The challenging interpretation of instrumental variable estimates under monotonicity. Int J Epidemiol. 2017:1;47(4):1289-97.

98. Burgess S, Labrecque JA. Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. Eur J Epidemiol. 2018;33(10):947-52.

99. Gill D, Georgakis MK, Laffan M, Sabater-Lleal M, Malik R, Tzoulaki I, et al. Genetically Determined FXI (Factor XI) Levels and Risk of Stroke. Stroke. 2018;49(11):2761-3.

100. Staiger D, Stock JH. Instrumental variables regression with weak instruments. Econometrica. 1997;65(3):557-86.

101. Li BB, Martin EB. An approximation to the F distribution using the chi-square distribution. Comput Stat Data An. 2002;40(1):21-6.

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

103. Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. Int J Epidemiol. 2016:1;45(6):1961-74.

104. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. Nat Genet. 2001;29(3):306-9.

105. Gkatzionis A, Burgess S. Contextualizing selection bias in Mendelian randomization: how bad is it likely to be? Int J Epidemiol 2018;48(3):691-701.

106. Day FR, Loh PR, Scott RA, Ong KK, Perry JR. A robust example of collider bias in a genetic association study. Am J Hum Genet. 2016;98(2):392-3.

107. Vansteelandt S, Dukes O, Martinussen T. Survivor bias in Mendelian randomization analysis. Biostatistics. 2018;19(4):426-43.

108. Smit RA, Trompet S, Dekkers OM, Jukema JW, le Cessie S. Survival bias in Mendelian randomization studies: a threat to causal inference. Epidemiology. 2019:10.1097/EDE.0000000000001072. [Epub ahead of print].

109. Noyce AJ, Kia DA, Hemani G, Nicolas A, Price TR, De Pablo-Fernandez E, et al. Estimating the causal influence of body mass index on risk of Parkinson disease: A Mendelian randomisation study. PLoS Med. 2017;14(6):e1002314.

110. Macias-Kauffer LR, Villamil-Ramirez H, Leon-Mimila P, Jacobo-Albavera L, Posadas-Romero C, Posadas-Sanchez R, et al. Genetic contributors to serum uric acid levels in Mexicans and their effect on premature coronary artery disease. Int J Cardiol. 2019;279:168-73.

111. Yan DD, Wang J, Jiang F, Zhang R, Wang T, Wang SY, et al. A causal relationship between uric acid and diabetic macrovascular disease in Chinese type 2 diabetes patients: A Mendelian randomization analysis. Int J Cardiol. 2016;214:194-9.

112. Testa A, Prudente S, Leonardis D, Spoto B, Sanguedolce MC, Parlongo RM, et al. A genetic marker of hyperuricemia predicts cardiovascular events in a meta-analysis of three cohort studies in high risk patients. Nutr Metab Cardiovasc Dis. 2015;25(12):1087-94.

113. Kleber ME, Delgado G, Grammer TB, Silbernagel G, Huang J, Kramer BK, et al. Uric Acid and Cardiovascular Events: A Mendelian Randomization Study. J Am Soc Nephrol. 2015;26(11):2831-8.

114. Mallamaci F, Testa A, Leonardis D, Tripepi R, Pisano A, Spoto B, et al. A genetic marker of uric acid level, carotid atherosclerosis, and arterial stiffness: a family-based study. Am J Kidney Dis. 2015;65(2):294-302.

115. Palmer TM, Nordestgaard BG, Benn M, Tybjaerg-Hansen A, Davey Smith G, Lawlor DA, et al. Association of plasma uric acid with ischaemic heart disease and blood pressure: mendelian randomisation analysis of two large cohorts. BMJ. 2013;347:f4262.

116. Oikonen M, Wendelin-Saarenhovi M, Lyytikainen LP, Siitonen N, Loo BM, Jula A, et al. Associations between serum uric acid and markers of subclinical atherosclerosis in young adults. The cardiovascular risk in Young Finns study. Atherosclerosis. 2012;223(2):497-503.

117. Keenan T, Zhao W, Rasheed A, Ho WK, Malik R, Felix JF, et al. Causal Assessment of Serum Urate Levels in Cardiometabolic Diseases Through a Mendelian Randomization Study. J Am Coll Cardiol. 2016;67(4):407-16.

118. White J, Sofat R, Hemani G, Shah T, Engmann J, Dale C, et al. Plasma urate concentration and risk of coronary heart disease: a Mendelian randomisation analysis. Lancet Diabetes Endocrinol. 2016;4(4):327-36.

119. Gill D, Benyamin B, Moore LSP, Monori G, Zhou A, Koskeridis F, et al. Associations of genetically determined iron status across the phenome: A mendelian randomization study. PLoS Med. 2019;16(6):e1002833.

120. GBD. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016;388(10053):1545-602.

121. Low MS, Speedy J, Styles CE, De-Regil LM, Pasricha SR. Daily iron supplementation for improving anaemia, iron status and health in menstruating women. Cochrane Database Syst Rev. 2016;4:CD009747.

122. Wright JM, Musini VM, Gill R. First-line drugs for hypertension. Cochrane Database Syst Rev. 2018;4:CD001841.

123. Walker VM, Davey Smith G, Davies NM, Martin RM. Mendelian randomization: a novel approach for the prediction of adverse drug events and drug repurposing opportunities. Int J Epidemiol. 2017;46(6):2078-89.

124. Sofat R, Hingorani AD, Smeeth L, Humphries SE, Talmud PJ, Cooper J, et al. Separating the mechanism-based and off-target actions of cholesteryl ester transfer protein inhibitors with CETP gene polymorphisms. Circulation. 2010;121(1):52-62.

125. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. eLife. 2018;7.

126. Wei WQ, Bastarache LA, Carroll RJ, Marlo JE, Osterman TJ, Gamazon ER, et al. Evaluating phecodes, clinical classification software, and ICD-9-CM codes for phenome-wide association studies in the electronic health record. PLoS One. 2017;12(7):e0175508.

127. Benyamin B, Esko T, Ried JS, Radhakrishnan A, Vermeulen SH, Traglia M, et al. Novel loci affecting iron homeostasis and their effects in individuals at risk for hemochromatosis. Nat Commun. 2014;5:4926.

128. Germain M, Chasman DI, de Haan H, Tang W, Lindstrom S, Weng LC, et al. Meta-analysis of 65,734 individuals identifies TSPAN15 and SLC44A2 as two susceptibility loci for venous thromboembolism. Am J Hum Genet. 2015;96(4):532-42.

129. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLoS Med. 2015;12(3):e1001779.

130. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet. 2018;50(8):1112-21.

131. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700,000 individuals of European ancestry. bioRxiv. 2018.

132. Wootton RE, Richmond RC, Stuijfzand BG, Lawn RB, Sallis HM, Taylor GMJ, et al. Causal effects of lifetime smoking on risk for depression and schizophrenia: Evidence from a Mendelian randomisation study. bioRxiv. 2018.

133. Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, 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.

Chapter 2: Implications of iron status on cardiovascular and thrombotic disease

All of the work presented in this chapter is my own, unless otherwise indicated in the text.

Related publications

- **Gill D**, Benyamin B, Moore LSP, Monori G, Zhou A, Koskeridis F, Evangelou E, Laffan M, Walker AP, Tsilidis KK, Dehghan A, Elliott P, Hypponen E and Tzoulaki I. Associations of genetically determined iron status across the phenome: A mendelian randomization study. PLoS Med. 2019;16:e1002833.
- **Gill D**, Brewer CF, Monori G, Tregouet DA, Franceschini N, Giambartolomei C, Consortium I, Tzoulaki I and Dehghan A. Effects of Genetically Determined Iron Status on Risk of Venous Thromboembolism and Carotid Atherosclerotic Disease: A Mendelian Randomization Study. J Am Heart Assoc. 2019;8:e012994.
- **Gill D**, Monori G, Tzoulaki I and Dehghan A. Iron Status and Risk of Stroke: A Mendelian Randomization Study. Stroke. 2018;49:2815-2821.
- **Gill D**, Del Greco M F, Walker AP, Srai SKS, Laffan MA and Minelli C. The effect of iron status on risk of coronary artery disease: a mendelian randomization study. Arterioscler Thromb Vasc Biol. 2017;37:1788-1792.

Data sources

- CARDIoGRAMplusC4D Consortium
- CHARGE Consortium
- GIS Consortium
- INVENT Consortium
- MEGASTROKE Consortium
- UK Biobank

2.1 Introduction

Iron is involved in a number of fundamental biological processes, including red blood cell production and metabolism (1). Systemic iron status can vary considerably in individuals, with this having associated clinical implications (2, 3). Relative iron deficiency can be caused by an inadequate diet, abnormal gastrointestinal absorption, and increased requirements or losses (4). This can in turn result in iron deficiency anaemia, which currently affects approximately 1.2 billion people worldwide and causes 35 million years lived with disability per annum (5). In comparison, iron overload is most commonly related to medical treatments such as blood transfusion for severe anaemia, thalassemia, or haemochromatosis (6). The ability to modify systemic iron status in both the healthy population and in those with abnormally low or high levels makes understanding the effects of fluctuations in iron status an important research pursuit.

Thrombosis is the common underlying aetiology for many forms of cardiovascular disease, including coronary artery disease (CAD), ischemic stroke (IS) and venous thromboembolism (VTE), and also represents the primary cause of mortality and disability worldwide (7-10). Iron status has been implicated in thrombosis, and specifically in the formation of blood clots mediated by endothelial oxidative stress and increased blood viscosity (11, 12). However, the available evidence investigating the association between iron status and risk of thrombotic diseases is mixed. Supporting a harmful effect of higher iron levels, a lower risk of heart disease was observed in premenopausal women as compared to men and postmenopausal women, which was attributed to iron and blood loss during menstrual bleeding (13). Observational study has also positively associated higher iron stores with risk factors for cardiovascular disease, including type 2 diabetes mellitus (14). Mutations that cause hereditary haemochromatosis have been related to increased risk of cardiovascular disease (CVD) (15). However, this contradicts the findings of a meta-analysis of observational studies that supports a protective effect of iron status in CAD (16). Additionally, some observational studies have linked lower iron status to an increased risk of stroke (17-19), while others find an opposite relationship (19-21), or no association (22-24). In regard to atherosclerosis and VTE, these disease processes have been associated with both iron deficiency and iron overload (11, 25-30).

Research investigating the observed associations of iron status may be biased by unmeasured or unknown environmental confounding factors, or by reverse causation where the outcomes being studied themselves affect iron status. By utilising genetic variants related to systemic iron status for the study of its effects, these limitations can be overcome because their random distribution during meiosis limits confounding from environmental factors and minimises reverse causation bias (31, 32). Such a Mendelian randomization (MR) approach has already been applied to investigate the effect of iron status on risk of Parkinson's disease (33). In the same way that traditional observational studies explore the association between a measured exposure and an outcome, MR considers the link between a genetic variant that is related to the exposure of interest, and the outcome (31). Importantly, this allows MR to draw causal inferences on the nature of these relationships.

The MR principle can be similarly applied a range of outcome traits. When agnostically investigating traits across the human phenome, the term MR-phenome-wide association study (MR-PheWAS) can be used to describe the approach (34). Such a strategy allows for the rapid study of any potential clinical implications related to varying an exposure (for example, systemic iron status), and offers a means to inform the direction of future research (35).

The genetic instrumental variables used to proxy the exposure of interest in an MR study should be associated with that exposure, which is systemic iron status in this current work. There are a number of known biomarkers of iron status that provide a quantifiable measure of systemic iron levels, including serum iron, transferrin, transferrin saturation and ferritin (36). Genetic instrumental variables that proxy systemic iron status would be expected to have concordant associations with all of these markers in a pattern reflecting an overall association with systemic iron status. Specifically, increasing iron status would be related to increasing serum iron, transferrin saturation, and ferritin and decreasing transferrin levels (36). Genetic instrumental variables used in MR to investigate the effect of systemic iron status on a phenotypic outcome of interest should also only relate to that outcome through effects on the exposure, and not by some other pleiotropic pathway. Such effects of the variants on the outcome that are independent of the exposure are described as pleiotropy, and represent violations of a requisite assumption of MR, potentially biasing the consequent estimates.

Genome-wide association studies (GWASs) on a range of cardiovascular and thrombotic diseases have been performed, thus providing genetic association estimates for consequent MR analyses. This current work considers the effect of systemic iron status on CAD, IS and its subtypes, carotid plaque and atherosclerosis, and VTE. In this context, CAD relates to atherosclerotic disease of the coronary arteries, resulting in disruption of the cardiac blood supply, with associated ischaemia and increased risk of thrombosis. In contrast, IS is defined as a neurological deficit of the central nervous system caused by ischaemia (37), and may be further categorised by the underlying aetiology - large artery stroke (LAS), cardioembolic stroke (CES), and small vessel stroke (SVS) (38). Carotid artery intima-media thickness (cIMT) and carotid plaque are measures of carotid artery narrowing, with cIMT associated with carotid artery thickening following shear stress, and carotid plaque more representing inflammatory atherosclerotic changes in the vessel wall (39-41). In addition, MR-PheWAS of genetically determined iron status was performed using data from the UK Biobank to measure genetic associations with a broad range of health outcomes. In view of the established role of iron in pivotal physiological processes (3, 42), and the potential for therapeutic manipulation of systemic levels to optimise health outcomes, the purpose of these analyses was to decipher the effect of varying iron status on cardiovascular and thrombotic disease, as well its wider implications across the human phenome.

2.2 Methods

Genetic association estimates

Single-nucleotide polymorphism (SNP)-iron status biomarker association estimates

Higher systemic iron status is associated with higher serum iron, ferritin and transferrin saturation, and lower transferrin (36). Therefore, genetic instrumental variables for systemic iron status were selected as SNPs that had genome-wide significant associations ($P<5\text{x}10^{-8}$) with increased serum iron, ferritin and transferrin saturation, and decreased transferrin levels, thus proxying increased systemic iron status (43). A GWAS performed by the Genetics of Iron Status (GIS) Consortium aggregated data from eleven discovery cohorts and eight replication cohorts (Table 2.1) to obtain association estimates between SNPs and biomarkers of iron status for a total of 48,972 European subjects, adjusting for age and principal component scores (2).

The GIS Consortium study identified 12 SNPs across 11 loci that were associated with any of these four biomarkers of systemic iron status at genome-wide significance (Table 2.2) (2). Of these 12 SNPs, only 3 had genome-wide significant associations with all four biomarkers in a pattern consistent with an increase in overall iron status, specifically increasing serum iron, transferrin saturation and ferritin, but decreasing transferrin. These SNPs were rs1800562 and rs1799945 in the hemochromatosis (*HFE*) gene, and rs855791 in the transmembrane protease, serine 6 (*TMPRSS6*) gene (2). These three SNPs offered viable instruments for systemic iron status, and were taken forward to MR analysis. There was low linkage disequilibrium (LD) correlation between the two SNPs (rs1800562 and rs1799945) in the HFE gene (LD $r^2<0.01$) (2), consistent with there being negligible correlation between them (44).

Considering the biological effects of the HFE and TMPRSS6 proteins to justify their relation to iron status, HFE is a membrane protein that modulates iron absorption by competitively inhibiting the TRF1 tranferrin receptor (45). When saturation of transferrin and by relation, systemic iron status, is high, the HFE protein is able to activate a transferrin receptor 2 (TFR2) protein complex to enhance expression of hepcidin, an iron transporter regulator (46). Hepcidin in turn limits production of the iron export protein ferroportin in enterocytes and macrophages, thus reducing uptake of iron through the hepatic portal system (47, 48). In this way, hepcidin reduces iron absorption. Through a contrasting mechanism, the transmembrane serine protease TMPRSS6 inhibits hepcidin production when systemic iron levels are depleted, to consequently increase uptake (49).

The first-stage regression (F) statistic was calculated for each variant as a measure of its strength as an instrument, using the formula $F = (R^2)/((1-R^2)/(N-2))$, with R^2 being the proportion of the iron status biomarker variance explained by the variant and N the total sample size (44).

Table 2.1. Cohort details for the studies contributing to the Genetics of Iron Status Consortium genome-wide association study. PMID: PubMed identifier; SD: standard deviation.

Cohort	Study	Discovery/ Replication	References (PMID)	Sample size	Sex	Mean age +/- SD (years)	Covariates
Australia-Adult	QIMR Berghofer Adult	Discovery	19820699; 21151130; 20802479	3432	MALE	47.5 +/- 12.3	Age, 5 PCs
				5716	FEMALE	46.0 +/- 12.8	
Australia-Adolescent	QIMR Berghofer Adolescent	Discovery	17539372	1230	MALE	14.6 +/- 2.0	Age, 5 PCs
				1314	FEMALE	14.9 +/- 2.3	
Estonia (original)	Estonian Genome Project	Discovery	24518929	440	MALE	37.3 +/- 15.4	Age, sex, 5 PCs
				453	FEMALE	37.5 +/- 15.7	
Val Borbera	Val Borbera Study	Discovery	19847309	733	MALE	54.4 +/- 18.4	Age, 5 PCs
				926	FEMALE	54.8 +/- 18.7	
NBS	Nikmegen Biomedial Study	Discovery	16254196; 18794855	889	MALE	66.3 +/- 7.1	
				902	FEMALE	56.6 +/- 10.8	
Cambridge	UK Blood Services (UKBS) Common Controls panel	Discovery	17554300	1198	MALE	45.1 +/- 11.9	
				1221	FEMALE	42.1 +/- 12.7	
Micros/EURAC	Micros/EURAC	Discovery	17550581	528	MALE	45.5 +/- 15.8	
				690	FEMALE	46.0 +/- 16.7	
ERF/Rotterdam	ERF/Rotterdam	Discovery	15054401; 16877869	342	MALE	54.6 +/- 14.1	Age
				529	FEMALE	52.8 +/- 15.1	
KORA F3	Kooperative Gesundheitsforschung in der Region Augsburg	Discovery	16032513; 16032514	809	MALE	63.0 +/- 10.1	Age
				825	FEMALE	62.1 +/- 10.1	
KORA F4	Kooperative Gesundheitsforschung in der Region Augsburg	Discovery	16032513; 16032514	882	MALE	61.2 +/- 8.9	Age
				927	FEMALE	60.6 +/- 8.8	
BHS	Busselton Health Study	Discovery	19643935	397	MALE	54.0 +/- 15.4	
				480	FEMALE	55.5 +/- 14.9	
Estonia (replication)	Estonian Genome Project	Replication	24518929	547	MALE	54.4 +/- 16.1	Age, sex, 5 PCs
				470	FEMALE	53.4 +/- 15.9	
InCHIANTI	InCHIANTI study	Replication	19880490	536	MALE	67.1 +/- 15.3	Age, sex, centre
				670	FEMALE	69.1 +/- 15.6	
SardiNIA	SardiNIA study on aging	Replication	16934002	2051	MALE	43.7 +- 18.1	Age, age-squared, sex
				2643	FEMALE	43.1 +/- 17.3	
CoLAUS	Cohorte Lausanne	Replication	18366642	2550	MALE	52.9 +/- 10.8	Age, sex, first 5 ancestry PCs
				2869	FEMALE	52.9 +/- 10.8	
PREVEND	Prevention of Renal and Vascular Endstage Disease	Replication		1875	MALE	50.9 +/- 12.8	Age, sex, first 5 PCs
				1769	FEMALE	48.2 +/- 12.0	
FENLAND	Fenland Study	Replication	21248185	615	MALE	44.5 +/- 7.4	Age, sex, 4 PCs
				787	FEMALE	45.4 +/- 7.2	
INTERACT (cases)	InterAct (cases)	Replication	21717116	2087	MALE	54.7 +/- 8.0	Age, sex, centre, 5 PCs
				2251	FEMALE	55.6 +/- 8.3	
INTERACT (subcohort)	InterAct (controls)	Replication	21717116	1816	MALE	52.2 +/- 9.2	Age, sex, centre, 5 PCs
				3140	FEMALE	51.7 +/- 9.6	

*Table 2.2. Genetic variants associated with iron status biomarkers in the Genetics of Iron Status Consortium genome-wide association study. The effect estimates (beta) are provided in standard deviation units. A1: effect allele; A1 Freq: effect allele frequency; SE: standard error; * denotes SNPs used in the Mendelian randomization analysis.*

				Iron			Transferrin			Transferring Saturation			Log10 Ferritin		
SNP	Gene	A1	A1 Freq	Beta	SE	P	Beta	SE	P	Beta	SE	P	Beta	SE	P
rs744653	*WDR75–SLC40A1*	T	0.854	0.004	0.010	0.702	0.068	0.010	1.35E−11	−0.028	0.011	0.008	−0.089	0.010	8.37E−19
rs8177240	*TF*	T	0.669	−0.066	0.007	6.65E−20	−0.380	0.007	8.43E−610	0.100	0.008	7.24E−38	0.021	0.007	0.004
rs9990333	*TFRC*	T	0.460	0.017	0.007	0.014	−0.051	0.007	1.95E−13	0.039	0.007	7.28E−8	0.001	0.007	0.878
rs1800562*	*HFE (C282Y)*	A	0.067	0.328	0.016	2.72E−97	−0.479	0.016	8.90E−196	0.577	0.016	2.19E−270	0.204	0.016	1.54E−38
rs1799945*	*HFE (H63D)*	C	0.850	−0.189	0.010	1.10E−81	0.114	0.010	9.36E−30	−0.231	0.010	5.13E−109	−0.065	0.010	1.71E−10
rs7385804	*TFR2*	A	0.621	0.064	0.007	1.36E−18	−0.003	0.007	0.728	0.054	0.008	6.07E−12	0.015	0.007	0.039
rs4921915	*NAT2*	A	0.782	0.004	0.009	0.633	0.079	0.009	7.05E−19	−0.026	0.009	0.004	0.001	0.009	0.886
rs651007	*ABO*	T	0.202	−0.004	0.009	0.611	−0.001	0.009	0.916	−0.006	0.009	0.498	−0.050	0.009	1.31E−8
rs6486121	*ARNTL*	T	0.631	−0.009	0.007	0.202	−0.046	0.007	3.89E−10	0.015	0.008	0.048	0.006	0.007	0.424
rs174577	*FADS2*	A	0.330	0.001	0.007	0.878	0.062	0.007	2.28E−17	−0.025	0.008	0.002	−0.012	0.007	0.098
rs411988	*TEX14*	A	0.564	−0.002	0.007	0.770	0.014	0.007	0.052	−0.012	0.007	0.115	−0.044	0.007	1.59E−10
rs855791*	*TMPRSS6 (V736A)*	A	0.446	−0.181	0.007	1.32E−139	0.044	0.007	1.98E10−9	−0.190	0.008	6.41E−137	−0.055	0.007	1.38E−14

Coronary artery disease genetic association estimates

Publicly available summary data from the CARDIoGRAMplusC4D 1000 Genomes GWAS (CARDIoGRAMplusC4D 1000G) and CARDIoGRAMplusC4D Metabochip were used to derive SNP-CAD genetic association estimates (50, 51). Specifically, CARDIoGRAMplusC4D 1000G included 60,801 CAD cases and 123,504 controls, with adjustment made for population stratification using the genomic control method (51). Cases and controls were of European (approximately 75%), East and South Asian, Hispanic and African American ancestry (51). The definition for CAD varied between studies, typically including individuals with a documented history of acute coronary syndrome, coronary artery bypass grafting, percutaneous coronary revascularization, coronary stenosis greater than 50% in at least one coronary vessel, or angina pectoris (51). The CARDIoGRAMplusC4D Metabochip study meta-analysed data on 63,746 CAD cases and 130,681 controls, which were genotyped using either the Metabochip array or GWAS data imputed using HapMap (50). CARDIoGRAMplusC4D Metabochip participants were of either European (approximately 95%) or South Asian ancestry (50). The CAD cases were diagnosed using criteria similar to those as for CARDIoGRAMplusC4D 1000G, with corrections made for age, sex and population stratification in the association analysis (50). Overall SNP-CAD association estimates were obtained by meta-analysing results from CARDIoGRAMplusC4D 1000G and CARDIoGRAMplusC4D Metabochip summary genetic association data after accounting for participant overlap between the studies, which consisted of 34,997 cases and 49,512 controls (52), using a ‘decoupling’ approach that transforms the covariance structure of the data from the studies so that consequent meta-analysis is able to assume independence (52).

Stroke genetic association estimates

Genetic association estimates for stroke were obtained from a GWAS meta-analysis performed by the MEGASTROKE Consortium that combined data from 67,162 stroke cases and 454,450 controls. Participants were of European (approximately 87%), East, South and mixed Asian, African, and Latin American ancestry. Considering IS subtypes, there were 9,006 CES cases, 6,688 LAS cases, and 11,710 SVS cases (53). All studies included in the meta-analysis adjusted for age and sex as covariates, and stroke was defined using the World Health Organization definition of sudden onset neurological deficit related to a vascular cause lasting more than 24 hours, with IS further categorised into subtypes as per the Trial of Org 10172 in Acute Stroke Treatment criteria (53, 54).

Venous thromboembolism genetic association estimates

The SNP-VTE genetic association estimates were obtained from a GWAS meta-analysis performed by the International Network on Venous Thrombosis Consortium (55), which consisted of 12 studies including participants of European descent, providing a total of 7,507 VTE cases and 52,632 controls. The VTE diagnoses included deep vein thrombosis or pulmonary embolism, and were made by a physician following relevant evaluation.

Carotid intima media thickness and carotid plaque genetic association estimates

The Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium performed a GWAS meta-analysis from which summary data was used to obtain SNP-cIMT and SNP-carotid plaque genetic association estimates (56). This meta-analysis consisted of 31 studies and 71,128 participants for cIMT, and 17 studies with 21,540 cases and 26,894 controls for carotid plaque. All participants for both traits were European heritage. Carotid traits were diagnosed using high-resolution B-mode ultrasonography (57), with carotid plaque defined as atherosclerotic thickening of the carotid artery wall or luminal stenosis greater than 25%, and cIMT measured as the mean of maximal values from several common carotid artery measurements, taken in millimetres.

Phenome-wide association study genetic association estimates

Genetic association estimates for PheWAS were obtained from the UK Biobank, a prospective cohort study consisting of approximately half a million participants in the United Kingdom that were aged 40-69 years at recruitment between 2006 and 2010 (58). Genetic data was obtained from blood samples, and were linked to National Health Service Hospital Episode Statistics (HES) from April 1995 to March 2016 (58). Analysis was only performed in participants of self-reported European descent, and one participant from each pair of relatives was randomly excluded based on a kinship coefficient of >0.0884 in order to minimise bias from related individuals. The International Classification of Diseases (ICD) versions 9 and 10 were used to ascertain cases from the HES data, including both incident and prevalent cases (as allocation of genetic instruments occurs at conception, preventing reverse causation bias). The 'phecode grouping system' was used to categorise HES diagnoses into groups based on clinically-related phenotypes (59). Cases were selected as having one documented event, with controls identified as all individuals with no record of any related outcomes (within that phecode category) (60). To generate genetic association estimates, case and control groups were created for each respective phecode, and logistic regression was used to estimate associations with each instrument SNP after adjusting for age, sex, the type of genotyping array used and the first four genetic principal components. Only phecodes where there were 200 or more cases were considered to improve statistical power (61, 62).

Mendelian randomization

For all outcomes, the ratio method was used to generate MR estimates for individual instrument SNPs, with fixed-effect inverse-variance weighted (IVW) meta-analysis performed to pool MR estimates across instruments (44, 63). The Delta method (second order weights) was used to estimate standard errors (63). MR estimates were scaled to the effect on serum iron (measured in standard deviation units, 6.1μmol/L) for the main analysis, with sensitivity analyses also performed with MR estimates scaled to effects on transferrin saturation, ferritin and transferrin (all also in standard deviation units). A statistical significance threshold of $P<0.05$ was used for directed analysis considering cardiovascular and thrombotic outcomes. No adjustment was made for multiple testing of the four iron status biomarkers, as they were each serving as a proxy for systemic iron status. Adjustment for multiple testing of various traits considered in the directed analysis of cardiovascular and thrombotic outcomes was also not made because each of these were specifically selected based on existing supportive evidence. Multiple outcomes were also investigating the same underlying mechanism – CAD and carotid plaque exploring effects on atherosclerosis, and VTE and CES exploring effects on thrombosis related to stasis of blood. For outcome traits throughout the phenome identified in PheWAS, statistical significance of MR estimates was established using the false discovery rate (FDR) method, with a 5% cut off applied (64).

In MR, pleiotropy refers to the scenario where the genetic instrumental variables affect the outcome under consideration through a pathway that is independent of the exposure under study, and can create bias (32, 65). Excess heterogeneity (beyond that expected by chance) between MR estimates arising from different instrument SNPs chance can be used to identify the presence of such a phenomenon (66). This was investigated using the Cochran's Q test, with $P<0.05$ interpreted as evidence of heterogeneity and pleiotropy. To further explore for possible pleiotropy arising in the MR analysis of all considered outcomes, including the targeted analysis of the cardiovascular and thrombotic diseases, the PhenoScanner online curated database of genetic association estimates (http://www.phenoscanner.medschl.cam.ac.uk/phenoscanner) was accessed on 1 March 2019 to search for secondary phenotypes that are known to be associated with the three selected iron status instruments at genome-wide significance (67).

All analyses are performed using the statistical programme R (version 3.4.2). The TwoSampleMR and MendelianRandomization packages were used to conduct the MR analyses (68, 69). All data used in this work were obtained from previous studies that had already obtained the relevant ethical approval and participant consent, and no further review was required.

Ethical approval

Ethical approval and participant consent for use of all data in this work had been previously obtained in their respective primary studies, and therefore was not required to be sought again here. The primary studies from which data were obtained have been cited on first introduction. UK Biobank data were accessed through application 236.

2.3 Results

Instruments

The three instrument SNPs used had F statistics ranging from 47 to 2,127 for the 4 iron status biomarkers (Table 2.3), consistent with marked weak instrument bias in the analyses being unlikely (44). Investigating possible bias related to pleiotropy, the three SNPs associated with iron status biomarkers and red blood cell traits as expected (67). Additionally, potentially pleiotropic effects for rs1799945 were identified through the association of its iron status increasing allele with blood pressure (67, 70, 71). Similarly, the iron status increasing allele of rs1800562 was inversely associated with total cholesterol and low density lipoprotein cholesterol levels (67, 70, 72).

Table 2.3. Variance explained and F statistics for the instrument single-nucleotide polymorphisms (SNPs) and iron status biomarkers. EA: effect allele; EAF: effect allele frequency.

			SNP-iron status associations (n=48,972)							
			Iron		Transferrin saturation		Log$_{10}$ Ferritin		Transferrin	
SNP	EA	EAF	R^2	F	R^2	F	R^2	F	R^2	F
rs1800562	A	0.07	1.3	668	4.2	2127	0.5	256	2.9	1446
rs1799945	G	0.15	0.9	450	1.4	676	0.1	53	0.3	163
rs855791	G	0.55	1.6	806	1.8	889	0.1	73	0.1	47

Mendelian randomization

The IVR MR result for the effect of iron status on CAD risk (reported as odds ratio [OR] scaled per SD increase in serum iron), was 0.94 (95% confidence interval [CI] 0.88 to 1.00, *P*=0.04). Similar results supporting a protective effect of iron stats on CAD risk were obtained when scaling the estimates to standard deviation (SD) change in the other iron status biomarkers - transferrin saturation OR 0.95 (95% CI 0.91 to 0.99, *P*=0.03) and (log-transformed) ferritin OR 0.85 (95% CI 0.73 to 0.98, *P*=0.02). Consistent with a protective effect of iron status, the estimate for transferrin (OR 1.08, 95% CI 1.01 to 1.16, *P*=0.03) was in the opposite direction, with higher transferrin levels reflecting lower systemic iron levels.

In contrast, when considering stroke the analysis identified a detrimental effect of genetically increased iron status (serum iron OR 1.07, 95% CI 1.01-1.14, *P*=0.03; [log-transformed] ferritin OR 1.18, 95% CI 1.02-1.36, *P*=0.03; transferrin saturation OR 1.06, 95% CI 1.01-1.11, *P*=0.02). In keeping with this pattern, higher genetically determined transferrin associated with a lower stroke risk (OR 0.92, 95% CI 0.86-0.99, *P*=0.02). Genetic association for the rs1800562 SNP with SVS was not available in the MEGASTROKE summary data, and there were no proxies available with LD r^2>0.3. Investigating ischemic stroke subtypes found the detrimental effect of genetically determined iron status was related to effects particularly on the CES subtype (serum iron OR 1.16, 95% CI 1.01-1.32, *P*=0.03; [log-transformed] ferritin OR 1.46, 95% CI 1.07-2.00, *P*=0.02; transferrin saturation OR 1.13, 95% CI 1.02-1.25, *P*=0.02; transferrin OR 0.82, 95% CI 0.70-0.96, *P*=0.01). For LAS, there was no apparent effect of genetically determined iron status (serum iron OR 0.95, 95% CI 0.81-1.12, *P*=0.54; [log-transformed] ferritin OR 0.82, 95% CI 0.55-1.22, *P*=0.32; transferrin saturation OR 0.95, 95% CI 0.84-1.08, *P*=0.41; transferrin OR 1.12, 95% CI 0.91-1.38, *P*=0.28). Furthermore, when performing IVW MR using the two SNPs available for SVS, no effect of iron status was identified (serum iron OR 0.98, 95% CI 0.84-1.15, *P*=0.79; [log-transformed] ferritin OR 0.94, 95% CI 0.57-1.55, *P*=0.81; transferrin saturation OR 0.98, 95% CI 0.85-1.14, *P*=0.82; transferrin OR 1.00, 95% CI 0.66-1.52, *P*=1.00).

The IVW MR also demonstrated a detrimental effect higher genetically determined systemic iron status on risk of VTE (serum iron OR 1.37, 95% CI 1.14-1.66, *P*=1x10^{-3}; transferrin saturation OR 1.25, 95% CI 1.09-1.43, *P*=1x10^{-3}; [log-transformed] ferritin OR 1.92, 95% CI 1.28-2.88, *P*=2x10^{-3}; transferrin OR 0.76, 95% CI 0.63-0.92, *P*=0.01).

In contrast, IVW MR analyses demonstrated a protective effect of iron status on risk of carotid plaque for serum iron (OR 0.85, 95% CI 0.73-0.99, *P*=0.04) and transferrin saturation (OR 0.89, 95% CI 0.80-1.00, *P*=0.05). The other biomarkers reflected a protective role of higher iron status on carotid plaque, although their effect estimates did not reach significance ([log-transformed] serum ferritin OR 0.72, 95% CI, 0.51-1.01, *P*=0.06; serum transferrin OR 1.15, 95% CI 0.97-1.35, *P*=0.11).

IVW MR did not identify any apparent effect of genetically determined iron status on cIMT measured in millimetre increase per SD higher iron biomarker (serum iron 0.00, 95% CI -0.01-0.01, *P*=0.90; transferrin saturation 0.00, 95% CI -0.01–0.01, *P*=0.75; [log-transformed] serum ferritin 0.01, 95% CI -0.02–0.03, *P*=0.58; serum transferrin -0.01, 95% CI -0.01–0.01, *P*=0.32).

Figure 2.1 summarises the results of the MR analysis (scaled to SD increase in genetically determined serum iron levels) for the binary cardiovascular and thrombotic disease outcomes considered, and depicts the trend for a protective effect of on diseases related to atherosclerosis but a detrimental effect on traits related to thrombosis attributable to stasis of blood.

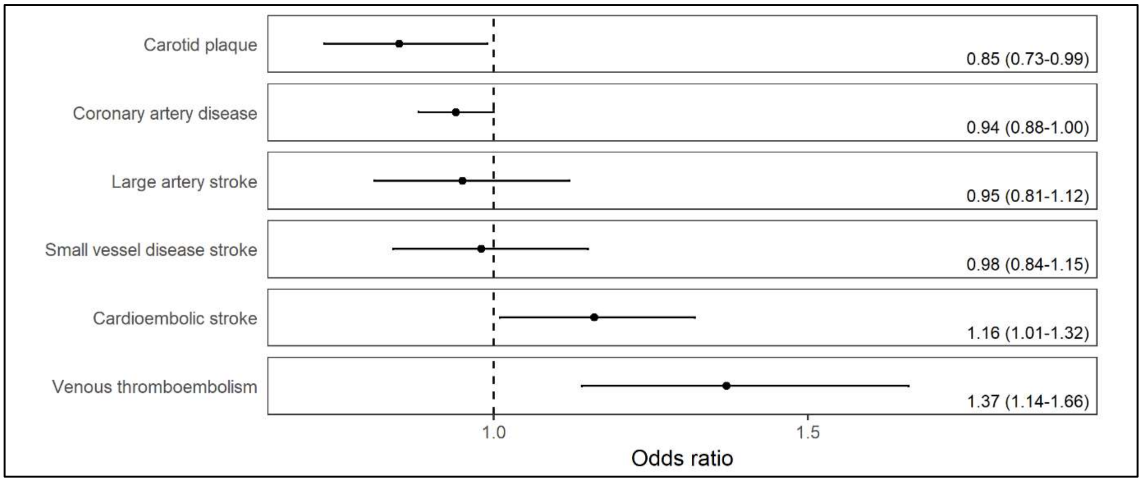


Figure 2.1. A forest plot summarising the results of the Mendelian randomization analyses investigating the effect of genetically determined iron status on dichotomous cardiovascular and thrombotic outcomes. Estimates are provided in odds ratio scaled per standard deviation increase in serum iron.

Table 2.4 summarises the Cochran's Q P-value for each of the cardiovascular and thrombotic phenotypes investigated in targeted MR. Only for cIMT was there evidence of heterogeneity between the three instrument SNPs (*P*=0.02).

Table 2.4. Cochran's Q test for heterogeneity between the instruments used to generate Mendelian randomization estimates. cIMT: carotid intima-media thickness.

Outcome	Cochran's Q P Value
Coronary artery disease	0.18
Stroke	0.23
Cardioembolic stroke	0.31
Large artery stroke	0.13
Small vessel disease stroke	0.58
cIMT	0.02
Carotid plaque	0.99
Venous thromboembolism	0.76

Mendelian randomization-phenome-wide association study

The characteristics of UK Biobank participants considered in PheWAS analysis, as well as the numbers of distinct phenotypes and cases included for each of the disease categories are detailed in Table 2.5.

Table 2.5. Part A: Characteristics of UK Biobank participants analysed in phenome-wide association study (N=424,439). Part B: Number of phenotypes and cases included for each disease category. BMI: body mass index; DBP: diastolic blood pressure; SD: standard deviation; SBP: systolic blood pressure.

Part A					
Age, years (SD)	Sex, female (%)	BMI (SD)	SBP, mmHg (SD)	DBP, mmHg (SD)	Current smoker (%)
56.8 (8.0)	229,239 (54.0%)	27.4 (4.8)	138.1 (18.6)	82.2 (10.13)	43,928 (10.4%)
Part B					
Disease Category	Phenotypes (N)	Subjects			
		Minimum	Median	Mean	Maximum
Circulatory System	98	202	1048	6308	133749
Congenital Anomalies	19	211	442	557	1823
Dermatologic	43	218	799	4765	82669
Digestive	116	228	1455	4817	79488
Endocrine/Metabolic	49	208	773	4076	45303
Genitourinary	106	203	1376	4153	103829
Hematopoietic	22	201	569	2690	12759
Infectious Diseases	25	219	1012	2237	10752
Injuries & Poisonings	59	222	536	1513	16683
Mental Disorders	36	202	710	3280	29405
Musculoskeletal	57	213	925	4164	53823
Neoplasms	82	215	1124	4261	90826
Neurological	44	204	567	2286	40703
Pregnancy Complications	17	208	1113	1854	9534
Respiratory	56	200	1124	3837	62168
Sense Organs	64	210	774	2443	39998
Symptoms	16	304	2341	7036	42311

After having excluded related and non-European individuals, there were 904 distinct phecodes for which genetic association estimates for all three instrument SNPs were available in MR-PheWAS. The IVW MR estimates for phenotypic outcomes that were significant at the 5% FDR threshold (equivalent to $P<1.1\text{x}10^{-3}$) are provided in Table 2.6, which also includes their Cochran's Q test *P*-value for heterogeneity between the three instrument SNPs. Only 14 traits showed no evidence of heterogeneity in the MR estimates produced from the three genetic instruments. Consistent estimates supporting an effect of overall systemic iron status were obtained for these traits when scaling MR estimates to SD changes in different biomarkers of iron status (Table 2.7).

Table 2.6. Inverse-variance weighted Mendelian randomization estimates, odds ratio (OR) per standard deviation (SD) increase in serum iron, for the outcomes reaching 5% false discovery rate (FDR) significance. CI: confidence interval; NOS; not otherwise specified.

