Year : 2023

Robust Causal Inference Methods to Assess Risk Factors for Common Diseases

Darrous Liza

UNIL | Université de Lausanne

Faculté de biologie
et de médecine

Unisanté
Département épidémiologie et systèmes de santé

Robust Causal Inference Methods to Assess Risk Factors for Common Diseases

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

Faculté de biologie et de médecine
de l'Université de Lausanne

par

Liza DARROUS

Master de l'Université de Lausanne

Jury

Prof. Curdin Conrad, Président
Prof. Zoltán Kutalik, Directeur de thèse
Prof. Valentin Rousson, Expert
Prof. Frank Dudbridge, Expert

Lausanne
(2023)

UNIL | Université de Lausanne

Faculté de biologie
et de médecine

Ecole Doctorale

Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président·e	Monsieur	Prof.	Curdin	**Conrad**
Directeur·trice de thèse	Monsieur	Prof.	Zoltán	**Kutalik**
Expert·e·s	Monsieur	Prof.	Valentin	**Rousson**
	Monsieur	Prof.	Frank	**Dudbridge**

le Conseil de Faculté autorise l'impression de la thèse de

Liza Darrous

Master - Maîtrise universitaire ès Sciences en sciences moléculaires du vivant, Université de Lausanne

intitulée

Robust causal inference methods to assess risk factors for common diseases

Lausanne, le 24 novembre 2023

pour le Doyen
de la Faculté de biologie et de médecine

Prof. Curdin Conrad

The more I learn,
the less I know.

– Unknown

Acknowledgements

In my twenty-seven years of life, I have journeyed many times and to several places, but no journey has been as filled with triumphs, touched by occasional defeats, and consistently supported by so many helping hands as this one. Looking back, many feelings course through me–pride, satisfaction, and above all, gratitude. I would like to take this opportunity to thank those who shared this journey with me and supported me along the way.

First and foremost, I would like to offer my sincere gratitude to my supervisor, **Zoltán**. Thank you for taking a chance on the young and wide-eyed student I was six years ago. For seeing potential in me, even when I found it difficult to. For teaching me, sharing your expertise with me, and encouraging me to pursue various opportunities. For being supportive, understanding, and providing a diverse, fair, and stimulating environment for me and everyone in the group. I will always be in awe of your brilliance and feel privileged to have been guided on this journey by you.

An immense thank you goes to the frankly remarkable members of the **Statistical Genetics Group**, both former and current. The SGG was imbued with the spirit of unconditional support, camaraderie, and diversity when I first joined, and I hope it continues long after I leave. **Eleonora** and **Sina**, thank you for welcoming me with such open and supportive arms. Your kindness, encouragement to pursue my goals, and earnest belief in me lingers to this very day. **Jonathan**, thank you for always having time to spare for my questions, both scientific and those about your pets. I valued your input and manuscript comments immensely. **Chiara**, you are a brilliant and passionate scientist, and I am both inspired by you and honoured to have shared most of my PhD journey with you.

One of the nicest surprises I got along the way was your friendship **Marie**, you are genuine and kind-hearted in every way. I will always treasure our many many many discussions, (road)trips, and your delicious baked sweets. **Sven** and **Marion**, although the running joke is that you were adopted by the SGG, in truth, you adopted me into your group even before then. Thank you for all your help and support, both academically and personally, I so appreciate all the shared drinks and laughs. **Leona**, thank you for always being compassionate, uplifting, and a source of inspiration in your determination to always get to the finish line.

Last but certainly not least, I would like to sincerely thank **Ninon**. Your earnest kindness took me by surprise, and your tireless patience and unending support built the strongest foundation for me to start this journey. You have been a mentor to me, and a colleague, but I am most grateful to have you as a dear friend at the end of it all.

Allow me this cliche of saying that I could not have done this without the support and encouragement of my family. Every infinitesimal decision I have made knowingly or not, that has led me here, was based on the very first virtues that my parents have instilled in me; curiosity and the thirst for knowledge. My Mother, **Sabah**, gifted me her endless patience, unwavering faith, and the knowledge that I can accomplish things on my own, although I am never alone. My father, **Michel**, lavished me with his pride, gentle kindness, and unconditional support that still echoes to this day. This is but a drop in the ocean of debt I owe them.

شايف البحر شو كبير، كبر البحر بحبَّك..

My sisters, **Laura** and **Luna**, have proven to me once and again, that distance does indeed make the heart grow fonder. During all the lows that I have experienced, their words were a soothing balm to my tired psyche, and during all the highs, they were the wind that carried my laughter loud and far. Additionally, my aunt **Najah**, uncle **Georges** and cousins **Ziad** and **Nada** have been staunch in their support, love, and pride from all corners of this earth.
I couldn't be more fortunate to have my family, and my gratitude for them will forever continue to grow.

Furthermore, I owe so much to **Yamane**, who has gifted me a home here in Lausanne, and who in all regards but biological, is a sister to me. It is difficult to imagine the past eight years without your warm heartedness, confidence in me, and generosity in sharing the good things in life. I am lucky I don't have to.
I would also like to thank **Germain** and **Elie**, who besides their advice, encouragement, and enthusiasm for my whims, have offered me a much-needed escape into fantasy with our accidental book-club. Other friends, both near and far, old and new: **Yara**, **George**, **Fayez**, **John**, **Joseph**, **Chris**, I thank you for all the joyful moments you have brought me over the years, and look forward to sharing many more with you all.

Lastly, a special thank you goes to **Lausanne**. My friends and I have so often spoken fondly of Lausanne as if it were a real person that I have come to think of it that way. My journey has been all the more special because it took place here; I have spent many afternoons reading manuscripts by its lake shore, many sunny days ruing my laptop screen for not letting me work outdoors, and many nights recovering under the marvel of its stars.

Liza Darrous
November 16, 2023

Abstract

The study of complex traits, those influenced by multiple genetic and environmental factors, has long been a cornerstone of genetic research, where scientists have sought to untangle this complexity. These traits include a vast array of human characteristics, from molecular phenotypes to diseases.
The advent of Genome-Wide Association Studies (GWAS) following human genome sequencing marked an essential moment in this pursuit. These studies, characterised by their large sample size and examination of millions of genetic variants, have significantly advanced our understanding of the genetic architecture underlying complex traits. GWAS have unearthed numerous genetic markers associated with various traits, providing vital clues for further exploration.

GWAS have not only identified genetic associations to complex traits, but have also helped researchers explore the relationships between these traits. Understanding the causal relationships among traits is essential due to its potential to improve medical practices and public health interventions. In response, Mendelian Randomisation (MR) emerged as a genetically-informed version of previous causal inference methods, such as Randomised Control Trials (RCTs). MR uses genetic variants as instrumental variables to elucidate causal relationships between traits, distinguishing true causation from mere correlation. As a statistical method, MR comes with several assumptions that must hold for accurate estimation. However, validating some of these assumptions can be challenging, potentially introducing bias in the estimation of causal effects.