Description	Cases	Controls	OR	Lower 95% CI	Upper 95% CI	*P*	5% FDR significant	Evidence of heterogeneity	Cochran's Q *P*
Disorders of iron metabolism	681	319180	10.30	6.65	15.95	1.58E-25	YES	YES	1.65E-60
Aplastic anaemia	12485	302401	0.68	0.62	0.74	3.90E-17	YES	NO	1.65E-01
Other anaemias	11586	302401	0.67	0.61	0.74	9.14E-16	YES	NO	3.18E-01
Polycythemia vera	399	311049	3.88	2.50	6.02	1.55E-09	YES	YES	1.08E-05
Iron deficiency anaemias, unspecified or not due to blood loss	7340	302401	0.72	0.64	0.81	3.57E-08	YES	NO	7.19E-01
Varicose veins of lower extremity	11323	281673	1.28	1.17	1.40	1.03E-07	YES	YES	1.45E-02
Other deficiency anaemia	8605	302401	0.75	0.67	0.83	1.76E-07	YES	NO	7.07E-01
Hypercholesterolemia	33268	285396	0.88	0.83	0.93	2.07E-05	YES	NO	8.15E-01
Other local infections of skin and subcutaneous tissue	10784	309738	1.22	1.11	1.34	5.06E-05	YES	NO	2.13E-01
Disorder of skin and subcutaneous tissue NOS	41334	280000	1.10	1.05	1.15	1.11E-04	YES	NO	3.46E-01
Chronic liver disease and cirrhosis	530	311623	2.09	1.42	3.08	1.94E-04	YES	YES	1.97E-03
Glossitis	298	315742	2.64	1.56	4.46	2.92E-04	YES	NO	8.50E-01
Poisoning by antibiotics	3446	293867	0.74	0.62	0.87	4.18E-04	YES	NO	9.15E-01
Cellulitis and abscess of arm/hand	5671	309738	1.25	1.10	1.42	5.56E-04	YES	NO	1.34E-01
Cellulitis and abscess of foot, toe	5635	309738	1.25	1.10	1.42	5.56E-04	YES	NO	1.34E-01
Cellulitis and abscess of leg, except foot	5679	309738	1.25	1.10	1.42	5.79E-04	YES	NO	1.62E-01
Cholesterolosis of gallbladder	459	299761	0.45	0.28	0.72	9.06E-04	YES	NO	8.18E-01
Acute post-haemorrhagic anaemia	262	302401	0.35	0.19	0.65	9.89E-04	YES	NO	2.93E-01
Arthropathy NOS	52689	268139	1.08	1.03	1.14	1.06E-03	YES	YES	4.00E-02

Table 2.7. Inverse-variance weighted Mendelian randomization (MR) estimates when scaling results to changes in the levels of the different iron status biomarkers (log odds ratio per standard deviation increase in biomarker). NOS: not otherwise specified; SE: standard error.

			Serum iron			Ferritin			Transferrin saturation			Transferrin		
Description	Cases	Controls	MR	SE	*P*	MR	SE	P	MR	SE	*P*	MR	SE	*P*
Aplastic anaemia	12485	302401	-0.39	0.05	3.90E-17	-0.71	0.11	2.23E-10	-0.26	0.03	8.12E-15	0.27	0.05	5.06E-08
Other anaemias	11586	302401	-0.40	0.05	9.14E-16	-0.75	0.12	9.96E-11	-0.28	0.03	1.20E-15	0.29	0.05	1.94E-08
Iron deficiency anaemias, unspecified or not due to blood loss	7340	302401	-0.33	0.06	3.57E-08	-0.63	0.14	3.78E-06	-0.22	0.04	2.31E-07	0.25	0.06	6.58E-05
Other deficiency anaemia	8605	302401	-0.29	0.06	1.76E-07	-0.56	0.13	8.67E-06	-0.20	0.04	4.20E-07	0.22	0.06	1.41E-04
Hypercholesterolemia	33268	285396	-0.13	0.03	2.07E-05	-0.26	0.07	7.68E-05	-0.09	0.02	8.00E-06	0.11	0.03	6.84E-04
Other local infections of skin and subcutaneous tissue	10784	309738	0.20	0.05	5.06E-05	0.47	0.11	9.21E-06	0.15	0.03	4.13E-06	-0.22	0.05	8.62E-06
Disorder of skin and subcutaneous tissue NOS	41334	280000	0.10	0.02	1.11E-04	0.22	0.06	1.72E-04	0.07	0.02	1.57E-04	-0.10	0.03	1.68E-04
Glossitis	298	315742	0.97	0.27	2.92E-04	1.96	0.57	5.86E-04	0.66	0.19	3.48E-04	-0.82	0.26	1.72E-03
Poisoning by antibiotics	3446	293867	-0.31	0.09	4.18E-04	-0.66	0.20	7.57E-04	-0.22	0.06	4.69E-04	0.28	0.09	2.46E-03
Cellulitis and abscess of arm/hand	5671	309738	0.23	0.07	5.56E-04	0.53	0.14	2.51E-04	0.17	0.05	2.17E-04	-0.25	0.07	1.89E-04
Cellulitis and abscess of foot, toe	5635	309738	0.23	0.07	5.56E-04	0.53	0.14	2.64E-04	0.17	0.05	2.36E-04	-0.25	0.07	2.05E-04
Cellulitis and abscess of leg, except foot	5679	309738	0.23	0.07	5.79E-04	0.52	0.14	2.97E-04	0.17	0.05	2.74E-04	-0.24	0.07	2.84E-04
Cholesterolosis of gallbladder	459	299761	-0.80	0.24	9.06E-04	-1.59	0.55	3.59E-03	-0.55	0.18	1.75E-03	0.61	0.26	1.92E-02
Acute post-haemorrhagic anaemia	262	302401	-1.05	0.32	9.89E-04	-1.97	0.73	6.84E-03	-0.70	0.23	2.46E-03	0.69	0.35	4.75E-02

Higher genetically determined iron status was most negatively associated with acute post-haemorrhagic anaemia, while in the other direction was most positively associated with glossitis, followed by skin and soft tissue infections at various body sites (Tables 2.6 and 2.7). Figure 2.2 is a forest plot summarising MR results for all the outcomes associated genetically determined systemic iron status at 5% FDR significance.

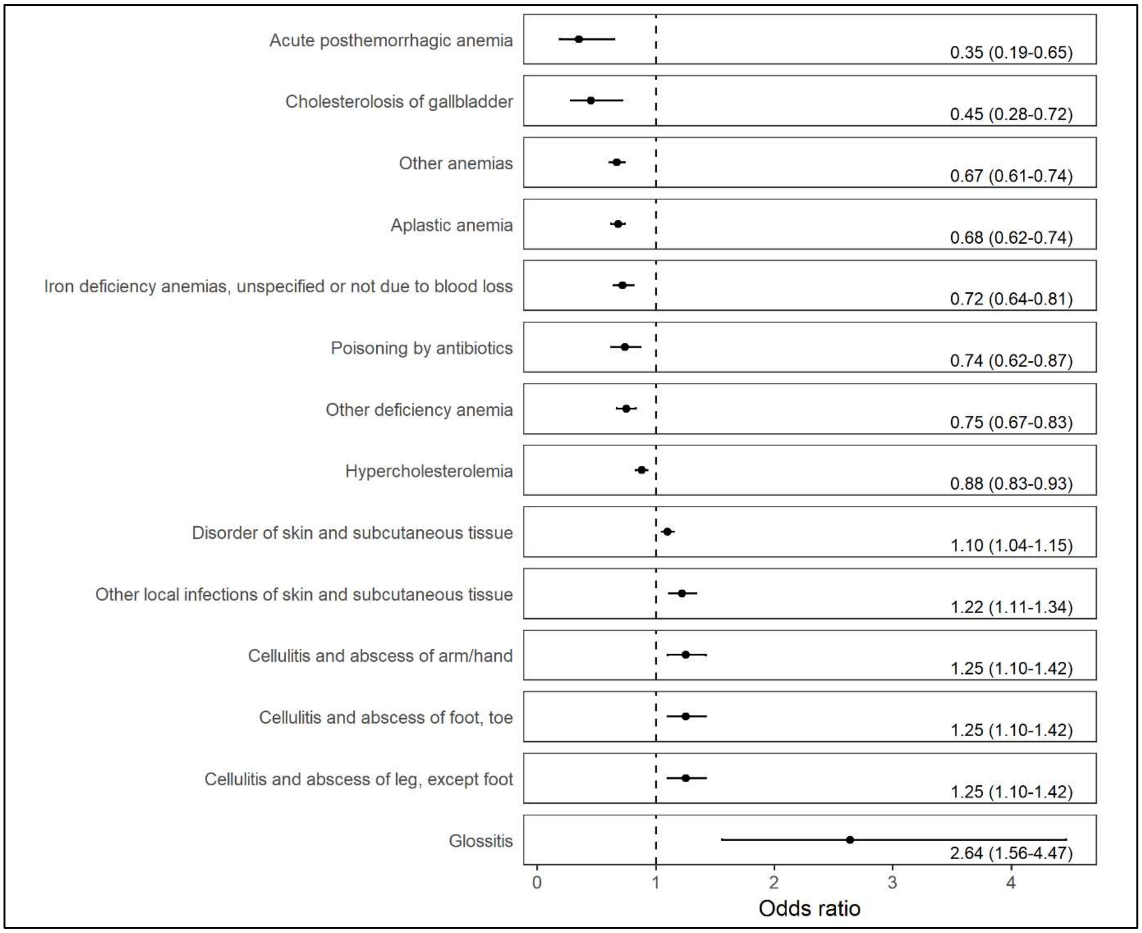


Figure 2.2. Mendelian randomization results for all outcomes achieving 5% false discovery rate significance in the phenome-wide association study of genetically determined systemic iron status. Units are in odds ratio per standard deviation increase in serum iron.

2.4 Discussion

Summary of main findings

The array of MR analyses performed in this work offer important insight into the role of iron status in cardiovascular and thrombotic disease pathogenesis, as well as offering a broader overview of the effect of iron status on outcomes throughout the phenome. The pattern of results offers overall support for a protective effect of higher iron status on atherosclerotic traits, including CAD and carotid plaque. In contrast, the findings support a detrimental effect of higher iron status on risk of thrombosis related to stasis of blood, such as that predisposing to CES and VTE. In the MR-PheWAS, higher iron status was shown to protect against hypercholesterolemia, potentially explaining part of the reduced risk of atherosclerotic traits, but was also associated with increased risk of skin and skin structure infections. The MR-PheWAS also identified a protective effect of higher iron status on a number of anaemia related traits, thus serving as a positive control for the validity of this approach.

Research in context

The association of heart disease with iron storage disorders and post-menopausal status in women has previously been attributed to an effect of higher systemic iron status (13). However, this was not supported in consequent observational research (16). While a randomised trial identified a protective effect of heavy metal chelation on heart disease, this finding may not be generalizable to a wider population, with the effects potentially specific to patients that have suffered recent myocardial infarction or independent to effects on systemic iron status (73). Consistent with the findings of the current MR analysis, a systematic review and meta-analysis of prospective observational studies on the association of body iron status and coronary heart disease risk found that the risk ratio for individuals with iron status biomarker levels in the top tertile compared with those in the bottom tertile was 0.80 (95% CI 0.73-0.87) for iron, 0.82 (95% CI 0.75-0.89) for transferrin saturation, 1.03 (95% CI 0.87-1.23) for ferritin, and 0.99 (95% CI 0.86-1.13) for transferrin (16). Both ferritin and transferrin levels are affected by inflammation, and it may be that the lack of significant findings for these biomarkers are related to confounding from this. Of relevance, all but one of the 17 studies included in the meta-analysis adjusted for smoking and major cardiovascular risk factors, which may represent possible sources of confounding (16).

In contrast to CAD, the MR analysis generated evidence that higher genetically determined iron status increases risk of IS, with this driven by increased CES risk. Many observational studies have supported a detrimental role of higher iron status on stroke risk (19-21), although there is also observational evidence in the other direction (17-19, 74-78). Such findings of increased stroke risk at both the higher and lower extremes of iron status may suggest a non-linear association with iron status (18, 19). Only two case-control studies have previously investigated the effect of iron status on risk of IS subtypes, with Chang et al. finding an increased risk of a prior iron deficiency diagnosis in individuals with both thrombotic and embolic stroke, as compared with non-stroke cases (17). However, the other case-control study that investigated ischaemic stroke subtypes did not find any significant association between SVS, LAS and CES with biomarkers of iron status (22). It may be that the distinct population investigated as well as unmeasured or unknown confounding factors contributed to the different findings, along with the possible play of non-linear effects of variation in systemic iron status on risk of stroke.

Numerous observational studies have also investigated the relationship between iron status and carotid atherosclerotic processes, although with inconsistent findings. There have been three studies that found gender-specific positive associations between serum ferritin and either carotid plaque (79, 80), or cIMT (81), with a further two studies supporting a positive association with carotid plaque when pooling individuals of either sexes (28, 82). However, there are also observational studies that produced no evidence of any association between serum ferritin and carotid atherosclerosis (83, 84), with two case-control studies also suggesting negative associations between ferritin and cIMT (85, 86). Given the susceptibility of such observational research to environmental confounding, this may we be the explanation for the discrepancy in findings between these studies.

Relatively few papers have explored the association between systemic iron levels and VTE risk. One case-control study identified an increased risk of VTE in individuals that had higher levels of hepcidin, a biomarker reflective of systemic iron status (87). This compared 390 patients with VTE with 802 age and sex-matched controls to identify a dose-dependent relationship between circulating hepcidin levels and VTE risk, independently of C-reactive protein, which was used as a biomarker of inflammation.

While recent work used MR to study the association of hereditary hemochromatosis genetic variants with risk of 11 outcomes implicated in iron overload (88), a notable advantage in the currently presented MR-PheWAS is that it explored the effects of iron status across over 900 human diseases captured in HES data using a hypothesis-free approach, and was thus able to identify potentially novel results, such as the possible effects of iron status on risk of cellulitis and hypercholesterolemia. The current MR-PheWAS also incorporated genetic instruments for systemic iron status, rather than narrowly focusing on only variants related to hereditary haemochromatosis (88), to thus allow for better exploration into the effects of fluctuations of iron status through any mechanism. Additionally, the use of strong instruments related to serum iron, ferritin, transferrin and transferrin saturation, all in a pattern supportive of their effect on overall systemic iron status, the MR-PheWAS was directed towards investigating systemic iron status specifically, rather than some other related trait.

Strengths and limitations

The MR analyses undertaken here overcome many of the limitations suffered in traditional observational epidemiology by minimising the play of bias related to environmental confounding (31). This is important because iron status biomarker levels are affected by other concurrent pathological processes, such as inflammation, liver disease, renal failure and malignancy, which can also affect risk of cardiovascular and thrombotic disease, thus acting as potential confounders. For example, ferritin levels rise and transferrin levels fall in the context of systemic inflammation (89). As a consequence, any observational studies that measure circulating levels of these biomarkers may suffer confounding related to unmeasured or unknown inflammation.

Another major advantage of the genetic epidemiological approaches undertaken in this work is that they allowed for a high degree of phenotypic resolution in disentangling the role of systemic iron status in distinct and heterogeneous cardiovascular and thrombotic disease processes. For example, it was only the cardioembolic subtype of IS for which there was an increase in risk in the context of higher genetically determined iron status. This finding was similar to that observed in the MR analysis of systemic iron status and VTE risk. Consistent with this, the pathophysiology of CES relates to stasis of blood, often in the left atrial appendage, with thrombus formation following a similar mechanism to that seen in VTE, where there is typically stasis of blood in the deep veins of the legs. In contrast, LAS, carotid plaque and CAD typically follow a different aetiology, more commonly related to atherosclerosis, and in-keeping with this, the MR analysis suggested detrimental effects of higher genetically determined iron status for all of these disease processes, though not reaching formal statistical significance for LAS.

A potential cause for bias in MR analyses is pleiotropy of the genetic variants used. In scenarios where these are many instrument variants available for MR analysis, statistical methods that are more robust to the inclusion of pleiotropic variants are available (90), as well as methods that can test for the presence of pleiotropy, and provide adjusted estimates (91, 92). However, with only three instrument SNPs, opportunities for such statistical methods are limited, with even the detection of heterogeneity for MR estimates produced by different variants subject to limited statistical power (65). Thus, alternative methods to investigate pleiotropy were incorporated – namely biological knowledge of potential pleiotropic pathways. Specifically, by searching for secondary phenotypes of the three instruments for systemic iron status, it was possible to identify potential pleiotropic pathways that could introduce bias into the MR analysis. The association of all three systemic iron status instruments with red blood cell traits would be expected given the well-established role of iron status in erythropoiesis (2), and furthermore this would be unlikely to be introducing bias into the MR considering cardiovascular and thrombotic outcomes, as red blood cells traits are not believed to directly affect these outcomes (93). In contrast, the observed association of the iron status increasing allele of rs1800562 (in the *HFE* gene) with lower low-density lipoprotein cholesterol (LDL-C) levels, and the iron status increasing allele of rs1799945 (also in the *HFE* gene) with higher systolic and diastolic blood pressures traits would likely introduce bias through pleiotropy in MR analyses considering atherosclerotic outcomes, because these secondary phenotypes represent established cardiovascular risk factors. In this context, the described associations of the rs1800562 and rs1799945 variants with LDL-C and blood pressure respectively were considered as potential genetic confounding rather than mediation because these traits were each only associated with one of the three instruments variants. If they were serving as mediating pathways, it would be expected that they relate to all three genetic instruments of systemic iron status. Reassuringly however, the MR estimates produced by the rs855791 SNP (in the *TMPRSS6* gene) provided consistent MR estimates to the overall IVW MR, suggesting that any bias from pleiotropy was unlikely to be significant affecting the results or conclusions reached (94-96).

A strength of this work is that it only includes genetic variants as instruments where they have a known biological role in determining systemic iron status (97), and further are related to all four considered biomarkers of systemic iron status (2), thus offering strong reassurances towards their validity. The fact that instruments were identified from two distinct genes, on different chromosomes, whose corresponding proteins regulate iron status by distinct mechanisms offers some reassurance that the effects of these variants can be attributed to variations in systemic iron status generally, rather than through protein specific mechanisms, While more relaxed instrument selection criteria may be employed to increase the number of genetic variants available for analysis, this may do more harm than good. For example, by selecting instruments by their association with only one biomarker of iron status rather than all four, there is considerable risk of including SNPs that do not reflect systemic iron status, but rather are related to distribution of iron. This point is well demonstrated by the genome-wide significant variants associated with different biomarkers of iron status, but in directions of effect that are not compatible with their reflection of overall iron status (2).

The limitations of this work must also be fully appreciated. The MR approach instruments genetic variants to study the effect of varying an exposure of interest, which in this case is systemic iron status. However, inherent to this process is that the genetic variants have lifelong effects on iron status, which would be distinct to the effect of a discrete clinical intervention, particularly one that is applied in adolescence or adult life (98). Furthermore, the genetic variants used in MR generally, and indeed for the current analyses, only represent the effect of small variations in the exposure of interest, around the mean levels for the considered population, and within the normal range (2). The results from these MR analyses should therefore not be extrapolated to the extremes of iron status, namely iron overload or deficiency. This is particularly important as the MR model assumes a linear effect of iron status on risk of the outcomes considered, an assumption that likely breaks down at very high or low levels of iron status.

The MR analyses were undertaken using genetic data from predominantly European ancestry, and it is therefore not clear whether the findings on the effect of genetically determined iron status can be extrapolated to other ethnic groups. There was also a small degree of overlap, in the region of 5%, between the populations used to obtain genetic association estimates for the iron status biomarkers and those used for cardiovascular and thrombotic disease outcome traits (96). Given the limited nature of this participant overlap, it is unlikely for this to be a source of substantial bias in the consequent MR estimates (99), and it will therefore not affect the conclusions drawn.

Particular limitations of the MR-PheWAS approach relate to the HES data used. Although this provided a source of clinically relevant outcomes that were linked to the genetic data, there may have also been some introduction of misclassification bias (100). As one example of this, it may be that the identified protective effect of iron status on aplastic anaemia is related to misclassification of iron deficiency anaemia. As another example, where iron status does not cause aplastic anaemia, it may still contribute to its diagnosis in borderline cases that would otherwise fall below the requisite threshold for label allocation. The MR-PheWAS also generated findings that seem to contradict established clinical knowledge, such as the finding of increased glossitis risk arising from higher iron status. One explanation may relate to a weakness with the phecode grouping system used, perhaps relating to ambiguity between underlying atrophy (101), or super-imposed infection for this diagnosis, with the latter being consistent with the findings for the effect of iron status on risk of superficial infections. Insufficient statistical power may have also contributed to potential false negative results in the MR-PheWAS. For example, this analysis did not replicate the findings from the direct MR analysis of CAD, stroke or VTE, which were all included in the considered PheWAS diagnoses. Other causes for type II error and false negative results may be the exclusion of any results that produced evidence of heterogeneity in the MR estimates derived different instrument SNPs, when filtering out findings potentially susceptible to bias related to pleiotropy.

Underlying mechanisms

The protective effect of higher iron status on CAD and carotid plaque may relate to reduced circulating LDL-C levels and resultant effects on atherosclerosis (102). Consistent with this, the MR-PheWAS also supported an effect of higher iron status on reducing risk of hypercholesterolemia. The discrepancy in the MR findings for cIMT and carotid plaque may represent a distinct role for iron in different atherosclerotic processes, with cIMT more representing arterial hyperplasia related to hypertension, while carotid plaque represents fatty atherosclerotic injury (103). This is also supported by observational research identifying an association between circulating ferritin levels and carotid plaque, rather than cIMT (79, 80). However, there was also a significant Cochran's Q P-value (of 0.02) when assessing for heterogeneity between the MR estimates produced by individual instrument SNPs for the analysis investigating the effect of iron status on cIMT, suggesting that pleiotropy may also be responsible for the null result.

In contrast, the MR finding of higher iron status increasing CES risk is unlikely to be related to effects related to atherosclerosis. Of more potential relevance, iron-catalysed processes have been implicated in the coagulation of blood (104), with dense fibrin-like deposits identified in the blood of individuals with diabetes mellitus, attributable to prothrombotic actions of iron (104). Following on from this, one possibility is that excess free iron in blood results in production of fibrin-like debris that consequently results in thrombus generation (105). However, caution should be taken here, as these mechanisms are only suggested as plausible explanations, and further confirmatory work is of course required.

Considering possible mechanisms explaining the novel MR-PheWAS results, iron scavenging systems are present in bacteria that cause skin and soft tissue infections, including *Staphylococcus aureus* (106), and *Streptococcus pyogenes* (107). It has also been proposed that iron metabolism in these organisms is related to their virulence (106). This is also consistent with the sequestration of free iron from invading organisms through immune defence mechanisms (108). In diseases of iron overload, such as hereditary hemochromatosis, increased susceptibility to bacterial infections has been observed (109), with certain bacteria also demonstrating increased growth and division in human serum that is obtained after iron supplementation (110).

The MR-PheWAS also produced evidence that higher iron status reduced risk of both hypercholesterolaemia and gallbladder cholesterolosis, which relates to accumulation of cholesteryl esters in the gallbladder (111). The *HFE* gene variant rs1800562 that was incorporated as a genetic instrument for iron status has been associated with LDL-C in GWAS (112). The mechanism behind this is not known, but may relate to its effects on systemic iron levels. Consistent with this, all three genetic variants used to instrument systemic iron status, both in the *HFE* gene (rs1800562 and rs1799945) and in the *TMPRSS6* gene (rs855791) produced MR estimates to demonstrate consistent effects of systemic iron status on lowering both hypercholesterolaemia and gallbladder cholesterolosis risk (113). Further supporting this hypothesis, iron status has been implicated in affecting lipid metabolism in both rodents and humans (114, 115).

Clinical relevance

There are available treatments to both increase and decrease systemic iron levels. However, the effectiveness of oral iron supplementation is restricted by its low gastrointestinal absorption (116), as well as side-effects including abdominal cramps, nausea and bowel disturbance, which affect compliance (116, 117). As a public health measure, fortification of food with iron was effective for reducing the burden of iron deficiency anaemia (118, 119). In the case of anaemia that is refractory to oral supplementation, or when anaemia is severe, intravenous iron infusion may be required (116). At the other end of the spectrum, iron overload is mainly managed with venesection in the context of haemochromatosis (120), with iron chelation also used to enhance iron excretion in certain scenarios (73). Clinical trials and formal guidance on titration of systemic iron levels have mostly focused on the treatment of anaemia (116), such as related to chronic kidney disease (121), heavy menstruation (122), and pregnancy (123). More limited evidence is available for the manipulation of iron levels to treatment other clinical outcomes, such as stroke (124), malaria (125) and restless leg syndrome (126). To date however, there has not been any clinical trial investigating the impact of adjusting iron status to prevent or treat hypercholesterolaemia, cardiovascular or thrombotic disease, nor skin and skin structure infections. Given the results of the described genetic analyses, there is now a rationale to further investigate the possibility of manipulating systemic iron status in the prevention of these diseases. Importantly however, caution must be taken, as while genetic support increases the chances of success for clinical trials into therapeutic interventions (127), these forms of evidence are not equivalent.

Conclusions

The targeted MR and MR-PheWAS analyses performed into this work offer novel insight into the role of systemic iron status in both cardiovascular and thrombotic disease, as well as in clinically relevant outcomes throughout the human phenome. The analyses support a detrimental role of higher genetically determined iron status on stasis-mediated thrombotic processes, including CES and VTE, but a protective effect of on dyslipidaemia-related atherosclerotic disease processes such as carotid plaque and CAD. Hypothesis-free exploration of traits throughout the phenome verified the known associations of higher iron status with lower risk of anaemia, but also identified inverse associations with hypercholesterolaemia, and positive associations with skin and soft tissue infections. Taken together, this work offers clinically relevant insight into the implications of varying iron status within the normal range, and warrants further study in the setting of clinical intervention.

2.5 References

1. Muñoz M, Villar I, García-Erce JA. An update on iron physiology. World J Gastroenterol. 2009;15(37):4617-26.

2. Benyamin B, Esko T, Ried JS, Radhakrishnan A, Vermeulen SH, Traglia M, et al. Novel loci affecting iron homeostasis and their effects in individuals at risk for hemochromatosis. Nat Commun. 2014;5:4926.

3. Abbaspour N, Hurrell R, Kelishadi R. Review on iron and its importance for human health. J Res Med Sci. 2014;19(2):164-74.

4. DeLoughery TG. Iron Deficiency Anemia. Med Clin North Am. 2017;101(2):319-32.

5. GBD. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016;388(10053):1545-602.

6. Sebastiani G, Pantopoulos K. Disorders associated with systemic or local iron overload: from pathophysiology to clinical practice. Metallomics. 2011;3(10):971-86.

7. Mathers C, Stevens GA, Mahanani WR, Fat DM, Hogan D. WHO | Disease burden and mortality estimates. World Health Organization; 2018. p. 1-65.

8. Wendelboe AM, Raskob GE. Global Burden of Thrombosis. Circ Res. 2016;118(9):1340-7.

9. Katan M, Luft A. Global Burden of Stroke. Semin Neurol. 2018;38(02):208-11.

10. Raskob GE, Angchaisuksiri P, Blanco AN, Buller H, Gallus A, Hunt BJ, et al. Thrombosis. Arterioscler Thromb Vasc Biol. 2014;34(11):2363-71.

11. Franchini M, Targher G, Montagnana M, Lippi G. Iron and thrombosis. Ann Hematol. 2008;87(3):167-73.

12. Basuli D, Stevens RG, Torti FM, Torti SV. Epidemiological associations between iron and cardiovascular disease and diabetes. Front Pharmacol. 2014;5:117.

13. Sullivan JL. Iron and the sex difference in heart disease risk. Lancet. 1981;1(8233):1293-4.

14. Jiang R, Manson JE, Meigs JB, Ma J, Rifai N, Hu FB. Body iron stores in relation to risk of type 2 diabetes in apparently healthy women. JAMA. 2004;291(6):711-7.

15. Pardo Silva MC, Njajou OT, Alizadeh BZ, Hofman A, Witteman JC, van Duijn CM, et al. HFE gene mutations increase the risk of coronary heart disease in women. Eur J Epidemiol. 2010;25(9):643-9.

16. Das De S, Krishna S, Jethwa A. Iron status and its association with coronary heart disease: systematic review and meta-analysis of prospective studies. Atherosclerosis. 2015;238(2):296-303.

17. Chang YL, Hung SH, Ling W, Lin HC, Li HC, Chung SD. Association between ischemic stroke and iron-deficiency anemia: a population-based study. PLoS One. 2013;8(12):e82952.

18. Marniemi J, Alanen E, Impivaara O, Seppänen R, Hakala P, Rajala T, et al. Dietary and serum vitamins and minerals as predictors of myocardial infarction and stroke in elderly subjects. Nutr Metab Cardiovasc Dis. 2005;15(3):188-97.

19. Gillum RF, Sempos CT, Makuc DM, Looker AC, Chien CY, Ingram DD. Serum transferrin saturation, stroke incidence, and mortality in women and men. The NHANES I Epidemiologic Followup Study. National Health and Nutrition Examination Survey. Am J Epidemiol. 1996;144(1):59-68.

20. van der A DL, Grobbee DE, Roest M, Marx JJ, Voorbij HA, van der Schouw YT. Serum ferritin is a risk factor for stroke in postmenopausal women. Stroke. 2005;36(8):1637-41.

21. Kannel WB, Gordon T, Wolf PA, McNamara P. Hemoglobin and the risk of cerebral infarction: the Framingham Study. Stroke. 1972;3(4):409-20.

22. Ekblom K, Hultdin J, Stegmayr B, Johansson I, Van Guelpen B, Hallmans G, et al. Iron stores and HFE genotypes are not related to increased risk of ischemic stroke. A prospective nested case-referent study. Cerebrovasc Dis. 2007;24(5):405-11.

23. Quintana Pacheco DA, Sookthai D, Wittenbecher C, Graf ME, Schübel R, Johnson T, et al. Red meat consumption and risk of cardiovascular diseases-is increased iron load a possible link? Am J Clin Nutr. 2018;107(1):113-9.

24. Shovlin CL, Chamali B, Santhirapala V, Livesey JA, Angus G, Manning R, et al. Ischaemic strokes in patients with pulmonary arteriovenous malformations and hereditary hemorrhagic telangiectasia: associations with iron deficiency and platelets. PLoS One. 2014;9(2):e88812.

25. Keung Y-K, Owen J. Iron deficiency and thrombosis: literature review. Clin Appl Thromb Hemost. 2004;10(4):387-91.

26. Hung S-H, Lin H-C, Chung S-D. Association between venous thromboembolism and iron-deficiency anemia. Blood Coagul Fibrinolysis. 2015;26(4):368-72.

27. Xie YG, Lillicrap DP, Taylor SA. An association between the common hereditary hemochromatosis mutation and the factor V Leiden allele in a population with thrombosis. Blood. 1998;92(4):1461-2.

28. Ahluwalia N, Genoux A, Ferrieres J, Perret B, Carayol M, Drouet L, et al. Iron status is associated with carotid atherosclerotic plaques in middle-aged adults. J Nutr. 2010;140(4):812-6.

29. Kraml P. The role of iron in the pathogenesis of atherosclerosis. Physiol Res. 2017;66:S55-S67.

30. Grammer TB, Kleber ME, Silbernagel G, Pilz S, Scharnagl H, Tomaschitz A, et al. Hemoglobin, iron metabolism and angiographic coronary artery disease (The Ludwigshafen Risk and Cardiovascular Health Study). Atherosclerosis. 2014;236(2):292-300.

31. Davey Smith G, Ebrahim S. What can mendelian randomisation tell us about modifiable behavioural and environmental exposures? BMJ. 2005;330(7499):1076-9.

32. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ. 2018;362:k601.

33. Pichler I, Del Greco MF, Gogele M, Lill CM, Bertram L, Do CB, et al. Serum iron levels and the risk of Parkinson disease: a Mendelian randomization study. PLoS Med. 2013;10(6):e1001462.

34. Millard LAC, Davies NM, Timpson NJ, Tilling K, Flach PA, Smith GD. MR-PheWAS: hypothesis prioritization among potential causal effects of body mass index on many outcomes, using Mendelian randomization. Sci Rep. 2015;5.

35. Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat Biotechnol. 2013;31(12):1102-10.

36. Wish JB. Assessing iron status: beyond serum ferritin and transferrin saturation. Clin J Am Soc Nephrol. 2006;1 Suppl 1:S4-8.

37. Sacco RL, Kasner SE, Broderick JP, Caplan LR, Connors JJ, Culebras A, et al. An updated definition of stroke for the 21st century: a statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2013;44(7):2064-89.

38. Ay H, Benner T, Arsava EM, Furie KL, Singhal AB, Jensen MB, et al. A computerized algorithm for etiologic classification of ischemic stroke: the Causative Classification of Stroke System. Stroke. 2007;38(11):2979-84.

39. Plichart M, Celermajer DS, Zureik M, Helmer C, Jouven X, Ritchie K, et al. Carotid intima-media thickness in plaque-free site, carotid plaques and coronary heart disease risk prediction in older adults. The Three-City Study. Atherosclerosis. 2011;219(2):917-24.

40. Spence JD. Measurement of intima-media thickness vs. carotid plaque: uses in patient care, genetic research and evaluation of new therapies. Int J Stroke. 2006;1(4):216-21.

41. Rundek T, Gardener H, Della-Morte D, Dong C, Cabral D, Tiozzo E, et al. The relationship between carotid intima-media thickness and carotid plaque in the Northern Manhattan Study. Atherosclerosis. 2015;241(2):364-70.

42. Puig S, Askeland E, Thiele DJ. Coordinated remodeling of cellular metabolism during iron deficiency through targeted mRNA degradation. Cell. 2005;120(1):99-110.

43. Taylor AE, Davies NM, Ware JJ, VanderWeele T, Smith GD, Munafo MR. Mendelian randomization in health research: using appropriate genetic variants and avoiding biased estimates. Econ Hum Biol. 2014;13:99-106.

44. Palmer TM, Lawlor DA, Harbord RM, Sheehan NA, Tobias JH, Timpson NJ, et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. Stat Methods Med Res. 2012;21(3):223-42.

45. Feder JN, Gnirke A, Thomas W, Tsuchihashi Z, Ruddy DA, Basava A, et al. A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. Nat Genet. 1996;13(4):399-408.

46. Gao J, Chen J, Kramer M, Tsukamoto H, Zhang AS, Enns CA. Interaction of the hereditary hemochromatosis protein HFE with transferrin receptor 2 is required for transferrin-induced hepcidin expression. Cell Metab. 2009;9(3):217-27.

47. Nemeth E, Tuttle MS, Powelson J, Vaughn MB, Donovan A, Ward DM, et al. Hepcidin regulates cellular iron efflux by binding to ferroportin and inducing its internalization. Science. 2004;306(5704):2090-3.

48. Nemeth E, Ganz T. Regulation of iron metabolism by hepcidin. Annu Rev Nutr. 2006;26:323-42.

49. Zhao N, Nizzi CP, Anderson SA, Wang J, Ueno A, Tsukamoto H, et al. Low intracellular iron increases the stability of matriptase-2. J Biol Chem. 2015;290(7):4432-46.

50. Consortium CD, Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet. 2013;45(1):25-33.

51. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet. 2015;47(10):1121-30.

52. Han B, Duong D, Sul JH, de Bakker PI, Eskin E, Raychaudhuri S. A general framework for meta-analyzing dependent studies with overlapping subjects in association mapping. Hum Mol Genet. 2016;25(9):1857-66.

53. Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. Nat Genet. 2018;50(4):524-37.

54. Adams HP, Jr., Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. Stroke. 1993;24(1):35-41.

55. Germain M, Chasman DI, de Haan H, Tang W, Lindström S, Weng L-C, et al. Meta-analysis of 65,734 individuals identifies TSPAN15 and SLC44A2 as two susceptibility loci for venous thromboembolism. Am J Hum Genet. 2015;96(4):532-42.

56. Franceschini N, Giambartolomei C, de Vries PS, Finan C, Bis JC, Huntley RP, et al. GWAS and colocalization analyses implicate carotid intima-media thickness and carotid plaque loci in cardiovascular outcomes. Nat Commun. 2018;9(1):5141-.

57. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, et al. Heart Disease and Stroke Statistics—2016 Update. Circulation. 2016;133(4):e38-360.

58. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLoS Med. 2015;12(3):e1001779.

59. Wei WQ, Bastarache LA, Carroll RJ, Marlo JE, Osterman TJ, Gamazon ER, et al. Evaluating phecodes, clinical classification software, and ICD-9-CM codes for phenome-wide association studies in the electronic health record. PLoS One. 2017;12(7):e0175508.

60. Li X, Meng XR, Spiliopoulou A, Timofeeva M, Wei WQ, Gifford A, et al. MR-PheWAS: exploring the causal effect of SUA level on multiple disease outcomes by using genetic instruments in UK Biobank. Ann Rheum Dis. 2018;77(7):1039-47.

61. Verma A, Bradford Y, Dudek S, Lucas AM, Verma SS, Pendergrass SA, et al. A simulation study investigating power estimates in phenome-wide association studies. BMC Bioinformatics. 2018;19(1):120.

62. Brion MJ, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. Int J Epidemiol. 2013;42(5):1497-501.

63. Thompson JR, Minelli C, Del Greco M F. Mendelian Randomization using Public Data from Genetic Consortia. Int J Biostat. 2016;12(2).

64. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met. 1995;57(1):289-300.

65. Del Greco M F, Minelli C, Sheehan NA, Thompson JR. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. Stat Med. 2015;34(21):2926-40.

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

67. Staley JR, Blackshaw J, Kamat MA, Ellis S, Surendran P, Sun BB, et al. PhenoScanner: a database of human genotype-phenotype associations. Bioinformatics. 2016;32(20):3207-9.

68. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. eLife. 2018;7.

69. Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. Int J Epidemiol. 2017;46(6):1734-9.

70. Allen CL, Bayraktutan U. Risk factors for ischaemic stroke. Int J Stroke. 2008;3(2):105-16.

71. Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature. 2011;478(7367):103-9.

72. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. Nat Genet. 2013;45(11):1274-83.

73. Poggiali E, Cassinerio E, Zanaboni L, Cappellini MD. An update on iron chelation therapy. Blood Transfus. 2012;10(4):411-22.

74. Swann IL, Kendra JR. Severe iron deficiency anaemia and stroke. Clin Lab Haematol. 2000;22(4):221-3.

75. Munot P, De Vile C, Hemingway C, Gunny R, Ganesan V. Severe iron deficiency anaemia and ischaemic stroke in children. Arch Dis Child. 2011;96(3):276-9.

76. Mehta PJ, Chapman S, Jayam-Trouth A, Kurukumbi M. Acute ischemic stroke secondary to iron deficiency anemia: a case report. Case Rep Neurol Med. 2012;2012:487080.

77. Naito H, Naka H, Kanaya Y, Yamazaki Y, Tokinobu H. Two cases of acute ischemic stroke associated with iron deficiency anemia due to bleeding from uterine fibroids in middle-aged women. Intern Med. 2014;53(21):2533-7.

78. Ellervik C, Tybjaerg-Hansen A, Appleyard M, Sillesen H, Boysen G, Nordestgaard BG. Hereditary hemochromatosis genotypes and risk of ischemic stroke. Neurology. 2007;68(13):1025-31.

79. Wolff B, Völzke H, Lüdemann J, Robinson D, Vogelgesang D, Staudt A, et al. Association Between High Serum Ferritin Levels and Carotid Atherosclerosis in the Study of Health in Pomerania (SHIP). Stroke. 2004;35(2):453-7.

80. Rossi E, McQuillan BM, Hung J, Thompson PL, Kuek C, Beilby JP. Serum ferritin and C282Y mutation of the hemochromatosis gene as predictors of asymptomatic carotid atherosclerosis in a community population. Stroke. 2000;31(12):3015-20.

81. Xu H, Song Y, Xu J, Gu Y, Zhang Q, Liu L, et al. Increased serum ferritin levels are independently associated with carotid atherosclerosis in women. 2017:117(11):1623-30.

82. Kiechl S, Aichner F, Gerstenbrand F, Egger G, Mair A, Rungger G, et al. Body iron stores and presence of carotid atherosclerosis. Results from the Bruneck Study. Arterioscler Thromb. 1994;14(10):1625-30.

83. Vergnaud AC, Bertrais S, Zureik M, Galan P, Blacher J, Hercberg S, et al. Dietary iron intake and serum ferritin in relation to 7.5 years structure and function of large arteries in the SUVIMAX cohort. Diabetes Metab. 2007;33(5):366-71.

84. Yunker LM, Parboosingh JS, Conradson HE, Faris P, Bridge PJ, Buithieu J, et al. The effect of iron status on vascular health. Vasc Med. 2006;11(2):85-91.

85. Moore M, Folsom AR, Barnes RW, Eckfeldt J. No Association between Serum Ferritin and Asymptomatic Carotid Atherosclerosis. Am J Epidemiol. 1995;141(8):719-23.

86. Raumaraa R, Vaisanen S, Mecuri M, Raniken T, Penttila I, Bond MG. Association of risk factors and body iron status to carotid atherosclerosis in middle-aged Eastern Finnish men. Eur Heart J. 1994;15(8):1020-7.

87. Ellingsen TS, Lappegård J, Ueland T, Aukrust P, Brækkan SK, Hansen J-B. Plasma hepcidin is associated with future risk of venous thromboembolism. Blood Advances. 2018;2(11):1191-7.

88. Pilling LC, Tamosauskaite J, Jones G, Wood AR, Jones L, Kuo CL, et al. Common conditions associated with hereditary haemochromatosis genetic variants: cohort study in UK Biobank. BMJ. 2019;364:k5222.

89. Ross AC. Impact of chronic and acute inflammation on extra- and intracellular iron homeostasis. Am J Clin Nutr. 2017;106(Suppl 6):1581S-7S.

90. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol. 2016;40(4):304-14.

91. Verbanck M, Chen CY, 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.

92. 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.

93. Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, et al. The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. Cell. 2016;167(5):1415-29 e19.

94. Gill D, Del Greco M F, Walker AP, Srai SKS, Laffan MA, Minelli C. The Effect of Iron Status on Risk of Coronary Artery Disease: A Mendelian Randomization Study-Brief Report. Arterioscler Thromb Vasc Biol. 2017;37(9):1788-92.

95. Gill D, Monori G, Tzoulaki I, Dehghan A. Iron Status and Risk of Stroke: A Mendelian Randomization Study. Stroke. 2018;49(12):2815-21.

96. Gill D, Brewer CF, Monori G, Tregouet DA, Franceschini N, Giambartolomei C, et al. Effects of Genetically Determined Iron Status on Risk of Venous Thromboembolism and Carotid Atherosclerotic Disease: A Mendelian Randomization Study. J Am Heart Assoc. 2019;8(15):e012994.

97. Swerdlow DI, Kuchenbaecker KB, Shah S, Sofat R, Holmes MV, White J, et al. Selecting instruments for Mendelian randomization in the wake of genome-wide association studies. Int J Epidemiol. 2016;45(5):1600-16.

98. Davey Smith G, Ebrahim S. Mendelian randomization: can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003;32(1):1-22.

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

100. Padmanabhan S, Carty L, Cameron E, Ghosh RE, Williams R, Strongman H. Approach to record linkage of primary care data from Clinical Practice Research Datalink to other health-related patient data: overview and implications. Eur J Epidemiol. 2018:34(1):91-9.

101. Wu YC, Wang YP, Chang JY, Cheng SJ, Chen HM, Sun A. Oral manifestations and blood profile in patients with iron deficiency anemia. J Formos Med Assoc. 2014;113(2):83-7.

102. Ozdemir A, Sevinc C, Selamet U, Turkmen F. The relationship between iron deficiency anemia and lipid metabolism in premenopausal women. Am J Med Sci. 2007;334(5):331-3.

103. Touboul PJ, Hennerici MG, Meairs S, Adams H, Amarenco P, Bornstein N, et al. Mannheim Carotid Intima-Media Thickness and Plaque Consensus (2004–2006–2011). Cerebrovasc Dis. 2012;34(4):290-6.

104. Lipinski B, Pretorius E. Novel pathway of ironinduced blood coagulation: implications for diabetes mellitus and its complications. Pol Arch Med Wewn. 2012;122(3):115-22.

105. Lipinski B, Pretorius E. Iron-induced fibrin in cardiovascular disease. Curr Neurovasc Res. 2013;10(3):269-74.

106. Dale SE, Doherty-Kirby A, Lajoie G, Heinrichs DE. Role of siderophore biosynthesis in virulence of Staphylococcus aureus: identification and characterization of genes involved in production of a siderophore. Infect Immun. 2004;72(1):29-37.

107. Bates CS, Montanez GE, Woods CR, Vincent RM, Eichenbaum Z. Identification and characterization of a Streptococcus pyogenes operon involved in binding of hemoproteins and acquisition of iron. Infect Immun. 2003;71(3):1042-55.

108. Parrow NL, Fleming RE, Minnick MF. Sequestration and scavenging of iron in infection. Infect Immun. 2013;81(10):3503-14.

109. Khan FA, Fisher MA, Khakoo RA. Association of hemochromatosis with infectious diseases: expanding spectrum. Int J Infect Dis. 2007;11(6):482-7.

110. Cross JH, Bradbury RS, Fulford AJ, Jallow AT, Wegmuller R, Prentice AM, et al. Oral iron acutely elevates bacterial growth in human serum. Sci Rep. 2015;5:16670.

111. Koga A. Fine-Structure of the Human Gallbladder with Cholesterosis with Special Reference to the Mechanism of Lipid-Accumulation. Brit J Exp Pathol. 1985;66(5):605-11.

112. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature. 2010;466(7307):707-13.

113. Gill D, Benyamin B, Moore LSP, Monori G, Zhou A, Koskeridis F, et al. Associations of genetically determined iron status across the phenome: A mendelian randomization study. PLoS Med. 2019;16(6):e1002833.

114. Stangl GI, Kirchgessner M. Different degrees of moderate iron deficiency modulate lipid metabolism of rats. Lipids. 1998;33(9):889-95.

115. Ahmed U, Latham PS, Oates PS. Interactions between hepatic iron and lipid metabolism with possible relevance to steatohepatitis. World J Gastroenterol. 2012;18(34):4651-8.

116. Jimenez K, Kulnigg-Dabsch S, Gasche C. Management of Iron Deficiency Anemia. Gastroenterol Hepatol (NY). 2015;11(4):241-50.

117. Cancelo-Hidalgo MJ, Castelo-Branco C, Palacios S, Haya-Palazuelos J, Ciria-Recasens M, Manasanch J, et al. Tolerability of different oral iron supplements: a systematic review. Curr Med Res Opin. 2013;29(4):291-303.

118. De-Regil LM, Jefferds MED, Peña-Rosas JP. Point-of-use fortification of foods with micronutrient powders containing iron in children of preschool and school-age. Cochrane Database Syst Rev. 2017;11:CD009666.

119. Arcanjo FPN, da Costa Rocha TC, Arcanjo CPC, Santos PR. Micronutrient Fortification at Child-Care Centers Reduces Anemia in Young Children. J Diet Suppl. 2018:1-10.

120. Assi TB, Baz E. Current applications of therapeutic phlebotomy. Blood Transfus. 2014;12:s75-83.

121. Padhi S, Glen J, Pordes BA, Thomas ME, Guideline Development G. Management of anaemia in chronic kidney disease: summary of updated NICE guidance. BMJ. 2015;350:h2258.

122. Low MS, Speedy J, Styles CE, De-Regil LM, Pasricha SR. Daily iron supplementation for improving anaemia, iron status and health in menstruating women. Cochrane Database Syst Rev. 2016;4:CD009747.

123. Pena-Rosas JP, De-Regil LM, Garcia-Casal MN, Dowswell T. Daily oral iron supplementation during pregnancy. Cochrane Database Syst Rev. 2015(7):CD004736.

124. Ma J, You C, Hao L. Iron chelators for acute stroke. Cochrane Database Syst Rev. 2012(9):CD009280.

125. Neuberger A, Okebe J, Yahav D, Paul M. Oral iron supplements for children in malaria-endemic areas. Cochrane Database Syst Rev. 2016;2:CD006589.

126. Trotti LM, Bhadriraju S, Becker LA. Iron for restless legs syndrome. Cochrane Database Syst Rev. 2012(5):CD007834.

127. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, et al. The support of human genetic evidence for approved drug indications. Nat Genet. 2015;47(8):856-60.

Chapter 3: Mediators of the effect of educational attainment on cardiovascular disease risk

All of the work presented in this chapter is my own, unless otherwise indicated in the text.

Related publication

- Carter AR, **Gill D**, Davies NM, Taylor AE, Tillmann T, Vaucher J, Wootton RE, Munafo MR, Hemani G, Malik R, Seshadri S, Woo D, Burgess S, Davey Smith G, Holmes MV, Tzoulaki I, Howe LD and Dehghan A. Understanding the consequences of education inequality on cardiovascular disease: mendelian randomisation study. BMJ. 2019;365:l1855.
- **Gill D**, Efstathiadou A, Cawood K, Tzoulaki I and Dehghan A. Education protects against coronary heart disease and stroke independently of cognitive 5 function: evidence from Mendelian randomization. Int J Epidemiol. 2019. doi: 10.1093/ije/dyz200 [Epub ahead of print].

Data sources

- CARDIoGRAMplusC4D Consortium
- COGENT Consortium
- GIANT Consortium
- MEGASTROKE Consortium
- National Institute of Neurological Disorders and Stroke-Stroke Genetics Network
- UK Biobank

3.1 Introduction

Cardiovascular disease (CVD) makes up the world's single greatest cause mortality, and is responsible for more than 17 million deaths every year (1). Social factors such as educational attainment are major determinants of CVD (2-4), with previous observational and Mendelian randomization (MR) evidence suggesting that for every additional 3.6 years spent in full time education, the probability of suffering from coronary artery disease (CAD) falls by approximately 33% (2). However, educational attainment is not easily modifiable, and is typically related to complex environmental and personal factors. Furthermore, intelligence and educational attainment are closely related, with evidence of bi-directional causal effects (5), as well as a shared genetic aetiology (6). Disentangling the effect of education on cardiovascular risk, and its mediators could have important public health implications, as it would allow for resources to be appropriately allocated into targeting the relevant determinants of health. Furthermore, for scenarios where educational attainment itself cannot be directly modified, downstream modifiable mediators of its effect may still be targeted.

Previous work has suggested that body mass index (BMI), systolic blood pressure (SBP) and cigarette smoking account for some of the discrepancy observed in CVD risk related to varying levels of education (7-9). However, such evidence largely originates from observational methods that are subject to notable limitations. Specifically, measurement error can be introduced because these approaches only study a cross-sectional snapshot of risk factor profiles, which are in fact dynamic across life course (10). Similarly, unmeasured or unknown confounding is a major limitation in observational epidemiology (11). Randomised, controlled trials that vary time spent in education are also not appropriate or practical.

The use of randomly allocated genetic variants to proxy an exposure of interest in MR allows for some of the limitations of observational research related to bias from environmental confounding and measurement error to be overcome (12). Furthermore, both network MR and multivariable MR (MVMR) approaches can be used in mediation analyses to quantify the degree to which the effect of an exposure on an outcome is attributable to a particular mediating phenotype (13, 14). Similarly, MVMR can also be used to estimate the effects of an exposure on an outcome that are not mediated through traits closely related to the exposure under consideration (15). In the case of education, it would be important to know whether any effects on CVD outcomes are mediated through cognition, for example, so that causal exposure can be directly targeted in public health interventions. The availability of large-scale genome-wide association studies (GWAS) for education (6), cognition (6), BMI (16), SBP (17), smoking (18), CAD (19), and stroke (20, 21) have made it possible to disentangle the effects of education on

CVD risk, as well as identify mediators, and its effects that are not mediated through cognitive function.

While MR has previously been employed to identify an effect of educational attainment on BMI, SBP and smoking, with these traits each also affecting CVD risk (22-27), these efforts have not formally quantified the degree of any mediation towards the protective effect of education on CVD. The following MR analyses performed in this work investigate the degree to which the effect of education on CAD and stroke risk is mediated by BMI, SBP and lifetime smoking, as well as the three risk factors together, and further the effect of education on cardiovascular risk that arises through pathways unrelated to cognitive function.

3.2 Methods

Genetic association estimates

Single-nucleotide polymorphisms (SNPs) were used as genetic instruments in the MR analyses, and their association with the phenotypes of interest were obtained from summary data produced in existing GWASs. For the analyses investigating the mediators of education on CVD risk, summary genetic association estimates for education were obtained from the Social Science Genetic Association Consortium (SSGAC) GWAS meta-analysis performed in 1,131,881 European ancestry individuals (28), and instruments were 1,271 independent (pairwise linkage disequilibrium [LD] r^2<0.1) genome-wide significant (P<5x10^{-8}) SNPs obtained from the analysis of the full sample in the published study (28). For the MVMR adjusting the effect of education for cognition (and vice versa), the summary data from this study that excluded the 23andMe cohort were used, consisting of 766,345 individuals of European ancestry (6). Instruments were again selected based on their genome wide-significance and LD r^2<0.1. For all analyses, educational attainment was defined as years of full time education completed, with the International Standard Classification of Education (ISCED) system used to compare educational achievement where there was discrepancy in the qualification systems used between the cohorts (29). For demonstrative purposes, Table 3.1 represents the equivalent number of years of education in the UK for each ISCED category based on UK Biobank data (18). Estimates are scaled to standard deviation (SD) units, which corresponds to a 3.6 year change in full time education.

Table 3.1. International Standard for Classification of Education (ISCED) codes mapped to the UK Biobank data, to relate qualifications and years of education estimates

Qualification (as reported in UK Biobank)	ISCED	Years of education
College or University degree	5	20
NVQ or HND or HNC or equivalent	5	19
Other prof. qual. e.g. nursing, teaching	4	15
A levels/AS levels or equivalent	3	13
O levels/GCSEs or equivalent	2	10
CSEs or equivalent	2	10
None of the above	1	7
Prefer not to answer	Excluded	

Genetic association estimates for cognitive function were taken from a GWAS meta-analysis performed in the UK Biobank and Cognitive Genomics Consortium (COGENT) Consortium, including 257,841 individuals of European ancestry (6). The same selection criteria was used to identify instruments as for educational attainment (i.e. genome-wide significance and $r^2<0.1$). Cognitive function was measured using verbal-numerical reasoning tests in UK Biobank, and neuropsychological assessment in the COGENT Consortium analyses (6), with association estimates provided in SD units.

Genetic association estimates for BMI were obtained from the Genetic Investigation of ANthropometric Traits (GIANT) Consortium's 2018 GWAS meta-analysis of 681,275 European ancestry individuals (30). Genetic association estimates for SBP (18), and smoking (31), were from a GWAS of 318,417 White British ancestry participants in the UK Biobank, as previously described. Blood pressure was measured automatically at the baseline assessment centre, with readings taken twice, and separated by two minutes (18). The GWAS used the second reading, with missing data replaced with the first measure, or any follow up recordings. For individuals on any antihypertensive medication, 10mmHg was added to the SBP reading (32). Lifetime smoking was estimated for UK Biobank participants using self-reported age at initiation, age at cessation and smoking intensity (cigarettes smoked per day), also accounting for an exponential decrease in the effect of cigarettes on health over time (31). It had a scoring unit that ranged from 0 (non-smokers) to 4.17, with a mean 0.35 and SD of 0.69 (31). BMI, SBP and smoking instruments for MR analysis were selected genome-wide significant SNPs clumped to an LD threshold $r^2<0.001$ and distance >10,000kb based on a 1000 genomes European reference panel.

For CAD genetic association estimates, the CARDIoGRAMplusC4D 1000 Genomes-based multi-ethnic (approximately 75% European) GWAS meta-analysis of 60,801 cases and 123,504 controls was used (19). A broad definition was applied for cases, including acute coronary syndrome, myocardial infarction and angina (19). In the analyses investigating mediators of the effect of education on cardiovascular disease, publicly available genetic association estimates from the MEGASTROKE Consortium's multi-ethnic GWAS meta-analysis of 67,162 stroke cases (comprising of ischaemic stroke, intracerebral haemorrhage and stroke of unknown type) and 406,111 controls were used (21). Participants were of European (approximately 87%), East, South and mixed Asian, African, and Latin American ancestry (21). However, the publicly available MEGSATROKE GWAS data was not approved for the study of intelligence, so all MR analyses considering the effect of cognition, or education adjusted for cognition, used genetic association estimates from a GWAS of 37,792 ischemic stroke cases and 397,209 controls performed by the National Institute of Neurological Disorders and Stroke-Stroke Genetics Network (20), which were downloaded from the Cerebrovascular Disease Knowledge Portal (33). Participants were of European (94%), Hispanic and African ancestry (20). The World Health Organisation definition of stroke was used, defined as rapidly evolving clinical signs of impaired cerebral lasting 24 hours or more or causing death, with no cause other than of vascular origin.

Mendelian randomization

Two-sample, ratio method MR analysis was performed to estimate the effect of education and cognition on CAD and stroke risk, respectively, with standard errors estimated using the Delta method (second order weights, to account for imprecision in both the exposure and outcome genetic association estimates) (34). MR estimates across individual SNPs were pooled using fixed-effect inverse-variance weighted (IVW) meta-analysis (35).

For MR mediation analysis, the network method was used to estimate the effect of BMI, SBP and smoking individually, on mediating the effect of education on the CVD outcomes (13). Specifically, IVW MR was used to first measure the effect of education on each risk factor separately (i.e. BMI, SBP and smoking), with regression-based MVMR implemented to then estimate the effect of each risk factor on risk of the considered CVD outcome (i.e. CAD or stroke), whilst adjusting for genetic effect of the instruments on education (36). The indirect effect of education on risk of each cardiovascular outcome mediated by each considered risk factor was estimated by multiplying results from the aforementioned IVW and MVMR analyses (13). The two stages of network MR are depicted in Figure 3.1, which details BMI as the mediator and CAD as the outcome, for demonstrative purposes. The estimated indirect effect was finally divided by the total effect to obtain a proportion of the total effect of education mediated through the considered risk factor.

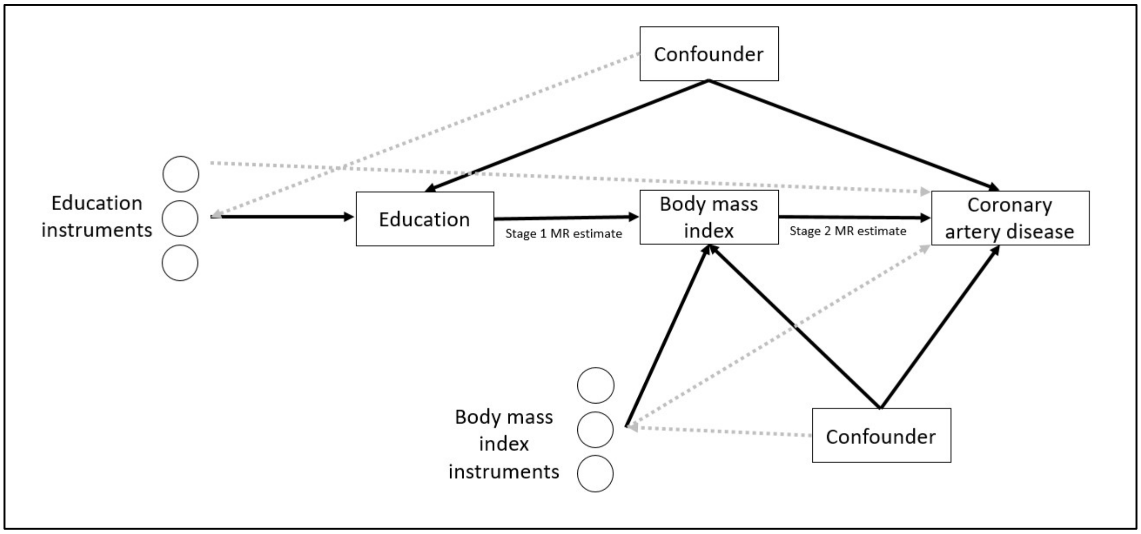


Figure 3.1. A schematic figure detailing the two stages of network Mendelian randomization (MR) in the scenario where education is the exposure, body mass index is the mediator and coronary artery disease risk is the outcome. To estimate the indirect effect of education on coronary artery disease risk in this scenario, the MR estimate from stage 1 is multiplied by that from stage 2. The dotted grey lines represent associations that would violate the requisite assumptions of the model.

To estimate the total effect of education mediated through all three risk factors together, MVMR was first used to estimate the direct effect of education on each respective CVD outcome after adjusting for genetic associations of the instrument SNPs with the three risk factors. This is demonstrated in Figure 3.2 for the scenario where CAD is the outcome under consideration. This direct effect estimate was then divided by the total effect of education on the considered CVD outcome (estimated using IVW MR) and subtracted from one. For all mediation analysis, standard errors were estimated using the propagation of error method.

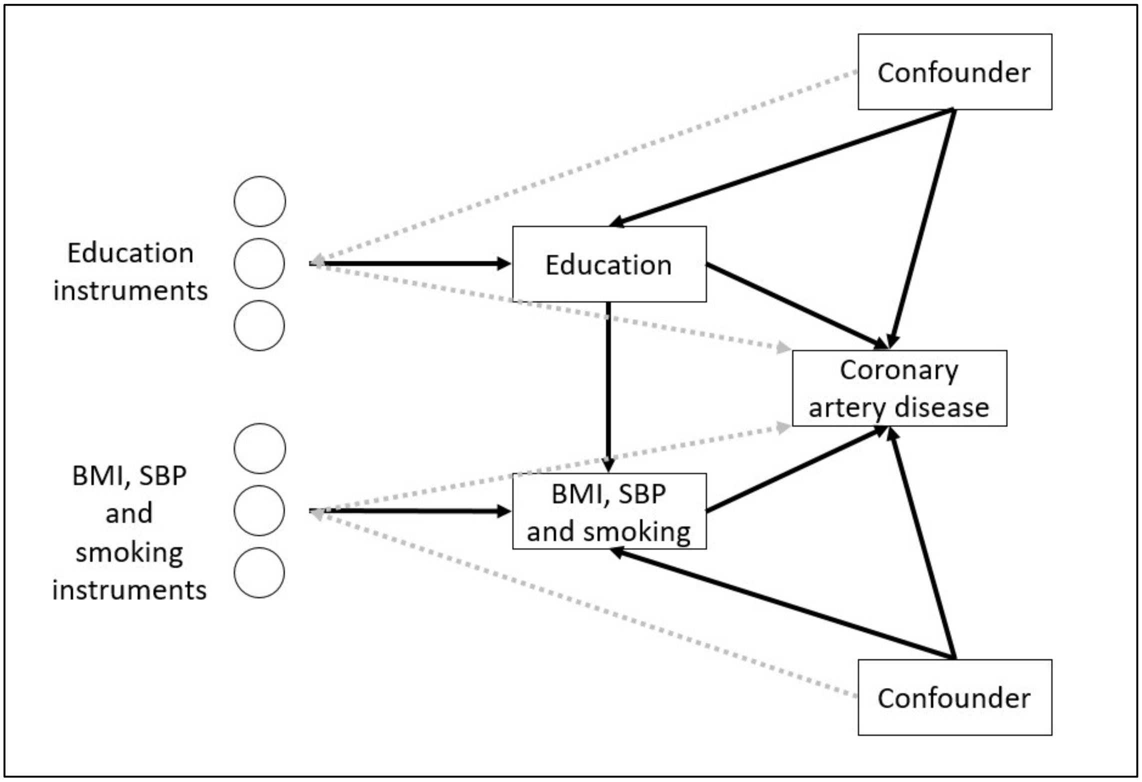


Figure 3.2. A schematic figure detailing the principles of multivariable Mendelian randomization for estimating the direct effect of education on coronary artery disease risk, which is not mediated through body mass index (BMI), systolic blood pressure (SBP) and lifetime smoking. The dotted grey lines represent associations that would violate the requisite assumptions of the model.

MR estimates can be biased if the necessary assumption of the model are not met (37). Horizontal pleiotropy describes the scenario where a genetic variant is related to the outcome under consideration through some pathway independent of the exposure being studied (37). To investigate for this, statistical sensitivity analyses that are more robust to the inclusion of pleiotropic variants were performed, namely MR-Egger and weighted median MR (36, 38). MR-Egger performs a linear regression of the SNP-outcome association estimates on the SNP-exposure association estimates, weighted for the precision of the SNP-outcome association estimates (36). This generates MR estimates that are adjusted for potential horizontal pleiotropy, with a non-zero regression intercept further serving as a test for directional pleiotropy (36). MR-Egger is valid when any direct effect of the variants on the outcome (i.e. not mediated through the exposure) are not correlated to the association of the variants with the exposure (36). In contrast, weighted median MR orders the MR estimates obtained from individual instrument SNPs by their magnitude, weighted for their precision, and consequently identifies the median value as the overall MR estimate (38). Confidence intervals can be calculated by bootstrapping (38), and the approach is usually robust when more than half of the analysis data is obtained from valid instruments.

Regression-based multivariable MR using summary data was performed to investigate the direct effects of education and cognition (i.e. not mediated through each other) on CAD and stroke risk respectively. With this approach, adjustment was made for the genetic associations of the education instruments for cognition, and vice versa (36). Specifically, a regression is performed of the SNP-outcome genetic association estimates on the SNP-exposure genetic association estimates and the SNP-genetic confounder genetic association estimate, weighted for the inverse standard error (precision) of the SNP-outcome estimates, with the a fixed zero intercept. Only instruments for the exposure under consideration were included in each analysis. The modelling assumptions made for IVW MR also apply to MVMR, including those relating to pleiotropic effects of the instrument SNPs. To explore this, MVMR-Egger was also performed as a sensitivity analysis (39). This method is similar to standard regression-based multivariable MR (36), but additionally does not fix the intercept of the regression line to zero, instead using it as a test for directional pleiotropy and additionally generating pleiotropy-adjusted effect estimates in a similar fashion to MR-Egger (36). MVMR median regression sensitivity analysis was also performed. This technique estimates the median of the SNP-outcome genetic association estimates, when conditioned on the SNP-exposure genetic association estimates and the SNP-genetic confounder genetic association estimates, with a zero intercept and weighted for the inverse standard error (precision) of the SNP-outcome genetic association estimates. As with conventional median-based MR approaches, standard errors for multivariable MR median regression can be estimated by bootstrapping (38).

As a further sensitivity analysis of the MVMR analyses investigating the direct effects of education and cognition on CAD and stroke risk respectively, all analyses were performed 1,000 times after sampling, at random and without replacement, a subset of only 200 of the available instrument SNPs for education and cognition in the respective analyses. This also additionally served as a sensitivity analysis into whether any differences in the findings for the effect of education and cognition on CAD or stroke risk related to a discrepancy in the number of available instruments.

All statistical analysis was performed using R 3.4.3 (The R Foundation for Statistical Computing), with the TwoSample MR package used to conduct IVW, MR-Egger and weighted median MR analyses (40). Ethical approval was not required because only summary data, obtained from existing studies that had each received relevant ethical approval and participant consent, were used.

Ethical approval

Ethical approval and participant consent for use of all data in this work had been previously obtained in their respective primary studies, and therefore was not required to be sought again here. The primary studies from which data were obtained have been cited on first introduction.

3.3 Results

Mediators of education

For the populations used to obtain the respective genetic association estimates, the standard deviation (SD) of educational attainment was 3.6 years, BMI was 4.69kg/m^2 and SBP was 18.68mmHg. For smoking, the 1-SD unit increase was used to scale effects related to lifetime smoking intensity, with a mean of 0.35 and SD of 0.69. The units cannot be directly translated to the number of cigarettes smoked (31). Details for instrument SNPs used for all MR analyses have previously been published (18, 41). All instruments had F statistics greater than 10, suggesting that any bias of MR estimates derived from individual instrument variants will be less than 10% of the bias arising in observational analysis (18, 41, 42).

Results of the univariable MR analyses are detailed in Table 3.2. There was evidence of a causal effect of education on all the considered cardiovascular risk factors (BMI, SBP and smoking), as well as risk on CAD and stroke. Consistent estimates were produced in MR sensitivity analyses that are more robust to the inclusion of pleiotropic variants, with the MR-Egger intercept not providing evidence of directional pleiotropy for any outcome.

Table 3.2. Univariable Mendelian randomization (MR) analyses investigating the effect of educational attainment. Coronary artery disease (CAD) and stroke are in odds ratio (OR) units, while body mass index (BMI), systolic blood pressure (SBP) and smoking are in standard deviation (SD) units, per 1-SD increase in education. CI: confidence interval.

Exposure-outcome	MR Method	Estimate	95% lower CI	95% upper CI	P value
Education-CAD	IVW	0.63	0.60	0.67	<0.001
	MR-Egger	0.68	0.54	0.85	0.001
	Intercept				0.370
	Weighted median	0.62	0.57	0.67	<0.001
Education-stroke	IVW	0.71	0.68	0.75	<0.001
	MR-Egger	0.72	0.60	0.87	0.001
	Intercept				0.757
	Weighted median	0.71	0.66	0.76	<0.001
Education-BMI	IVW	-0.22	-0.24	-0.20	<0.001
	MR-Egger	-0.28	-0.49	-0.07	0.009
	Intercept				0.989
	Weighted median	-0.27	-0.30	-0.23	<0.001
Education-SBP	IVW	-2.86	-3.12	-2.60	<0.001
	MR-Egger	-2.41	-3.93	-0.89	0.002
	Intercept				0.325
	Weighted median	-3.39	-3.82	-2.95	<0.001
Education-smoking	IVW	-0.32	-0.33	-0.31	<0.001
	MR-Egger	-0.29	-0.36	-0.22	<0.001
	Intercept				0.057
	Weighted median	-0.35	-0.37	-0.33	<0.001

For the mediation analysis, the considered risk factors of BMI, SBP and smoking each mediated 8 to 33% of the effect of education on CVD (CAD and stroke) risk (Table 3.3), with this rising to 36 to 41% when considering all three risk factors together.

Table 3.3. Mendelian randomization (MR) estimates for the effects of education on coronary artery disease (CAD) and stroke that are mediated through body mass index (BMI), systolic blood pressure (SBP) and smoking. CI: confidence interval

Outcome	Mediator	Percentage of the effect of education mediated through each risk factor (95% CI)
Coronary artery disease	BMI	18 (14 – 23)
	SBP	21 (15 – 27)
	Smoking	33 (17 – 49)
	BMI, SBP and smoking together	36 (5 – 68)
Stroke	BMI	8 (4 – 13)
	SBP	28 (21 – 35)
	Smoking	20 (4 – 36)
	BMI, SBP and smoking together	41 (7 – 75)

Direct effects of education and cognition on cardiovascular disease risk

Results of the MR analyses investigating the effects of education and cognition on CAD and stroke risk respectively are provided in Figure 3.3. There was consistent evidence of a protective effect of education on CAD and stroke risk respectively, even after adjustment was made for cognition. However, cognition only demonstrated a protective effect on risk of CAD, but not stroke, when no adjustment was made for education. However, the protective effect of cognition on CAD risk was not present after adjusting for education using MVMR methods. The MR-Egger and MVMR-Egger intercept *P* values did not support the presence of directional pleiotropy in any of the analyses (Table 3.4).

Table 3.4. Mendelian randomization (MR)-Egger and multivariable MR-Egger (MVMR-Egger) intercept P value results. These intercept values act as a test for directional pleiotropy (with statistical significance defined as P<0.05).

Analysis	*P*
Education-coronary artery disease MR-Egger	0.60
Education-coronary artery disease MVMR-Egger (adjusted for cognition)	0.46
Cognition-coronary artery disease MR-Egger	0.61
Cognition-coronary artery disease MVMR-Egger (adjusted for education)	0.83
Education-ischaemic stroke MR-Egger	0.48
Education-ischaemic stroke MVMR-Egger (adjusted for cognition)	0.74
Cognition-ischaemic stroke MR-Egger	0.65
Cognition-ischaemic stroke MVMR-Egger (adjusted for education)	0.73

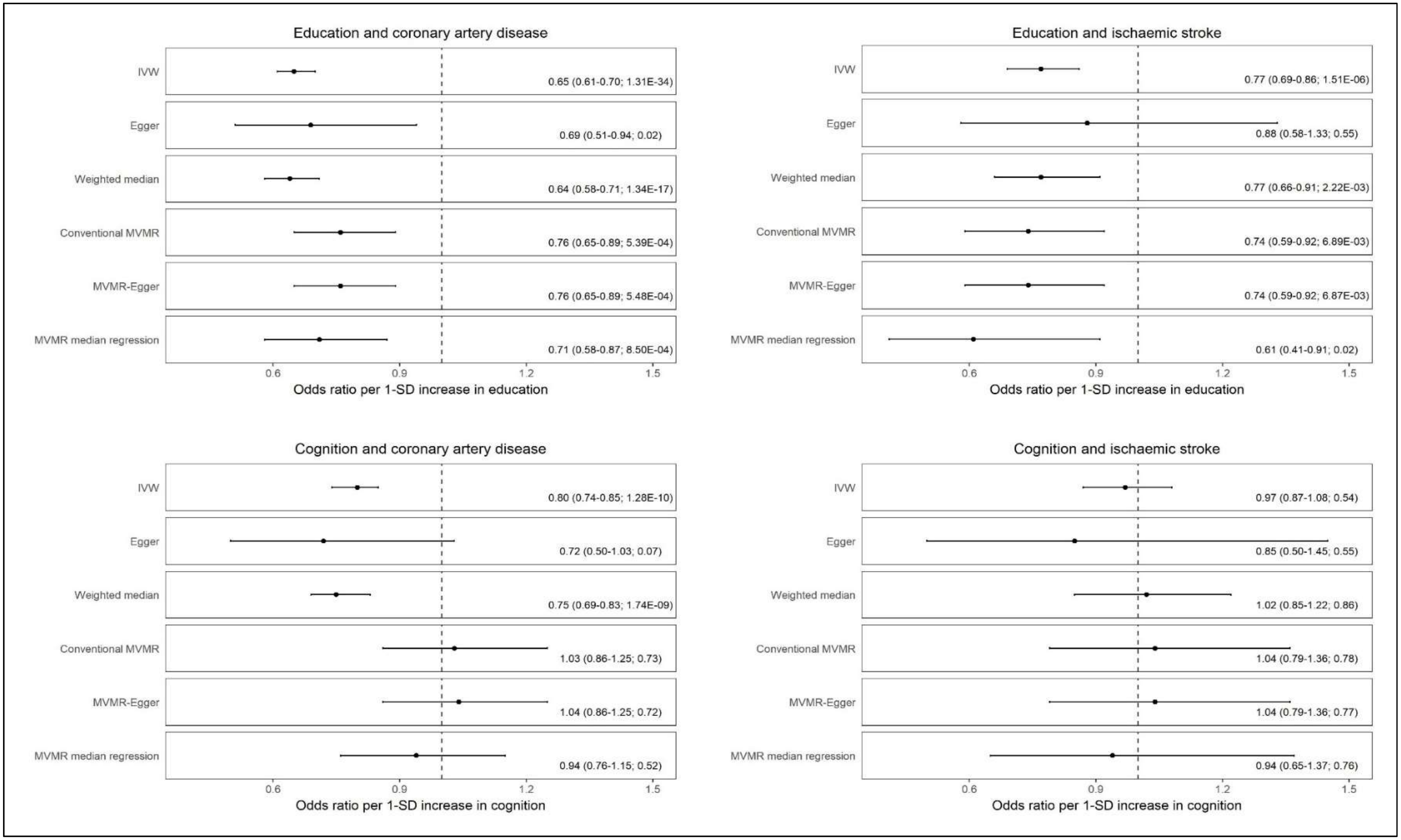


Figure 3.3. Univariable and multivariable Mendelian randomization (MR) results for the analysis investigating the effects of education and cognition on coronary artery disease and ischaemic stroke risk, respectively (41). The univariable analyses estimate the total effect of the exposure on the outcome under study. Considering education as the exposure, multivariable MR (MVMR) analyses are adjusted for cognition. When cognition is the exposure, multivariable MR analyses are adjusted for education. For both univariable and multivariable MR, Egger and median statistical sensitivity analyses that make distinct assumptions regarding the inclusion of pleiotropic

variants are provided. Odds ratios are detailed, with 95% confidence intervals and P values in brackets. SD: standard deviation. The standard deviation of education is 3.6 years, When randomly selecting only 200 instrument SNPs from the total available pool (625 for education and 226 for cognitive function, respectively (41), and repeating the standard MVMR analysis 1,000 times (i.e. examining the effect of education adjusted for cognition, and vice versa), results consistent with the main MVMR analyses that included all instrument SNPs were produced (Figure 3.4).

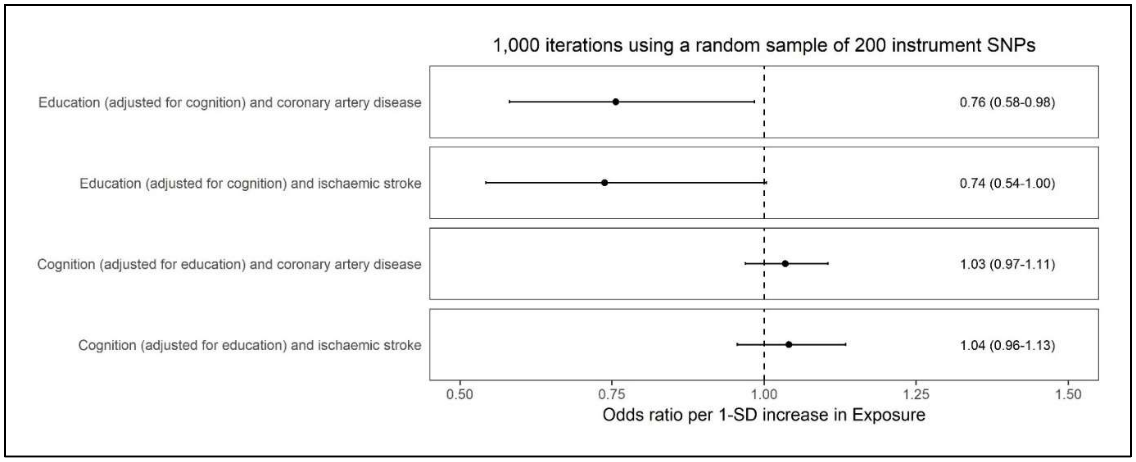


Figure 3.4. Results of the multivariable Mendelian randomization performed 1,000 times after randomly sampling (without replacement) 200 instruments from the available pool of 625 for education and 226 for cognition (41). The mean odds ratios (and 95% confidence intervals) are detailed. SNP: single-nucleotide polymorphism; SD: standard deviation. The standard deviation of education is 3.6 years,

3.4 Discussion

Main findings

The findings of the MR analyses support that approximately 8% to 33% of the effect of education on CAD and stroke risk is mediated through any one of BMI, SBP and smoking. Interestingly, this estimate rose to 36% and 41% for CAD and stroke respectively when considering all three of these risk factors together, which is considerably lower than would be expected if their effects were independent and additive. Therefore, these findings also suggest that more than 50% of the effect of education in protecting against CVD is not related to these risk factors and occurs through other mechanisms. While the study did not consider other traditional cardiovascular risk factors, including those related to exercise, diet, dyslipidaemia and glycaemic control (43-49), it is likely that these would have at least partially overlapped with the considered BMI, SBP and smoking phenotypes. The remaining effects of education may therefore be related to other factors, such as engagement with healthcare services and compliance with medical advice.

Furthermore, exploring whether it is education or cognition per se that has implications for cardiovascular health, the univariable MR analyses supported that education has a protective effect on risk of both CAD and stroke risk, and that cognition only showed evidence of having a protective effect on CAD risk, but not stroke risk. However, the MVMR that investigated the direct effects of education and cognition (i.e. not mediate through each other) supported that it is in fact education that has a direct protective effect on cardiovascular risk. In converse, there was no apparent effect of cognition on CAD (or stroke) risk when adjusting for education. Consistent findings were obtained when using MR methods that are more robust to the inclusion of pleiotropic SNPs, suggesting that bias from pleiotropy was unlikely to be affecting these conclusions. Furthermore, the same pattern of results was obtained when randomly sampling 200 SNPs (from the full pool of instruments), in the MVMR that adjusted education for cognition, and vice versa, suggesting that the results of the main analyses were not related to differences in the number of instrument SNPs used in the various analyses.

Research in context

The findings of the MR analyses, in terms of the proportion of the effect of education mediated by BMI, SBP and smoking, are consistent with traditional observational analyses (18), thus triangulating evidence across distinct methodological approaches. Indeed, numerous previous observational studies have performed mediation analyses to investigate the degree to which BMI, SBP and smoking mediate the effect of education on CVD risk (7, 8, 50, 51), to generate consistent results irrespective of whether education was measured using time spent in school, or through academic qualifications. In a Dutch study, approximately 27% of the association between education and CAD was related to smoking, with 10% and 5% attributed to adiposity and high blood pressure respectively (50). Another study estimated that 7% and 14% of the association between education and CVD could be explained by BMI and raised blood pressure respectively, although with no evidence that smoking was involved in the association (8). Other studies that have stratified by sex also found consistent evidence of SBP and smoking mediating the association (51).

While a protective effect of education on CVD risk has been described previously in observational work (52-54), as well as using MR (2, 18), these studies did not adjust for cognition or make any attempt to disentangle the direct contributions of education and cognition in this relationship. The attenuation of the protective effect of cognition on risk of CAD after having adjusted for education has previously been identified in conventional observational research (55-57), although in the case of stroke, cognition was described to affect risk of stroke independently of education (58, 59). This discrepancy may relate to differences in the definition used for cognitive function, as well as the broad range of measures and domains used in its assessment.

For the various causal effects considered in this work, the MR analysis results are larger in magnitude than corresponding observational estimates (18), possibly related to their instrumenting the lifetime effect of the exposure under consideration, rather than representing a single snapshot as in traditional observational epidemiology. Furthermore, MR does not suffer the same potential confounding that observation approaches are susceptible to, perhaps also explaining some of the discrepancy in the point estimates achieved between approaches (18).

MR has previously been used to investigate the causal effect of education on CAD, BMI, SBP and smoking (2, 22-24), as well as the effect of BMI and smoking on CVD (26, 27). However, these previous studies did not perform formal mediation analysis, as was done in this work. The current MR analysis also makes other notable advances over these previous studies in that it uses larger GWAS data to increase statistical power, particularly for the education instruments, which here may explain approximately 12% of the phenotypic variation, as compared to the 3% reportedly explained in the previous GWAS meta-analysis (28, 60). Furthermore, the current work also separately considers CAD and stroke, thus confirming consistency in the effect of education on both these CVD outcomes. With regard to consideration of smoking, this has typically been modelled as a binary trait in previous work to potentially introduce bias in MR (61), in contrast to the current approach where it is modelled as a continuous variable, and also accounts for lifetime smoking intensity and time since cessation, where applicable (31).