During my thesis, I investigated assumption violations that MR often faces, particularly in two scenarios: (i) the presence of unmeasured heritable confounding factors introducing spurious causal relationships and (ii) the heterogeneity of causal effects due to potential underlying pleiotropic pathways or confounder mechanisms.
To address the first assumption violation, I developed an extension to the MR model known as **LHC-MR**, which accounts for the presence of a **L**atent **H**eritable **C**onfounder. LHC-MR is applicable to association summary statistics of trait pairs, allowing simultaneous estimation of bi-directional causal effects, direct heritabilities, and confounder effects on the pair.
For the second assumption violation, I proposed an approach, **PWC-MR**, that leverages **P**henome-**W**ide association data across several traits to perform informative **C**lustering of the focal trait instruments. PWC-MR revealed that for body mass index (BMI), distinct clusters of instruments exist with heterogeneous causal effects on educational attainment.
Lastly, I explored indirect genetic effects using individual-level genetic data of sibling pairs. The aim was to estimate the causal effect of the parental environment/rearing on offspring traits in later life, using MR.

In summary, this journey from the study of complex traits to the emergence of GWAS and MR as tools for causal inference has reshaped our understanding of genetics. While MR offers great promise, its often-violated assumptions necessitate careful consideration, and my work aimed to address some of these challenges.

Résumé

L'étude des traits complexes, qui sont influencés par de multiples facteurs génétiques et environnementaux, a toujours été un pilier de la recherche en génétique, où les scientifiques ont cherché à démêler cette complexité. Ces traits englobent une vaste gamme de caractéristiques humaines, comme des phénotypes moléculaires mais aussi certaines maladies courantes.
L'avènement des études d'association pangénomique (GWAS) à la suite du séquençage du génome humain, a marqué un moment essentiel dans cette quête. Ces études, caractérisées par leur grande taille d'échantillon et l'analyse de millions de variants génétiques, ont considérablement avancé notre compréhension de l'architecture génétique des traits complexes, en permettant d'identifier de nombreux marqueurs génétiques associés à divers traits, fournissant ainsi des indices essentiels pour de futures explorations.

Les GWAS ont non seulement permis d'identifier des associations génétiques, mais elles ont également aidé les chercheurs à explorer les relations entre ces traits. Il est essentiel de comprendre les relations de cause à effet entre les traits pour pouvoir améliorer les pratiques médicales et les interventions de santé publique. En réponse, la Randomisation Mendélienne (MR), version génétiquement informée des méthodes précédentes d'inférence causale, telles que les Essais Contrôlés Randomisés, a émergé. La MR utilise des variants génétiques en tant que variables instrumentales pour élucider les relations de cause à effet entre les traits, distinguant ainsi véritable causalité et simple corrélation. C'est une méthode statistique qui repose sur plusieurs hypothèses qui doivent être respectées afin d'obtenir une estimation précise. Cependant, la validation de certaines de ces hypothèses peut s'avérer difficile et leur violation peut introduire un biais dans l'estimation des effets de causalité.

Au cours de ma thèse, j'ai examiné les violations d'hypothèses auxquelles la MR est souvent confrontée, en particulier dans deux scénarios : (i) la présence de facteurs confondants héréditaires non mesurés introduisant des relations de causalité fallacieuses et (ii) l'hétérogénéité des effets de causalité due à d'éventuels effets pléiotropiques ou à des facteurs confondants.
Concernant la première violation d'hypothèse, j'ai développé une extension du modèle MR appelée **LHC-MR**, qui prend en compte la présence d'un facteur **C**onfondant **H**éréditaire **L**atent. LHC-MR utilise des statistiques synthétiques issues des GWAS pour étudier la relation entre deux traits, via l'estimation simultanée d'effets de causalité bidirectionnels, d'héritabilités directes et des effets du facteur confondant sur chacun des traits.
Pour aborder le deuxième scénario, j'ai proposé une approche, **PWC-MR**, qui permet d'effectuer un regroupement informatif des instruments, sélectionnés pour leur association avec le facteur de risqué d'intérêt, en exploitant des données d'association génétique avec plusieurs autres traits. PWC-MR a révélé que, pour l'indice de masse corporelle (IMC), il existe des groupes distincts d'instruments avec des effets de causalité hétérogènes sur le niveau d'éducation. Enfin, j'ai exploré les effets génétiques indirects en utilisant des données génétiques d'individus issus d'une même fratrie. L'objectif était d'utiliser la MR pour estimer l'effet de causalité de l'environnement parental sur les traits des enfants à un stade ultérieur de leur vie.

En résumé, l'étude des traits complexes, depuis l'émergence des GWAS jusqu'à l'utilisation de la MR en tant qu'outil pour l'inférence de causalité, a remodelé notre compréhension de la génétique. Bien que la MR offre de grandes promesses, ses hypothèses souvent violées nécessitent une réflexion minutieuse, et mon travail de doctorat a permis de proposer des solutions pour relever certains de ces défis.

Contents

Acknowledgements v

Abstract vii

Résumé ix

Contents xi

1 Introduction 1
- 1.1 Genetics of complex traits . . . 1
- 1.2 Genome Wide Association Studies . . . 5
- 1.3 Biases in GWAS: how are they (thus far) addressed? . . . 11
- 1.4 Beyond association and correlation: Causation . . . 14

2 Chapter 2 21
- 2.1 Accounting for latent heritable confounding in Mendelian Randomisation . . . 21
- 2.2 R package: `lhcMR` . . . 23
- 2.3 Estimating the causal effect of stratified physical activity on cognitive functioning 24

3 Chapter 3 27
- 3.1 PheWAS-based clustering of Mendelian Randomisation instruments . . . 27
- 3.2 Application: clustering of obesity, a composite trait . . . 29

4 Chapter 4 35
- 4.1 Ongoing: Estimating family-to-offspring causal effects from genetic data of first-degree relatives . . . 35
- 4.2 Methods . . . 36
- 4.3 Results . . . 39
- 4.4 Discussion . . . 42

5 Chapter 5 45
- 5.1 Minor contributions to other publications . . . 45

6 Discussion 47
- 6.1 The golden thread: confounding . . . 47
- 6.2 Moving forward: MVMR and its caveats . . . 50
- 6.3 Future works: Using haplotype data for indirect genetic effect estimation . . . 54
- 6.4 Conclusion . . . 56

Bibliography 59

Appendix 67

List of Figures

1.1 **Sequencing of genetic variants.** The 3-billion base pair genome contains about 88 million SNVs, with over 15 million variants having MAF > 1%, as evidenced by projects like HapMap and the 1000 Genomes Project. Sequencing reads organised into *contigs* are typically used in genome assembly. GWAS use genotyping arrays to analyse genetic variation within a population. Further variant data can be imputed by inferring LD structure from a reference panel (see next section). Exon sequencing focuses on exons that make up genes. 3