Strengths and limitations

To avoid the introduction of collider bias through the adjustment for BMI that is made in many large scale GWAS of SBP (62, 63), these genetic association estimates were obtained from a GWAS conducted in the UK Biobank that did not make this adjustment. This was particularly important as BMI was one of the mediators under investigation. A major potential limitation of MR is related to bias through pleiotropic effects of the genetic variants employed as instruments on the outcome through pathways at least partly independent of the exposure. In anticipation of this possibility, statistical sensitivity analyses that are more robust to the inclusion of such pleiotropic variants were fully incorporated. The consistent findings in these analyses provided support that pleiotropy was unlikely to be a cause of major bias, and would be unlikely to be affecting the conclusions drawn. Of relevance, the wide 95% CIs of the MR-Egger approach are expected given it's notoriously low statistical power (64). Furthermore, the IVW, MR-Egger, multivariable MR and multivariable MR-Egger approaches all rely on the 'instrument strength independent of direct effect' (InSIDE) assumption being maintained, which requires that the genetic associations of the instruments with the exposure are not correlated to any direct effect that they have on the outcome (36). As it is possible that the associations of the education and cognition instruments are proportional to any direct effect that they have on CAD or stroke risk, this assumption may break down, rendering the MR estimates from these methods biased. Reassuringly however, consistent results were obtained when using the weighted median approaches, making this possibility less likely to affect the conclusions drawn.

A strength of mediation analysis performed using MR is its robustness to mediator measurement error. This is important because measurement error in a mediator can result in underestimation of mediation, possibly also explaining the discrepancy between observational and MR analyses investigating this (10, 18). For example, SBP varies temporally, even within the same day, and as such investigation of mediation through SBP using traditional observational approaches may be susceptible to underestimation. Furthermore, the SBP instruments measure average adult SBP, and may not capture blood pressure variability, which is an independent risk factor for stroke (65). Other sources of measurement error include participants underreporting undesirable traits, such as smoking, and over-reporting desirable traits, such as education (66).

Another theoretical source of bias in two-sample MR analysis relates to participant overlap in the populations sued to obtain genetic association estimates between the exposure and the outcome (67). However, it is unlikely that there was greater than 10% overlap in these populations for the current work, so the implications of this are unlikely to be marked (18, 41). In addition, precautions were taken in all MR analyses to minimise any implications of participant overlap in the genetic association estimates used within any given analysis. Firstly, only relatively strong instruments that are less susceptible to such bias were incorporated (68). Secondly, all IVW MR analyses used second-order weights (i.e. derived using the Delta method, the second order expansion of the Taylor series), which can reduce the rate of false-positive results in this context, as compared to the use of first-order weight. A separate consideration is participant overlap between the populations used to obtain genetic association estimates for the exposure and the mediator (or genetic confounder) in MVMR. The majority of participants involved in the studies used for obtaining genetic association estimates for education (58%) and cognition (86%) were from the UK Biobank (28). However, this does not introduce bias in the context of MVMR.

It is important to note that in the investigation of the direct effects of education and cognition on CAD and stroke risk respectively, although the instruments used for the univariable MR were all considered relatively strong, with F-statistics > 30 (41), there is not currently any available method to measure instrument strength in the MVMR setting when using only summary data (69). It is therefore theoretically possible that weak instrument bias was affecting these analyses.

The UK Biobank cohort is made up of a select population that may not be representative of the UK in general (70), and selection bias related to this can bias MR analyses (71). Discrepancies in the genetic ancestry of populations considered in MR is another potential source of bias, because for example, some of the genetic variants may have different frequencies or effect sizes in different ethnic groups. The education and cognition GWAS meta-analyses considered only participants of European ancestry, in contrast to the multi-ethnic populations included in the study of CAD and stroke. However, greater than 75% of the participants of these multi-ethnic GWAS meta-analyses consisted of European ancestry participants, thus limiting the potential implications of this.

In regard to the GWAS meta-analyses used for education and cognition, the estimates used are likely to have been inflated due to the effects of parental rearing on these traits, independently of inherited variants (6). In addition, the relationship between the genetic variants and the traits considered might vary in relation to different environmental contexts (6), thus representing an additional source of bias in the MR. Cognition was assessed in the respective GWAS analyses using verbal-numerical reasoning tests and neuropsychological assessment (6), which may only pick up on some aspects of cognitive function. Finally, education was measured as the number of years spent in an academic institution (using academic qualifications to infer this in some contexts). However, education can also be considered as an on-going learning process that is not restricted to such a definition. For example, obtaining skills and life experience through other means may also offer similar benefits, and warrants further study.

Clinical implications

Interventions to increase the minimum duration of compulsory education typically require both social and political reform, which may not be easy to achieve. The finding that approximately 40% of the effect of education on reducing cardiovascular risk occurs through BMI, SBP and smoking is of relevance to policymakers because it offers an opportunity to target the downstream mediators of education, and thus also minimise healthcare inequalities related to difference in educational attainment. Furthermore, these findings provide quantitative estimates of the mediators of education in reducing CVD risk, thus allowing for calculated decisions to be made, which consider both the economic and social costs of disparities in education. It is relevant that this work finds that BMI, SBP and smoking together account for less than half of the total effect of educational attainment on CVD risk. Further work that investigates other possible mediators of education's effect on reducing risk of these diseases, as well as the interplay between them, will be important for understanding the remaining effect of education. Similarly, further work might also aim to explore whether these findings can be extrapolated to diverse populations, including different ethnic groups and educational systems.

Regarding the effects of education and cognition on CVD risk, these traits are closely related (6), with previous work suggesting bi-directional effects (5). People with higher cognitive function are likely to spend more time in education, and in the other direction, spending more time in education can also improve cognition (72). The MR analyses investigating the independent effects of education and cognition on CVD risk therefore offer important insight towards understanding which trait should be targeted to improve clinical outcomes and optimise population health. Despite education and cognition overlapping (5, 6), the MVMR approaches used in this work provide evidence that it is education rather than cognition that is causal for CVD, and further that any effect of cognition is only causal through effects on education. These findings would therefore suggest that increasing the minimum compulsory time that an individual must spend in education would protect against CVD risk, irrespective of whether cognition is affected. Consistent with this, previous interventions through educational policy have reduced morbidity and mortality from a range of chronic diseases, including CAD and stroke (73). In this example, the UK has increased the age that individuals must remain in full time education from 16 to 18 years, offering a case study to highlight that such changes are feasible, and beneficial to health (73). A further consideration is that the health and economic benefits of such interventions may not be fully apparent for some years after their first introduction, so sufficient follow up is required for accurate evaluation. Furthermore, it is also unclear whether similar benefits might be achieved through education that is provided in a different format, such as through vocational training or apprenticeships, and further research will be required to explore this.

The findings from the various MR analyses performed in this work can also be aggregated to offer insight into why it is education rather than cognition that protects against CVD risk. The effects of education mediated through BMI, SBP and smoking are consistent with the finding that greater education is associated with a healthier lifestyle, including with lower BMI, SBP and smoking incidence (74-76). Additionally, education protects against CVD through improved socioeconomic status, including association with professions that have safer working conditions, and life situations that offer better healthcare access (74-76). Furthermore, more educated people may also better understand modifiable risk factors related to health, offering them greater opportunity to optimise these (74-76). In all this, cognition may be secondary to education, as it could be the additional knowledge afforded by educational settings that provides the majority of the described benefits, rather than improved cognition itself.

Conclusions

The MR analyses performed here provide evidence to suggest that approximately a third of the effect of education on reducing CVD risk is mediated through BMI, SBP and smoking, with these three traits together explaining approximately 40% of the effect of education. Thus, interventions that aim to reduce these risk factors would negate some of the effect of disparities in educational attainment in a European population settings. However, more than half of the protective effect of education on cardiovascular risk is mediated through alternative pathways, and further work is required to identify these. This study also provides evidence to support that it is a direct effect of education that protects against CVD risk. This is in-keeping with the proposed mechanisms by which education is likely to be exerting its beneficial effects, and highlights educational attainment as a potentially modifiable exposure that may be targeted to improve population cardiovascular health. The findings of this work are consistent with the existing body of literature in this field, and add important novel insight towards preventing CVD.

3.5 References

1. Roth GA, Johnson C, Abajobir A, Abd-Allah F, Abera SF, Abyu G, et al. Global, Regional, and National Burden of Cardiovascular Diseases for 10 Causes, 1990 to 2015. J Am Coll Cardiol. 2017;70(1):1-25.

2. Tillmann T, Vaucher J, Okbay A, Pikhart H, Peasey A, Kubinova R, et al. Education and coronary heart disease: mendelian randomisation study. BMJ. 2017;358:j3542.

3. Davies NM, Dickson M, Davey Smith G, van den Berg GJ, Windmeijer F. The causal effects of education on health outcomes in the UK Biobank. Nat Hum Behav. 2018;2(2):117-25.

4. Di Chiara T, Scaglione A, Corrao S, Argano C, Pinto A, Scaglione R. Association between low education and higher global cardiovascular risk. J Clin Hypertens. 2015;17(5):332-7.

5. Anderson EL, Howe LD, Wade KH, Ben-Shlomo Y, Hill WD, Deary IJ, et al. Education, intelligence and Alzheimer's disease: Evidence from a multivariable two-sample Mendelian randomization study. bioRxiv. 2018:401042.

6. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet. 2018;50(8):1112-21.

7. Nordahl H, Rod NH, Frederiksen BL, Andersen I, Lange T, Diderichsen F, et al. Education and risk of coronary heart disease: assessment of mediation by behavioral risk factors using the additive hazards model. Eur J Epidemiol. 2013;28(2):149-57.

8. Degano IR, Marrugat J, Grau M, Salvador-Gonzalez B, Ramos R, Zamora A, et al. The association between education and cardiovascular disease incidence is mediated by hypertension, diabetes, and body mass index. Sci Rep. 2017;7(1):12370.

9. Lynch JW, Kaplan GA, Cohen RD, Tuomilehto J, Salonen JT. Do cardiovascular risk factors explain the relation between socioeconomic status, risk of all-cause mortality, cardiovascular mortality, and acute myocardial infarction? Am J Epidemiol. 1996;144(10):934-42.

10. Blakely T, McKenzie S, Carter K. Misclassification of the mediator matters when estimating indirect effects. J Epidemiol Community Health. 2013;67(5):458-66.

11. Richiardi L, Bellocco R, Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias. Int J Epidemiol. 2013;42(5):1511-9.

12. Smith GD, Ebrahim S. Mendelian randomization: can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003;32(1):1-22.

13. Burgess S, Daniel RM, Butterworth AS, Thompson SG, Consortium E-IA. Network Mendelian randomization: using genetic variants as instrumental variables to investigate mediation in causal pathways. Int J Epidemiol. 2015;44(2):484-95.

14. Burgess S, Thompson DJ, Rees JMB, Day FR, Perry JR, Ong KK. Dissecting Causal Pathways Using Mendelian Randomization with Summarized Genetic Data: Application to Age at Menarche and Risk of Breast Cancer. Genetics. 2017;207(2):481-7.

15. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. Am J Epidemiol. 2015;181(4):251-60.

16. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015;518(7538):197-206.

17. Neale Lab. Accessed 2019 January 16. Rapid GWAS of thousands of phenotypes for 337,000 samples in the UK Biobank 2018. http://www.nealelab.is/blog/2017/7/19/rapid-gwas-of-thousands-of-phenotypes-for-337000-samples-in-the-uk-biobank.

18. Carter AR, Gill D, Davies NM, Taylor AE, Tillmann T, Vaucher J, et al. Understanding the consequences of education inequality on cardiovascular disease: mendelian randomisation study. BMJ. 2019;365:l1855.

19. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet. 2015;47(10):1121-30.

20. Pulit SL, McArdle PF, Wong Q, Malik R, Gwinn K, Achterberg S, et al. Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. Lancet Neurol. 2016;15(2):174-84.

21. Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. Nat Genet. 2018;50(4):524-37.

22. Bockerman P, Viinikainen J, Pulkki-Raback L, Hakulinen C, Pitkanen N, Lehtimaki T, et al. Does higher education protect against obesity? Evidence using Mendelian randomization. Prev Med. 2017;101:195-8.

23. Hagenaars SP, Gale CR, Deary IJ, Harris SE. Cognitive ability and physical health: a Mendelian randomization study. Sci Rep. 2017:7(1):2651.

24. Gage SH, Bowden J, Davey Smith G, Munafo MR. Investigating causality in associations between education and smoking: a two-sample Mendelian randomization study. Int J Epidemiol. 2018;47(4):1131-40.

25. Holmes MV, Lange LA, Palmer T, Lanktree MB, North KE, Almoguera B, et al. Causal effects of body mass index on cardiometabolic traits and events: a Mendelian randomization analysis. Am J Hum Genet. 2014;94(2):198-208.

26. Lyall DM, Celis-Morales C, Ward J, Iliodromiti S, Anderson JJ, Gill JMR, et al. Association of Body Mass Index With Cardiometabolic Disease in the UK Biobank: A Mendelian Randomization Study. JAMA Cardiol. 2017;2(8):882-9.

27. Asvold BO, Bjorngaard JH, Carslake D, Gabrielsen ME, Skorpen F, Smith GD, et al. Causal associations of tobacco smoking with cardiovascular risk factors: a Mendelian randomization analysis of the HUNT Study in Norway. Int J Epidemiol. 2014;43(5):1458-70.

28. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet. 2018:50(8):1112-21.

29. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, et al. Genome-wide association study identifies 74 loci associated with educational attainment. Nature. 2016;533(7604):539-42.

30. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. Hum Mol Genet. 2018;27(20):3641-9.

31. Wootton RE, Richmond RC, Stuijfzand BG, Lawn RB, Sallis HM, Taylor GMJ, et al. Causal effects of lifetime smoking on risk for depression and schizophrenia: Evidence from a Mendelian randomisation study. bioRxiv. 2018.

32. Tobin MD, Sheehan NA, Scurrah KJ, Burton PR. Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. Stat Med. 2005;24(19):2911-35.

33. Cerebrovascular Disease Knowledge Portal, NINDS grant # 1R24NS092983. 2018 April 02; http://cerebrovascularportal.org/.

34. Thompson JR, Minelli C, Del Greco MF. Mendelian Randomization using Public Data from Genetic Consortia. Int J Biostat. 2016;12(2).

35. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013;37(7):658-65.

36. 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.

37. Sheehan NA, Didelez V, Burton PR, Tobin MD. Mendelian randomisation and causal inference in observational epidemiology. PLoS Med. 2008;5(8):e177.

38. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol. 2016;40(4):304-14.

39. Rees JMB, Wood AM, Burgess S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. Stat Med. 2017;36(29):4705-18.

40. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. eLife. 2018;7.

41. Gill D, Efstathiadou A, Cawood K, Tzoulaki I, Dehghan A. Education protects against coronary heart disease and stroke independently of cognitive 5 function: evidence from Mendelian randomization. Int J Epidemiol 2019:doi: 10.1093/ije/dyz200 [Epub ahead of print].

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

43. Luepker RV, Rosamond WD, Murphy R, Sprafka JM, Folsom AR, McGovern PG, et al. Socioeconomic status and coronary heart disease risk factor trends. The Minnesota Heart Survey. Circulation. 1993;88(5):2172.

44. Garrison RJ, Gold RS, Wilson PWF, Kannel WB. Educational Attainment and Coronary Heart Disease Risk: The Framingham Offspring Study. Prev Med. 1993;22(1):54-64.

45. Mayer O, Šimon J, Heidrich J, Cokkinos DV, De Bacquer D. Educational level and risk profile of cardiac patients in the EUROASPIRE II substudy. J Epidemiol Community Health. 2004;58(1):47-52.

46. Jacobsen BK, Thelle DS. Risk factors for coronary heart disease and level of education. The Tromso Heart Study. Am J Epidemiol. 1988;127(5):923-32.

47. Matthews KA, Kelsey SF, Meilahn EN, Kuller LH, Wing RR. Educational attainment and behavioral and biologic risk factors for coronary heart disease in middle-aged women. Am J Epidemiol. 1989;129(6):1132-44.

48. Lynch J, Smith GD, Harper S, Hillemeier M. Is income inequality a determinant of population health? Part 2. US national and regional trends in Income inequality and age-and cause-specific mortality. Milbank Q. 2004;82(2):355-400.

49. Loucks EB, Gilman SE, Howe CJ, Kawachi I, Kubzansky LD, Rudd RE, et al. Education and Coronary Heart Disease Risk Potential Mechanisms Such as Literacy, Perceived Constraints, and Depressive Symptoms. Health Educ Behav. 2015;42(3):370-9.

50. Kershaw KN, Droomers M, Robinson WR, Carnethon MR, Daviglus ML, Monique Verschuren WM. Quantifying the contributions of behavioral and biological risk factors to socioeconomic disparities in coronary heart disease incidence: the MORGEN study. Eur J Epidemiol. 2013;28(10):807-14.

51. Veronesi G, Ferrario MM, Kuulasmaa K, Bobak M, Chambless LE, Salomaa V, et al. Educational class inequalities in the incidence of coronary heart disease in Europe. Heart. 2016;102(12):958-65.

52. Qureshi AI, Suri MFK, Saad M, Hopkins LN. Educational attainment and risk of stroke and myocardial infarction. Med Sci Monitor. 2003;9(11):CR466-73.

53. Chang C-L, Marmot MG, Farley TMM, Poulter NR. The influence of economic development on the association between education and the risk of acute myocardial infarction and stroke. J Clin Epidemiol. 2002;55(8):741-7.

54. Wang H, Yuan Y, Song L, Qiu G, Lai X, Yang L, et al. Association between education and the risk of incident coronary heart disease among middle-aged and older Chinese: the Dongfeng-Tongji Cohort. Sci Rep. 2017;7(1):776.

55. Ariansen I, Mortensen L, Igland J, Tell GS, Tambs K, Graff-Iversen S, et al. The educational gradient in coronary heart disease: the association with cognition in a cohort of 57 279 male conscripts. J Epidemiol Community Health. 2015;69(4):322-9.

56. Hemmingsson T, Essen Jv, Melin B, Allebeck P, Lundberg I. The association between cognitive ability measured at ages 18–20 and coronary heart disease in middle age among men: A prospective study using the Swedish 1969 conscription cohort. Soc Sci Med. 2007;65(7):1410-9.

57. Lawlor DA, Batty GD, Clark H, McIntyre S, Leon DA. Association of Childhood Intelligence with Risk of Coronary Heart Disease and Stroke: Findings from the Aberdeen Children of the 1950s Cohort Study. Eur J Epidemiol. 2008;23:695-706.

58. Wiberg B, Lind L, Kilander L, Zethelius B, Sundelöf JE, Sundström J. Cognitive function and risk of stroke in elderly men. Neurology. 2010;74(5):379-85.

59. Elkins JS, Knopman DS, Yaffe K, Johnston SC. Cognitive function predicts first-time stroke and heart disease. Neurology. 2005;64(10):1750-5.

60. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, et al. Genome-wide association study identifies 74 loci associated with educational attainment. Nature. 2016;533(7604):539-42.

61. Burgess S, Labrecque JA. Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. Eur J Epidemiol. 2018;33(10):947-52.

62. Warren HR, Evangelou E, Cabrera CP, Gao H, Ren M, Mifsud B, et al. Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. Nat Genet. 2017;49(3):403-15.

63. Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, 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.

64. Slob EAW, Burgess S. A Comparison Of Robust Mendelian Randomization Methods Using Summary Data. bioRxiv. 2019:577940.

65. Rothwell PM, Howard SC, Dolan E, O'Brien E, Dobson JE, Dahlof B, et al. Effects of beta blockers and calcium-channel blockers on within-individual variability in blood pressure and risk of stroke. Lancet Neurol. 2010;9(5):469-80.

66. Pulcu E. Self-report distortions of puffing topography in daily smokers. J Health Psychol. 2016;21(8):1644-54.

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

68. Pierce BL, Burgess S. Efficient Design for Mendelian Randomization Studies: Subsample and 2-Sample Instrumental Variable Estimators. Am J Epidemiol. 2013;178(7):1177-84.

69. Sanderson E, Davey Smith G, Windmeijer F, Bowden J. An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. Int J Epidemiol 2018:48(3):713-27.

70. Haworth S, Mitchell R, Corbin L, Wade KH, Dudding T, Budu-Aggrey A, et al. Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. Nat Commun. 2019;10(1):333.

71. Taylor AE, Jones HJ, Sallis H, Euesden J, Stergiakouli E, Davies NM, et al. Exploring the association of genetic factors with participation in the Avon Longitudinal Study of Parents and Children. Int J Epidemiol 2018;47(4):1207-16.

72. Deary IJ, Johnson W. Intelligence and education: causal perceptions drive analytic processes and therefore conclusions. Int J Epidemiol. 2010;39(5):1362-9.

73. Hahn RA, Truman BI. Education Improves Public Health and Promotes Health Equity. Int J Health Serv. 2015;45(4):657-78.

74. Hoeymans N, Smit HA, Verkleij H, Kromhout D. Cardiovascular risk factors in relation to educational level in 36 000 men and women in The Netherlands. Eur Heart J. 1996;17(4):518-25.

75. Kilander L, Berglund L, Boberg M, Vessby B, Lithell H. Education, lifestyle factors and mortality from cardiovascular disease and cancer. A 25-year follow-up of Swedish 50-year-old men. Int J Epidemiol. 2001;30(5):1119-26.

76. Woolf SH, Braveman P. Where health disparities begin: the role of social and economic determinants--and why current policies may make matters worse. Health Aff (Millwood). 2011;30(10):1852-9.

Chapter 4: Application of genetic variants to study antihypertensive drug efficacy, side-effects and repurposing potential

In this chapter, the genomic loci of the genes corresponding to antihypertensive drug targets were extracted by Marios K Georgakis (Ludwig-Maximilians Universität LMU, Germany). The phenome-wide association analysis in the UK Biobank was performed by Fotios Koskeridis (University of Ioannina, Greece), and that in the Vanderbilt University Biobank was performed by Lan Jiang, Qiping Feng, Wei-Qi Wei and Joshua C. Denny (Vanderbilt University Medical Center, United States of America). The remainder of the work presented in this chapter is my own, unless otherwise indicated in the text.

Related publication

- **Gill D**, Georgakis MK, Koskeridis F, Jiang L, Feng Q, Wei WQ, Theodoratou E, Elliott P, Denny JC, Malik R, Evangelou E, Dehghan A, Dichgans M and Tzoulaki I. Use of Genetic Variants Related to Antihypertensive Drugs to Inform on Efficacy and Side Effects. Circulation. 2019;140:270-279.

Data sources

- CARDIoGRAMplusC4D Consortium
- International Consortium of Blood Pressure
- MEGASTROKE
- UK Biobank
- Vanderbilt University Biobank

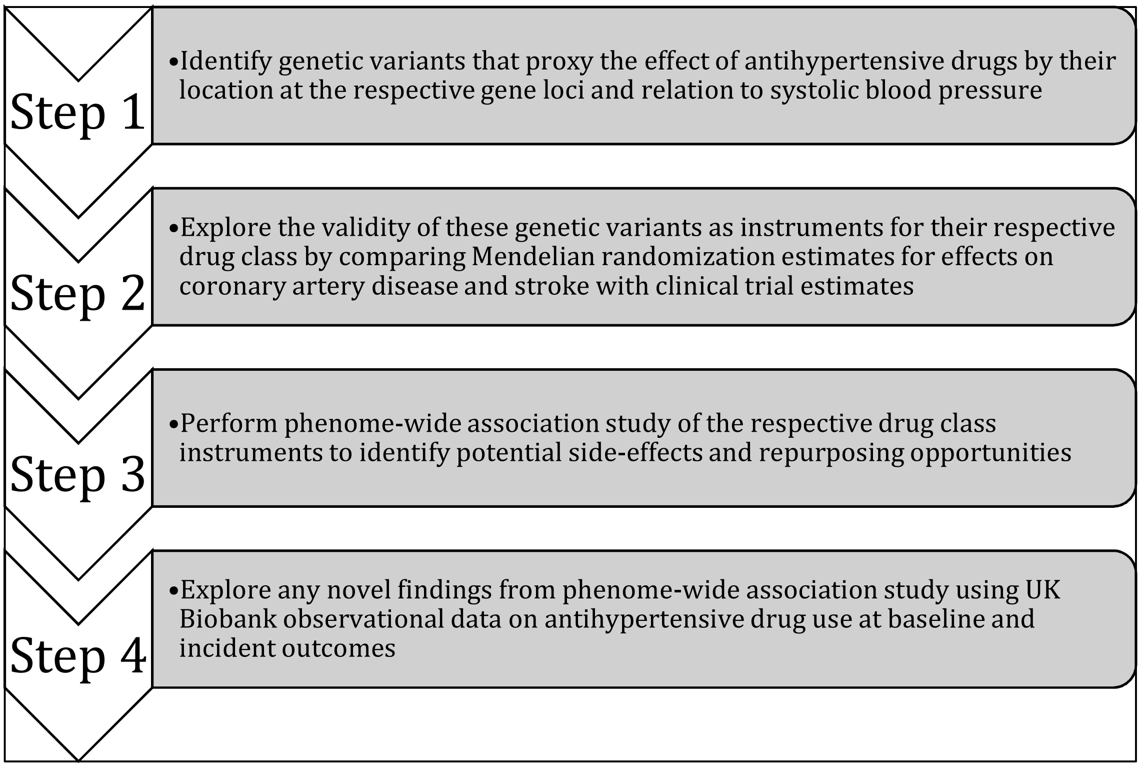
4.1 Introduction

Hypertension is a major risk factor for cardiovascular disease (CVD), and thus represents a leading cause of morbidity and mortality worldwide. There are approximately 900 million individuals around the world estimated to have a systolic blood pressure (SBP) greater than 140mmHg, which in turn is responsible for 106 deaths per 100,000 in the population every year, and loss of 143 million disability-adjusted life years (1). Interventions to lower blood pressure can reduce CVD risk, and every 10mmHg reduction in SBP has been estimated to lower all-cause mortality by 13% (2).

There have been notable developments in the available pharmacological treatments for treating hypertension through randomized controlled trials (RCTs), with several drug classes identified as being effective and safe for blood pressure lowering (3). However, there are also constraints to the RCT design (4). To increase their economic efficiency (i.e. reduce their overall cost and duration), they often focus on individuals at higher risk of the disease outcomes under consideration, as well as a relatively short follow up period (5). This often makes it necessary in clinical practice to extrapolate their findings to a wider population than those considered in the actual studies. Furthermore, the RCT design does not prioritise the identification of side-effects or repurposing opportunities (6). Although traditional observational epidemiology has been more conducive towards these aims, such approaches are often limited by environmental confounding factors, indication biases and reverse causation (7).

Genome-wide association study (GWAS) of blood pressure traits has resulted in the availability of genetic association estimates related to genetic variants at genes corresponding to the protein targets of antihypertensive drugs (7). Such variants can serve as instrumental variables for the effect of varying systolic blood pressure (SBP) through the drug target corresponding to the particular locus under consideration (7). This approach has previously been extensively employed to study the effect of genetic variants related to lipid lower drug targets (8-11). In support of such a strategy, trials of drugs for which there is supportive genetic evidence are more likely to be successful than for those for which there is no such supportive evidence (12). Furthermore, genetic data can be utilised to identify and validate targets for drug development (13), as well as increase the probability of successful drug discovery (14).

In this work, genetic variants located at genes corresponding to the protein targets of common antihypertensive drugs that can serve as instrumental variables for studying these medications were first identified. To confirm the validity of the identified variants as proxies for their respective antihypertensive drug class, Mendelian randomization (MR) estimates for their SBP lowering their effects on coronary artery disease (CAD) and stroke risk were compared to corresponding estimates obtained for pharmacological therapy in RCTs against placebo. Finally, phenome-wide association study (PheWAS) was performed using genetic risk scores constructed for the genetic variants that proxy the effect of each respective antihypertensive drug class, with the aim of identifying potential side-effects and repurposing opportunities. The various stages of this work are summarised schematically in Figure 4.1.



Step 1

- Identify genetic variants that proxy the effect of antihypertensive drugs by their location at the respective gene loci and relation to systolic blood pressure

Step 2

- Explore the validity of these genetic variants as instruments for their respective drug class by comparing Mendelian randomization estimates for effects on coronary artery disease and stroke with clinical trial estimates

Step 3

- Perform phenome-wide association study of the respective drug class instruments to identify potential side-effects and repurposing opportunities

Step 4

- Explore any novel findings from phenome-wide association study using UK Biobank observational data on antihypertensive drug use at baseline and incident outcomes

Figure 4.1. A schematic figure summarising the steps of the work presented in this Chapter.

4.2 Methods

Genetic variant selection

The antihypertensive drugs to be investigated in this work were selected from expert guidelines on pharmacological therapy for treating hypertension (6). The following drug classes were short-listed for consideration:

- Angiotensin-converting-enzyme inhibitor (ACEI)
- Angiotensin receptor blocker (ARB)
- Beta-blocker (BB)
- Calcium channel blocker (CCB)
- Thiazide diuretic (TD)

The genes corresponding to the blood pressure lowering protein targets of these drugs were identified using the DrugBank database (15), a publicly accessible online database of drug and drug target information. The corresponding genomic location of the identified genes, along with their known enhancers and promoters were retrieved using GeneHancer, a database within the GeneCards publicly accessible online platform (version 4.7) (16).

GWAS summary data for SBP were derived from a study of 757,601 European-ancestry individuals performed by the International Consortium of Blood Pressure (17), which also included UK Biobank participants. To increase the statistical power of GWAS, correction for antihypertensive medication use was made by adding 15mmHg to the SBP of individuals taking any antihypertensive medication (18). Additional adjustment was also made for body mass index (BMI) (17). As a sensitivity analysis to explore the possibility of introduction of collider bias related to correction for antihypertensive medication use or adjustment for BMI, genetic association estimates from an SBP GWAS performed on approximately 337,000 White British individuals from the UK Biobank were also used; this GWAS did not correct for antihypertensive medication use, or adjust for BMI (19).

As uncorrelated instrumental variables for studying the effect of the considered antihypertensive drugs, single-nucleotide polymorphisms (SNPs) at the gene, gene promoter or gene enhancers corresponding to the protein target of the respective drug class that associated with SBP at genome-wide significance ($P<5 \times 10^{-8}$) were identified, and clumped to a linkage disequilibrium (LD) threshold of $r^2<0.1$ based on the 1000 Genomes European reference panel. For all individual SNPs used as instrumental variables, R^2 statistics were calculated as measures of the variance in SBP that they explained, with F statistics calculated to assess their strength (20).

Mendelian randomization

The MR approach was used to study the effect of randomly allocated genetic variants related to SBP lowering through a particular antihypertensive drug target on risk of CAD and stroke. The CARDIoGRAMplusC4D Consortium's 1000 Genomes-based trans-ethnic GWAS meta-analysis considering 60,801 CAD cases and 123,504 controls was used to obtain summary genetic association estimates for CAD risk (21). The MEGASTROKE Consortium's trans-ethnic GWAS meta-analysis of 67,162 stroke cases (of any aetiology) and 454,450 controls was used to obtain genetic association estimates for stroke (22).

As the main MR analysis, the ratio method was used to produce individual MR estimates for each instrument SNP, with standard errors estimated using second order weights (23). Second order weights were preferred over first order weights because they also take into consideration possible measurement error in the exposure (as well as the outcome) (24). To produce overall MR estimates for antihypertensive drugs that had numerous instrument SNPs, individual MR estimates derived from single SNPs were pooled using fixed-effects inverse-variance weighted (IVW) meta-analysis (i.e. the IVW MR method) (24).

MR estimates were given for the approximate SBP lowering effect of the considered drug target in RCTs (3), to thus allow direct comparison between MR estimates and these RCT estimates for the effect of different antihypertensive drug classes on risk of CAD and stroke. This was for the purposes of exploring instrument validity, as the MR estimates in this context would be expected to corroborate RCT findings (25). MR estimates were originally produced in odds ratio (OR) units, and were further converted to relative risk (RR) units for comparison with RCT findings. For this conversion, the baseline risk of CAD and stroke were estimated as 0.042 and 0.041 respectively, based on a systematic review of 613,815 patients that were included in blood pressure lowering clinical trials (2). In a sensitivity analyses to explore the implications of varying baseline risk of CAD and stroke, MR RR estimates were also estimated assuming incidences of 1%, 5% and 10%.

The main analysis MR estimates for the effect of each antihypertensive drug on CAD and stroke risk were compared to corresponding estimates from a systematic review and meta-analysis of RCTs against placebo (3).

Investigation of potential pleiotropy

MR estimates can be biased if the genetic variants used affect the outcome through pathway that is independent of the exposure under consideration, in a phenomenon referred to as pleiotropy. Potential bias from pleiotropy can be investigated both biologically, through secondary associations of the genetic variants selected as instruments, and also by statistical methods that consider the MR estimates produced by different instrument SNPs (26).

To investigate potential bias from pleiotropic variants using biological knowledge, the PhenoScanner database of publicly available genetic association estimates was used to identify genome-wide significant (P<5 x 10^{-8}) associations of the instrument SNPs (or proxies with LD r^2>0.8) with secondary traits (other than blood pressure). Any SNPs identified to have such potentially pleiotropic associations were excluded in consequent MR sensitivity analyses using the IVW method.

Numerous statistical methods for assessing potential bias from pleiotropy were incorporated. Heterogeneity between MR estimates derived from individual instrument SNPs greater than would be expected by chance can indicate potential bias from pleiotropic variants, and was tested using the Cochran's Q test (with P<0.05 representing statistical evidence of pleiotropy) (27). MR techniques that are more robust to the inclusion of pleiotropic variants were also applied in statistical sensitivity analyses. Firstly, MR-Egger was used, which regresses the SNP-outcome associations by the SNP-exposure associations, weighing this for the precision of the SNP-outcome associations (28). Provided that the association of the instrument SNPs with the exposure are not correlated to any direct effects that they have on the outcome (i.e. independent of the exposure), MR-Egger is able to provide an overall MR estimate that is adjusted for any pleiotropic effects of the variants used, and further also offers a non-zero MR-Egger intercept as a test for directional pleiotropy (with P<0.05 used to identify statistical significance) (28). Secondly, MR-PRESSO was applied (29). This performs a regression of the SNP-outcome estimates against the SNP-exposure estimates with an intercept of zero, and uses the consequent residual errors to identify any outlier MR estimates (P<0.05) arising from individual instrument SNPs, and further whether removing these changes the overall MR estimate (29). Finally, the weighted median MR was incorporated. This method orders individual SNP MR estimates by their magnitude, weighted for their precision, and selects the median result as the overall MR estimate (30). Given the potentially low statistical power of MR sensitivity analyses (26, 31), a formal statistical significance threshold was not used for these, but rather consistency in the MR estimates with the main IVW MR approach was assessed.

Phenome-wide association study

The UK Biobank cohort of approximately 500,000 individuals was used for performing PheWAS (32). Genetic data was acquired from individuals through donated blood samples, with data on diagnoses obtained through linkage with Hospital Episode Statistics (HES). PheWAS analysis was limited to individuals of European ancestry, with one participant from each pair of relatives excluded for situations where the kinship coefficient was >0.0884. Genetic variants identified to proxy the effect of antihypertensive drugs were used to create genetic risk scores (GRSs) using PLINK software (33). These scores were weighted for the blood pressure lowering effect of each included variant SNP, and were standardized to have a mean of 0 and a standard deviation (SD) of 1 (33).

Clinical diagnoses recoded in the HES were grouped into clinically relevant groups using the phecode grouping system (34). For the PheWAS analyses, case-control groups were generated for each phecode, where controls were selected as individuals that did not have any diagnosis from the disease group under consideration, nor its related conditions (34). Logistic regression was then performed against the respective antihypertensive drug class GRS, with adjustments made for age, sex and first four genetic principal components. The PheWAS analysis was restricted to phecodes that had a minimum of 200 cases to increase the statistical power of the analyses (35), and a 5% false discovery rate (FDR) threshold was applied for ascertaining statistical significance of the associations.

Consistent to the MR analysis described above, PheWAS sensitivity analyses were also performed using SBP genetic association estimates that had not been corrected for antihypertensive medication use or adjusted for BMI, and also after excluding from the respective GRSs any SNPs that may be having pleiotropic associations (at genome-wide significance), as identified from PhenoScanner (36).

For 5% FDR significant associations with non-cardiovascular outcomes (that are therefore unlikely to be attributable to the blood pressure lowering effects of antihypertensive drugs) identified in PheWAS, the association of these outcomes to a GRS for SBP more generally (i.e. created by selecting SBP variants from throughout the genome rather than restricted to any given antihypertensive protein target gene) was also investigated. Specifically, a permutation-based approach was used generate a GRS for SBP by randomly sampling (without replacement) from the available pool of SBP instruments (that had genome-wide significant associations with SBP and at LD r2<0.001 using a 1000 Genomes European reference panel) an equal number of SNPs to that used in the GRS for the antihypertensive drug under consideration. This was repeated 1,000 times, with the mean and 95% confidence interval of the associations of the GRS for SBP with the outcome under consideration (measured using the same logistic regression model as in the main PheWAS) used to investigate whether this was an SBP-related effect, or specific to the drug class under study. Furthermore, the proportion of the 1,000 analyses using the GRS for SBP that had consistent directions of effect but with magnitude greater than that observed for the investigation with the GRS of the antihypertensive drug class would serve as an adjusted *P*-value of the null hypothesis (i.e. that the effect is due to SBP lowering generally, rather than specific to the drug target under consideration).

Any 5% FDR significant PheWAS associations with the GRSs of the antihypertensive drug targets with non-cardiovascular outcomes were also followed up in the Vanderbilt University Biobank (BioVU), a separate cohort that has genetic data on approximately 50,000 participants linked to their Electronic Health Records (37). As for the main PheWAS, the standardised GRS of the antihypertensive drug class was applied in a logistic regression model with the outcome under consideration (and controls were also identified using the same approach as in the main analysis), restricting to European ancestry individuals and adjusting for age, sex and first three principal components. Association estimates obtained using the UK Biobank and BioVU cohorts were pooled using a fixed-effects meta-analysis model.

Conventional observational analysis of antihypertensive drug use

PheWAS associations for non-cardiovascular conditions reaching 5% FDR significance were also explored by investigating actual antihypertensive drug use in the UK Biobank. This approach additionally allowed the dihydropyridine and non-dihydropyridine CCB subclasses to be distinguished, which was not possible in the GRS analysis because the protein targets of these drug subclasses are related to the same genes. In Cox regression analysis, the time to first incident event was compared between individuals orally taking different antihypertensive drug classes at recruitment to the UK Biobank.

Antihypertensive drug treatment was categorised into ACEIs monotherapy, ARBs monotherapy, BBs monotherapy, dihydropyridine CCBs monotherapy, non-dihydropyridine CCBs monotherapy, TDs monotherapy, a combination of medications from any two antihypertensive classes, and a combination of medications from three of more antihypertensive classes. In a separate analysis model, participants on any subclass of CCBs (i.e. dihydropyridine CCBs or non-dihydropyridine CCBs) were pooled into one category. For all analyses, adjustments were made for age, sex, BMI, Townsend Deprivation Index, smoking status, self-reported diagnosis of cancer, number of non-cancer diagnoses and number of surgical operations. Participants diagnosed with the condition under consideration before they were recruited to the UK Biobank were excluded and those that died during follow-up prior to receiving this diagnosis were censored.

Ethical approval and statistical software

The data used in this work were obtained from studies that had already obtained the necessary ethical approval and participant consent. UK biobank data were accessed through application 236. Statistical analyses were performed using the statistical software R, version 3.4.1 (The R Foundation for Statistical Computing).

4.3 Results

Instrument selection

The protein targets of ACEIs (*ACE* gene), ARBs (*AGTR1* gene), BBs (*ADRB1* gene) and TDs (*SLC12A3* gene) corresponded to single genes respectively, whereas the protein target of CCBs corresponded to 11 genes (*CACNA1D, CACNA1F, CACNA2D1, CACNA2D2, CACNA1S, CACNB1, CACNB2, CACNB3, CACNB4, CACNG1, CACNA1C*), which each encode different calcium channel subunits. The *CACNA1F* gene is located on the X chromosome and SNPs for this were not available.

Considering the gene, promotor and enhancer regions of the relevant genes, 1 instrument SNP was identified for ACEIs, 6 for BBs and 24 SNPs for CCBs (Table 4.1). The F-statistic for SNPs ranged from 54 to 534, suggesting a relatively low risk of bias from use of weak instrument (20). No instrument SNPs were identified for any of the other considered antihypertensive drug classes.

Mendelian randomization

MR estimates for each antihypertensive drug were scaled to the corresponding SBP lowering effect of that agent, in order to allow direct comparison with RCT estimates (against placebo) for effects on CAD and stroke risk. Therefore, MR estimates for ACEIs are given per 21.14mmHg decrease in SBP, for BBs are per 9.51mmHg decrease, and for CCBs are per 8.90mmHg decrease (3). In the main IVW MR analysis, estimates were also converted to RR units from OR units, assuming a CAD and stroke prevalence of 0.042 and 0.041 respectively.

For ACEI, MR using the single identified instrument SNP supported a protective effect on risk of stroke (RR 0.21, 95% confidence interval [CI] 0.06-0.72, *P*=0.01), but not CAD (RR 0.67, 95% CI 0.16-2.56, *P*=0.58). For BBs, the main IVW MR supported a protective effect on risk of CAD (RR 0.62, 95% CI 0.47-0.81, *P*=4x10^{-4}), but not stroke (RR 0.91, 95% CI 0.73-1.14, *P*=0.41). The main IVW MR analysis supported a protective effect of CCBs on risk of both CAD (RR 0.73, 95% CI 0.64-0.84, *P*=6x10^{-6}) and stroke (RR 0.75, 95% CI 0.66-0.84, *P*=1x10^{-6}). Consistent results were found when using OR units, or modelling the incidence of coronary artery disease and stroke as being 1%, 5% or 10% (Table 4.2).

MR analysis results for each antihypertensive drug class had overlapping 95% CIs to the corresponding RCT meta-analysis (against placebo) (3) (Figure 4.2). Figures 4.3-4.6 show individual MR estimates for each of the BB and CCB SNPs for the ratio method MR analysis considering CAD and stroke as outcomes.

Considering the sensitivity analyses that identified antihypertensive drug target instruments and their genetic association estimates with SBP from a GWAS that did not correct for antihypertensive medication use or adjusted for BMI (17), no genetic instrument SNPs were identified for ACEI, with 2 SNPs identified for BB, and 6 for CCB (Table 4.3). IVW MR using these instrument SNPs for BBs and CCBs produced results that were consistent with the main analysis, although with wider 95% CIs (Figures 4.7-4.10).

Searching PhenoScanner for potential pleiotropic effects of the instrument SNPs (36), there was 1 BB SNP and 5 CCB SNPs that may be exerting bias through effects on the considered outcomes through pathways independent of blood pressure lowering (Table 4.4). Performing the IVW MR analysis after excluding these variants similarly produced consistent estimates to the main analysis (Figures 4.7-4.10).

When using statistical methods to explore possible bias from pleiotropy, there was evidence of heterogeneity only in the MR analysis of BBs on risk of stroke (Cochran's Q P=0.03). The MR-Egger intercepts were not significant for directional pleiotropy in any of the analyses (BBs; CAD P=0.87, stroke P=0.89 and CCBs; CAD P=0.89, stroke P=0.51). MR-PRESSO detected 2 outlier SNPs in the analysis of BBs on stroke, and estimates were consistent with the main analysis after excluding these (Figure 4.8). MR-Egger, weighted median MR and MR-PRESSO also produced consistent estimates to the main IVW MR (Figures 4.7-4.10).

Table 4.1. The main instrument single-nucleotide polymorphisms (SNPs) for angiotensin-converting-enzyme inhibitors (ACEIs), beta-blockers (BBs) and calcium channel blockers (CCBs). The Effect estimate is given for change in systolic blood pressure (in mmHg units). The F statistics are provided as an indication of instrument strength.

Drug	SNP	Chromosome	Position	Effect allele	Other allele	Effect allele frequency	Effect	Standard error	P value	Sample size	R^2	F statistic
ACEI	rs4291	17	61554194	a	t	0.6155	-0.2839	0.0312	8.65E-20	745820	3.69E-04	276
BB	rs11196549	10	1.16E+08	a	g	0.0425	0.6884	0.0784	1.58E-18	738169	1.54E-04	114
	rs460718	10	1.16E+08	a	g	0.3266	-0.2764	0.0324	1.36E-17	738169	3.34E-04	247
	rs11196597	10	1.16E+08	a	g	0.1330	0.2858	0.0458	4.23E-10	737164	1.81E-04	134
	rs17875473	10	1.16E+08	t	c	0.0871	0.3283	0.0552	2.66E-09	738170	1.43E-04	106
	rs1801253	10	1.16E+08	c	g	0.7338	0.4626	0.0344	2.84E-41	738169	4.97E-04	367
	rs4359161	10	1.16E+08	a	g	0.1812	-0.2662	0.0391	9.46E-12	738168	2.17E-04	160
CCB	rs3821843	3	53558012	a	g	0.6808	0.3373	0.0335	6.56E-24	736049	4.03E-04	297
	rs114987861	3	53605712	a	g	0.0284	0.5289	0.0958	3.36E-08	737054	8.02E-05	59
	rs113210396	3	53612327	t	g	0.0451	-0.4338	0.0770	1.76E-08	737164	1.03E-04	76
	rs7340705	3	53734443	t	c	0.6732	-0.2425	0.0322	4.87E-14	738169	2.93E-04	217
	rs2488136	10	18334521	a	g	0.2875	0.2261	0.0334	1.22E-11	738169	2.55E-04	188
	rs1888693	10	18440444	a	g	0.3449	0.3858	0.0317	4.69E-34	736050	4.79E-04	353
	rs16916914	10	18457722	t	c	0.9631	-0.5636	0.0806	2.72E-12	737424	1.10E-04	81
	rs7076319	10	18459450	a	g	0.7339	-0.3210	0.0341	5.07E-21	737054	3.45E-04	254
	rs61278674	10	18481737	a	g	0.9062	-0.3298	0.0540	1.03E-09	737163	1.54E-04	114
	rs1779209	10	18514561	t	c	0.2876	0.2736	0.0336	4.23E-16	729448	3.08E-04	225
	rs10828399	10	18553968	a	g	0.5218	-0.1947	0.0302	1.10E-10	738168	2.67E-04	197
	rs10828452	10	18592450	a	t	0.7930	0.3046	0.0388	4.20E-15	737164	2.75E-04	203
	rs10828542	10	18627285	a	g	0.6137	0.1817	0.0311	5.18E-09	738170	2.37E-04	175
	rs12780039	10	18678987	c	g	0.1210	0.2852	0.0470	1.26E-09	738167	1.67E-04	123
	rs112133583	10	18695681	t	c	0.0299	-0.5546	0.0973	1.18E-08	737169	8.84E-05	65
	rs11014170	10	18710991	a	g	0.0206	-0.6701	0.1150	5.61E-09	732148	7.43E-05	54
	rs7923191	10	18727901	a	g	0.7918	-0.3690	0.0376	1.10E-22	737054	3.34E-04	246
	rs12258967	10	18727959	c	g	0.7047	0.6327	0.0337	1.08E-78	737165	7.24E-04	534
	rs72786098	10	18729855	a	g	0.0322	-0.5033	0.0883	1.18E-08	737055	8.62E-05	64
	rs1998822	10	18755664	a	g	0.7234	-0.1958	0.0343	1.15E-08	727331	2.15E-04	157
	rs4748474	10	18790727	a	g	0.5214	0.1946	0.0304	1.61E-10	729908	2.67E-04	195
	rs150857355	12	49209340	c	g	0.0217	0.9406	0.1122	5.20E-17	731300	1.10E-04	80
	rs2239046	12	2434419	a	g	0.6817	0.2082	0.0322	9.58E-11	745818	2.48E-04	185
	rs714277	12	2514270	t	c	0.2834	0.1986	0.0333	2.38E-09	745820	2.22E-04	165

Table 4.2. Mendelian randomization (MR) results, per change in systolic blood pressure observed in clinical trials of the corresponding drug (3), *and for different disease incidence rates, demonstrating how the conversion of MR estimates from odds ratio to relative risk units is affected by modelling different disease incidence rates (1%, 5% and 10%). ACEI: angiotensin-converting enzyme inhibitor; BB: beta-blocker; CCB: calcium channel blocker; CI: confidence interval.*

		Odds ratio			1% incidence			5% incidence			10% incidence		
Drug	Outcome	Odds ratio	Low 95% CI	Upper 95% CI	Relative risk	Low 95% CI	Upper 95% CI	Relative risk	Low 95% CI	Upper 95% CI	Relative risk	Low 95% CI	Upper 95% CI
ACEI	Coronary artery disease	0.66	0.16	2.75	0.66	0.16	2.70	0.67	0.16	2.53	0.68	0.17	2.34
ACEI	Stroke	0.21	0.06	0.72	0.21	0.06	0.72	0.22	0.06	0.73	0.23	0.07	0.74
BB	Coronary artery disease	0.61	0.46	0.80	0.61	0.46	0.80	0.62	0.47	0.81	0.63	0.49	0.82
BB	Stroke	0.91	0.72	1.15	0.91	0.72	1.14	0.91	0.73	1.14	0.92	0.74	1.13
CCB	Coronary artery disease	0.72	0.63	0.83	0.72	0.63	0.83	0.73	0.64	0.84	0.74	0.65	0.84
CCB	Stroke	0.74	0.66	0.84	0.74	0.66	0.84	0.75	0.67	0.84	0.76	0.68	0.85

Table 4.3. Instrument single-nucleotide polymorphisms (SNPs) for beta-blockers (BBs) and calcium channel blockers (CCBs) obtained using genetic association study summary data that were not corrected for antihypertensive medication use or adjusted for body mass index. These instruments were used in sensitivity analyses. The Effect estimate is given for change in systolic blood pressure (in mmHg units).

Drug	SNP	Chromosome	Position	Effect allele	Other allele	Effect	Standard error	*P* value	Sample size
BB	rs151597	10	115720514	c	g	0.0160	0.0026	1.23E-09	317754
	rs1801253	10	115805056	c	g	0.0182	0.0028	9.24E-11	317754
CCB	rs10741083	10	18790858	c	t	0.0142	0.0026	2.64E-08	317754
	rs12258967	10	18727959	g	c	-0.0314	0.0027	1.73E-31	317754
	rs17604757	10	18442940	g	a	0.0318	0.0050	1.78E-10	317754
	rs1779240	10	18476313	a	g	-0.0189	0.0029	6.89E-11	317754
	rs10828650	10	18691531	g	a	0.0185	0.0026	1.56E-12	317754
	rs35593046	3	53553923	t	g	-0.0164	0.0028	6.23E-09	317754

Table 4.4. Possible pleiotropic effects related to the instrument genetic variants, as identified using PhenoScanner (accessed 28 March 2018).

Drug	Instrument	Effect allele	Other allele	SNP	Effect allele	Other allele	Proxy	r^2	Trait	Study	PMID	Ancestry	Year	Beta	Standard error	P
BB	rs1801253	C	G	rs1801253	C	G	No	1.00	Birth weight and gestational age	EGGC	23202124	European	2013	NA	NA	4E-09
		C	G	rs1801253	C	G	No	1.00	Birth weight	EGGC	23202124	European	2013	-0.041	0.0070	4E-09
		C	G	rs1801253	C	G	No	1.00	Birth weight	Neale B	UKBB	European	2017	0.0286	0.0036	3E-15
		C	G	rs1801253	C	G	No	1.00	Height	Neale B	UKBB	European	2017	0.0125	0.0020	2E-10
		C	G	rs2484294	A	G	Yes	0.97	Birth weight	Neale B	UKBB	European	2017	0.0280	0.0036	8E-15
		C	G	rs2484294	A	G	Yes	0.97	Height	Neale B	UKBB	European	2017	0.0125	0.0020	2E-10
		C	G	rs740746	A	G	Yes	0.97	Birth weight	Neale B	UKBB	European	2017	0.0278	0.0036	1E-14
		C	G	rs740746	A	G	Yes	0.97	Height	Neale B	UKBB	European	2017	0.0125	0.0020	2E-10
		C	G	rs2773469	G	A	Yes	0.97	Birth weight	Neale B	UKBB	European	2017	0.0281	0.0036	6E-15
		C	G	rs2773469	G	A	Yes	0.97	Height	Neale B	UKBB	European	2017	0.0125	0.0020	2E-10
		C	G	rs7076938	T	C	Yes	0.96	Birth weight	Horikoshi	27680694	Mixed	2016	0.0349	0.0040	5E-18
		C	G	rs7076938	T	C	Yes	0.96	Birth weight	Neale B	UKBB	European	2017	0.0280	0.0036	8E-15
		C	G	rs7076938	T	C	Yes	0.96	Height	Neale B	UKBB	European	2017	0.0125	0.0020	2E-10
CCB	rs3821843	A	G	rs3821843	A	G	No	1.00	Impedance of	Neale B	UKBB	European	2017	-0.011	0.0020	4E-08
	rs10828399	A	G	rs10828399	A	G	No	1.00	Body mass index	Akiyama	28892062	East	2017	0.0215	0.0037	5E-09
		A	G	rs10764373	T	G	Yes	0.99	Body mass index	Akiyama	28892062	East	2017	0.0215	0.0037	5E-09
		A	G	rs2357928	A	G	Yes	0.94	Body mass index	Akiyama	28892062	East	2017	0.0198	0.0036	5E-08
	rs12780039	C	G	rs79586955	T	A	Yes	0.92	Pulse rate	Neale B	UKBB	European	2017	-0.021	0.0038	2E-08
	rs72786098	A	G	rs72786098	A	G	No	1.00	Small vessel stroke	Cheng YC	26732560	Mixed	2016	1.2730	0.2268	2E-08
	rs714277	C	T	rs714277	C	T	No	1.00	Haematocrit	Astle W	27863252	European	2016	-0.024	0.0039	3E-10
		C	T	rs714277	C	T	No	1.00	Red blood cell	Astle W	27863252	European	2016	-0.024	0.0039	8E-10
		C	T	rs714277	C	T	No	1.00	Schizophrenia	PGC	25056061	Mixed	2014	0.0686	0.0118	7E-09
		C	T	rs12823424	A	G	Yes	1.00	Haematocrit	Astle W	27863252	European	2016	-0.024	0.0039	3E-10
		C	T	rs12823424	A	G	Yes	1.00	Red blood cell	Astle W	27863252	European	2016	-0.024	0.0039	6E-10
		C	T	rs12823424	A	G	Yes	1.00	Schizophrenia	Goes FS	26198764	European	2015	0.0677	0.0116	5E-09
		C	T	rs12823424	A	G	Yes	1.00	Schizophrenia	PGC	25056061	Mixed	2014	0.0690	0.0118	5E-09
		C	T	rs2239063	A	C	Yes	1.00	Haematocrit	Astle W	27863252	European	2016	-0.024	0.0039	3E-10
		C	T	rs2239063	A	C	Yes	1.00	Red blood cell	Astle W	27863252	European	2016	-0.024	0.0039	7E-10
		C	T	rs2239063	A	C	Yes	1.00	Schizophrenia	PGC	25056061	Mixed	2014	0.0690	0.0118	5E-09
		C	T	rs758117	C	T	Yes	1.00	Haematocrit	Astle W	27863252	European	2016	-0.024	0.0039	3E-10
		C	T	rs758117	C	T	Yes	1.00	Red blood cell	Astle W	27863252	European	2016	-0.024	0.0039	6E-10
		C	T	rs758117	C	T	Yes	1.00	Schizophrenia	Li Z	28991256	Mixed	2017	0.0608	0.0108	2E-08
		C	T	rs758117	C	T	Yes	1.00	Schizophrenia	PGC	25056061	Mixed	2014	0.0686	0.0118	6E-09
		C	T	rs10491964	G	A	Yes	0.92	Haematocrit	Astle W	27863252	European	2016	-0.025	0.0040	7E-11
		C	T	rs10491964	G	A	Yes	0.92	Haemoglobin	Astle W	27863252	European	2016	-0.022	0.0040	2E-08
		C	T	rs10491964	G	A	Yes	0.92	Red blood cell	Astle W	27863252	European	2016	-0.025	0.0040	4E-10
		C	T	rs10491964	G	A	Yes	0.92	Schizophrenia	PGC	25056061	Mixed	2014	0.0656	0.0118	3E-08

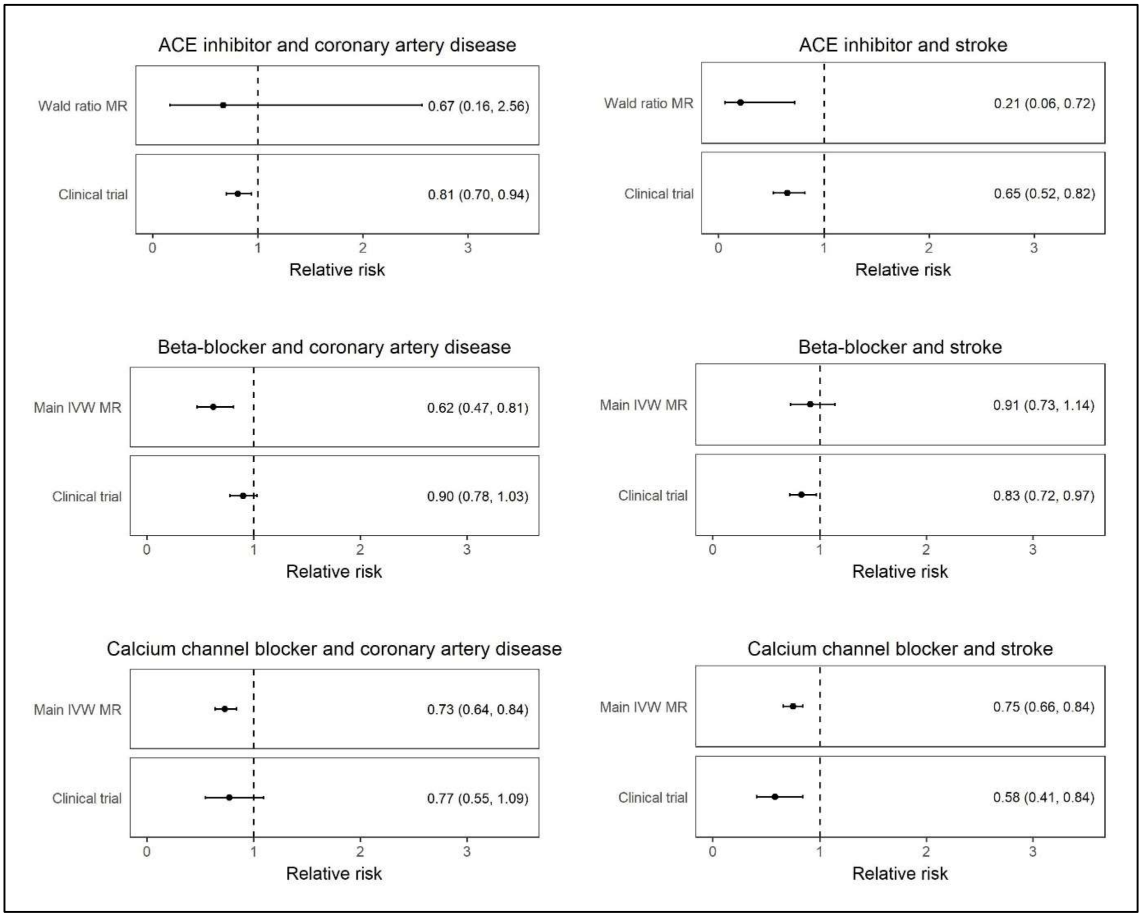


Figure 4.2. The main Mendelian randomization (MR) results compared to clinical trial meta-analyses against placebo (38). The MR estimates have been converted from odds ratios to relative risk estimates to allow comparison with trail estimates, as detailed in the Methods section. There was only one instrument variant for the angiotensin-converting enzyme inhibitor drug class, and so the confidence intervals for related estimates were wider than for other drug classes where more instrument variants were identified. ACE: angiotensin-converting enzyme inhibitor; IVW: inverse-variance weighted.

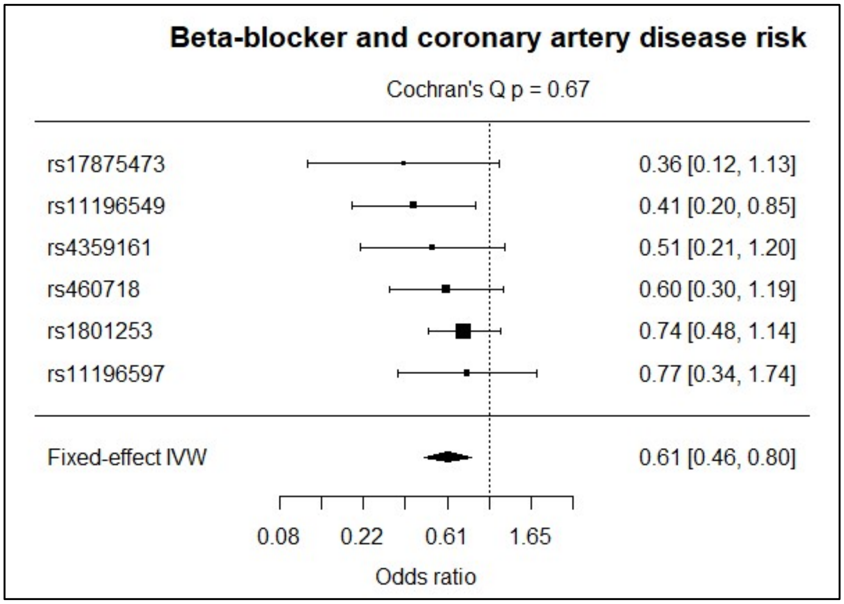


Figure 4.3. Individual instrument Mendelian randomization estimates in the analysis of beta-blockers and coronary artery disease risk. Units are odds ratio units per change in systolic blood pressure observed in clinical trials of beta-blockers (9.51mmHg) (3). *IVW: inverse-variance weighted.*

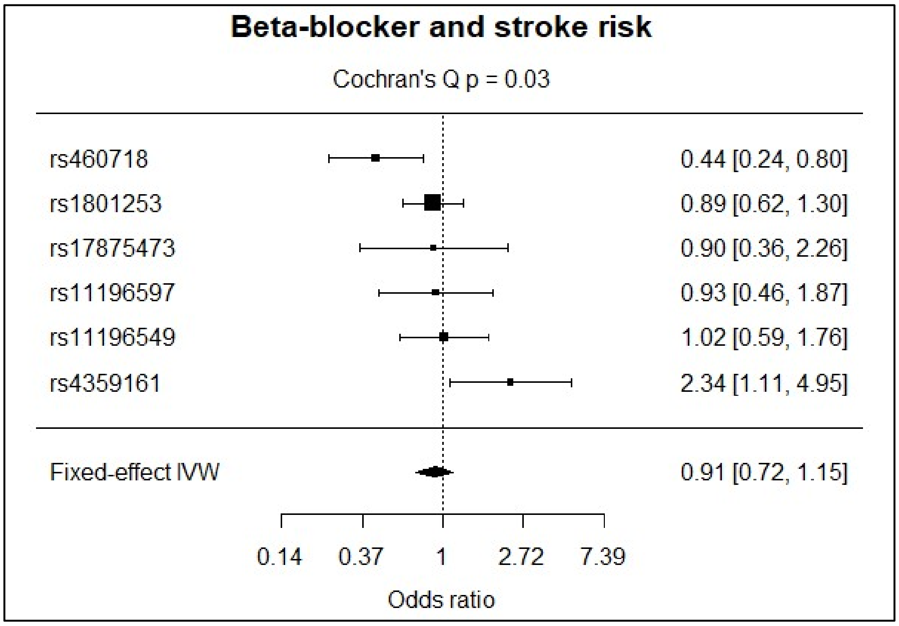


Figure 4.4. Individual instrument Mendelian randomization estimates in the analysis of beta-blockers and stroke risk. Units are odds ratio units per change in systolic blood pressure observed in clinical trials of beta-blockers (9.51mmHg) (3). *IVW: inverse-variance weighted.*

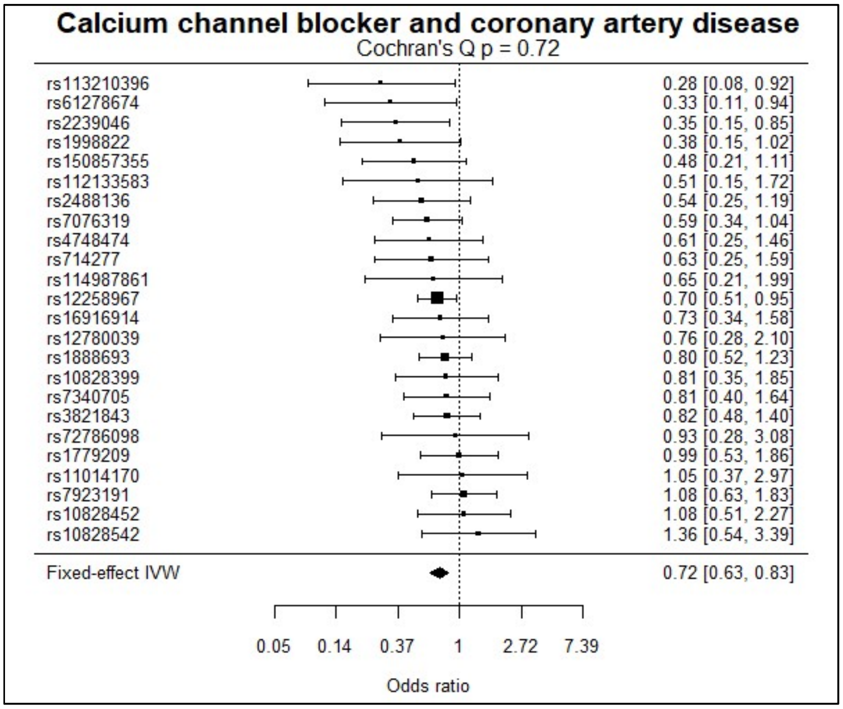


Figure 4.5. Individual instrument Mendelian randomization estimates in the analysis of calcium channel blockers and coronary artery disease risk. Units are odds ratio units per change in systolic blood pressure observed in clinical trials of calcium channel blockers (8.9mmHg) (3). *IVW: inverse-variance weighted.*

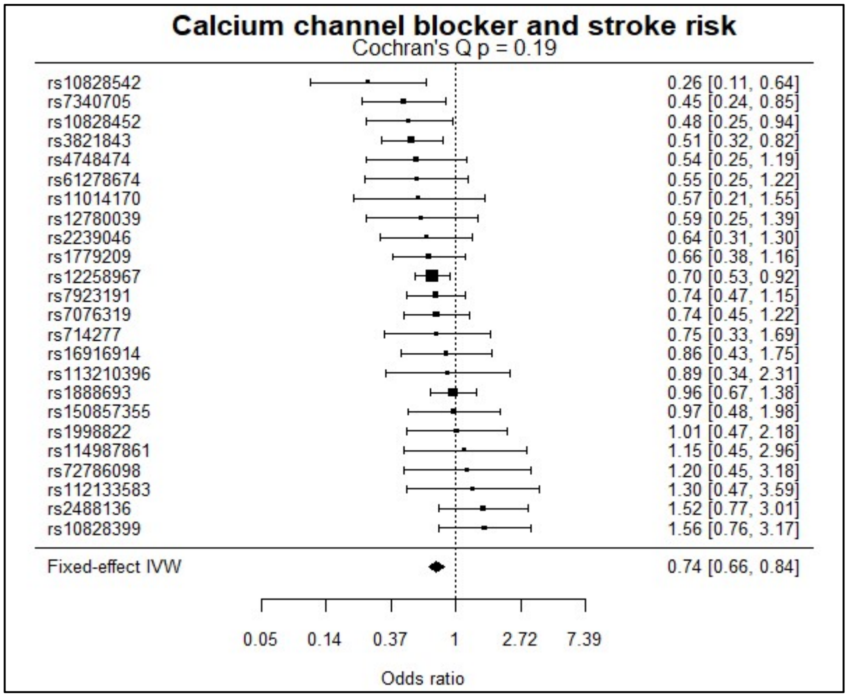


Figure 4.6. Individual instrument Mendelian randomization estimates in the analysis of calcium channel blockers and stroke risk. Units are odds ratio units per change in systolic blood pressure observed in clinical trials of calcium channel blockers (8.9mmHg) (3). *IVW: inverse-variance weighted.*

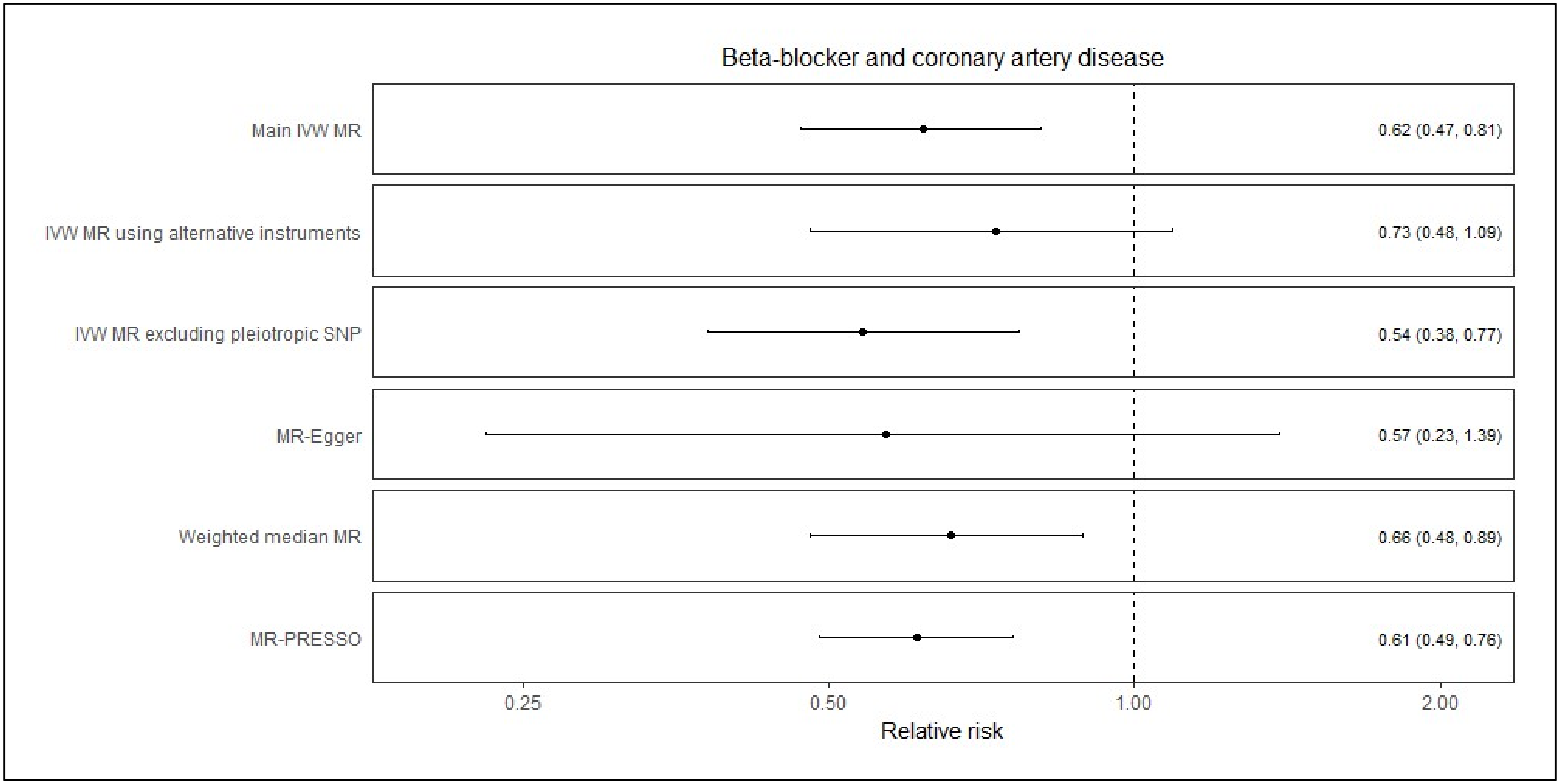


Figure 4.7. Mendelian randomization (MR) sensitivity analyses for the investigation of beta-blockers and coronary artery disease risk. Units are odds ratio units per change in systolic blood pressure observed in clinical trials of beta-blockers (9.51mmHg) (3). *IVW: inverse-variance weighted.*

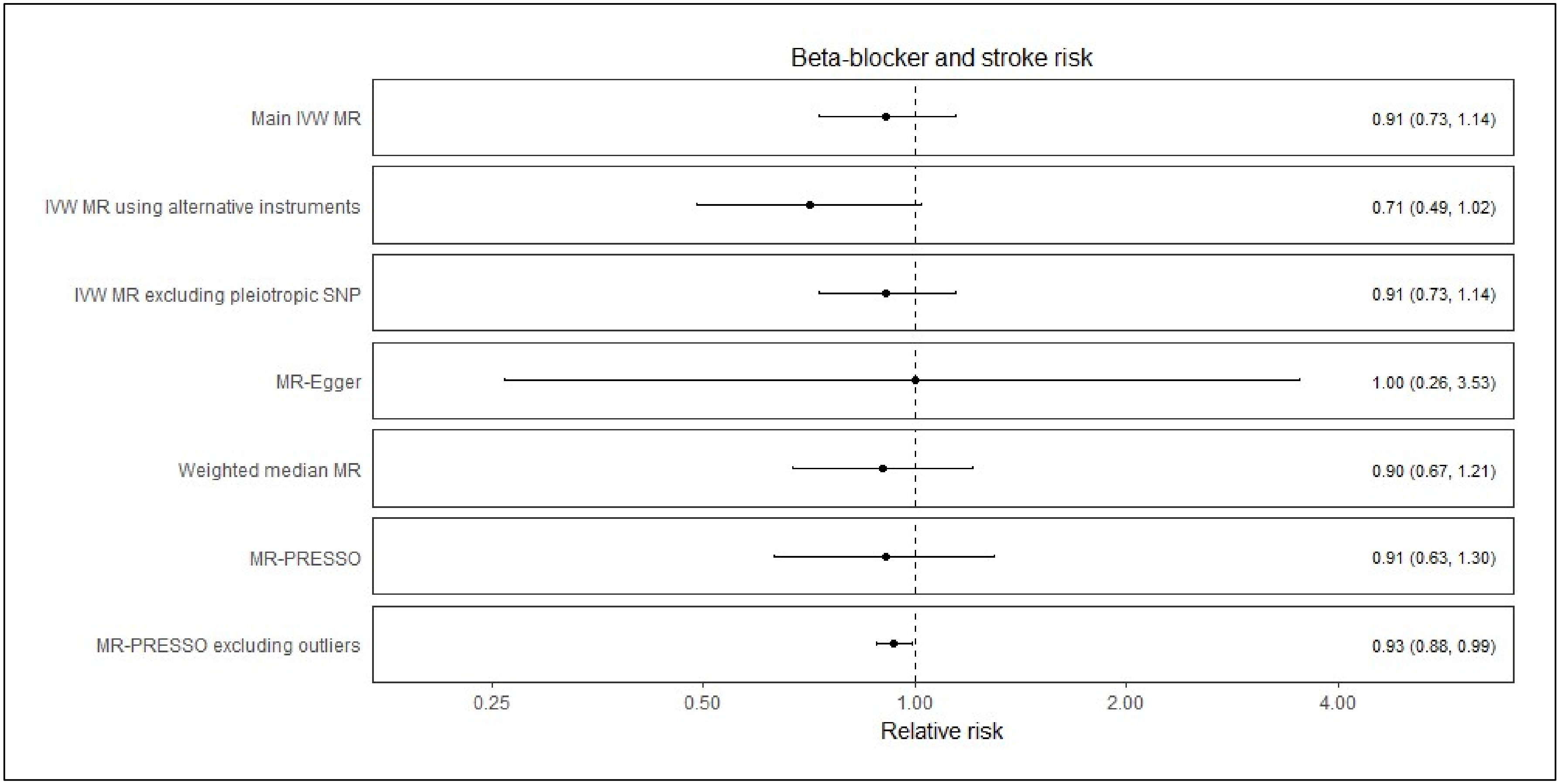


Figure 4.8. Mendelian randomization (MR) sensitivity analyses for the investigation of beta-blockers and stroke risk. Units are odds ratio units per change in systolic blood pressure observed in clinical trials of beta-blockers (9.51mmHg) (3). *IVW: inverse-variance weighted.*

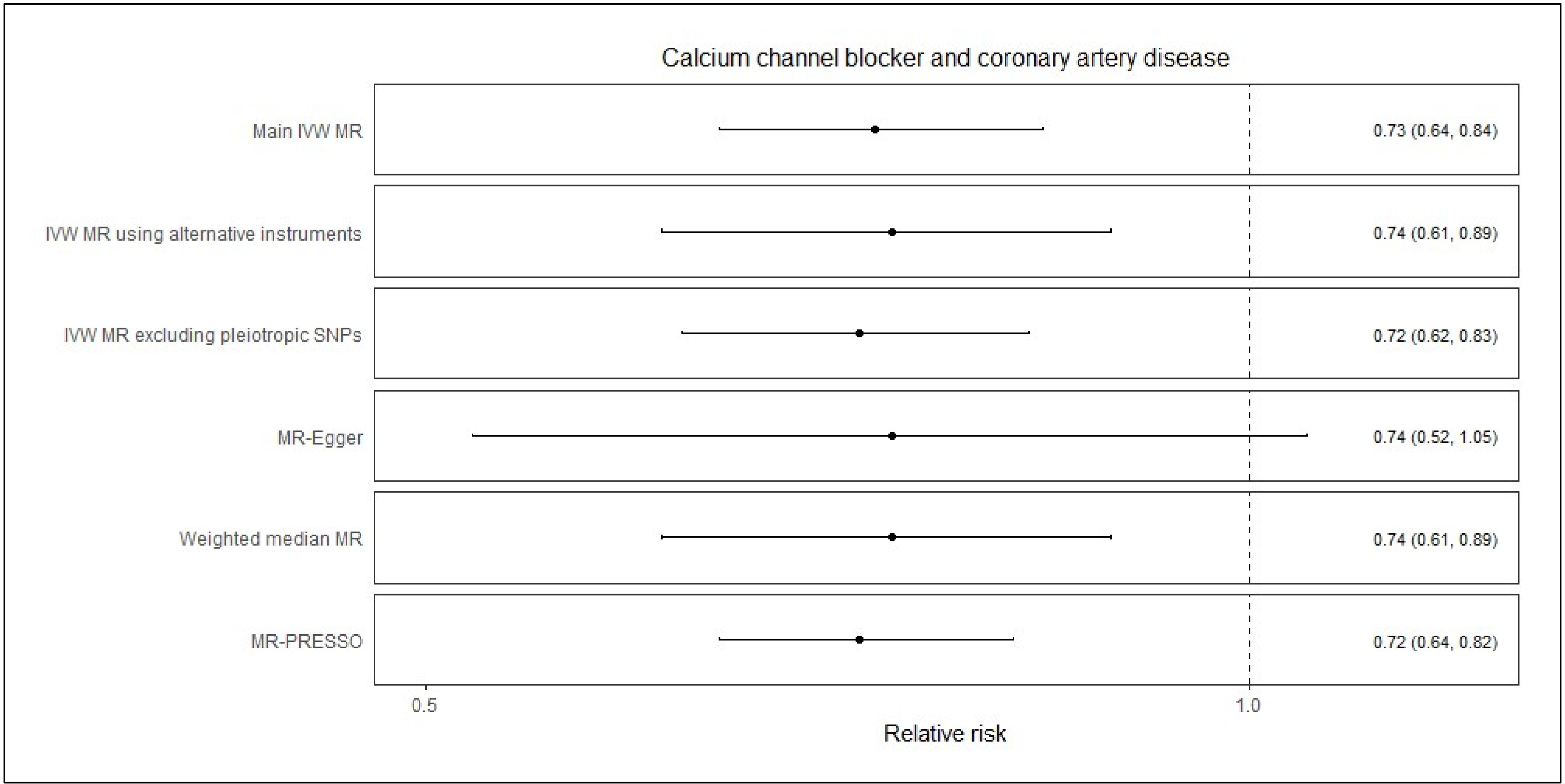


Figure 4.9. Mendelian randomization (MR) sensitivity analyses for the investigation of calcium channel blockers and coronary artery disease risk. Units are odds ratio units per change in systolic blood pressure observed in clinical trials of calcium channel blockers (8.9mmHg) (3). *IVW: inverse-variance weighted.*

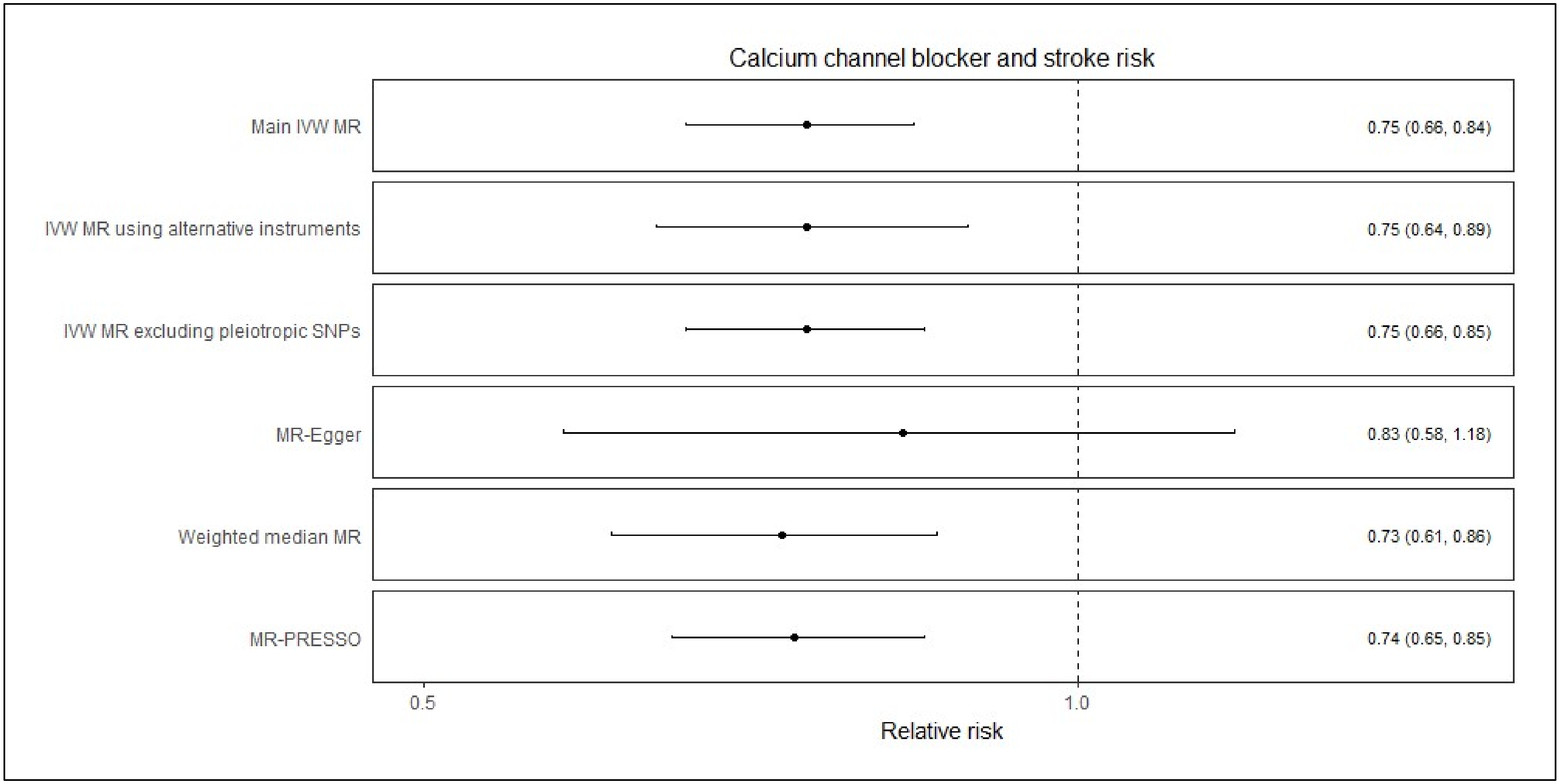


Figure 4.10. Mendelian randomization (MR) sensitivity analyses for the investigation of calcium channel blockers and stroke risk. Units are odds ratio units per change in systolic blood pressure observed in clinical trials of calcium channel blockers (8.9mmHg) (3). *IVW: inverse-variance weighted.*

Phenome-wide association study

After excluding non-European and related participants, 424,439 participants were included in the PheWAS analysis, with the ICD-9 and ICD-10 diagnoses mapped to 909 distinct phecodes. Table 4.5 provides details of the number of phenotypes and cases included in each disease category.