1.2 **Individual-level genotype data to GWAS summary statistics. 1.** Individual-level genotype data for M individuals across K SNVs. Note that for SNV k there is missing genotype data. **2.** Association summary statistic estimated per SNV. By regressing phenotypic values on genotypic information, we estimate the per allele effect size and standard error. **3.** Aggregated data, also called GWAS summary statistics, listing variant IDs, effect sizes, their standard errors, effect allele frequency, sample size amongst other variant information. Imputation scores are reported for SNVs with missing genotypes that are imputed using reference panels. 7

1.3 **Schematic representation of allele frequency vs. effect size.** The proposed relationship between allele frequency and effect size (or penetrance) based on our genetic understanding of complex traits and disease. Most GWAS findings of complex traits and common diseases correspond to common variants with relatively small effect sizes/low penetrance. Detecting rare variants with low penetrance requires large sample sizes to distinguish these rare variants and to have enough power to estimate their effects. Monogenic/mendelian diseases on the other hand, have variants with high penetrance despite being rare. Adapted from Bush and Moore [24]. 8

1.4 **Illustration of the genetic correlation between traits A and B.** Here, the product of the standardised effects/Z-score of traits A and B (Y-axis) is regressed onto the LD score (X-axis). The genetic correlation can be derived from the regression slope. The yellow slope shows an example of two positively correlated traits, the dark blue slope shows an example of anti-correlated traits, and cyan, an example of uncorrelated traits. 10

1.5 **Core assumptions of Mendelian randomisation.** Genetic marker G representing an IV has a direct effect on exposure/risk factor X denoted as γ. It also has an effect on outcome/disease Y through X. The measured effect of G on Y is denoted as τ. U represents potential confounding factors such as population stratification or sex. The dashed lines illustrate potentially violated MR assumptions. β represents the MR causal effect. Adapted from Bowden, Davey Smith, and Burgess [48]. 16

1.6 **The different types of pleiotropy.** G represents an instrumental variable with a direct effect on exposure/risk factor X. It can also have an effect on outcome/disease Y through C, a latent heritable trait shown in red. U represents potential confounding factors such as population stratification or sex. The dashed arrows illustrate potential relationships between traits, whereas red arrows illustrate relationships that will bias the causal effect estimation between X and Y. 19

2.1 The manuscript *Simultaneous estimation of bi-directional causal effects and heritable confounding from GWAS summary statistics* and its supplementary materials can be found on *Nature communications* here, or in Appendix A. This work was selected as one of SIB's remarkable outputs for 2021. 21

2.2 **Illustration of causal effects estimated from SNP-outcome and SNP-exposure associations.** The regression of the SNP-outcome association $\hat{\tau}$, onto the SNP-exposure association $\hat{\gamma}$ reveals in reality two separate SNP clusters, in the presence of a heritable confounder: those truly associated with the exposure with an effect on the outcome shown in light blue and having a causal effect β, and those that are primarily associated with the confounder having an effect equivalent to the ratio of the confounder's effect on outcome to exposure (q_y/q_x). Running standard IVW would result in an underestimated causal effect represented by the slope of the dashed grey line, taking into account all SNPs. LHC-MR aims to disentangle these separate slopes/effects. 22

2.3 The manuscript *Genetic insights into the causal relationship between physical activity and cognitive functioning* can be found on *Scientific Reports - Nature* here, or in Appendix B. 24

3.1 The manuscript *PheWAS-based clustering of Mendelian Randomisation instruments reveals distinct mechanism-specific causal effects between obesity and educational attainment* and its supplementary materials can be found on *medRxiv* here. A revised version of the manuscript is found in Appendix C. This manuscript has been submitted to *Nature Communications*, where it is under a second round of revision. 27

3.2 **A simplified graph representation of a systematic confounder search.** Around 400 traits were each used as an exposure (represented by T), and their bi-directional causal effect estimate on BMI and on EDU were separately estimated. Traits with significant causal effects on both BMI and EDU were labelled as putative confounder traits. 28

3.3 **Heat map of the trait enrichment ratios in each cluster for PCs of the obesity composite trait. a** PC1 representing body size clustered into 6 groups. **b** PC2 representing adiposity clustered into 8 groups. **c** PC3 representing abdominal fat clustered into 5 groups. **d** PC4 representing lean mass clustered into 6 groups. The darker the colour, the stronger the enrichment ratio (ER). 30

3.4 **IVW causal effect estimates of all GWS SNPs as well as cluster-specific SNPs for each PC of the obesity composite trait. a** Causal effect estimate of PC1 representing body size on EDU. **b** Causal effect estimate of PC2 representing adiposity on EDU. **c** Causal effect estimate of PC3 representing abdominal fat on EDU. **d** Causal effect estimate of PC4 representing lean mass on EDU. 32

4.1 **MR representation of estimating the causal effect of parental rearing environment on offspring phenotype (β), using unstransmitted parental/indirect effects.** . . . 35

4.2 **Graph representation of direct and indirect effects of parental genotype onto offspring genotype and subsequently, phenotype.** G_p represents the parental genotype, where a single allele is inherited by the index individual's genotype (G_i). In this scenario, the other allele is inherited by the sibling (G_j). This untransmitted effect, which we aim to estimate, is hypothesised to act via a rearing/environmental effect on the offspring phenotype. 36

4.3 **Illustration of allele transmission from parents to offspring** Three scenarios are shown of differing allele transmission leading to different allele status for sibling 2. Consequently, the allele dosage differs, and so does the estimated untransmitted allele with respect to sibling 1. 37

4.4 **Forest plot of the genetic correlation of both BMI and EDU across 12 other selected traits.** Effects of BMI and EDU come from 3 different sources of GWAS: population-based, within-sibship, sibling meta-analysed indirect effects (res_sib). The effects of the other traits all come from population-based GWAS. 95% confidence intervals are shown as error bars. Points that are not filled indicate a non-significant genetic correlation estimate. 39

4.5 **Nominally significant IVW causal effect estimates of various traits on both untransmitted BMI and EDU (meta-analysed sibling effects).** 40

6.1 **A simplified multivariable Mendelian randomisation (MVMR) model with two exposures.** G represents a group of IVs, each associated with at least one of the two exposures. The line between the first exposure X_1, and the second X_2 is bidirectional and dashed as no assumptions are made about this relationship in the estimation of their respective causal effects, β_1 and β_2, on the outcome (Y). Confounders U_1 and U_2 are assumed to be unknown. 50

6.2 **Graph representation of transmitted and untransmitted parental alleles onto siblings. a** shows scenario 1, where the T allele of sibling 1 is inherited from parent 1, leaving the A allele to be untransmitted. **b** shows scenario 2 where the A allele of parent 1 is inherited by sibling 1, leaving the T allele to be untransmitted. 54

6.3 **Graphical representation of pedigree haplotype phasing and reverse pedigree haplotype phasing.** Top panel shows haplotype phasing of offspring genomes using parental genomes. Bottom panel shows reverse haplotype phasing using offspring genome to extract untransmitted alleles from the parents (highlighted in red and blue). Adapted from Robin Hofmeister. 55