Table 4.5: Phenotypes and cases in each disease category for the phenome-wide association study in UK Biobank.

Category	Phenotypes (number)	Cases (number)			
		Min	Median	Mean	Max
Circulatory System	98	202	1048	6308	133749
Congenital Anomalies	19	211	442	557	1823
Dermatologic	43	218	799	4765	82669
Digestive	116	228	1455	4817	79488
Endocrine/Metabolic	49	208	773	4076	45303
Genitourinary	106	203	1376	4153	103829
Hematopoietic	22	201	569	2690	12759
Infectious Diseases	25	219	1012	2237	10752
Injuries & Poisonings	59	222	536	1513	16683
Mental Disorders	36	202	710	3280	29405
Musculoskeletal	57	213	925	4164	53823
Neoplasms	82	215	1124	4261	90826
Neurological	44	204	567	2286	40703
Pregnancy Complications	17	208	1113	1854	9534
Respiratory	56	200	1124	3837	62168
Sense Organs	64	210	774	2443	39998
Symptoms	16	304	2341	7036	42311

Performing PheWAS for the ACEI, BB and CCB standardized GRSs respectively, the results highlighted associations with hypertension and related diseases (Figures 4.11-4.13). The CCB analysis further showed an association with increased risk of diverticulosis (OR per SD increase in standardized GRS 1.02, 95% CI 1.01-1.04, P=2x10^{-4}). Similar results were obtained in PheWAS sensitivity analyses (OR 1.02, 95% CI 1.01-1.03 when using SBP genetic association estimates that were not corrected for antihypertensive medication use or adjusted for BMI; OR

1.02, 95% CI 1.01-1.04 when using a GRS that excluded potentially pleiotropic SNPs based on their identified secondary associations in PhenoScanner).

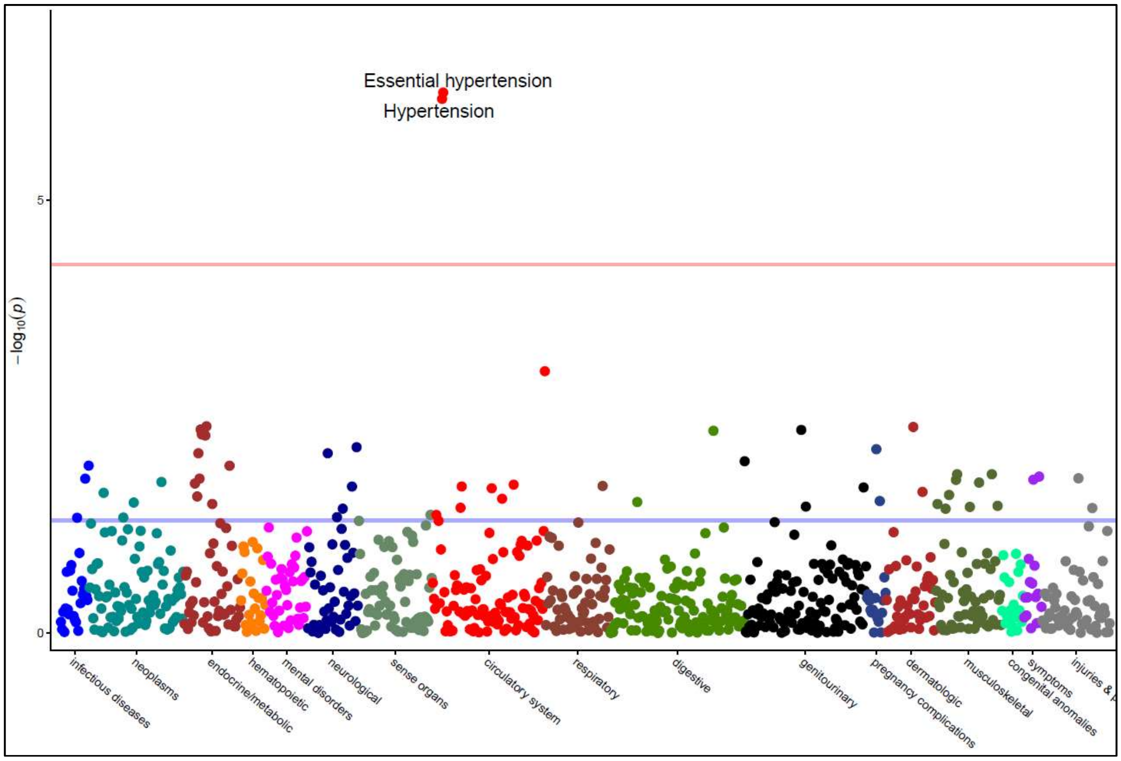


Figure 4.11. Phenome-wide association study plot for the angiotensin-converting enzyme inhibitor genetic risk score.

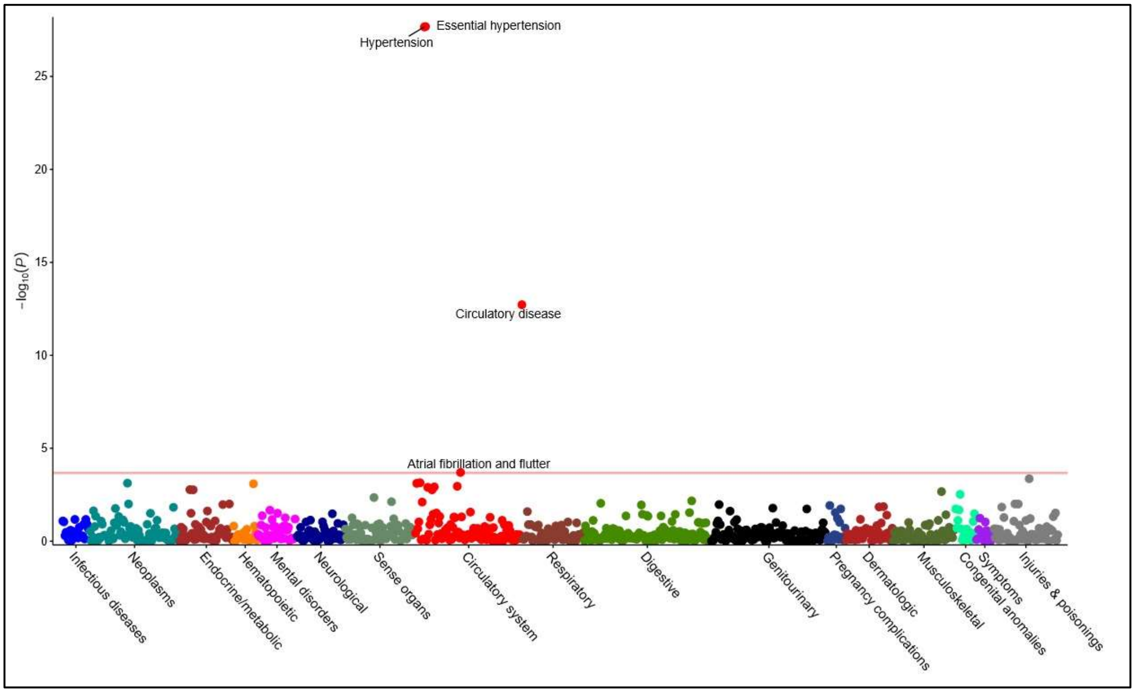


Figure 4.12. Phenome-wide association study plot for the beta-blocker genetic risk score.

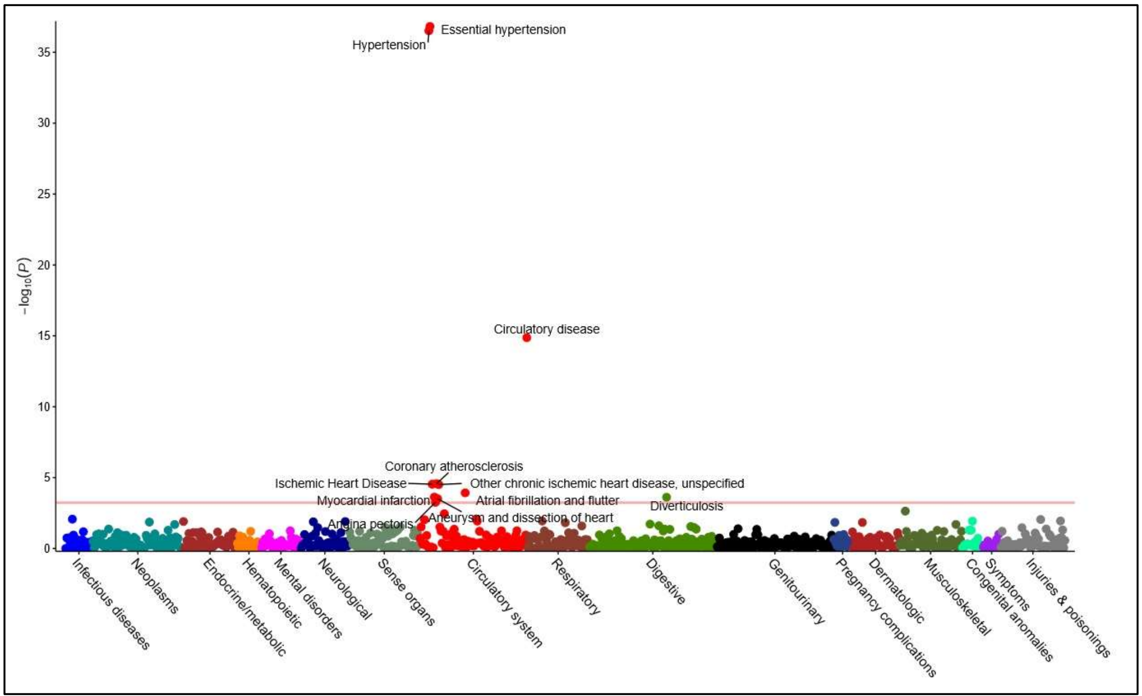


Figure 4.13. Phenome-wide association study plot for the calcium channel blocker genetic risk score.

Creating a standardized GRS by randomly sampling 24 SBP instrument SNPs from throughout the genome and estimating associations with diverticulosis risk over 1,000 iterations produced effect estimates close to the null (mean OR per SD increase in standardized GRS 1.00, 95% CI 0.98-1.02, *P*=0.79; Figure 4.14). Only 10 of the 1,000 analyses (1%) had a consistent direction of effect and a *P*-value lower than found when investigating the association of the standardized CCB GRS with diverticulosis, to thus produce adjusted *P*-value=0.01.

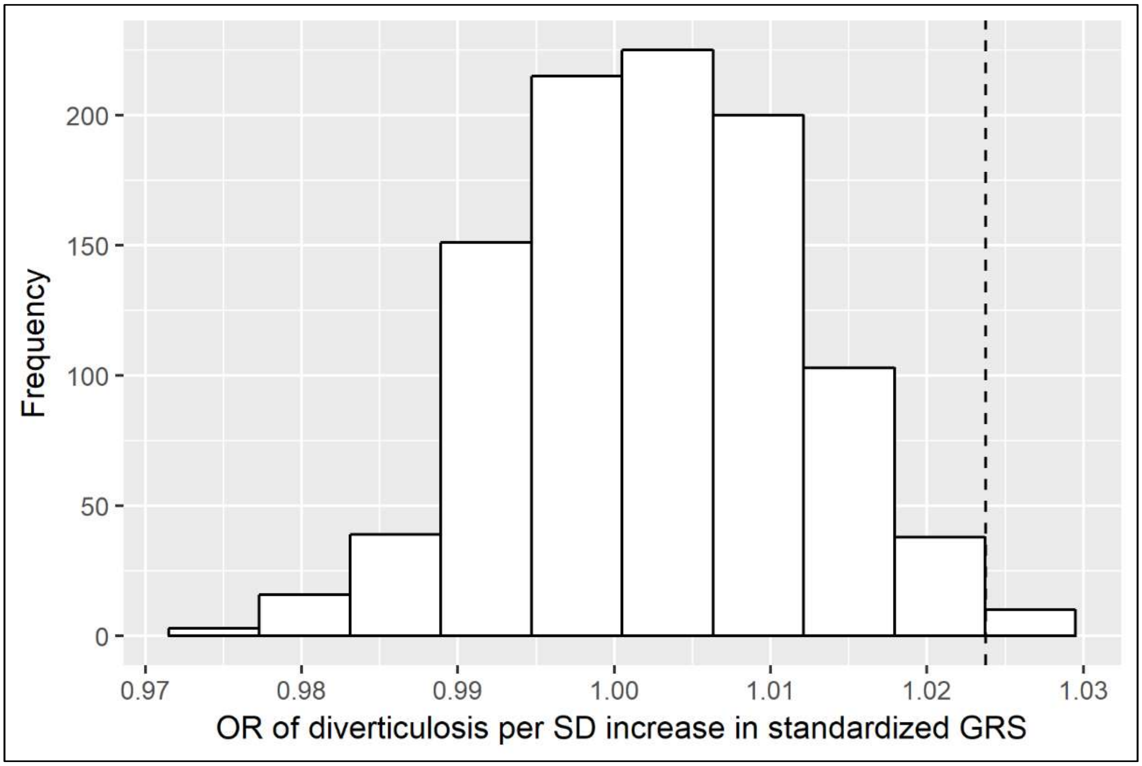


Figure 4.14. Associations with diverticulosis risk of the standardized genetic risk score (GRS) produced after randomly sampling 24 systolic blood pressure instrument variants from throughout the genome 1,000 times. The dashed line is result obtained when using the standardized calcium channel blocker GRS. OR: odds ratio; SD: standard deviation.

Linked genetic and Electronic Health Record data for 45,517 participants were available in BioVU, with cohort characteristics for the considered populations from the UK Biobank and BioVU provided in Table 4.6. Diverticulosis prevalence was 10% in the UK Biobank, and 12% in BioVU. The CCB standardized GRS association with diverticulosis had an OR per SD increase 1.01 (95% CI 1.00-1.02, *P*=0.17), with the meta-analysis of UK Biobank and BioVU estimates having OR 1.02 (95% CI 1.01-1.03, *P*=3x10^{-4}).

Observational analysis of drug use

From recruitment (13 March 2006 to 1 October 2010) to the UK Biobank, there were 1,408 incident diverticulosis cases up to 13 February 2016 amongst the 54,612 participants taking any of the considered antihypertensive drug classes at baseline, providing a mean follow-up period of 2,538 days. Adjusted Cox regression considering TD antihypertensive medication use alone as the reference category did not provide evidence for an association between use of any CCBs and risk of diverticulosis (hazard ratio [HR] 1.10, 95% CI 0.88-1.35, *P*=0.43). However, when investigating CCB subclasses, there was evidence supporting an association between risk of diverticulosis with non-dihydropyridine CCB use (HR 1.49, 95% CI 1.03-2.14, *P*=0.03), but not dihydropyridine CCB use (HR 1.01, 95% CI 0.80-1.28, *P*=0.91) or any other drug class (Table 4.7).

Table 4.6. Characteristics of the UK Biobank and BioVU populations. Mean (standard deviation)/number (%) estimates are provided. BMI: Body Mass Index, SBP: systolic blood pressure, DBP: diastolic blood pressure.

Cohort	Total number	Age, years	Sex, female	BMI	SBP, mmHg	Current smoker
UK Biobank	424,439	57 (8)	229,239 (54%)	27 (5)	138 (19)	43,928 (10%)
BioVU	45,517	61 (21)	25,148 (62%)	29 (8)	125 (13)	13,701 (30%)

Table 4.7. Cox regression results for the association between antihypertensive medication use and incident diverticulosis. Thiazide diuretics are the reference category (N=5501, diverticulosis cases=138). The calcium channel blocker (CCB) category analysis was performed in a separate model to the CCB subclass analysis. ACEI: angiotensin-converting enzyme inhibitor; ARB: angiotensin receptor blocker; BB: beta-blocker; HR: hazard ratio.

Antihypertensive drug class	Number	Diverticulosis cases	HR	L95% CI	U95% CI	*P*
ACEI	7210	162	1.00	0.79	1.26	0.99
ARB	4021	111	1.11	0.86	1.43	0.42
BB	6908	191	1.14	0.91	1.42	0.24
CCB	6756	180	1.09	0.88	1.35	0.43
Dihydropyridine CCB	5961	142	1.01	0.80	1.28	0.91
Non-dihydropyridine CCB	782	37	1.49	1.03	2.14	0.03
Two drug classes	18488	465	0.95	0.78	1.15	0.61
Greater than two drug classes	5741	162	1.02	0.81	1.28	0.90

4.4 Discussion

Summary of findings and clinical relevance

Using GWAS summary data from over 750,000 individuals, this work has identified genetic variants that serve as instrumental variables for the effect of the ACEI, BB and CCB classes of antihypertensive drug, which represent three of the most commonly prescribed medications in the world. MR estimates for risk of CAD and stroke generated using these respective instruments were comparable RCT meta-analyses (against placebo), to support the validity of this approach for studying the effect of these drugs. PheWAS performed to explore potential side-effects and repurposing opportunities for these medications across over 900 clinically relevant outcomes further supported the known efficacy of these drugs in preventing cardiovascular outcomes related to hypertension, further corroborating the validity of the genetic instruments for studying the effects of these antihypertensive drug classes. PheWAS in the UK Biobank also provided evidence to support an increased risk of diverticulosis for the CCB standardized GRS, with similar associations identified when considering the BioVU cohort. However, a similar association was not observed when considering GRSs for SBP using genetic variants from throughout the genome, suggesting that the association with diverticulosis is unlikely to be attributable to effects of lower SBP generally.

The CCB association with diverticulosis was also observed when investigating antihypertensive drug use and new diverticulosis diagnoses in the UK Biobank, although it was only use of the non-dihydropyridine CCB subclass that was associated with increased risk of consequent diverticulosis, with no such association found for dihydropyridine CCBs. Consistent with this, dihydropyridine and non-dihydropyridine CCB subclasses have different pharmacological effects (39). For example, constipation is a known side-effect of non-dihydropyridine CCBs that arises due to their effect on reducing bowel contractility (40), and it could be through a similar mechanism that diverticulosis risk is increased. Another potential mechanism might be through drug effects on the vasa recta vessels penetrating the colon wall, in turn leading to weaknesses where diverticulae form (41). Diverticulosis can lead to a number of complications that require admission to hospital (42), and is rising in incidence (43). Over 10% of the world's adult population are estimated to have hypertension and non-dihydropyridine CCBs in particular are recommended for individuals that have concurrent atrial fibrillation (1, 6). These findings could therefore have clinical implications, and for example it may be that individuals suffering with, or at increased risk of developing, diverticulosis are better suited to pharmacological treatments for hypertension other than non-dihydropyridine CCBs.

It is also worth noting that the PheWAS analyses using GRSs for ACEIs, BBs and CCBs did not identify detrimental associations with any of the other traits considered in PheWAS. Although this does not necessarily serve as evidence to support that these drugs do not have side-effects related to any of these considered outcomes, it does offer some reassurance towards the general safety of long-term use of these drugs, with serious adverse effects either being relatively rare or not being so serious as to require hospital admission.

Strengths and limitations

A major advantage of the approach taken in this work is that it uses existing genetic summary data to rapidly and efficiently investigate the efficacy, side-effects and repurposing potential of commonly prescribed antihypertensive drug classes. The initial MR analyses were able to help support the validity of the identified genetic instruments for the various antihypertensive drug classes, with PheWAS allowing hypothesis-free exploration of over 900 disease outcomes. In particular, this strategy does not suffer the time and resource constraints that often limit investigation in the form of RCT (4), while also overcoming the potential confounding and reverse causation biases that make it difficult to infer causality from conventional observational research methods (7). A range of sensitivity analyses were incorporated to assess the robustness of findings in the context of the various assumptions made by the genetic methods employed, with further study of drug use and consequent diverticulosis diagnoses in the UK Biobank cohort serving to triangulate the findings using methods that make distinct assumptions.

This work also has limitations. Both the MR and PheWAS approaches estimate the cumulative effect of lifetime exposure to genetic variants, which is not the same as a discrete clinical intervention. The genetic variants incorporated as instruments may also have unknown pleiotropic effects that affect the outcomes under study through pathways independent of SBP to bias the consequent MR and PheWAS estimates (31). Use of a less stringent criteria for instrument selection (e.g. a more relaxed *P*-value cut off for the SBP association, or a more relaxed LD threshold for clumping) may have identified more variants for use as genetic instruments, but could potentially also have had adverse implications for the sensitivity and specificity of the analyses because of greater susceptibility to incorporate weak or invalid instruments, respectively. Information relating to gene expression was not used to identify genetic instruments in this work. While such a strategy has been used previously (7), such data are restricted to the particular cells and tissues where gene expression is measured, and so may not necessarily be extrapolated to explore systemic drug effects more generally. Finally, the observational analysis exploring antihypertensive drug use in the UK Biobank may be susceptible to some residual confounding or ascertainment bias, particularly as diverticulosis itself can often be asymptomatic, only being diagnosed in the context of complications, or incidentally during interactions with the healthcare service for other reasons.

Conclusions

To summarise, the analyses undertaken in this work have identified genetic variants to serve as instrumental variables for exploring the effects of the ACEI, BB and CCB types of antihypertensive drug. Both in MR and PheWAS instrumental variable analyses, the findings supported known associations of these drug classes with disease outcomes related to hypertension. The hypothesis-free PheWAS investigation of potential drug side-effects and repurposing opportunities additionally identified a previously unreported detrimental association of the CCB GRS with risk of diverticulosis, a finding that was supported for the non-dihydropyridine CCB drug class when investigating antihypertensive drug use and consequent diverticulosis diagnoses in the UK Biobank. Although this finding could be of clinical relevance, it requires further validation before it should change clinical practice. There was no other evidence produced to support potential side-effects or lack of long-term safety for the antihypertensive drug classes considered. The approach taken in this work highlights that the use of genetic variants can offer a complementary approach to existing RCT and observational research techniques for investigating the clinical effects of antihypertensive drugs.

4.5 References

1. Forouzanfar MH, Liu P, Roth GA, Ng M, Biryukov S, Marczak L, et al. Global burden of hypertension and systolic blood pressure of at least 110 to 115 mm Hg, 1990-2015. JAMA. 2017;317(2):165-82.

2. Ettehad D, Emdin CA, Kiran A, Anderson SG, Callender T, Emberson J, et al. Blood pressure lowering for prevention of cardiovascular disease and death: a systematic review and meta-analysis. Lancet. 2016;387(10022):957-67.

3. Wright JM, Musini VM, Gill R. First-line drugs for hypertension. Cochrane Database Syst Rev. 2018;4:CD001841.

4. Frieden TR. Evidence for Health Decision Making - Beyond Randomized, Controlled Trials. N Engl J Med. 2017;377(5):465-75.

5. Tsang R, Colley L, Lynd LD. Inadequate statistical power to detect clinically significant differences in adverse event rates in randomized controlled trials. J Clin Epidemiol. 2009;62(6):609-16.

6. Williams B, Mancia G, Spiering W, Agabiti Rosei E, Azizi M, Burnier M, et al. 2018 ESC/ESH Guidelines for the management of arterial hypertension: The Task Force for the management of arterial hypertension of the European Society of Cardiology and the European Society of Hypertension: The Task Force for the management of arterial hypertension of the European Society of Cardiology and the European Society of Hypertension. J Hypertens. 2018;36(10):1953-2041.

7. Walker VM, Davey Smith G, Davies NM, Martin RM. Mendelian randomization: a novel approach for the prediction of adverse drug events and drug repurposing opportunities. Int J Epidemiol. 2017;46(6):2078-89.

8. Ference BA, Kastelein JJP, Ginsberg HN, Chapman MJ, Nicholls SJ, Ray KK, et al. Association of Genetic Variants Related to CETP Inhibitors and Statins With Lipoprotein Levels and Cardiovascular Risk. JAMA. 2017;318(10):947-56.

9. Ference BA, Majeed F, Penumetcha R, Flack JM, Brook RD. Effect of naturally random allocation to lower low-density lipoprotein cholesterol on the risk of coronary heart disease mediated by polymorphisms in NPC1L1, HMGCR, or both: a 2 x 2 factorial Mendelian randomization study. J Am Coll Cardiol. 2015;65(15):1552-61.

10. Ference BA, Robinson JG, Brook RD, Catapano AL, Chapman MJ, Neff DR, et al. Variation in PCSK9 and HMGCR and Risk of Cardiovascular Disease and Diabetes. N Engl J Med. 2016;375(22):2144-53.

11. Sofat R, Hingorani AD, Smeeth L, Humphries SE, Talmud PJ, Cooper J, et al. Separating the mechanism-based and off-target actions of cholesteryl ester transfer protein inhibitors with CETP gene polymorphisms. Circulation. 2010;121(1):52-62.

12. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, et al. The support of human genetic evidence for approved drug indications. Nat Genet. 2015;47(8):856-60.

13. Finan C, Gaulton A, Kruger FA, Lumbers RT, Shah T, Engmann J, et al. The druggable genome and support for target identification and validation in drug development. Sci Transl Med. 2017;9(383).

14. Hingorani AD, Kuan V, Finan C, Kruger FA, Gaulton A, Chopade S, et al. Flipping the odds of drug development success through human genomics. bioRxiv. 2017:170142.

15. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res. 2006;34(Database issue):D668-72.

16. Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database (Oxford). 2017;10.1093/database/bax028.

17. Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, 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.

18. Tobin MD, Sheehan NA, Scurrah KJ, Burton PR. Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. Stat Med. 2005;24(19):2911-35.

19. Neale Lab. Accessed 2019 January 16. Rapid GWAS of thousands of phenotypes for 337,000 samples in the UK Biobank 2018. http://www.nealelab.is/blog/2017/7/19/rapid-gwas-of-thousands-of-phenotypes-for-337000-samples-in-the-uk-biobank.

20. Palmer TM, Lawlor DA, Harbord RM, Sheehan NA, Tobias JH, Timpson NJ, et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. Stat Methods Med Res. 2012;21(3):223-42.

21. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet. 2015;47(10):1121-30.

22. Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. Nat Genet. 2018;50(4):524-37.

23. Thompson JR, Minelli C, Del Greco MF. Mendelian Randomization using Public Data from Genetic Consortia. Int J Biostat. 2016;12(2).

24. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013;37(7):658-65.

25. Gill D, Walker VM, Martin RM, Davies NM, Tzoulaki I. Comparison with randomized controlled trials as a strategy for evaluating instruments in Mendelian randomization. Int J Epidemiol. 2019.

26. Slob EAW, Burgess S. A Comparison Of Robust Mendelian Randomization Methods Using Summary Data. bioRxiv. 2019:577940.

27. Del Greco M F, Minelli C, Sheehan NA, Thompson JR. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. Stat Med. 2015;34(21):2926-40.

28. 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.

29. Verbanck M, Chen CY, 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.

30. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol. 2016;40(4):304-14.

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

32. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLoS Med. 2015;12(3):e1001779.

33. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559-75.

34. Li X, Meng XR, Spiliopoulou A, Timofeeva M, Wei WQ, Gifford A, et al. MR-PheWAS: exploring the causal effect of SUA level on multiple disease outcomes by using genetic instruments in UK Biobank. Ann Rheum Dis. 2018;77(7):1039-47.

35. Verma A, Bradford Y, Dudek S, Lucas AM, Verma SS, Pendergrass SA, et al. A simulation study investigating power estimates in phenome-wide association studies. BMC Bioinformatics. 2018;19(1):120.

36. Staley JR, Blackshaw J, Kamat MA, Ellis S, Surendran P, Sun BB, et al. PhenoScanner: a database of human genotype-phenotype associations. Bioinformatics. 2016;32(20):3207-9.

37. Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat Biotechnol. 2013;31(12):1102-10.

38. Gill D, Georgakis MK, Koskeridis F, Jiang L, Feng Q, Wei WQ, et al. Use of Genetic Variants Related to Antihypertensive Drugs to Inform on Efficacy and Side Effects. Circulation. 2019;140(4):270-9.

39. Frishman WH. Calcium channel blockers: Differences between subclasses. Am J Cardiovasc Drug. 2007;7:17-23.

40. Morris CR, Harvey IM, Stebbings WSL, Speakman CTM, Kennedy HJ, Hart AR. Do calcium channel blockers and antimuscarinics protect against perforated colonic diverticular disease? A case control study. Gut. 2003;52(12):1734-7.

41. Brian West A. The pathology of diverticulosis: classical concepts and mucosal changes in diverticula. J Clin Gastroenterol. 2006;40 Suppl 3:S126-31.

42. Everhart JE, Ruhl CE. Burden of Digestive Diseases in the United States Part II: Lower Gastrointestinal Diseases. Gastroenterology. 2009;136(3):741-54.

43. Etzioni DA, Mack TM, Beart RW, Jr., Kaiser AM. Diverticulitis in the United States: 1998-2005: changing patterns of disease and treatment. Ann Surg. 2009;249(2):210-7.

Chapter 5: Discussion, conclusions and future perspectives

All of the work presented in this chapter is my own, unless otherwise indicated in the text.

5.1 Introduction

The work in the preceding chapters has highlighted the broad applicability of the Mendelian randomization (MR) approach to investigate underlying mechanisms and therapeutic targets in cardiovascular disease (CVD). In this regard, Chapter 1 offered an introductory overview of CVD and MR, and also explored existing epidemiological evidence supporting diet, education and blood pressure as modifiable behavioural, social and metabolic risk factors respectively that have been shown to have important effects on CVD risk. The themes of systemic iron status, mediators of education, and antihypertensive drugs were consequently investigated in the proceeding chapters as demonstrative examples to apply the breadth of Mendelian randomization approaches available for exploring causal mechanisms in CVD.

The work in Chapter 2 first identified robust genetic instruments for systemic iron status as variants located at genes known to have biologically plausible roles in maintaining iron levels that also have associations with the four main biomarkers of iron status (serum iron, transferrin, transferrin saturation and ferritin) in a pattern consistent with their relation to systemic iron status. These were then employed in targeted MR analyses to investigate the effect of systemic iron levels on CVD subtypes, namely coronary artery disease (CAD), stroke and its subtypes, carotid plaque and intima media thickness, and venous thromboembolism. This provided evidence to support a potential contrasting effect of iron status on CVD subtypes – reducing risk of outcomes related to lipid-mediated atherosclerosis, while increasing risk of those related to stasis-mediated thrombosis. Furthermore, phenome-wide association study (PheWAS) analysis was undertaken to investigate the broad health implications of iron status, identifying the expected protective effect on risk of anaemia, but also novel protective effects on risk of hypercholesterolaemia, as well as detrimental effects on risk of skin and soft tissue infections. Given the variability in systemic iron levels and the potential to modify this through diet and pharmacological intervention, these findings are clinically relevant, and warrant further investigation.

Chapter 3 focused on the degree to which the traditional cardiovascular risk factors of systolic blood pressure (SBP), body mass index (BMI) and smoking mediate the effect of educational attainment on CVD risk, as well as the relation of any effects of educational attainment with cognitive function. The results support that approximately 20% of the effect of education on reducing CVD risk is mediated through each of the three considered mediator traits, and that due to their overlap, they together explain approximately 40% of the mediated effect. Furthermore, multivariable MR (MVMR) provided evidence to support that the effects of education on CVD risk are not related to cognitive function. In summary, these findings have important public health and policy implications, as they highlight that education, rather than cognition, should be the target of intervention to improve cardiovascular health, and further that where educational attainment cannot be directly modified, its downstream consequences can be optimised through modifying blood pressure, obesity and smoking. Potential limitations in the study of education and cognition include that they span across many domains and can therefore be difficult to measure and quantify. Furthermore, both of these traits can be difficult to modify, typically requiring simultaneous political, social and cultural reform for this.

Raised blood pressure is a well-established risk factor for CVD (1-3), with numerous pharmacological interventions currently available as treatment options (4, 5). However, investigation of the efficacy, side-effects and repurposing potential of such therapies within the framework of randomised, controlled trial (RCT) can be expensive and slow (6). Chapter 4 identified genetic instruments for common antihypertensive drugs as variants at the loci for the gene coding the corresponding target protein of the drug, which also relate to SBP. Diastolic blood pressure could similarly have been used as a trait by which to identify instruments, given its high phenotypic and genetic correlation with SBP (7). The validity of the identified instruments within the MR framework was explored by comparing MR estimates for their effects on CVD outcomes to those measured in RCT meta-analyses against placebo. Following successful validation of the instruments, PheWAS was performed to explore potential side-effects and repurposing opportunities, with the novel association of calcium channel blockers with diverticulosis later replicated for the non-dihydropyridine class in analysis of incident events for those taking the medication (versus other antihypertensive drug classes) in UK Biobank. This work thus highlights the potential of genetic approaches to rapidly and cost-effectively identify novel side-effects of commonly prescribed drugs, and pending further validation, could have important clinical implications.

The breadth of research themes and analyses presented so far in this thesis has allowed for thorough exploration of the current potential of MR. This current chapter summarises the described methodological strategies in more detail, including instrument selection, sensitivity analysis, distinguishing association from causation, scalability and reproducibility, mediation analysis and MVMR, paying particular attention to how they can be applied more widely in other contexts. The final sections consider future perspectives, including emerging techniques and applications.

5.2 Methodological strategies in Mendelian randomization

Instrument selection

Instrument selection is perhaps the most important stage of MR analysis. Although previous work has offered comprehensive practical suggestions for selecting instruments in MR analyses investigating the effects of disease biomarkers (8), the field has continued to evolve since (9). The increased availability of genetic data, which include epigenetic factors, gene, metabolite and protein expression, as well as phenotypic traits, is now coupled with improved efficiency in performing MR studies using automated software and online tools (9, 10). Additionally, MR analyses are no longer typically restricted to the investigation of biomarkers, and often explore dietary factors, social factors and drug targets, for example (10).

Instrument selection for MR should relate to the specific exposure being investigated in that given MR study. For practical purposes, exposures may be considered as "proximal" or "distal". Proximal exposures relate to a gene, such as for circulating proteins (e.g. C-reactive protein (11)) or protein drug targets (e.g. beta-blocker anti-hypertensive drugs (12)). In contrast, distal exposures typically relate to multiple gene effects that cannot be attributed to one mechanism or pathway (e.g. age at menarche (13), and time spent in education (14)).

It is because proximal exposures correspond to protein-coding genes that greater confidence is afforded in the validity of genetic instruments located at the gene locus for the protein of interest (i.e. *cis*-acting variants) (8). Of note though, is that there is no consensus on how such variants should relate to the corresponding gene locus. For example, studies considering the same exposure vary in the distance on either side of the gene that they include (15, 16), as well as whether enhancer or promotor regions for the gene are considered (12). A similar consideration is the permissible degree of correlation between variants through linkage disequilibrium (LD), while still assuming independent, additive effects. In practice, a range of LD thresholds are applied where adjustment is not made for this, such as $r^2<0.1$ (12), or $r^2<0.3$ (15, 16). Methods that can adjust for genetic correlation through LD are available (17, 18), and can be applied, such as in sensitivity analyses that aim to confirm the robustness of the main MR findings.

For distal exposures, where there is generally no corresponding gene locus, it is typically preferable to select instruments from throughout the genome. A low LD threshold (e.g. $r^2<0.001$) can therefore be used minimise correlation between variants. For some distal exposures, it is still possible to identify underlying genes with biological relevance, thus allowing for MR analyses that use instruments restricted to these loci. One such example of this is consideration of variants at genes for neuronal pathways when studying the effect of appetite-mediated obesity on smoking behaviour (19). Inclusion of a large number of genetic variants will likely introduce some pleiotropy, with potential violation of the requisite MR assumptions and related bias (20). However, if the instruments together explain a relatively large proportion of the exposure variance, this bias could be relatively small, to still allow for robust conclusions.

The *P*-value for association with the exposure used to select instruments varies in MR studies. For example, some analyses have included all variants that relate to the exposure at $P<0.05$ (15), while others incorporate a more conservative threshold of $P<5x10^{-8}$ (12). There are advantages and disadvantages to both relaxed and conservative criteria. More relaxed thresholds can allow for incorporation of a larger number of variants that cumulatively explain a greater proportion of the exposure variance to increase the statistical power of the consequent MR analysis (assuming that all the included variants are valid instruments) (21, 22). On the contrary, more relaxed selection criteria increase the potential to include invalid instruments that increase type I and type II error in MR analysis. Furthermore, more lenient significance thresholds for selecting instruments can also increase weak instrument bias, distorting MR estimates towards the confounded observational association in one sample-MR, and towards the null hypothesis in two-sample MR where there is no population overlap (18, 23). An additional concern is that bias in MR related to pleiotropic effects of the variants will be relatively greater when their association with the exposure is weaker.

Associations with secondary traits can also be used to support (or refute) the validity of instruments. As a practical example, when investigating the effect of systemic iron status using MR, genetic association estimates for four biomarkers of systemic iron status are available, including serum iron, transferrin, transferrin saturation and ferritin (24). Instruments for systemic iron status would be expected to relate to all four of these biomarkers – others that only relate to one biomarker may be more reflective of iron distribution for example (25-28). Secondary traits that can be used to explore the validity of instruments do not have to be phenotypic, and may include measures of gene, metabolite or protein expression, for example. In this way, instruments for antihypertensive drugs can be identified by their relation to expression of genes corresponding to their protein targets, as well as their relation to SBP (29).

Randomised, controlled trials (RCTs) represent a robust study design for estimating the clinical effects of an exposure on an outcome. Although they can often be time-consuming and costly, they continue to be considered as the gold-standard for informing clinical practice (6). Where RCT data are available on exposure-outcome effects, these can be compared against corresponding MR estimates to either support or refute the validity of selected genetic instruments (12). Any discrepancy between RCT and MR findings may serve as evidence that the chosen instruments do not appropriately proxy the clinical exposure under study.

The contrasting methods that can used to select instruments for the same exposure are well-exampled by two recent studies that aimed to instrument antihypertensive drug effects (12, 30). One study identified genetic instruments for antihypertensive classes as variants located at the corresponding target protein's gene, promotor or enhancer region that also associated with SBP at P<5x10^{-8} (12). For the same exposures, a separate study by Walker et al. selected instruments as variants at the corresponding protein target's gene locus that had the strongest association with expression of that gene in any tissue within the Genotype-Tissue Expression data (30, 31). These genetic variants were then only included in the final analyses if they also had an MR effect on SBP in a two-sample setting (30). While the separate approaches used in these two studies did produce different instruments, consistent MR estimates were obtained when considering the same exposure and outcome, suggesting that distinct selection strategies may be successfully employed to identify valid instruments (30).

Sensitivity analysis

A range of MR sensitivity analyses can be incorporated to investigate the robustness of findings to possible violations of the modelling assumptions. Such approaches might vary the criteria used to select instruments that are described above, including the *P*-value threshold for association to the exposure, LD r^2 criteria, and (for *cis*-acting variants) the specific genomic region. Consistency across MR estimates produced with variations in these selection criteria would support that the MR findings are more robust. Additionally, both biological and statistical sensitivity analysis strategies can be incorporated.

Biological sensitivity analyses consider the existing information on the variants. For example, this could consist of a review of available associations, and where there is suggestion that particular variants are exerting pleiotropic effects that could be introducing bias in MR, these can be excluded to explore that consistent MR estimates are obtained in their absence. For such exploration, numerous online databases now provide information on known biological effects and associations of genetic variants, with examples including the PhenoScanner and MR-Base curated databases of genetic association data (32, 33).

When numerous variants are used as instruments, excess heterogeneity (greater than would be expected by chance) across the MR estimates produced can indicate pleiotropy (34). In such a situation, various statistical methods are also available that make distinct assumptions on the nature of any underlying pleiotropy (35-37). While such statistical methods have been discussed in the preceding chapters, it is again worth noting here that where genetic association estimates are available for a known pleiotropic pathway relating the variants to the outcome independently of the exposure, MVMR can be applied to adjust for this genetic confounding (38). As an example, this method has been applied to measure the independent effect of different lipid traits on coronary artery disease risk (38). Discrepancy in MR estimates produced when applying different statistical sensitivity methods would suggest some violation of the requisite assumptions such as through the presence of pleiotropy, and should encourage greater caution and consideration in interpreting the results.