6.4 **Graphical representation of IBD mapping between sibling pairs.** Chromosomal areas in grey are not shared between either sibling. Taking sibling 1 as reference, grey areas on either chromosome for sibling 2 that are not shared could be either diploid for unknown alleles or haploid. Adapted from Robin Hofmeister. 56

0.1 **QQ-plot and Manhattan plot for untransmitted BMI and EDU effects (meta-analysed sibling effects). a** QQ-plot of GWAS P-values. **b** Manhattan plot of selected SNPs with $-log_{10}(P) \geq 2$. 69

0.2 **Forest plot of the genetic correlation of both BMI and EDU across** ~190 **other traits.** The effects of both BMI and EDU come from 3 different sources of GWAS: population-based, within-sibship, sibling meta-analysed indirect effects. The effects of other traits all come from population-based GWAS. 95% confidence intervals are shown as error bars. Points that are not filled indicate a non-significant genetic correlation estimate. 74

List of Tables

1.1 **Table of some MR methods and extensions that handle various assumption violations by relaxing the targeted assumption.** PWC-MR and LHC-MR are discussed further in the next chapter. 20

4.1 **Stepwise-MVMR causal effect estimates on offspring BMI.** One parental trait that survives stepwise-MVMR has its causal effect on untransmitted BMI estimated using MVMR. 41

4.2 **Stepwise MVMR causal effect estimates on offspring EDU.** Two parental traits that survive stepwise-MVMR have their conditional causal effect on untransmitted EDU estimated using MVMR. 41

Introduction 1

1.1 Genetics of complex traits 1
1.2 Genome Wide Association Studies 5
1.3 Biases in GWAS: how are they (thus far) addressed? 11
1.4 Beyond association and correlation: Causation 14

The following document details the work I have accomplished during my PhD studies under the supervision of Professor Zoltán Kutalik. Titled *"Robust Causal Inference Methods to Assess Risk Factors for Common Diseases"*, my thesis work focused on improving methods for causal inference between pairs of traits, such as risk factors and various diseases. My work on improving these methods was through developing more robust ways to account for the various assumption violations that could occur, and thus was not specific to any single trait pair and their subsequent potential causal relationship.
However, understanding the various ways in which traits interact with each other was imperative to comprehend the potential sources of these violations and how to better account for, or utilise them.

This introduction provides a summary of our current understanding of the genetics underlying many of our (complex) traits, how these genetic-trait associations are typically estimated, and the different sources of estimation bias that could exist. Lastly, it discusses various methods to infer causality between phenotypes, with an emphasis on one in particular that uses genetic data.
The following chapters summarise my contributions to the development of various method extensions aimed at making causal inference more robust in the face of assumption violations. They also include ongoing work related to estimating genetic effects that are not directly associated with the trait of interest. The final chapter then discusses the relevance of these findings, persistent challenges, and possible avenues for future research.

Trait, *phenotype*, and *disease* will be used interchangeably in this document

1.1 Genetics of complex traits

The common denominator amongst all living things is the **Deoxyribonucleic acid (DNA)**, a polymer consisting of two polynucleotide chains that coil around each other to form a double helix. Comprising approximately 3 billion nucleotide pairs, the human DNA houses all the genetic instructions necessary for growth, development, functioning, repair, and reproduction.

Duality is a common aspect of the human genome, first seen in the double strands that make up the DNA molecule, in the nucleotide pairings (base-pairings of A-T or C-G) that glue together the two strands, and in homologous chromosome pairs, each inherited from a parent, resulting in 22 autosomal pairs and one sex pair.

The genome is composed of several crucial building blocks that are called **genes** - segments of DNA that code for **proteins**. Sections of the genome are referred to as **loci**, where a locus can be of any length, ranging from a single nucleotide (1 bp) up to 10 million base pairs (10 Mbp).

At any point in the genome, variations can arise due to mutations, random mating, recombination between homologous chromosomes during meiosis, or various other factors. If such variations occur within a gene or its regulatory region, they can lead to changes in the physical structure of the resulting protein or alter the timing and location of protein production, subsequently resulting in phenotypic variations.

Genetic mutations can take various forms, each with distinct consequences. *Point mutations* involve changes in individual base pairs, where a single nucleotide can be substituted, deleted, or inserted. *Chromosomal mutations*, on the other hand, occur in the form of numerical abnormalities, where there is an atypical number of chromosomes, or as structural abnormalities.

The latter encompasses events such as inversion (where a segment is reversed), translocation (where a segment moves to another location), as well as duplication (where a segment is copied), and deletion (where a segment is lost), both of which are also known as *copy number variations*. The number of duplication or deletion repeats varies between individuals.

While point mutations can result in subtle changes at the gene level, chromosomal mutations and copy number variations can have more extensive effects, impacting larger portions of an individual's genetic material.

For a more detailed overview on genetics, refer to *Introduction to Quantitative Genetics* by Falconer and Mackay [1], and *Genetic architecture: the shape of the genetic contribution to human traits and disease* by Timpson et al. [2]

A point-mutation is called a **single-nucleotide variant (SNV)**, and its two versions (one paternal and the other maternal) are called **alleles**. An individual having different alleles in their genome is said to be **heterozygous** at that locus, and an individual having the same allele is said to be **homozygous** in comparison. If both alleles are common in the population, then the **minor allele frequency (MAF)**[1] of that variant is > 1%, and the variant is called a **single-nucleotide polymorphism (SNP)**.

The Human Genome Project [3, 4] (launched in 1990 and completed in 2003) advanced genetic studies by producing the

[1] *Minor allele frequency (MAF)*: the allele frequency of the less common allele. Minor alleles and their frequencies can vary across populations, which is why large-scale studies often report the allele frequency of the genotyped allele, termed the *effect allele*

first sequence of the human genome (92% of it, accounting for the advancements in genetic sequencing at the time), and making every part of the draft human genome sequence publicly available shortly after production. This international and open collaboration led to several other projects that aimed at publicly cataloguing genetic variations such as The **HapMap project** [5] (2002-2009), that produced a haplotype[2] map of the human genome to describe the common patterns of human DNA sequence variation.

[2] *Haplotype*: a DNA sequence along a single chromosome with variations that tend to be inherited together

As sequencing technologies improved, increasingly larger catalogues were formed like the **1000 Genomes Project** [6, 7] (2008-2015) which sequenced the genomes of over 2′500 unrelated volunteers from 26 populations around the world, allowing it to better distinguish common variants and their allele frequencies across different population.

Focus then shifted to exons, the coding regions of genes which make up 2% of the genome and that are translated into proteins. The **Exome Aggregation Consortium** [8, 9] (2014-2016) began this initiative by sequencing the exomes of 60′000 individuals, and was built upon by the **Genome Aggregation Database (gnomAD)** [10] which also includes whole genome sequencing information.