Where there are only a few (such as less than ten) variants available to serve as instruments, statistical methods for investigating bias related to pleiotropy are less feasible (37). This may be more commonly encountered with proximal exposures because *cis*-acting variants are identified from relatively small genomic regions where relatively few independent variants might be available. In addition, when variants are in close proximity, pleiotropic associations are more likely to be shared, in turn having implications for the ability of MR statistical sensitivity analyses to detect or adjust for this.

Distinguishing causation from association

Violations of the modelling assumptions of MR may in some cases also limit its ability to distinguish causation from association. In 1965, Austin Bradford Hill described a series of considerations for distinguishing causation from association in the context of epidemiological associations (39). These eventually resulted in the formation of 'Hill's criteria', which are discussed below in the context of how they can also be related to the MR approach. They are divided into two sections; those which relate to the MR modelling assumptions themselves, and those that apply to the context of the MR analysis within the larger framework of academic pursuit (39, 40).

Strength. Hill originally described that stronger associations were more likely to be suggestive of causation, and a similar phenomenon may also be applicable in the context of MR. This is particularly relevant as a major source of bias in MR analyses relates to pleiotropic effects of the genetic variants directly on the outcome, independently of the exposure under consideration. Such pleiotropy is likely to have relatively greater influence on the interpretation of MR analyses when the exposure-outcome association estimate itself is weaker. Another consideration is that MR, like all other analysis methods, has limited statistical power, and will be less likely to produce evidence to support a causal effect when the strength of the effect is smaller. However, this would not affect the MR approach's ability to distinguish between association and causation, but rather its ability to identify a true causal effect (i.e. minimise type II error).

Consistency. Hill's consistency criterion required that the association be observed across different populations, times and places. Within the context of MR, this may involve performing analyses using both exposure and outcome genetic association estimates taken from different populations, at different times and places. One difficulty with implementing this may relate to the availability of sufficiently large cohorts for which such data are available. Considering this criterion from another perspective, consistency of MR estimates could also be explored across different genetic variants – heterogeneity in MR estimates greater than expected by chance would indicate the presence of bias, such as due to pleiotropic effects (34).

Specificity. Hill argued that specificity of the exposure with the outcome provided support for a causal effect, rather than mere association. An example of this might be the effect of iron status on risk of particular subtypes of ischaemic stroke, rather than all types of cardiovascular disease generally (28). Where genetic association estimates that distinguish between specific exposures and outcomes are available, then such MR analysis may also be feasible.

Temporal sequence. A key advantage of the MR approach is that it is often able to overcome the reverse causation bias that limits interpretation of causal effects in traditional epidemiological research (41). It is because the genetic variants used as instrumental variables for the exposure are randomly allocated during conception that MR is typically able to delineate the temporal sequence between exposure and outcome. However, in some scenarios such as where the exposure and outcome being considered are closely related, it may be that the genetic variants used to proxy the exposure better serve as instruments for the outcome (42). Under these conditions, the exposure-outcome direction of effect may be reversed. Fortunately, statistical techniques such as the Steiger test are also available to disentangle this (42), with bi-directional MR a further option.

Dose response. In MR, the genetic variants used as instrumental variables for studying the effect of varying the exposure typically relate to small changes in the exposure around the population average. A linear effect of the exposure on the outcome is consequently assumed (in most analyses), with extrapolation of the exposure-outcome relationship to a scale relevant for meaningful interpretation. Where multiple genetic variants are available to instrument the exposure, the MR estimates from these are typically also fitted assuming a linear effect of the exposure on the outcome. Thus, while a dose-response relationship is inherent to MR using multiple genetic variants, potential limitations of this may relate to the assumption of a linear effect, with extrapolation of this beyond the range of measured variation in the considered exposure. It should be appreciated that such assumptions may not be valid if a linear model does not apply throughout the exposure-outcome dose-response relationship. More recently, non-linear MR methods have also been applied, such as to explore the J-shaped relation between BMI and all-cause mortality (43).

Analysis in context. While the above five Hill's criteria relate to details of the MR analysis and modelling assumptions, the remaining four criteria more relate to the context of the scientific inquiry, namely other supportive evidence. Specifically, these include the presence of experimental evidence, biological plausibility, coherence with known details about the phenomenon under consideration, and analogous examples that demonstrate similar relationships.

To therefore draw holistic evidence to support exposure-outcome causal effects, MR analysis should be performed with the above considerations in mind, within the design of the MR analysis itself, as well in its interpretation in the wider context.

Scalability and reproducibility

The scalability of MR also allows for an array of traits to be simultaneously investigated. Where the effect of one exposure with multiple outcomes is the primary objective of the study, PheWAS approaches may be undertaken to identify associations of the genetic instruments to outcomes throughout the phenome (44), with MR similarly applied using the resultant genetic association estimates of the instrument variants with the considered outcomes (25). In a similar way, the effect of multiple health exposure on a single outcome may be investigated with MR (45). The traits considered with MR also extend beyond phenotypic measures, and can include epigenetic modification (46), gene expression (31), and serum metabolite (47), cytokine, growth factor (48), and protein concentrations (49). Using such data that represents the cross-section from basic genetic elements to distal physiological and behavioural traits, it is also possible to incorporate MR mediation analyses to quantitatively measure causal mediators at each level (50, 51). Such a strategy not only offers mechanistic insight towards underlying causal mechanisms, but can also offer interventional targets where modification of the primary exposure is not feasible (14). With the availability of genetic association estimates relating to cell and tissue specific traits, MR may now also be focused to study such specific causal mechanisms (31).

The availability of publicly accessible genetic association summary estimates has allowed for MR analyses using these data to be semi-automated on online platforms or linked software packages (33). In this way, this information can be used for obtaining genetic association estimates related to either the exposure or outcome of interest, potential mediators of this relationship, as well as possible genetic confounders (i.e. pleiotropic pathways) (33, 52). It is possible that in the future, such resources may allow for automated, machine-learning algorithms to make use of the vast data that is available to inform, refine and optimise the conduct of MR, such as through neural networks (33).

Given the tremendous and continued increase in popularity of MR analyses, there is corresponding variation in the methods used and quality of reporting (53, 54). To this end, efforts have been made to provide advice for researchers performing (8, 53, 55), reporting (56, 57), reading (58) and reviewing MR analyses (56), with the aim of maintaining scientific standards and consistency. Such standards are important for assessing the robustness of evidence, particularly given the known potential for reporting biases in the scientific literature (59), which has also been shown relevant to MR work (53, 54). The umbrella review methodological approach is used to systematically assess evidence on the association between a trait with multiple outcomes, across distinct methodological approaches, to produce high level conclusions on the strength of evidence for association (60), and is now applied to include MR analyses (61).

Mediation analysis

Mediation methods have been adapted to the MR framework (50, 51), and have been successfully incorporated in applied examples that estimate mediating effects consistently across traditional observational, one-sample MR and two-sample MR approaches (14). Along with the advantages of univariable MR over traditional observation research, an additional strength of MR mediation analysis is its relative robustness to measurement error in the mediator, which would otherwise result in underestimation of any mediating effect (14, 62). However, MR mediation analysis methods are still in their relative infancy and make distinct assumptions beyond those required for conventional univariable MR.

The two approaches typically used to perform MR mediation analysis are network MR and multivariable MR (MVMR). Network MR is a two-stage analysis method, where the exposure-mediator association estimate is estimated in the first step, and the mediator-outcome association estimate adjusted for the genetic effect of the instruments on the exposure is calculated in the second step using MVMR (Figure 5.1). Thus, network MR also requires MVMR analysis in its second stage. To calculate the effect of the exposure on the outcome occurring through the mediator, the MR estimates from the two stages are multiplied together. Standard errors can be calculated using a number of approaches, including bootstrapping or the propagation of error method. The proportion of effect occurring through the mediator is then calculated by dividing this estimate by the total effect of the exposure on the outcome that can be estimated using univariable MR approaches such as inverse variance weighted (IVW) MR (18), and again using bootstrapping or the propagation of error method to estimate standard errors.

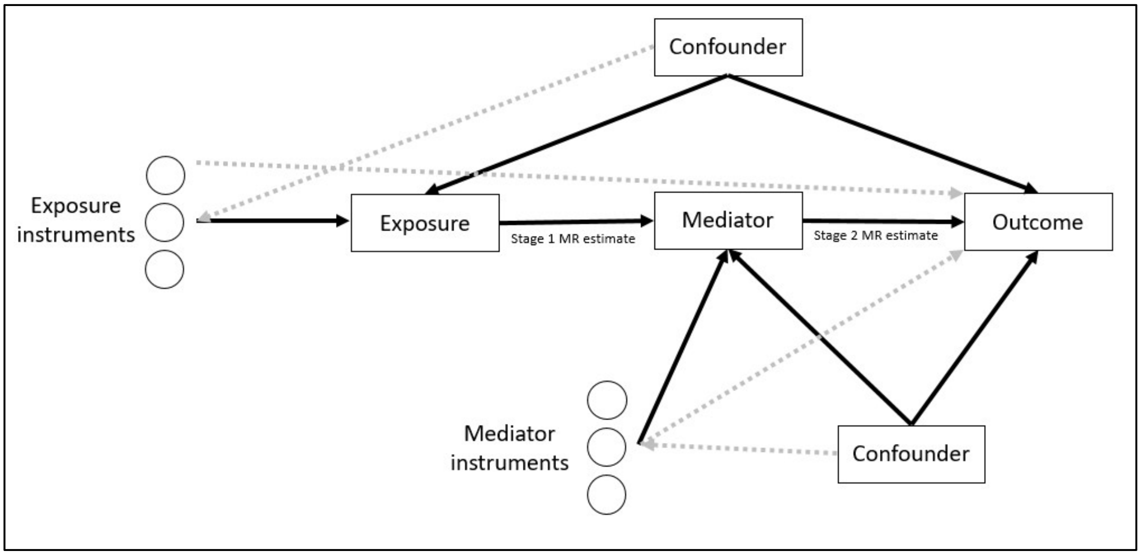


Figure 5.1. A schematic diagram depicting network Mendelian randomization. The solid black lines represent causal effects, while the grey dashed lines represent causal effects that would violate the requisite modelling assumptions. The exposure-mediator effects and the mediator-outcome effects (adjusted for the exposure) are estimated in two separate stages.

In MVMR mediation analysis, all the instruments specific for the exposure and mediator respectively are applied into the same model (Figure 5.2). Clumping may be required to ensure their LD independence (33), as such instruments for different traits are typically selected from distinct GWAS summary data and may therefore not necessarily be in low LD. A summary data regression based approach may then be used to regress the instrument-outcome genetic association estimates on the instrument-exposure and instrument-mediator genetic association estimates, weighted for the precision of the variant-outcome genetic association estimates, and with the intercept fixed at zero (51, 63). This gives the direct effect of the exposure on the outcome that is not mediated via the mediator. By subtracting this from the total effect of the exposure on the outcome estimated using univariable MR approaches (18), the effect of the exposure on the outcome arising through the mediator can be estimated. As for network MR, this may be divided by the total effect of the exposure on the outcome (obtained such as with IVW MR) to estimate the proportion of the exposure effect on the outcome arising through the mediator. Standard errors for each step may again be estimated by bootstrapping or the propagation of error method.

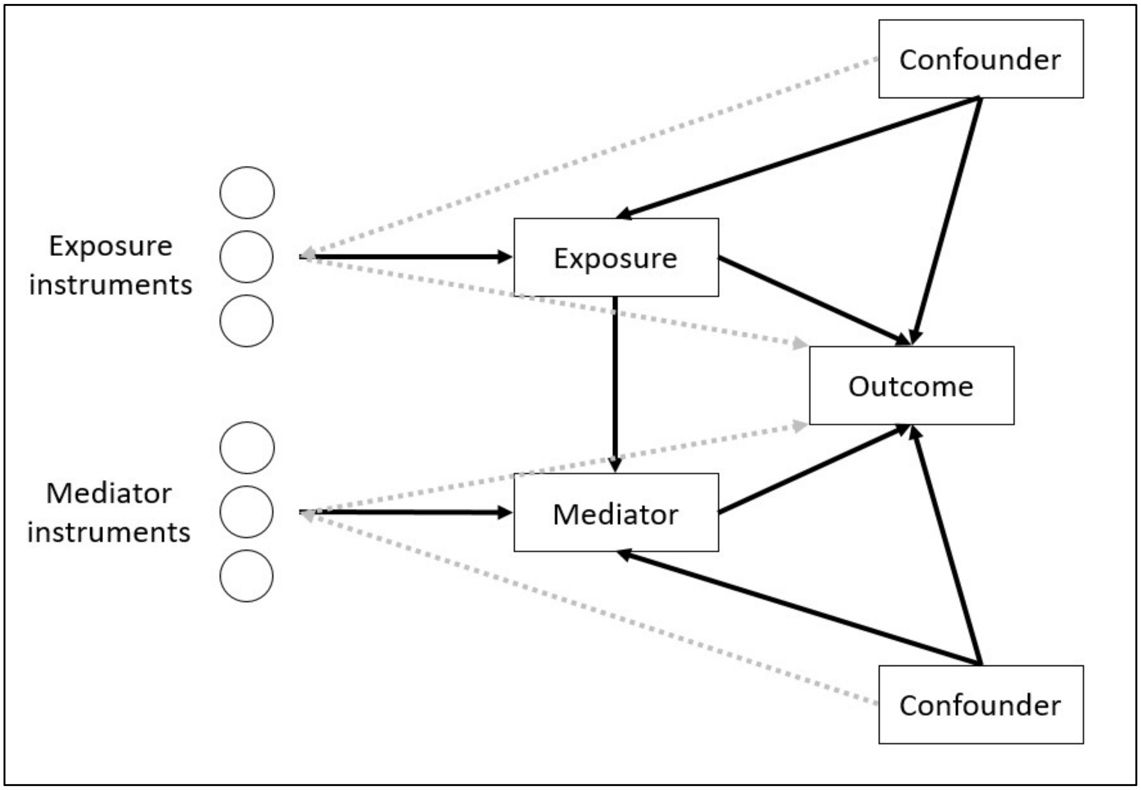


Figure 5.2. A schematic diagram depicting multivariable Mendelian randomization. The solid black lines represent causal effects, while the grey dashed lines represent causal effects that would violate the requisite modelling assumptions. For mediation analysis, the instrument variants should be specific to the exposure and mediator respectively.

Multivariable Mendelian randomization

As described above, both the network MR and MVMR approaches for mediation analysis incorporate MVMR analysis. Even outside mediation analysis, MVMR can be used to investigate the effect of an exposure on an outcome that is not attributable to genetic confounding or mediation through a known pathway for which genetic association estimates are also available (63, 64). Other general considerations are required for performing MVMR however, and are discussed here.

The measures used to estimate instrument strength in univariable MR do not directly translate to the MVMR setting, and it may be the case that instruments considered as strong in univariable MR are weak in the multivariable setting (65). At a practical level, if the instruments selected for the exposure and mediator respectively in MVMR mediation analyses are specific to these traits and do not overlap, such attenuation of instrument strength when moving from a univariable to multivariable MR setting should be minimal. In univariable MR analyses that obtain exposure and outcome genetic association estimates from distinct populations, any weak instrument bias is towards the null (18). For MVMR, this bias is more complicated and can vary depending on the context. Currently, consideration of instrument strength in MVMR when using summary data requires an estimate of the covariance of the error term for genetic association estimates of the instruments with the exposure and mediator (65), which may not necessarily be available in all studies. When this is not available, an estimate of the phenotypic correlation between the exposure and mediator may offer a suitable proxy, and future work is needed to allow this to be used to estimate instrument strength in MVMR that uses summary data from different populations.

Consideration of dichotomised exposure traits in univariable MR can produce severely biased estimates (66), with some bias also anticipated with a dichotomised outcome trait (and continuous exposure trait) (67). The same applies for MVMR, with dichotomised exposures and mediators likely to cause severe bias. For MVMR considering dichotomised outcomes, genetic associations are usually provided as log odds ratios, with such estimates typically sensitive to the choice of covariates used in the model. Thus, any such differences between the studies used to obtain genetic association estimates for the exposure, mediator and outcome in MVMR can also introduce bias into mediation estimates (51). At a practical level however, such bias is likely to be slight, and would typically still allow for interpretable estimates of the direct effect of the exposure on the outcome in MVMR analysis (51).

Pleiotropy of the genetic variants incorporated as instruments remains a source of bias in MVMR, as it does for univariable MR. Care should therefore be taken to select instruments that are specific for their association with the trait that they are instrumenting, and not some other potentially pleiotropic pathways. Publicly available databases of known genetic association estimates may be used to facilitate this process (32). While statistical sensitivity analysis methods for MVMR such as MVMR-Egger have been proposed (68), their implications within the context of mediation analysis have not yet been studied.

5.3 Future perspectives

The final sections of this chapter consider emerging methods and future perspectives, specifically focusing on investigation of non-linear effects and disease progression within the MR framework, colocalization and non-MR instrumental variable approaches. This is followed by concluding remarks on the current scope and future potential of MR in medical sciences.

Non-linear effects

The majority of published MR analyses assume a linear effect of the exposure on the outcome under study. In practice, this assumption may not hold, and methods are available for investigating non-linear effects with MR (69). The approach typically taken with such investigations is to stratify the population under study based on what their observed level of the exposure would be if they were not carrying any instrument variants (i.e. their residual exposure). It is important that the population is stratified on instrument-free exposure levels, as conditioning on the exposure directly would introduce collider bias (69). An MR estimate is consequently calculated for each stratum using conventional MR approaches such as the ratio method, and non-linearity can be investigated using statistical methods that compare MR estimates across different strata. Such methods include meta-regression with the fractional polynomial method, or estimation of a continuous piecewise linear function (70). Non-linear MR has previously been used to demonstrate that the J-shaped association between BMI and mortality has a causal aetiology in smokers (43).

Disease progression

MR has been applied to study disease progression as well as recovery after an event such as stroke (71). Incident event bias is an important consideration for such analyses, as any study of disease progression would inevitably stratify on disease occurrence. This has the potential to introduce collider bias into MR analyses of exposures that affect risk of disease onset (72). To deal with this, there is an available statistical approach that can adjust for any such incident event bias on the genetic association estimates for disease progression, which may in turn allow for unbiased MR analyses (73). In a similar way to the "instrument strength independent of direct effect" (InSIDE) assumption of MR-Egger however, this approach requires as a requisite that any direct effect of the variants under consideration on disease progression is not correlated to their effect on disease incidence (73). Currently, alternative methods that investigate differences between MR estimates produced by variants that superficially relate to disease incidence, disease progression or both are being developed to address such incident event bias without this described assumption.

Colocalization

MR approaches for studying causal effects often rely on the availability of instruments to serve as proxies for the exposure under consideration (29). The criteria used to select genetic instruments can vary, in turn affecting the consequent MR estimates generated (12, 30). Where appropriate instruments for MR are available, any exclusion of variants, which may otherwise be providing useful information, due to high LD may inadvertently reduce statistical power. Furthermore, MR methods that account for the genetic correlation due to LD between variants can become biased when too lenient an LD threshold is used (17, 18). In contrast to MR, the colocalization approach investigates whether any genetic correlation between variants at two loci can be attributed to them sharing the same causal variants (74). This approach employs a Bayesian statistical method, with principal component analysis used to summarize the information provided by genetic variants correlated by LD.

However, such colocalization methods also have limitations. Similarly to the difficulties encountered in achieving consistency for instrument selection in MR, the described colocalization approach can be affected by the choice of priors used in the Bayesian analysis, as well as the number of principal components incorporated (74). Furthermore, this method cannot be applied to summary genetic association data, such as are typically publicly released by consortia after performing GWAS meta-analyses. Existing colocalization methods that can use GWAS summary data are centered on the assumption that there is only one underlying causal variant that explains the genetic correlation between the two traits at the locus of interest (75, 76). A final limitation of existing colocalization methods is that, in scenarios where one trait is believed to be exerting a causal effect on the other, they are unable to quantify the magnitude of any such effect in the same way that MR approaches can (20).

While MR and colocalization have their respective strengths and weaknesses in different settings, methods for both approaches continue to develop rapidly. Rather than as competitors, these two strategies should be used synergistically where appropriate, to improve the overall arsenal of methods that may be directed to the available data for addressing the research question under study.

Other instrumental variable approaches

The success of MR, as measured by its popularity, has also encouraged a more general growth of instrumental variable approaches in biomedical sciences. Examples of such methods include discontinuity regression (77), study of doctors' prescribing preferences (78), and use of offspring phenotypes as instruments for parental phenotypes (79).

Discontinuity regression is based on the principle that individuals placed immediately on either side of an arbitrary cut off threshold used for allocation of an exposure would be unlikely to differ significantly in their baseline risk, including demographic profile or exposure to confounders for the exposure-outcome effect under study. An example of this is where an SBP of 140mmHg is used as the threshold for treatment with antihypertensive medications. Particularly given the degree of measurement error in such readings, it would be unlikely for an individual with an SBP of 139mmHg to differ markedly from one with an SBP of 141mmHg, yet the former would not receive treatment, while the latter would. Thus, restricting such analyses to those with a SBP close to 140mmHg, the measured reading can be used as an instrument to study the effect of antihypertensive medications.

Use of prescribing preferences in instrumental variable analysis is based on such preference being unrelated to the exposure, outcome and confounders of the exposure-outcome association. An example of this would be variation in choice of antihypertensive medications between different general medical practices (78). In this example, the practice would be the instrument, the choice of antihypertensive medication would be the exposure, and a range of different outcome traits might be considered.

Where linked data for parents and their offspring are available, observed traits in the offspring can be used as instruments for the same traits in the parents (79). This approach is unlikely to be biased by reverse causation, as parental traits typically affect offspring traits, rather than the other way around. Furthermore, such instrumental variable analysis is relatively protected from confounding, as the environmental determinants of the exposure in the offspring will not be the same as those in the parents. Such an approach has recently been used to study the effect of BMI on multiple health outcomes, where offspring BMI is used as an instrument for parental BMI (79).

5.4 Conclusion

The MR approach has gained tremendous popularity over the last decade, and despite the various assumptions implicit to the model, it continues to represent an informative source of evidence towards optimising clinical practice. Rather than its application to obtain precise causal effect estimates, the technique has found its niche within the wider framework for assessing causality, and serves as a powerful and efficient means of undertaking both exploratory and hypothesis-driven analyses. However, despite the numerous applications and successes of MR, the approach is also entirely fallible (53), being dependent on the availability of suitable genetic proxies for the exposure under consideration, as well as appropriate interpretation. Given the widespread pleiotropy throughout the human genome (80), the limited ability of genetic variants to accurately reflect the effect of a discrete intervention (20), and the other biases that limit the reliability of MR analyses (9), such findings should always be interpreted in context (39). To this end, there remains no substitute for careful study design and triangulation of all available evidence when interpreting research (9, 81). With continued growth in the availability of genetic data and corresponding analytical methods, the future offers incredible promise for the application of MR and genetic epidemiology more generally. Accordingly, the ambitions of what can be achieved also grow. The limits of MR continue to be expanded, with it now becoming common place to be able to rapidly investigate, for example, causal effects, their mediating mechanisms, linearity of effects and population specificity, in a large variety of research settings.

5.5 References

1. Yusuf S, Joseph P, Rangarajan S, Islam S, Mente A, Hystad P, et al. Modifiable risk factors, cardiovascular disease, and mortality in 155 722 individuals from 21 high-income, middle-income, and low-income countries (PURE): a prospective cohort study. Lancet. 2019:S0140-6736(19)32008-2. [Epub ahead of print].

2. GBD. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018;392(10159):1923-94.

3. Forouzanfar MH, Liu P, Roth GA, Ng M, Biryukov S, Marczak L, et al. Global burden of hypertension and systolic blood pressure of at least 110 to 115 mm Hg, 1990-2015. JAMA. 2017;317(2):165-82.

4. Whelton PK, Carey RM, Aronow WS, Casey DE, Jr., Collins KJ, Dennison Himmelfarb C, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. J Am Coll Cardiol. 2018;71(19):2199-269.

5. Wright JM, Musini VM, Gill R. First-line drugs for hypertension. Cochrane Database Syst Rev. 2018;4:CD001841.

6. Bothwell LE, Greene JA, Podolsky SH, Jones DS. Assessing the Gold Standard--Lessons from the History of RCTs. N Engl J Med. 2016;374(22):2175-81.

7. Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, 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.

8. Swerdlow DI, Kuchenbaecker KB, Shah S, Sofat R, Holmes MV, White J, et al. Selecting instruments for Mendelian randomization in the wake of genome-wide association studies. Int J Epidemiol. 2016;45(5):1600-16.

9. Burgess S, Davey Smith G. How humans can contribute to Mendelian randomization analyses. Int J Epidemiol 2019:48(3):661-4.

10. Zheng J, Baird D, Borges MC, Bowden J, Hemani G, Haycock P, et al. Recent Developments in Mendelian Randomization Studies. Curr Epidemiol Rep. 2017;4(4):330-45.

11. Casas JP, Shah T, Cooper J, Hawe E, McMahon AD, Gaffney D, et al. Insight into the nature of the CRP-coronary event association using Mendelian randomization. Int J Epidemiol. 2006;35(4):922-31.

12. Gill D, Georgakis MK, Koskeridis F, Jiang L, Feng Q, Wei WQ, et al. Use of Genetic Variants Related to Antihypertensive Drugs to Inform on Efficacy and Side Effects. Circulation. 2019;140(4):270-9.

13. Gill D, Brewer CF, Del Greco MF, Sivakumaran P, Bowden J, Sheehan NA, et al. Age at menarche and adult body mass index: a Mendelian randomization study. Int J Obes (Lond). 2018;42(9):1574-81.

14. Carter AR, Gill D, Davies NM, Taylor AE, Tillmann T, Vaucher J, et al. Understanding the consequences of education inequality on cardiovascular disease: mendelian randomisation study. BMJ. 2019;365:l1855.

15. Ference BA, Ray KK, Catapano AL, Ference TB, Burgess S, Neff DR, et al. Mendelian Randomization Study of ACLY and Cardiovascular Disease. N Engl J Med. 2019;380(11):1033-42.

16. Ference BA, Majeed F, Penumetcha R, Flack JM, Brook RD. Effect of naturally random allocation to lower low-density lipoprotein cholesterol on the risk of coronary heart disease mediated by polymorphisms in NPC1L1, HMGCR, or both: a 2 x 2 factorial Mendelian randomization study. J Am Coll Cardiol. 2015;65(15):1552-61.

17. Burgess S, Zuber V, Valdes-Marquez E, Sun BB, Hopewell JC. Mendelian randomization with fine-mapped genetic data: Choosing from large numbers of correlated instrumental variables. Genet Epidemiol. 2017;41(8):714-25.

18. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013;37(7):658-65.

19. Carreras-Torres R, Johansson M, Haycock PC, Relton CL, Davey Smith G, Brennan P, et al. Role of obesity in smoking behaviour: Mendelian randomisation study in UK Biobank. BMJ. 2018;361:k1767.

20. Davey Smith G, Ebrahim S. Mendelian randomization: can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003;32(1):1-22.

21. 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.

22. Brion MJ, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. Int J Epidemiol. 2013;42(5):1497-501.

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

24. Benyamin B, Esko T, Ried JS, Radhakrishnan A, Vermeulen SH, Traglia M, et al. Novel loci affecting iron homeostasis and their effects in individuals at risk for hemochromatosis. Nat Commun. 2014;5:4926.

25. Gill D, Benyamin B, Moore LSP, Monori G, Zhou A, Koskeridis F, et al. Associations of genetically determined iron status across the phenome: A mendelian randomization study. PLoS Med. 2019;16(6):e1002833.

26. Gill D, Brewer CF, Monori G, Tregouet DA, Franceschini N, Giambartolomei C, et al. Effects of Genetically Determined Iron Status on Risk of Venous Thromboembolism and Carotid Atherosclerotic Disease: A Mendelian Randomization Study. J Am Heart Assoc. 2019;8(15):e012994.

27. Gill D, Del Greco M F, Walker AP, Srai SKS, Laffan MA, Minelli C. The effect of iron status on risk of coronary artery disease: a mendelian randomization study. Arterioscler Thromb Vasc Biol. 2017;37(9):1788-92.

28. Gill D, Monori G, Tzoulaki I, Dehghan A. Iron Status and Risk of Stroke: A Mendelian Randomization Study. Stroke. 2018;49(12):2815-21.

29. Walker VM, Davey Smith G, Davies NM, Martin RM. Mendelian randomization: a novel approach for the prediction of adverse drug events and drug repurposing opportunities. Int J Epidemiol. 2017;46(6):2078-89.

30. Walker VM, Kehoe PG, Martin RM, Davies NM. Repurposing antihypertensive drugs for the prevention of Alzheimer's disease: a Mendelian randomization study. Int J Epidemiol 2019.

31. GTEx Consortium, Laboratory Data Analysis Coordinating Center Analysis Working Group, Statistical Methods Analysis Working Group, Enhancing GTEx Groups, NIH Common Fund BCSSN, Biospecimen Collection Source Site RPCI, et al. Genetic effects on gene expression across human tissues. Nature. 2017;550(7675):204-13.

32. Staley JR, Blackshaw J, Kamat MA, Ellis S, Surendran P, Sun BB, et al. PhenoScanner: a database of human genotype-phenotype associations. Bioinformatics. 2016;32(20):3207-9.

33. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. eLife. 2018;7.

34. Del Greco M F, Minelli C, Sheehan NA, Thompson JR. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. Stat Med. 2015;34(21):2926-40.

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

36. Slob EAW, Burgess S. A Comparison Of Robust Mendelian Randomization Methods Using Summary Data. bioRxiv. 2019:577940.

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

38. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. Am J Epidemiol. 2015;181(4):251-60.

39. Hill AB. The Environment and Disease: Association or Causation? Proc R Soc Med. 1965;58:295-300.

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

41. Davey Smith G, Ebrahim S. What can mendelian randomisation tell us about modifiable behavioural and environmental exposures? BMJ. 2005;330(7499):1076-9.

42. Hemani G, Tilling K, Smith GD. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. PLOS Genet. 2017;13(11).

43. Sun YQ, Burgess S, Staley JR, Wood AM, Bell S, Kaptoge SK, et al. Body mass index and all cause mortality in HUNT and UK Biobank studies: linear and non-linear mendelian randomisation analyses. BMJ. 2019;364:l1042.

44. Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat Biotechnol. 2013;31(12):1102-10.

45. Larsson SC, Traylor M, Malik R, Dichgans M, Burgess S, Markus HS, et al. Modifiable pathways in Alzheimer's disease: Mendelian randomisation analysis. BMJ. 2017;359:j5375.

46. Gaunt TR, Shihab HA, Hemani G, Min JL, Woodward G, Lyttleton O, et al. Systematic identification of genetic influences on methylation across the human life course. Genome Biol. 2016;17:61.

47. Kettunen J, Demirkan A, Wurtz P, Draisma HH, Haller T, Rawal R, et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. Nat Commun. 2016;7:11122.

48. Ahola-Olli AV, Wurtz P, Havulinna AS, Aalto K, Pitkanen N, Lehtimaki T, et al. Genome-wide Association Study Identifies 27 Loci Influencing Concentrations of Circulating Cytokines and Growth Factors. Am J Hum Genet. 2017;100(1):40-50.

49. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. Nature. 2018;558(7708):73-9.

50. Burgess S, Daniel RM, Butterworth AS, Thompson SG, Consortium E-IA. Network Mendelian randomization: using genetic variants as instrumental variables to investigate mediation in causal pathways. Int J Epidemiol. 2015;44(2):484-95.

51. Burgess S, Thompson DJ, Rees JMB, Day FR, Perry JR, Ong KK. Dissecting Causal Pathways Using Mendelian Randomization with Summarized Genetic Data: Application to Age at Menarche and Risk of Breast Cancer. Genetics. 2017;207(2):481-7.

52. Burgess S, Foley CN, Allara E, Staley JR, Howson JM. A robust and efficient method for Mendelian randomization with hundreds of genetic variants: unravelling mechanisms linking HDL-cholesterol and coronary heart disease. bioRxiv. 2019:566851.

53. Haycock PC, Burgess S, Wade KH, Bowden J, Relton C, Davey Smith G. Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies. Am J Clin Nutr. 2016;103(4):965-78.

54. Davies NM, Smith GD, Windmeijer F, Martin RM. Issues in the Reporting and Conduct of Instrumental Variable Studies A Systematic Review. Epidemiology. 2013;24(3):363-9.

55. Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan N, Thompson J. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. Stat Med. 2017:36(11):1783-802.

56. Davey Smith G, Davies NM, Dimou N, Egger M, Gallo V, Golub R, et al. STROBE-MR: Guidelines for strengthening the reporting of Mendelian randomization studies. https://doi.org/10.7287/peerj.preprints.27857v1. PeerJ Preprints. 2019;7:e27857v1.

57. Swanson SA, Hernan MA. How to Report Instrumental Variable Analyses (Suggestions Welcome). Epidemiology. 2013;24(3):370-4.

58. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ. 2018;362:k601.

59. Ioannidis JPA. Why most published research findings are false. PLoS Med. 2005;2(8):696-701.

60. Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. Int J Evid Based Healthc. 2015;13(3):132-40.

61. Li X, Meng X, Timofeeva M, Tzoulaki I, Tsilidis KK, Ioannidis JP, et al. Serum uric acid levels and multiple health outcomes: umbrella review of evidence from observational studies, randomised controlled trials, and Mendelian randomisation studies. BMJ. 2017;357:j2376.

62. Blakely T, McKenzie S, Carter K. Misclassification of the mediator matters when estimating indirect effects. J Epidemiol Community Health. 2013;67(5):458-66.

63. 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.

64. Burgess S, Freitag DF, Khan H, Gorman DN, Thompson SG. Using multivariable Mendelian randomization to disentangle the causal effects of lipid fractions. PLoS One. 2014;9(10):e108891.

65. Sanderson E, Davey Smith G, Windmeijer F, Bowden J. An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. Int J Epidemiol 2018:48(3):713-27.

66. Burgess S, Labrecque JA. Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. Eur J Epidemiol. 2018;33(10):947-52.

67. Harbord RM, Didelez V, Palmer TM, Meng S, Sterne JA, Sheehan NA. Severity of bias of a simple estimator of the causal odds ratio in Mendelian randomization studies. Stat Med. 2013;32(7):1246-58.

68. Rees JMB, Wood AM, Burgess S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. Stat Med. 2017;36(29):4705-18.

69. Burgess S, Davies NM, Thompson SG, Consortium EP-I. Instrumental variable analysis with a nonlinear exposure-outcome relationship. Epidemiology. 2014;25(6):877-85.

70. Staley JR, Burgess S. Semiparametric methods for estimation of a nonlinear exposure-outcome relationship using instrumental variables with application to Mendelian randomization. Genet Epidemiol. 2017;41(4):341-52.

71. Gill D, James NE, Monori G, Lorentzen E, Fernandez-Cadenas I, Lemmens R, et al. Genetically Determined Risk of Depression and Functional Outcome After Ischemic Stroke. Stroke. 2019;50(8):2219-22.

72. Paternoster L, Tilling K, Davey Smith G. Genetic epidemiology and Mendelian randomization for informing disease therapeutics: Conceptual and methodological challenges. PLOS Genet. 2017;13(10):e1006944.

73. Dudbridge F, Allen RJ, Sheehan NA, Schmidt AF, Lee JC, Jenkins RG, et al. Adjustment for index event bias in genome-wide association studies of subsequent events. Nat Commun. 2019;10(1):1561.

74. Wallace C, Rotival M, Cooper JD, Rice CM, Yang JH, McNeill M, et al. Statistical colocalization of monocyte gene expression and genetic risk variants for type 1 diabetes. Hum Mol Genet. 2012;21(12):2815-24.

75. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. PLOS Genetics. 2014;10(5).

76. Foley CN, Staley JR, Breen PG, Sun BB, Kirk PDW, Burgess S, et al. A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. bioRxiv. 2019:592238.

77. Bor J, Moscoe E, Mutevedzi P, Newell ML, Barnighausen T. Regression discontinuity designs in epidemiology: causal inference without randomized trials. Epidemiology. 2014;25(5):729-37.

78. Walker VM, Davies NM, Jones T, Kehoe PG, Martin RM. Can commonly prescribed drugs be repurposed for the prevention or treatment of Alzheimer's and other neurodegenerative diseases? Protocol for an observational cohort study in the UK Clinical Practice Research Datalink. BMJ Open. 2016;6(12):e012044.

79. Wade KH, Carslake D, Tynelius P, Davey Smith G, Martin RM. Variation of all-cause and cause-specific mortality with body mass index in one million Swedish parent-son pairs: An instrumental variable analysis. PLoS Med. 2019;16(8):e1002868.

80. Watanabe K, Stringer S, Frei O, Umicevic Mirkov M, de Leeuw C, Polderman TJC, et al. A global overview of pleiotropy and genetic architecture in complex traits. Nat Genet. 2019.

81. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. Int J Epidemiol. 2016;45(6):1866-86.

Appendix

Appendix Table 1. Mendelian randomization studies investigating the effect of circulating factors on cardiovascular disease.