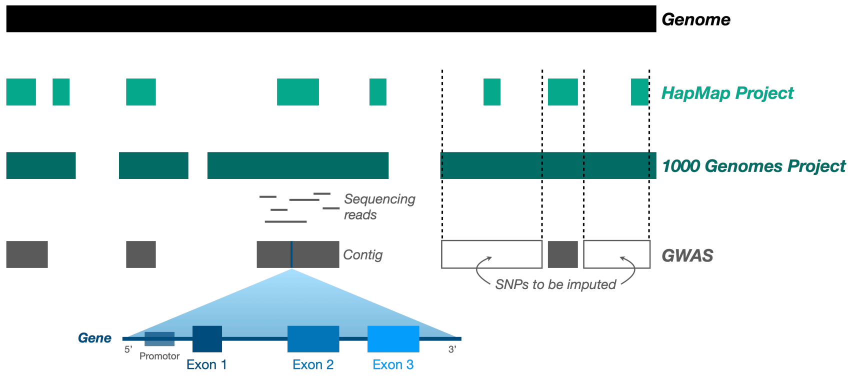


Figure 1.1: Sequencing of genetic variants. The 3-billion base pair genome contains about 88 million SNVs, with over 15 million variants having MAF > 1%, as evidenced by projects like HapMap and the 1000 Genomes Project. Sequencing reads organised into *contigs* are typically used in genome assembly. GWAS use genotyping arrays to analyse genetic variation within a population. Further variant data can be imputed by inferring LD structure from a reference panel (see next section). Exon sequencing focuses on exons that make up genes.

The final and complete human genome sequence [11] was released by the **Telomere to Telomere (T2T) consortium** (2020-2022), which addressed the remaining 8% of the genome to present a complete 3.055 billion base pair sequence.

The sequencing of the human genome led to many new discoveries about the make up and architecture of our genome. Contrary to previous beliefs, humans only have about 20′000 to 25′000 protein-coding genes, an estimate much lower than those of other organisms such as plants and insects. The 1000

Genomes Project [6] showed that of the 3 billion base pairs of the human genome, around 88 million only (< 3%) are SNVs and of those, only ~8 million have a frequency > 5%.
Furthermore, we learned that the human genome has a **haplotype block structure** [12] resulting from the recombination of chromosome segments during meiosis, at specific sites in the genome called recombination hot-spots. Consequently, phasing[3] can be accurately performed in areas of low recombination by using the genotyped data of multiple individuals and using a probabilistic model to estimate the more likely haplotypes of a given population [13].

Recombination, also known as crossing-over, is a fundamental step in meiosis and plays a crucial role in creating genetic diversity among offspring

We also learnt that SNPs on the same haplotype tag each other well, since each of the two possible alleles for a particular SNP can only belong to a limited number of haplotypes in that region. This feature of the genome; whereby genotyping a set of highly informative tag-SNPs[4] allows us to effectively encompass a significant portion of adjacent genetic variations that are not directly genotyped, has been very beneficial for designing genotyping microarrays.
Since haplotypes are inherited as large segments of the genome with few recombination events per chromosome and per generation, a SNP in a haplotype will be passed down to new generations with the same set of neighbouring alleles found on the original haplotype. If the alleles of two different loci are not independent from each other, we say that the loci are in **linkage disequilibrium (LD)** [14, 15].
In other words, if two SNPs are in LD, then by observing for a certain haplotype the allele of the first SNP, we can more accurately determine the allele of the second SNP compared to only knowing the population allele frequency of the second SNP.
LD can be crudely measured using the Pearson's correlation coefficient between the allele frequencies of two different SNPs on the same haplotype.

As we will soon see, traits can be classified based on their measurement types, but in genetics, traits are also classified based on the number of genes that influence them. **Monogenic traits**, also known as Mendelian traits are influenced by a single gene or locus, such as sickle cell anaemia or cystic fibrosis.
These monogenic traits of diseases follow a Mendelian pattern of inheritance, where the responsible gene or genetic variant is rare and highly penetrant[5], hence their study and understanding has been largely dependent on studying and sequencing large family pedigrees.
In contrast **polygenic phenotypes**, such as eye colour and height, are influenced by many genes and loci, and their association

[3] *Phasing*: the process of determining the specific combination of alleles on each chromosome of an individual within a pair of homologous chromosomes

[4] *Tag-SNP*: a representative SNP in a region of the genome that is strongly linked to a group of SNPs in a haplotype

[5] *Penetrance*: the proportion of people with a particular genetic variant (or gene mutation) who exhibit signs and symptoms of a genetic disorder.

study requires much larger sample sizes than family pedigrees to achieve statistical power, due to the polygenicity and low penetrance of the genetic variants involved. **Complex traits**, as their names suggests, are not only polygenic but could also be influenced by several environmental factors. The extent of the genetic contribution to the variability of a trait in a population is termed **heritability**, which is discussed in detail in the next section.

The advancement of genetic sequencing and the efficient design of genotyping microarrays have made our access to genetic data easier, quicker, and cheaper. In turn, this enabled scientists to investigate genome-wide common genetic variations that impact complex traits without any predetermined assumptions, using association studies and large cohorts of (unrelated) individuals.

1.2 Genome Wide Association Studies

Genome wide associations studies (GWAS) quantify the statistical association between a genetic variant and an observed phenotype [16]. Phenotypes can vary from being **quantitative traits** like standing height or blood levels of cholesterol, **qualitative traits** like sex or disease status, or **ordinal traits** like educational level or socio-economic status.

GWAS will be used hereinafter to refer to both the singular and plural term as opposed to other text's usage of *GWASs* for plural.

This association is beneficial for the study of possible biological mechanisms affecting the phenotype and for the prediction of phenotypes given genomic information. Future or downstream benefits include causal inference, personalised medicine, custom intervention techniques, and ancestry inference among others.

The association between each SNV and the phenotype of interest is often estimated using a **fixed effect linear regression model** as β, where the phenotype measure represents the outcome Y, and the dosage[6] of the effect allele represents the predictor X:

$$Y = \beta X + \epsilon \quad , \text{where } \epsilon \sim \mathcal{N}(0, \sigma^2) \tag{1.1}$$

The dosage effect of the SNV can be grouped into different genotype models, such as the additive, dominant or recessive model. Most frequently, the **additive model** - where each additional copy of the effect allele adds to the outcome association - is used.

In the case of qualitative phenotypes, especially binary ones like disease status, a **fixed effect logistic regression** model is used with a case-control study design. Cohorts in this scenario

Often, a linear model will still be fitted to binary outcomes, as larger case and control cohorts and a focus on variants with high MAF allow a comparable performance between linear and logistic regression models

[6] *Allele dosage*: the number of copies of the effect allele (0,1,2)

consist of cases (affected individuals) and controls (healthy individuals), and are often enriched for the former to increase the statistical power in order to detect associations between genetic markers and the disease.
Common **covariates** that are accounted for in GWAS to increase the accuracy of the estimation and correct for confounder effects include age, sex, and genetic principal components that reflect population structure[7]. A fixed effect logistic regression model with covariate K and probability of the binary outcome p is written as:

$$\log\left(\frac{p}{1-p}\right) = \beta X + \alpha K \tag{1.2}$$

Another method to account for population structure or relatedness, other than using fixed effect models with specific covariates, is using **mixed effect models**. These models take into account both fixed effects such as genetic variants, and random effects such as population structure.

$$\begin{gathered} Y = \beta X + \alpha K + u + \epsilon \quad \text{, where } u \sim \mathcal{N}(0, \tau^2\Sigma) \\ \text{with } \Sigma \text{ representing a relatedness matrix} \end{gathered} \tag{1.3}$$

In this case, the association test between the genetic variant and the outcome is calculated conditional on the **genetic relationship matrix (GRM)**. The GRM is estimated from genome-wide SNV data and represents the covariance structure resulting from genetic relatedness in the population. While the fixed-effect models assume that all individuals in the study population are unrelated and are drawn from a single homogeneous population, mixed-effect models are particularly useful when dealing with complex study populations that exhibit varying levels of relatedness, hence accommodating a larger sample size, or when analysing family-based or longitudinal data.

A GWAS (process illustrated in Figure 1.2) returns three main quantities for each variant:

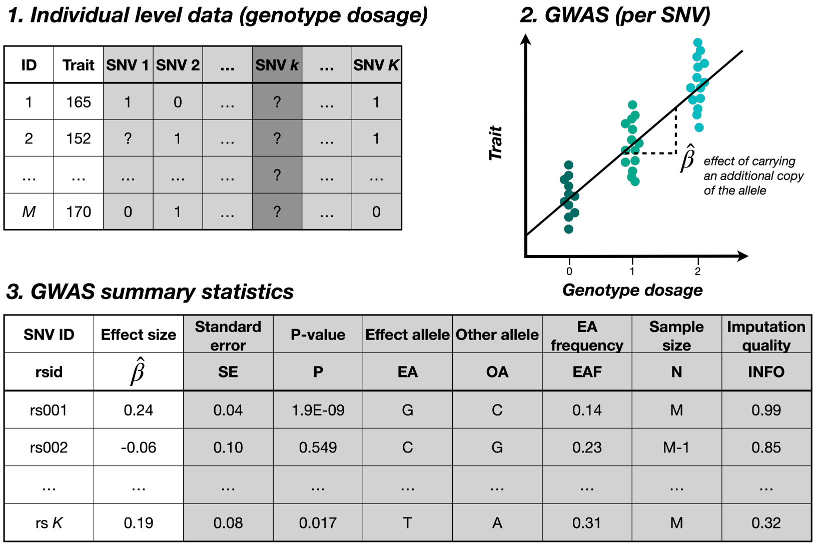
- $\hat{\beta}$, the effect size estimate for the effect allele
- SE of $\hat{\beta}$, the uncertainty if the $\hat{\beta}$ estimate
- P-value, the probability of getting an estimate as extreme as what has been observed, if the null hypothesis (the variant has an effect size of 0) was true.

These **summary test statistics** are further supplemented by other statistics such as the Z-score ($\hat{\beta}/SE$), and are often joined by variant information such as chromosome position, reference/effect and alternate/other alleles, allele frequencies in the studied population, sample size, and INFO imputation score if the variant is imputed[8].

Z-score or *t-statistic* is a statistical measure used to assess the significance of an association. It follows a t-distribution under the null hypothesis. However, with large enough sample sizes and when testing a single regression coefficient, the t-distribution can be well approximated by a standard normal distribution. The square of the t-statistic follows a chi-square distribution with 1 degree of freedom: $t^2 \sim \chi^2$

[7] *Population structure*: also known as *population stratification*, is the presence of a systematic difference in allele frequencies between subpopulations, often caused by non-random mating between groups. It is often affected by physical separation, migration, population bottlenecks and other similar effects

[8] *Imputation*: the process of inferring or predicting missing genetic information at specific positions in an individual's genome based on patterns observed in reference panels



1. Individual level data (genotype dosage)

ID	Trait	SNV 1	SNV 2	...	SNV *k*	...	SNV *K*
1	165	1	0	...	?	...	1
2	152	?	1	...	?	...	1
...	...	...	...	...	?	...	...
M	170	0	1	...	?	...	0

2. GWAS (per SNV)

3. GWAS summary statistics

SNV ID	Effect size	Standard error	P-value	Effect allele	Other allele	EA frequency	Sample size	Imputation quality
rsid	$\hat{\beta}$	SE	P	EA	OA	EAF	N	INFO
rs001	0.24	0.04	1.9E-09	G	C	0.14	M	0.99
rs002	-0.06	0.10	0.549	C	G	0.23	M-1	0.85
...	...	...	...	...	...	...	...	...
rs *K*	0.19	0.08	0.017	T	A	0.31	M	0.32

Figure 1.2: Individual-level genotype data to GWAS summary statistics. 1. Individual-level genotype data for *M* individuals across *K* SNVs. Note that for SNV *k* there is missing genotype data. **2.** Association summary statistic estimated per SNV. By regressing phenotypic values on genotypic information, we estimate the per allele effect size and standard error. **3.** Aggregated data, also called GWAS summary statistics, listing variant IDs, effect sizes, their standard errors, effect allele frequency, sample size amongst other variant information. Imputation scores are reported for SNVs with missing genotypes that are imputed using reference panels.

Thanks to many international efforts aimed at recruiting cohort volunteers, and collecting and documenting their data, GWAS across many phenotypes can be carried out with their resulting summary statistics shared publicly. One of the largest and most well-known biobanks is the **UK Biobank** [17], which has genetic and health information from over 500'000 participants. It includes extensive phenotypic data, such as medical history, lifestyle factors, and imaging data, making it valuable for a wide range of research. Summary statistic of the UK Biobank has been calculated and shared publicly to the benefit of many researchers (including myself) by the Neale Lab [18].
Based in Iceland, **deCODE Genetics** [19] has collected genetic and health data from a significant portion (> 50%) of the adult Icelandic population. Focused on the Finnish population, **FinnGen** [20] aims to collect genetic and health data from 500'000 participants. The project aims to advance understanding of diseases and treatment responses by combining genetic data with electronic health records. Other consortia have been made to combine data from different worldwide cohorts focusing on specific traits such as the **Psychiatric Genomics Consortium** [21] that focuses on the genetics of psychiatric diseases, the **GIANT consortium** [22] with various anthropometric traits and the **Global Lipids Genetics Consortium** [23] that is dedicated to the study of quantitative lipid traits to name a few.