Title	First Author	PMID	Journal	Date
Mendelian randomization evaluation of causal effects of fibrinogen on incident coronary heart disease.	Ward-Caviness CK	31075152	PLoS One	11/05/2019
Genetic Determinants of Circulating Glycine Levels and Risk of Coronary Artery Disease.	Jia Q	31070104	J Am Heart Assoc	10/05/2019
Estimation of the Required Lipoprotein(a)-Lowering Therapeutic Effect Size for Reduction in Coronary Heart Disease Outcomes: A Mendelian Randomization Analysis.	Lamina C	31017618	JAMA Cardiol	25/04/2019
Effect of glutamate and aspartate on ischemic heart disease, blood pressure, and diabetes: a Mendelian	Zhao JV	30949673	Am J Clin Nutr	06/04/2019

randomization study.				
Effect of linoleic acid on ischemic heart disease and its risk factors: a Mendelian randomization study.	Zhao JV	30866921	BMC Med	15/03/2019
Causal associations of blood lipids with risk of ischemic stroke and intracerebral hemorrhage in Chinese adults.	Sun L	30858617	Nat Med	13/03/2019
Association of genetically predicted testosterone with thromboembolism , heart failure, and myocardial infarction: mendelian randomisation study in UK Biobank.	Luo S	30842065	BMJ	08/03/2019
Serum magnesium and calcium levels in relation to ischemic stroke: Mendelian randomization study.	Larsson SC	30804065	Neurology	26/02/2019
Relative effects of LDL-C on ischemic stroke and	Valdes-Marquez E	30787162	Neurology	23/02/2019

coronary disease: A Mendelian randomization study.				
Homocysteine and small vessel stroke: A mendelian randomization analysis.	Larsson SC	30785218	Ann Neurol	21/02/2019
Evaluation of GDF15 as a therapeutic target of cardiometabolic diseases in human: A Mendelian randomization study.	Cheung CL	30772304	EBioMedicine	18/02/2019
Cardioprotective Properties Of HDL: Structural And Functional Considerations.	Pappa E	30714519	Curr Med Chem	05/02/2019
LDL triglycerides, hepatic lipase activity, and coronary artery disease: An epidemiologic and Mendelian randomization study.	Silbernagel G	30685440	Atherosclerosis	28/01/2019
A genome-wide association study identifies new loci for factor VII and implicates factor	de Vries PS	30642921	Blood	16/01/2019

VII in ischemic stroke etiology.				
Genome-Wide Association Transethnic Meta-Analyses Identifies Novel Associations Regulating Coagulation Factor VIII and von Willebrand Factor Plasma Levels.	Sabater-Lleal M	30586737	Circulation	28/12/2018
Genetically Determined Levels of Circulating Cytokines and Risk of Stroke.	Georgakis MK	30586705	Circulation	28/12/2018
Genome-wide meta-analysis identifies 3 novel loci associated with stroke.	Malik R	30383316	Ann Neurol	02/11/2018
Plasma C-Reactive Protein and Abdominal Aortic Aneurysm: A Mendelian Randomization Analysis.	Qin XY	30381605	Chin Med J (Engl)	02/11/2018
Circulating Vitamin K Levels in Relation to Ischemic Stroke and Its Subtypes: A Mendelian	Larsson SC	30366361	Nutrients	28/10/2018

Randomization Study.				
Serum 25-Hydroxyvitamin D Concentrations and Ischemic Stroke and Its Subtypes.	Larsson SC	30355092	Stroke	26/10/2018
Genetically Determined FXI (Factor XI) Levels and Risk of Stroke.	Gill D	30355187	Stroke	26/10/2018
Genetic contributors to serum uric acid levels in Mexicans and their effect on premature coronary artery disease.	Macias-Kauffer LR	30305239	Int J Cardiol	12/10/2018
Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease.	Yao C	30111768	Nat Commun	17/08/2018
ADAMTS-13 activity and ischemic heart disease: a Mendelian randomization study.	Schooling CM	30099840	J Thromb Haemost	14/08/2018
Adiponectin and coronary artery disease risk: A bi-	Au Yeung SL	30041791	Int J Cardiol	26/07/2018

directional Mendelian randomization study.				
Blood CSF1 and CXCL12 as Causal Mediators of Coronary Artery Disease.	Sjaarda J	30012324	J Am Coll Cardiol	18/07/2018
Are serum concentrations of vitamin B-12 causally related to cardiometabolic risk factors and disease? A Mendelian randomization study.	Moen GH	29982347	Am J Clin Nutr	10/07/2018
Association of LPA Variants With Risk of Coronary Disease and the Implications for Lipoprotein(a)-Lowering Therapies: A Mendelian Randomization Analysis.	Burgess S	29926099	JAMA Cardiol	22/06/2018
Genomic atlas of the human plasma proteome.	Sun BB	29875488	Nature	08/06/2018
Coagulation Factors and the Risk of Ischemic Heart Disease: A Mendelian	Zhao JV	29874180	Circ Genom Precis Med	07/06/2018

Randomization Study.				
Genetic predictors of testosterone and their associations with cardiovascular disease and risk factors: A Mendelian randomization investigation.	Schooling CM	29804699	Int J Cardiol	29/05/2018
Serum magnesium levels and risk of coronary artery disease: Mendelian randomisation study.	Larsson SC	29769070	BMC Med	18/05/2018
Evaluation of the relationship between plasma lipids and abdominal aortic aneurysm: A Mendelian randomization study.	Weng LC	29649275	PLoS One	13/04/2018
Triglyceride-Rich Lipoprotein Cholesterol and Risk of Cardiovascular Events Among Patients Receiving Statin Therapy in the TNT Trial.	Vallejo-Vaz AJ	29618599	Circulation	06/04/2018
Relationship of Familial	Beheshti S	29593013	Circulation	30/03/2018

Hypercholesterolemia and High Low-Density Lipoprotein Cholesterol to Ischemic Stroke.				
ET (Endothelin)-1 and Ischemic Heart Disease: A Mendelian Randomization Study.	Schooling CM	29555672	Circ Genom Precis Med	21/03/2018
Role of Blood Lipids in the Development of Ischemic Stroke and its Subtypes: A Mendelian Randomization Study.	Hindy G	29535274	Stroke	15/03/2018
Mendelian randomization analysis to assess a causal effect of haptoglobin on macroangiopathy in Chinese type 2 diabetes patients.	Wang S	29338727	Cardiovasc Diabetol	18/01/2018
Genetic Architecture of the Cardiovascular Risk Proteome.	Benson MD	29258991	Circulation	21/12/2017
Identification of a novel proinsulin-associated SNP and demonstration that proinsulin is	Strawbridge RJ	29040868	Atherosclerosis	19/10/2017

unlikely to be a causal factor in subclinical vascular remodelling using Mendelian randomisation.				
Genetically Determined Plasma Lipid Levels and Risk of Diabetic Retinopathy: A Mendelian Randomization Study.	Sobrin L	28951389	Diabetes	28/09/2017
High Lipoprotein(a) and Low Risk of Major Bleeding in Brain and Airways in the General Population: a Mendelian Randomization Study.	Langsted A	28877919	Clin Chem	08/09/2017
Vascular Endothelial Growth Factor and Ischemic Heart Disease Risk: A Mendelian Randomization Study.	Au Yeung SL	28765276	J Am Heart Assoc	03/08/2017
Association of Genetic Variants Related to Serum Calcium Levels With Coronary	Larsson SC	28742912	JAMA	26/07/2017

Artery Disease and Myocardial Infarction.				
Association of Triglyceride-Related Genetic Variants With Mitral Annular Cal cification.	Afshar M	28619195	J Am Coll Cardiol	18/06/2017
Oxidized Phospholipids and Risk of Calcific Aortic Valve Disease: The Copenhagen General Population Study.	Kamstrup PR	28572160	Arterioscler Thromb Vasc Biol	03/06/2017
Causal Effect of Plasminogen Activator Inhibitor Type 1 on Coronary Heart Disease.	Song C	28550093	J Am Heart Assoc	28/05/2017
The biomarker and causal roles of homoarginine in the development of cardiometabolic diseases: an observational and Mendelian randomization analysis.	Seppala I	28442717	Sci Rep	27/04/2017
Apolipoprotein(a) isoform size, lipoprotein(a) concentration, and coronary artery disease: a	Saleheen D	28408323	Lancet Diabetes Endocrinol	15/04/2017

mendelian randomisation analysis.				
Galectin-3 binding protein, coronary artery disease and cardiovascular mortality: Insights from the LURIC study.	Gleissner CA	28390290	Atherosclerosis	09/04/2017
Bilirubin and Stroke Risk Using a Mendelian Randomization Design.	Lee SJ	28389615	Stroke	09/04/2017
The -174 G>C Interleukin-6 Gene Polymorphism is Associated with Angiographic Progression of Coronary Artery Disease over a 4-Year Period.	Toutouzas K	28212870	Hellenic J Cardiol	19/02/2017
A Mendelian randomization study of the effect of calcium on coronary artery disease, myocardial infarction and their risk factors.	Xu L	28195141	Sci Rep	15/02/2017
A Non-Targeted Liquid Chromatographic-Mass Spectrometric	Zhang XZ	28151921	Med Sci Monit	06/02/2017

Metabolomics Approach for Association with Coronary Artery Disease: An Identification of Biomarkers for Depiction of Underlying Biological Mechanisms.				
Mendelian randomization estimates of alanine aminotransferase with cardiovascular disease: Guangzhou Biobank Cohort study.	Xu L	28007909	Hum Mol Genet	23/12/2016
Liver Enzymes and Risk of Ischemic Heart Disease and Type 2 Diabetes Mellitus: A Mendelian Randomization Study.	Liu J	27996050	Sci Rep	21/12/2016
Effect of L-arginine, asymmetric dimethylarginine, and symmetric dimethylarginine on ischemic heart disease risk: A	Au Yeung SL	27914500	Am Heart J	05/12/2016

Mendelian randomization study.				
Homocysteine-reducing B vitamins and ischemic heart disease: a separate-sample Mendelian randomization analysis.	Zhao JV	27901035	Eur J Clin Nutr	03/12/2016
Plasma levels of the anti-coagulation protein C and the risk of ischaemic heart disease. A Mendelian randomisation study.	Schooling CM	27882376	Thromb Haemost	25/11/2016
Is hyperhomocysteinemia a causal factor for heart failure? The impact of the functional variants of MTHFR and PON1 on ischemic and non-ischemic etiology.	Strauss E	27863359	Int J Cardiol	20/11/2016
Plasma Levels of Fatty Acid-Binding Protein 4, Retinol-Binding Protein 4, High-Molecular-Weight	Liu G	27609367	Arterioscler Thromb Vasc Biol	28/10/2016

Adiponectin, and Cardiovascular Mortality Among Men With Type 2 Diabetes: A 22-Year Prospective Study.				
Loss-of-function variants influence the human serum metabolome.	Yu B	27602404	Sci Adv	08/09/2016
Cystatin C and Cardiovascular Disease: A Mendelian Randomization Study.	van der Laan SW	27561768	J Am Coll Cardiol	27/08/2016
Endogenous androgen exposures and ischemic heart disease, a separate sample Mendelian randomization study.	Zhao JV	27526363	Int J Cardiol	16/08/2016
Association of Lipid Fractions With Risks for Coronary Artery Disease and Diabetes.	White J	27487401	JAMA Cardiol	04/08/2016
Mendelian Randomization Studies Do Not Support a Role for Vitamin D in Coronary Artery Disease.	Manousaki D	27418593	Circ Cardiovasc Genet	16/07/2016

Investigating the Causal Relationship of C-Reactive Protein with 32 Complex Somatic and Psychiatric Outcomes: A Large-Scale Cross-Consortium Mendelian Randomization Study.	Prins BP	27327646	PLoS Med	22/06/2016
A Genetic Biomarker of Oxidative Stress, the Paraoxonase-1 Q192R Gene Variant, Associates with Cardiomyopathy in CKD: A Longitudinal Study.	Dounousi E	27313824	Oxid Med Cell Longev	18/06/2016
Role of Adiponectin in Coronary Heart Disease Risk: A Mendelian Randomization Study.	Borges MC	27252388	Circ Res	03/06/2016
A causal relationship between uric acid and diabetic macrovascular disease in Chinese type 2 diabetes patients: A	Yan D	27064641	Int J Cardiol	12/04/2016

Mendelian randomization analysis.				
Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA.	Kettunen J	27005778	Nat Commun	24/03/2016
Observationally and Genetically High YKL-40 and Risk of Venous Thromboembolism in the General Population: Cohort and Mendelian Randomization Studies.	Kjaergaard AD	26988593	Arterioscler Thromb Vasc Biol	19/03/2016
A polymorphism in a major antioxidant gene (Kelch-like ECH-associated protein 1) predicts incident cardiovascular events in chronic kidney disease patients: an exploratory study.	Testa A	26974313	J Hypertens	15/03/2016
Harnessing publicly available genetic data to prioritize lipid modifying	Tragante V	26946290	Hum Genet	08/03/2016

therapeutic targets for prevention of coronary heart disease based on dysglycemic risk.				
Genome-wide association and Mendelian randomization study of NT-proBNP in patients with acute coronary syndrome.	Johansson A	26908625	Hum Mol Genet	26/02/2016
Genetically predicted 17beta-estradiol and cardiovascular risk factors in women: a Mendelian randomization analysis using young women in Hong Kong and older women in the Guangzhou Biobank Cohort Study.	Au Yeung SL	26907540	Ann Epidemiol	26/02/2016
Causal Assessment of Serum Urate Levels in Cardiometabolic Diseases Through a Mendelian Randomization Study.	Keenan T	26821629	J Am Coll Cardiol	30/01/2016

Genetic loci on chromosome 5 are associated with circulating levels of interleukin-5 and eosinophil count in a European population with high risk for cardiovascular disease.	McLeod O	26821299	Cytokine	29/01/2016
Evidence of a causal relationship between high serum adiponectin levels and increased cardiovascular mortality rate in patients with type 2 diabetes.	Ortega Moreno L	26817832	Cardiovasc Diabetol	29/01/2016
Plasma urate concentration and risk of coronary heart disease: a Mendelian randomisation analysis.	White J	26781229	Lancet Diabetes Endocrinol	20/01/2016
Lipoprotein-associated phospholipase A2 is related to risk of subclinical atherosclerosis but is not supported by Mendelian	Ueshima H	26775119	Atherosclerosis	18/01/2016

randomization analysis in a general Japanese population.				
Elevated Lipoprotein(a) Levels, LPA Risk Genotypes, and Increased Risk of Heart Failure in the General Population.	Kamstrup PR	26656145	JACC Heart Fail	15/12/2015
A genetic marker of hyperuricemia predicts cardiovascular events in a meta-analysis of three cohort studies in high risk patients.	Testa A	26607700	Nutr Metab Cardiovasc Dis	27/11/2015
Fetuin-A and risk of coronary heart disease: A Mendelian randomization analysis and a pooled analysis of AHSG genetic variants in 7 prospective studies.	Laugsand LE	26343871	Atherosclerosis	08/09/2015
Sex-Specific Effects of Adiponectin on Carotid Intima-Media Thickness and Incident Cardiovascular Disease.	Persson J	26276317	J Am Heart Assoc	16/08/2015

Iron and hepcidin as risk factors in atherosclerosis: what do the genes say?	Galesloot TE	26159428	BMC Genet	15/07/2015
Cystatin C Is Not Causally Related to Coronary Artery Disease.	Svensson-Farbom P	26057752	PLoS One	10/06/2015
No evidence that genetically reduced 25-hydroxyvitamin D is associated with increased risk of ischaemic heart disease or myocardial infarction: a Mendelian randomization study.	Brondum-Jacobsen P	25981321	Int J Epidemiol	20/05/2015
Genetically high plasma vitamin C, intake of fruit and vegetables, and risk of ischemic heart disease and all-cause mortality: a Mendelian randomization study.	Kobylecki CJ	25948669	Am J Clin Nutr	08/05/2015
Elevated Lipoprotein(a) Does Not Cause Low-Grade Inflammation Despite Causal	Langsted A	25938632	J Clin Endocrinol Metab	06/05/2015

Association With Aortic Valve Stenosis and Myocardial Infarction: A Study of 100,578 Individuals from the General Population.				
Genetic analysis of leukocyte type-I interferon production and risk of coronary artery disease.	Nelson CP	25882064	Arterioscler Thromb Vasc Biol	18/04/2015
Uric Acid and Cardiovascular Events: A Mendelian Randomization Study.	Kleber ME	25788527	J Am Soc Nephrol	20/03/2015
Circulating brain-derived neurotrophic factor concentrations and the risk of cardiovascular disease in the community.	Kaess BM	25762803	J Am Heart Assoc	13/03/2015
Cardiometabolic effects of genetic upregulation of the interleukin 1 receptor antagonist: a Mendelian randomisation analysis.	Interleukin 1 Genetics Consortium.	25726324	Lancet Diabetes Endocrinol	03/03/2015

Elevated plasma YKL-40, lipids and lipoproteins, and ischemic vascular disease in the general population.	Kjaergaard AD	25624368	Stroke	28/01/2015
Lipid and lipoprotein measurements and the risk of ischemic vascular events: Framingham Study.	Pikula A	25568296	Neurology	09/01/2015
Large-scale metabolomic profiling identifies novel biomarkers for incident coronary heart disease.	Ganna A	25502724	PLoS Genet	17/12/2014
Inactive matrix Gla protein is causally related to adverse health outcomes: a Mendelian randomization study in a Flemish population.	Liu YP	25421980	Hypertension	26/11/2014
The causal effect of vitamin D binding protein (DBP) levels on calcemic and cardiometabolic diseases: a Mendelian	Leong A	25350643	PLoS Med	29/10/2014

randomization study.				
Association of low-density lipoprotein cholesterol-related genetic variants with aortic valve calcium and incident aortic stenosis.	Smith JG	25344734	JAMA	27/10/2014
Using multivariable Mendelian randomization to disentangle the causal effects of lipid fractions.	Burgess S	25302496	PLoS One	11/10/2014
A genetic marker of uric acid level, carotid atherosclerosis, and arterial stiffness: a family-based study.	Mallamaci F	25301104	Am J Kidney Dis	11/10/2014
Distribution and medical impact of loss-of-function variants in the Finnish founder population.	Lim ET	25078778	PLoS Genet	01/08/2014
Lipoprotein (a) concentrations, apolipoprotein (a) phenotypes, and peripheral arterial disease in three	Laschkolnig A	24760552	Cardiovasc Res	25/04/2014

independent cohorts.				
Association of myeloperoxidase with total and cardiovascular mortality in individuals undergoing coronary angiography--the LURIC study.	Scharnagl H	24746542	Int J Cardiol	22/04/2014
Plasma levels of vitamin K and the risk of ischemic heart disease: a Mendelian randomization study.	Schooling CM	27061505	J Thromb Haemost	12/04/2016
Lipoprotein(a) levels, genotype, and incident aortic valve stenosis: a prospective Mendelian randomization study and replication in a case-control cohort.	Arsenault BJ	24704946	Circ Cardiovasc Genet	08/04/2014
The association between plasminogen activator inhibitor type 1 (PAI-1) levels, PAI-1 4G/5G polymorphism,	Nikolopoulos GK	24695040	Clin Chem Lab Med	04/04/2014

and myocardial infarction: a Mendelian randomization meta-analysis.				
A serum 25-hydroxyvitamin D concentration-associated genetic variant in DHCR7 interacts with type 2 diabetes status to influence subclinical atherosclerosis (measured by carotid intima-media thickness).	Strawbridge RJ	24663808	Diabetologia	26/03/2014
Novel genetic approach to investigate the role of plasma secretory phospholipase A2 (sPLA2)-V isoenzyme in coronary heart disease: modified Mendelian randomization analysis using PLA2G5 expression levels.	Holmes MV	24563418	Circ Cardiovasc Genet	25/02/2014
Mendelian randomization of blood lipids for coronary heart disease.	Holmes MV	24474739	Eur Heart J	30/01/2014

Genetically predicted testosterone and cardiovascular risk factors in men: a Mendelian randomization analysis in the Guangzhou Biobank Cohort Study.	Zhao J	24302542	Int J Epidemiol	05/12/2013
Elevated C-reactive protein, depression, somatic diseases, and all-cause mortality: a mendelian randomization study.	Wium-Andersen MK	24246360	Biol Psychiatry	20/11/2013
Elevated lipoprotein(a) and risk of aortic valve stenosis in the general population.	Kamstrup PR	24161338	J Am Coll Cardiol	29/10/2013
Common quantitative trait locus downstream of RETN gene identified by genome-wide association study is associated with risk of type 2 diabetes mellitus in Han Chinese: a Mendelian	Chung CM	24123702	Diabetes Metab Res Rev	15/10/2013

randomization effect.				
Ceruloplasmin and atrial fibrillation: evidence of causality from a population-based Mendelian randomization study.	Adamsson Eryd S	24118451	J Intern Med	15/10/2013
Elevated remnant cholesterol causes both low-grade inflammation and ischemic heart disease, whereas elevated low-density lipoprotein cholesterol causes ischemic heart disease without inflammation.	Varbo A	23926208	Circulation	09/08/2013
Secretory phospholipase A(2)-IIA and cardiovascular disease: a mendelian randomization study.	Holmes MV	23916927	J Am Coll Cardiol	07/08/2013
Association between C677T polymorphism of methylene tetrahydrofolate reductase and congenital heart	Mamasoula C	23876493	Circ Cardiovasc Genet	24/07/2013

disease: meta-analysis of 7697 cases and 13,125 controls.				
Association of plasma uric acid with ischaemic heart disease and blood pressure: mendelian randomisation analysis of two large cohorts.	Palmer TM	23869090	BMJ	23/07/2013
The shared allelic architecture of adiponectin levels and coronary artery disease.	Dastani Z	23664276	Atherosclerosis	15/05/2013
Vitamin D status, filaggrin genotype, and cardiovascular risk factors: a Mendelian randomization approach.	Skaaby T	23460889	PLoS One	06/03/2013
The relevance of the association between inflammation and atrial fibrillation.	Alegret JM	23397981	Eur J Clin Invest	13/02/2013
Genetic associations with valvular calcification and aortic stenosis.	Thanassoulis G	23388002	N Engl J Med	08/02/2013
Causal relevance of blood lipid fractions in the	Shah S	23275344	Circ Cardiovasc Genet	01/01/2013

development of carotid atherosclerosis: Mendelian randomization analysis.				
Remnant cholesterol as a causal risk factor for ischemic heart disease.	Varbo A	23265341	J Am Coll Cardiol	26/12/2012
Mendelian randomization suggests non-causal associations of testosterone with cardiometabolic risk factors and mortality.	Haring R	23258625	Andrology	22/12/2012
Genetically elevated non-fasting triglycerides and calculated remnant cholesterol as causal risk factors for myocardial infarction.	Jorgensen AB	23248205	Eur Heart J	19/12/2012
Elevated fibrinogen levels are associated with risk of pulmonary embolism, but not with deep venous thrombosis.	Klovaite J	23220916	Am J Respir Crit Care Med	12/12/2012

Interleukin-6 receptor pathways in abdominal aortic aneurysm.	Harrison SC	23111417	Eur Heart J	01/11/2012
Effect of long-term exposure to lower low-density lipoprotein cholesterol beginning early in life on the risk of coronary heart disease: a Mendelian randomization analysis.	Ference BA	23083789	J Am Coll Cardiol	23/10/2012
Genetically elevated bilirubin and risk of ischaemic heart disease: three Mendelian randomization studies and a meta-analysis.	Stender S	22805420	J Intern Med	19/07/2012
Associations between serum uric acid and markers of subclinical atherosclerosis in young adults. The cardiovascular risk in Young Finns study.	Oikonen M	22749515	Atherosclerosis	04/07/2012
Nonfasting glucose, ischemic heart disease, and	Benn M	22698489	J Am Coll Cardiol	16/06/2012

myocardial infarction: a Mendelian randomization study.				
Genetically elevated levels of circulating triglycerides and brachial-ankle pulse wave velocity in a Chinese population.	Yao WM	22648266	J Hum Hypertens	01/06/2012
Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study.	Voight BF	22607825	Lancet	23/05/2012
Genetic evidence that lipoprotein(a) associates with atherosclerotic stenosis rather than venous thrombosis.	Kamstrup PR	22516069	Arterioscler Thromb Vasc Biol	21/04/2012
Prospective study of methylenetetrahydrofolate reductase (MTHFR) variant C677T and risk of all-cause and cardiovascular disease mortality	Yang Q	22492374	Am J Clin Nutr	12/04/2012

among 6000 US adults.				
The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis.	Interleukin-6 Receptor Mendelian Randomisation Analysis (IL6R MR) Consortium.	22421340	Lancet	17/03/2012
Association between bilirubin and cardiovascular disease risk factors: using Mendelian randomization to assess causal inference.	McArdle PF	22416852	BMC Cardiovasc Disord	16/03/2012
Homocysteine and coronary heart disease: meta-analysis of MTHFR case-control studies, avoiding publication bias.	Clarke R	22363213	PLoS Med	01/03/2012
LCAT, HDL cholesterol and ischemic cardiovascular disease: a Mendelian randomization study of HDL cholesterol in 54,500 individuals.	Haase CL	22090275	J Clin Endocrinol Metab	18/11/2011

Type II secretory phospholipase A2 and prognosis in patients with stable coronary heart disease: mendelian randomization study.	Breitling LP	21799821	PLoS One	30/07/2011
Association between C reactive protein and coronary heart disease: mendelian randomisation analysis based on individual participant data.	C Reactive Protein Coronary Heart Disease Genetics Collaboration (CCGC).	21325005	BMJ	18/02/2011
Conventional and Mendelian randomization analyses suggest no association between lipoprotein(a) and early atherosclerosis: the Young Finns Study.	Kivimaki M	21078622	Int J Epidemiol	17/11/2010
Does elevated C-reactive protein increase atrial fibrillation risk? A Mendelian randomization of 47,000 individuals from the general population.	Marott SC	20797493	J Am Coll Cardiol	28/08/2010

Triglyceride-mediated pathways and coronary disease: collaborative analysis of 101 studies.	Triglyceride Coronary Disease Genetics Consortium and Emerging Risk Factors Collaboration.	20452521	Lancet	11/05/2010
Association of AHSG gene polymorphisms with fetuin-A plasma levels and cardiovascular diseases in the EPIC-Potsdam study.	Fisher E	20031641	Circ Cardiovasc Genet	25/12/2009
Genetic Loci associated with C-reactive protein levels and risk of coronary heart disease.	Elliott P	19567438	JAMA	02/07/2009
Integrated associations of genotypes with multiple blood biomarkers linked to coronary heart disease risk.	Drenos F	19336475	Hum Mol Genet	02/04/2009
Does high C-reactive protein concentration increase atherosclerosis? The Whitehall II Study.	Kivimaki M	18714381	PLoS One	21/08/2008
Lifelong reduction of LDL-cholesterol related to a	Linsel-Nitschke P	18714375	PLoS One	21/08/2008

common variant in the LDL-receptor gene decreases the risk of coronary artery disease--a Mendelian Randomisation study.				
Mendelian randomization suggests no causal association between C-reactive protein and carotid intima-media thickness in the young Finns study.	Kivimaki M	17377152	Arterioscler Thromb Vasc Biol	23/03/2007
Fibrinogen and coronary heart disease: test of causality by 'Mendelian randomization'.	Keavney B	16870675	Int J Epidemiol	28/07/2006
Insight into the nature of the CRP-coronary event association using Mendelian randomization.	Casas JP	16565153	Int J Epidemiol	28/03/2006
Homocysteine and stroke: evidence on a causal link from mendelian randomisation.	Casas JP	15652605	Lancet	18/01/2005

Appendix Table 2. Mendelian randomization studies investigating the effect of physiological traits and diseases on cardiovascular disease.

Title	First Author	PMID	Journal	Date
Genetic overlap of chronic obstructive pulmonary disease and cardiovascular disease-related traits: a large-scale genome-wide cross-trait analysis.	Zhu Z	30940143	Respir Res	04/04/2019
Shared mechanisms between coronary heart disease and depression: findings from a large UK general population-based cohort.	Khandaker GM	30886334	Mol Psychiatry	20/03/2019
Differential Association of Genetic Risk of Coronary Artery Disease With Development of Heart Failure With Reduced Versus Preserved Ejection Fraction.	Mordi IR	30742529	Circulation	12/02/2019
Thyroid Function and Dysfunction in Relation to 16 Cardiovascular Diseases.	Larsson SC	30702347	Circ Genom Precis Med	01/02/2019
Mendelian Randomization Analysis of Hemoglobin A(1c) as a Risk Factor for Coronary Artery Disease.	Leong A	30659074	Diabetes Care	20/01/2019
Iron Status and Risk of Stroke.	Gill D	30571402	Stroke	21/12/2018
Clinical and Genetic Determinants of Varicose Veins.	Fukaya E	30566020	Circulation	20/12/2018
Assessing the causal role of body mass index on cardiovascular health in young adults: Mendelian randomization and recall-by-genotype analyses.	Wade KH	30524135	Circulation	14/12/2018
Adult height and risk of 50 diseases: a combined epidemiological and genetic analysis.	Lai FY	30355295	BMC Med	26/10/2018
Genetic Association of Albuminuria with Cardiometabolic Disease and Blood Pressure.	Haas ME	30220432	Am J Hum Genet	18/09/2018
A comprehensive evaluation of the genetic architecture of sudden cardiac arrest.	Ashar FN	30169657	Eur Heart J	01/09/2018

Mendelian randomisation analysis of clustered causal effects of body mass on cardiometabolic biomarkers.	Conde S	30066639	BMC Bioinformatics	02/08/2018
Exploring shared genetic bases and causal relationships of schizophrenia and bipolar disorder with 28 cardiovascular and metabolic traits.	So HC	30045777	Psychol Med	27/07/2018
The Impact of Glycated Hemoglobin (HbA(1c)) on Cardiovascular Disease Risk: A Mendelian Randomization Study Using UK Biobank.	Au Yeung SL	29950300	Diabetes Care	29/06/2018
Genetically driven adiposity traits increase the risk of coronary artery disease independent of blood pressure, dyslipidaemia, glycaemic traits.	Lv WQ	29891878	Eur J Hum Genet	13/06/2018
Birthweight, Type 2 Diabetes Mellitus, and Cardiovascular Disease: Addressing the Barker Hypothesis With Mendelian Randomization.	Zanetti D	29875125	Circ Genom Precis Med	08/06/2018
Assessing causal estimates of the association of obesity-related traits with coronary artery disease using a Mendelian randomization approach.	Zhang X	29739994	Sci Rep	10/05/2018
Causal Impact of Type 2 Diabetes Mellitus on Cerebral Small Vessel Disease: A Mendelian Randomization Analysis.	Liu J	29686024	Stroke	25/04/2018
Association of Genetic Instrumental Variables for Lung Function on Coronary Artery Disease Risk: A 2-Sample Mendelian Randomization Study.	Au Yeung SL	29650766	Circ Genom Precis Med	14/04/2018
Childhood BMI and Adult Type 2 Diabetes, Coronary Artery Diseases, Chronic Kidney Disease, and Cardiometabolic Traits: A Mendelian Randomization Analysis.	Geng T	29483184	Diabetes Care	28/02/2018
Liver fat content, non-alcoholic fatty liver disease, and ischaemic heart	Lauridsen BK	29228164	Eur Heart J	12/12/2017

disease: Mendelian randomization and meta-analysis of 279,013 individuals.				
Bone mineral density and risk of type 2 diabetes and coronary heart disease: A Mendelian randomization study.	Gan W	28989980	Wellcome Open Res	11/10/2017
Causal effects of cardiovascular risk factors on onset of major age-related diseases: A time-to-event Mendelian randomization study.	He L	28964830	Exp Gerontol	02/10/2017
The role of glycaemic and lipid risk factors in mediating the effect of BMI on coronary heart disease: a two-step, two-sample Mendelian randomisation study.	Xu L	28889241	Diabetologia	11/09/2017
Thyroid function and ischemic heart disease: a Mendelian randomization study.	Zhao JV	28819171	Sci Rep	19/08/2017
The Effect of Iron Status on Risk of Coronary Artery Disease: A Mendelian Randomization Study-Brief Report.	Gill D	28684612	Arterioscler Thromb Vasc Biol	08/07/2017
Association of Body Mass Index With Cardiometabolic Disease in the UK Biobank: A Mendelian Randomization Study.	Lyall DM	28678979	JAMA Cardiol	06/07/2017
Type 2 diabetes, glucose, insulin, BMI, and ischemic stroke subtypes: Mendelian randomization study.	Larsson SC	28667182	Neurology	02/07/2017
Age at menarche and cardiovascular risk factors using Mendelian randomization in the Guangzhou Biobank Cohort Study.	Au Yeung SL	28601624	Prev Med	12/06/2017
Assessing the causal relationship between obesity and venous thromboembolism through a Mendelian Randomization study.	Lindstrom S	28528403	Hum Genet	22/05/2017
Causal Associations of Adiposity and Body Fat Distribution With Coronary Heart Disease, Stroke Subtypes, and Type 2 Diabetes Mellitus: A Mendelian Randomization Analysis.	Dale CE	28500271	Circulation	14/05/2017

Taller height as a risk factor for venous thromboembolism: a Mendelian randomization meta-analysis.	Roetker NS	28445597	J Thromb Haemost	27/04/2017
Genetic variants associated with type 2 diabetes and adiposity and risk of intracranial and abdominal aortic aneurysms.	van 't Hof FN	28378816	Eur J Hum Genet	06/04/2017
Genetic Analysis of Venous Thromboembolism in UK Biobank Identifies the ZFPM2 Locus and Implicates Obesity as a Causal Risk Factor.	Klarin D	28373160	Circ Cardiovasc Genet	05/04/2017
Genetically Driven Hyperglycemia Increases Risk of Coronary Artery Disease Separately From Type 2 Diabetes.	Merino J	28298470	Diabetes Care	17/03/2017
Genetic Association of Waist-to-Hip Ratio With Cardiometabolic Traits, Type 2 Diabetes, and Coronary Heart Disease.	Emdin CA	28196256	JAMA	15/02/2017
Relationships of Measured and Genetically Determined Height With the Cardiac Conduction System in Healthy Adults.	Kofler T	28039282	Circ Arrhythm Electrophysiol	01/01/2017
Genetic Obesity and the Risk of Atrial Fibrillation: Causal Estimates from Mendelian Randomization.	Chatterjee NA	27974350	Circulation	16/12/2016
Birth weight and risk of ischemic heart disease: A Mendelian randomization study.	Au Yeung SL	27924921	Sci Rep	08/12/2016
Mendelian Randomisation study of the influence of eGFR on coronary heart disease.	Charoen P	27338949	Sci Rep	25/06/2016
Obesity and peripheral arterial disease: A Mendelian Randomization analysis.	Huang Y	26945778	Atherosclerosis	08/03/2016
Increased genetic risk for obesity in premature coronary artery disease.	Cole CB	26220701	Eur J Hum Genet	30/07/2015
A Mendelian randomization study of the effect of type-2 diabetes on coronary heart disease.	Ahmad OS	26017687	Nat Commun	29/05/2015

Adiposity as a cause of cardiovascular disease: a Mendelian randomization study.	Hagg S	26016847	Int J Epidemiol	29/05/2015
Adult height, coronary heart disease and stroke: a multi-locus Mendelian randomization meta-analysis.	Nuesch E	25979724	Int J Epidemiol	17/05/2015
Mendelian randomization analysis supports the causal role of dysglycaemia and diabetes in the risk of coronary artery disease.	Ross S	25825043	Eur Heart J	01/04/2015
Age- and sex-specific causal effects of adiposity on cardiovascular risk factors.	Fall T	25712996	Diabetes	26/02/2015
Remnant cholesterol, low-density lipoprotein cholesterol, and blood pressure as mediators from obesity to ischemic heart disease.	Varbo A	25411050	Circ Res	21/11/2014
Obesity as a causal risk factor for deep venous thrombosis: a Mendelian randomization study.	Klovaite J	25161014	J Intern Med	28/08/2014
Clinical effect of naturally random allocation to lower systolic blood pressure beginning before the development of hypertension.	Ference BA	24591335	Hypertension	05/03/2014
Causal effects of body mass index on cardiometabolic traits and events: a Mendelian randomization analysis.	Holmes MV	24462370	Am J Hum Genet	28/01/2014
The role of adiposity in cardiometabolic traits: a Mendelian randomization analysis.	Fall T	23824655	PLoS Med	05/07/2013
The effect of elevated body mass index on ischemic heart disease risk: causal estimates from a Mendelian randomisation approach.	Nordestgaard BG	22563304	PLoS Med	09/05/2012
Genes associated with adult cerebral venous thrombosis.	Marjot T	21350198	Stroke	26/02/2011
Association of a fasting glucose genetic risk score with subclinical atherosclerosis: The Atherosclerosis Risk in Communities (ARIC) study.	Rasmussen-Torvik LJ	21036910	Diabetes	03/11/2010

Lifetime body mass index and later atherosclerosis risk in young adults: examining causal links using Mendelian randomization in the Cardiovascular Risk in Young Finns study.	Kivimaki M	18550552	Eur Heart J	14/06/2008

Appendix Table 3. Mendelian randomization studies investigating the effect of social and behavioural traits on cardiovascular disease.

Title	First Author	PMID	Journal	Date
Conventional and genetic evidence on alcohol and vascular disease aetiology: a prospective study of 500,000 men and women in China.	Millwood IY	30955975	Lancet	09/04/2019
Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways.	Jansen PR	30804565	Nat Genet	26/02/2019
Evaluation of the causal effects between subjective wellbeing and cardiometabolic health: mendelian randomisation study.	Wootton RE	30254091	BMJ	27/09/2018
Alcohol Intake and Risk of Ischemic and Haemorrhagic Stroke: Results from a Mendelian Randomisation Study.	Christensen AI	29886720	J Stroke	12/06/2018
Education and coronary heart disease: mendelian randomisation study.	Tillmann T	28855160	BMJ	01/09/2017
Alcohol Consumption, Aldehyde Dehydrogenase 2 Gene Polymorphisms, and Cardiovascular Health in Korea.	Shin MJ	28540979	Yonsei Med J	26/05/2017
Effect of handgrip on coronary artery disease and myocardial infarction: a Mendelian randomization study.	Xu L	28424468	Sci Rep	21/04/2017
Genetically predicted milk consumption and bone health, ischemic heart disease and type 2 diabetes: a Mendelian randomization study.	Yang Q	28225053	Eur J Clin Nutr	23/02/2017
Coffee intake, cardiovascular disease and all-cause mortality: observational and Mendelian randomization analyses in 95,000-223,000 individuals.	Nordestgaard AT	28031317	Int J Epidemiol	30/12/2016
Habitual coffee consumption and risk of type 2 diabetes, ischemic heart disease, depression and Alzheimer's disease: a Mendelian randomization study.	Kwok MK	27845333	Sci Rep	16/11/2016
Associations of the MCM6-rs3754686 proxy for milk intake in Mediterranean and	Smith CE	27624874	Sci Rep	15/09/2016

American populations with cardiovascular biomarkers, disease and mortality: Mendelian randomization.				
Alcohol intake and cardiovascular risk factors: A Mendelian randomisation study.	Cho Y	26687910	Sci Rep	22/12/2015
Evaluation of Moderate Alcohol Use With QT Interval and Heart Rate Using Mendelian Randomization Analysis Among Older Southern Chinese Men in the Guangzhou Biobank Cohort Study.	Au Yeung SL	26153479	Am J Epidemiol	15/07/2015
Milk intake is not associated with ischaemic heart disease in observational or Mendelian randomization analyses in 98,529 Danish adults.	Bergholdt HK	26085675	Int J Epidemiol	19/06/2015
Testing for non-linear causal effects using a binary genotype in a Mendelian randomization study: application to alcohol and cardiovascular traits.	Silverwood RJ	25192829	Int J Epidemiol	07/09/2014
Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data.	Holmes MV	25011450	BMJ	12/07/2014
Causal associations of tobacco smoking with cardiovascular risk factors: a Mendelian randomization analysis of the HUNT Study in Norway.	Asvold BO	24867305	Int J Epidemiol	29/05/2014
Moderate alcohol use and cardiovascular disease from Mendelian randomization.	Au Yeung SL	23874492	PLoS One	23/07/2013
Exploring causal associations between alcohol and coronary heart disease risk factors: findings from a Mendelian randomization study in the Copenhagen General Population Study.	Lawlor DA	23492672	Eur Heart J	16/03/2013

Appendix Table 4. Mendelian randomization studies investigating the effect of cellular traits on cardiovascular disease.

Title	First Author	PMID	Journal	Date
No Causal Effect of Telomere Length on Ischemic Stroke and Its Subtypes: A Mendelian Randomization Study.	Cao W	30769869	Cells	17/02/2019
Prioritizing putative influential genes in cardiovascular disease susceptibility by applying tissue-specific Mendelian randomization.	Taylor K	30704512	Genome Med	02/02/2019
Genetically Determined Platelet Count and Risk of Cardiovascular Disease.	Gill D	30571169	Arterioscler Thromb Vasc Biol	21/12/2018
Blood Eosinophil Count and Metabolic, Cardiac and Pulmonary Outcomes: A Mendelian Randomization Study.	Amini M	29506594	Twin Res Hum Genet	07/03/2018
Deep molecular phenotypes link complex disorders and physiological insult to CpG methylation.	Zaghlool SB	29325019	Hum Mol Genet	13/01/2018
Mendelian Randomization Analysis Identifies CpG Sites as Putative Mediators for Genetic Influences on Cardiovascular Disease Risk.	Richardson TG	28985495	Am J Hum Genet	07/10/2017
Large-Scale Identification of Common Trait and Disease Variants Affecting Gene Expression.	Hauberg ME	28552197	Am J Hum Genet	30/05/2017
Exploring the Causal Pathway From Telomere Length to Coronary Heart Disease: A Network Mendelian Randomization Study.	Zhan Y	28515044	Circ Res	19/05/2017
Association Between Telomere Length and Risk of Cancer and Non-Neoplastic Diseases: A Mendelian Randomization Study.	Telomeres Mendelian Randomization Collaboration.	28241208	JAMA Oncol	28/02/2017
The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease.	Astle WJ	27863252	Cell	20/11/2016

The effect of hematocrit and hemoglobin on the risk of ischemic heart disease: A Mendelian randomization study.	Zhong Y	27609746	Prev Med	10/09/2016
Predicting gene targets from integrative analyses of summary data from GWAS and eQTL studies for 28 human complex traits.	Pavlides JM	27506385	Genome Med	11/08/2016
Telomere length and health outcomes: A two-sample genetic instrumental variables analysis.	Hamad R	27321645	Exp Gerontol	21/06/2016
Short Telomere Length and Ischemic Heart Disease: Observational and Genetic Studies in 290 022 Individuals.	Scheller Madrid A	27259814	Clin Chem	05/06/2016

Appendix Table 5. Mendelian randomization studies investigating the effect of existing drugs on cardiovascular disease.

Title	First Author	PMID	Journal	Date
Mendelian Randomization Study of ACLY and Cardiovascular Disease.	Ference BA	30865797	N Engl J Med	14/03/2019
Association of Triglyceride-Lowering LPL Variants and LDL-C-Lowering LDLR Variants With Risk of Coronary Heart Disease.	Ference BA	30694319	JAMA	30/01/2019
Genetic Regulation of PCSK9 (Proprotein Convertase Subtilisin/Kexin Type 9) Plasma Levels and Its Impact on Atherosclerotic Vascular Disease Phenotypes.	Pott J	29748315	Circ Genom Precis Med	12/05/2018
CETP (Cholesteryl Ester Transfer Protein) Concentration: A Genome-Wide Association Study Followed by Mendelian Randomization on Coronary Artery Disease.	Blauw LL	29728394	Circ Genom Precis Med	08/05/2018
Genetic Association of Lipids and Lipid Drug Targets With Abdominal Aortic Aneurysm: A Meta-analysis.	Harrison SC	29188294	JAMA Cardiol	01/12/2017
Mendelian randomization analysis of cholesteryl ester transfer protein and subclinical atherosclerosis: A population-based study.	Christen T	29174438	J Clin Lipidol	28/11/2017
Differential effects of PCSK9 variants on risk of coronary disease and ischaemic stroke.	Hopewell JC	29020353	Eur Heart J	12/10/2017
Association of Genetic Variants Related to CETP Inhibitors and Statins With Lipoprotein Levels and Cardiovascular Risk.	Ference BA	28846118	JAMA	29/08/2017
Investigating Real-World Clopidogrel Pharmacogenetics in Stroke Using a Bioresource Linked to Electronic Medical Records.	Tornio A	28653333	Clin Pharmacol Ther	28/06/2017
Effect of Bile Acid Sequestrants on the Risk of Cardiovascular Events: A Mendelian Randomization Analysis.	Ross S	26043746	Circ Cardiovasc Genet	06/06/2015
Genetic variation in the cholesterol transporter NPC1L1, ischaemic vascular disease, and gallstone disease.	Lauridsen BK	25841872	Eur Heart J	07/04/2015

Effect of naturally random allocation to lower low-density lipoprotein cholesterol on the risk of coronary heart disease mediated by polymorphisms in NPC1L1, HMGCR, or both: a 2 x 2 factorial Mendelian randomization study.	Ference BA	25770315	J Am Coll Cardiol	17/03/2015
Association of cholesteryl ester transfer protein (CETP) gene polymorphism, high density lipoprotein cholesterol and risk of coronary artery disease: a meta-analysis using a Mendelian randomization approach.	Wu Z	25366166	BMC Med Genet	05/11/2014