Phenotype associations are often conceptualised across two dimensions (illustrated in Figure 1.3); allele frequency and effect size or penetrance. GWAS findings, thus far, have been associations of common variants with small to moderate effect sizes. More rare variants are difficult to identify especially if

For an outlook on GWAS and its progress, refer to *15 years of GWAS discovery: Realizing the promise* by Abdellaoui et al. [25]

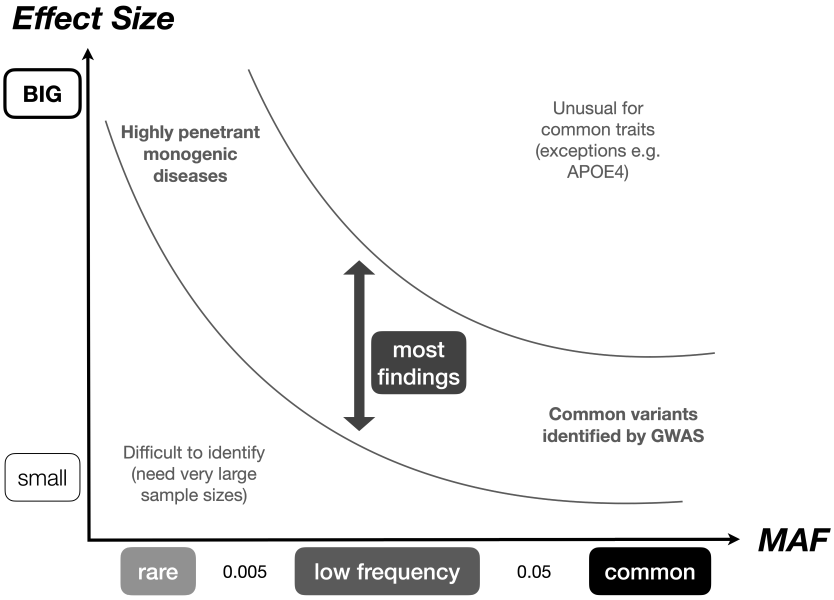

Effect Size
BIG
small
Highly penetrant monogenic diseases
Unusual for common traits (exceptions e.g. APOE4)
most findings
Common variants identified by GWAS
Difficult to identify (need very large sample sizes)
MAF
rare
0.005
low frequency
0.05
common

Figure 1.3: Schematic representation of allele frequency vs. effect size. The proposed relationship between allele frequency and effect size (or penetrance) based on our genetic understanding of complex traits and disease. Most GWAS findings of complex traits and common diseases correspond to common variants with relatively small effect sizes/low penetrance. Detecting rare variants with low penetrance requires large sample sizes to distinguish these rare variants and to have enough power to estimate their effects. Monogenic/mendelian diseases on the other hand, have variants with high penetrance despite being rare.
Adapted from Bush and Moore [24].

they have low penetrance as larger sample sizes are required to (i) observe these variants, and (ii) have enough statistical power for their association. Conversely, some rare variants have large penetrance/effect sizes and these are associated to Mendelian disorders. It is unusual to find common variants with large effects, however, some exceptions include the APOE4 variant of the APOE gene, where possessing one or two copies of the APOE4 allele significantly increases the risk of developing Alzheimer's disease.

In a Genome-Wide Association Study (GWAS), we estimate the observed association for a *single* genetic variant, which is also known as the **marginal effect**. This estimate reveals the relationship between a specific genetic variant and a phenotype of interest, considering that variant in isolation, without accounting for other factors. On the other hand, the **true causal effect** represents the authentic cause-and-effect relationship between a genetic variant and a phenotype. It implies that changes in the genetic variant directly lead to alterations in the phenotype. However, establishing true causality is a rigorous process that involves a combination of statistical evidence, biological knowledge, and experimental validation.
Conversely, a joint model takes into account multiple genetic variants simultaneously, often within the same genomic region or pathway, and assesses their combined effects on the outcome, providing a **multivariable effect** for each variant.

In cases where the genotyped variant in a GWAS is not in LD with any other causal variant, the marginal effect is equal to the

causal effect. This causal effect could be null if the genotyped variant is not truly causal. On the contrary, if the genotyped variant is in LD with other causal variants, the marginal effect combines the causal effect of that SNP with the causal effects of all other neighbouring variants, weighted by their LD score.
This situation presents a challenge when the SNP in question is non-causal but is in LD with one or more causal variants. In such cases, its marginal effect is estimated as non-zero, even when there is no true association with the phenotype.
Similarly, for a genotyped SNP that is genuinely causal, its marginal effect can be overestimated if it tags other causal SNPs in the neighbouring region. In contrast, its multivariable effect is more accurately estimated due to the nature of the joint model, which considers the SNP in the context of other variables, such as covariates or other genetic variants. It's worth noting that these joint models require more computational resources and are, therefore, less commonly used.

Another feature of genetic architecture is **heritability** [26]. In a particular population, heritability measures the proportion of **phenotypic variance** that is explained by **genetic variation** between individuals (including additive and non-additive genetic effects, such as dominant and epistatic effects). Heritability can vary across populations and across time, as the roles of the environment and genetics change.
Broad-sense heritability (H^2**)**, where both additive and dominance genetic effects are considered in relation to phenotypic variation, is typically estimated from twin or family studies. **Narrow sense heritability (**h^2**)**, in contrast, is the phenotypic variance explained by additive effects only. Furthermore, heritability estimated from GWAS top hits is often referred to as h^2_{GWAS}.

Phenotypic variance can be decomposed into *genetic variance*, *environmental variance*, and their *interaction*

The discrepancy between H^2 observed in family studies and h^2_{GWAS}, termed *missing heritability*, can be attributed to multiple reasons including GWAS neglecting non-additive effects (gene-by-gene interactions, gene-by-environment interactions), or other genetic variation such as rare variants or copy number variations

The two main methods to estimate narrow-sense heritability using individual-level data are **GCTA** [27] and **LDAK** [28, 29]. These methods both model all SNVs simultaneously using linear mixed effect model to estimate the total explained variance. They also both account for genetic relationships among individuals: GCTA constructs a GRM to capture genetic relatedness between individuals, while LDAK uses LD-pruned relatedness matrices. However, they differ in that GCTA expects each SNP to contribute equally to heritability, whereas LDAK assumes that the expected heritability of each SNP to vary with LD levels.
The most common method to estimate heritability from readily available summary statistics is called **LD-score regression (LDSC)** [30]. LDSC uses the patterns of LD among genetic vari-

ants within a population to estimate the overall heritability of a trait. By fitting a regression model of the effect sizes of genetic variants against their LD scores, LDSC provides an estimate of the heritability of the trait based on the genetic architecture captured in the summary statistics.

Another concept that is fundamental in genetics is **genetic correlation (r_G)**. It is a statistical measure that quantifies the degree of shared genetic influences between two different phenotypes. Consequently, it can provide insights into the underlying biological relationships between traits and can help us understand the genetic architecture of complex traits and diseases.
It is often computed using GWAS summary statistics for both traits, where the correlation coefficient between the effect sizes of genetic variants associated with each trait is calculated.
LDSC [30] is also used to calculate genetic correlation by leveraging patterns of LD among genetic variants. More specifically, using summary statistics of both traits, LDSC estimates r_G by regressing the product of the observed effect sizes across variants for both traits, against their cross-trait LD score (see Figure 1.4). This helps to disentangle the shared genetic components from other confounding factors of the two traits. It is important to note that although a genetic correlation indicates a statistical association between traits at the genetic level, it does not provide information about the specific genes involved or the direction of causality between them.

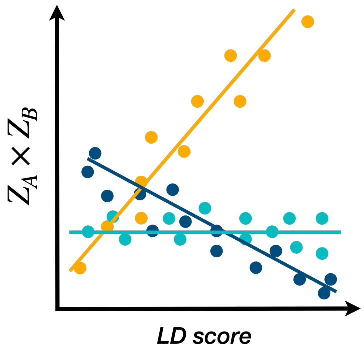


Figure 1.4: Illustration of the genetic correlation between traits *A* and *B*. Here, the product of the standardised effects/Z-score of traits A and B (Y-axis) is regressed onto the LD score (X-axis). The genetic correlation can be derived from the regression slope. The yellow slope shows an example of two positively correlated traits, the dark blue slope shows an example of anti-correlated traits, and cyan, an example of uncorrelated traits.

In conclusion, the growing number of consortia, the existence of semi-publicly accessible large-scale genetic data and biobanks, as well as the increase in publicly available GWAS summary statistics has made a variety of multi-trait analyses possible and increased our understanding of the genetics of complex traits.

1.3 Biases in GWAS: how are they (thus far) addressed?

GWAS are powerful tools for identifying genetic variants associated with complex traits and diseases. However, they can be prone to various sources of bias that might lead to falsely estimated association effects with misleading or inaccurate downstream results.
These biases arise from multifaceted sources, spanning study design, population characteristics, genotyping methods, and statistical approaches. By understanding and accounting for these biases, we can ensure the robustness and reliability of GWAS findings while making informed interpretations of genetic associations and downstream applications [31–34].

The most common source of bias is **population stratification**, as previously mentioned, this occurs when the study population has sub-populations with different genetic backgrounds either due to non-random mating[9] or migration amongst other physical causes. If not accounted for properly, it can lead to spurious associations between genetic variants and phenotypes. **Population admixture**[10] can cause false associations if not properly controlled for. Furthermore, **cryptic relatedness** (when individuals in the study are related to each other but this relationship is not apparent) leading to shared genetic factors can cause inflated association signals and false positive associations.
On the other hand, **dynastic effects** (or familial correlations), occur when shared genetic and environmental factors within families lead to correlations between relatives' traits, which in turn can lead to inflated effect sizes and false positive associations.
In case-control cohorts, **unequal sample sizes** or **systematic differences** in the ascertainment between cases and controls can lead to biased associations. Genotyping design also comes with its sources of bias, such as non-random missing genotype data, biased SNP selection, or simple genotyping errors.

Mitigating bias in GWAS requires a comprehensive approach, from careful study design to meticulous data analysis. Some techniques have been previously discussed, such as adjusting for population substructure using **principal component analysis (PCA)**, or using linear mixed models to account for relatedness or admixture.
However, there is increasing interest in methods that make use of relatedness and shared genetic factors within families to better understand genetic components and interactions. These studies are known as **family-based GWAS** [35], and involve

[9] *Assortative mating*: a mating pattern where individuals with similar phenotypes or genotypes mate with one another more frequently than would be expected under a random mating pattern

[10] *Admixture*: a phenomenon that occurs when individuals from two or more previously isolated populations interbreed, resulting in the introduction of new genetic lineages into a population

recruiting and studying families, which typically include parents, siblings, and sometimes extended relatives. These family structures provide a platform to investigate genetic influences within a shared genetic background.

Family members share a significant portion of their genetic makeup due to inheritance (i.e. similar ancestry). This relatedness enhances the method's ability to identify (rare) genetic variants associated with traits and diseases, as it reduces genetic heterogeneity and the design inherently accounts for population stratification.

Trios consist of two parents and an (affected) offspring

Family based designs include **parent-offspring trios**, which provide an opportunity to analyse the inheritance patterns of genetic variants from parents to offspring and assess their association with the trait, or **sibling-pairs (sib-pair)** analysis that involves comparing the genetic similarity between siblings who share the same parents. This method helps identify genetic markers that segregate with a trait in affected sibling pairs. Family-based studies however, tend to have smaller sample sizes compared to population-based studies which can affect their statistical power [36].

TDT provides a statistic, usually a chi-square test, to assess the deviation from expected transmission

FBAT calculates differences in allele transmission from parents to affected and unaffected offspring within each family

Common statistical methods used in family-based GWAS are the **Transmission Disequilibrium Test (TDT)** and its extension, the **family-based association test (FBAT)** [35, 37]. TDT is often used in linkage analysis of parent-offspring trios (consisting of affected offspring and their parents), where the co-segregation of a chromosome region marked by SNPs with a trait is tested. FBAT, on the other hand, is a more general family-based method that can handle various family structures, including trios, sibships, and extended pedigrees. It assesses the association between allele transmission and the trait of interest while accounting for within-family correlation and covariates.

Other family-based designs employ within-family tests to either adjust for or leverage parental genotypes, such as using **structural equation modelling** (SEM)[11] to estimate maternal and offspring genetic effects [38].

[11] *Structural equation modelling (SEM)*: a multivariate statistical analysis technique that is used to analyse structural relationships, by modelling said relations between measured and latent variables, or between multiple latent variables

Despite the stated advantages of within-family designs, the adoption of such approaches in contemporary genetic studies has remained limited. This limitation largely stems from the scarcity of genomic data gathered from families at a scale sufficient for suitably powered analysis.

However, the advent of large-scale biobanks and extensive twin studies has allowed researches to combine said data for extensive analysis. For example, the largest within-sibship GWAS conducted to date by Howe et al. [39] combined the data of 178'076 siblings from 19 studies and estimated associations across 25 different phenotypes.

Typically, within-sibship GWAS model the outcome as a function of the genetic variants, the within-pair effect (captures genetic variation that contributes to differences within sibling pairs), the between-pair effect (similar, but between sibling pairs instead), covariates and noise or shared environment variables. In Howe et al., they extended the population based GWAS model to include the mean genotype of siblings in each sibship as a covariate to account for family structure, as shown below: For individual i in sibship w with n siblings,

$$\begin{aligned} &Y_{iw} \sim G^C_{iw} + G^F_i + age_{iw} + sex_{iw} + PC1_{iw} + \ldots + PC20_{iw} \\ &\text{where} \quad G^F_i = \frac{\sum_1^n G_{iw}}{n} \quad \text{, and} \quad G^C_{iw} = G_{iw} - G^F_i \end{aligned} \tag{1.4}$$

G_{iw} represents the genotype of sibling i in sibship w, and G^F_i the mean family genotype for sibship w over n siblings.

By centring each individual's genotype around the mean sibship/family genotype (G^C_{iw}), the model estimates the **direct individual genetic effect**[12] as well as the **indirect genetic effect**[13] independently.

Their findings suggested that GWAS results and downstream analyses of behavioural phenotypes (e.g. educational attainment, smoking) and some anthropometric phenotypes (e.g. height, body mass index) are affected by demographic and indirect genetic effects. However, most analyses of molecular phenotypes, such as lipids, were not strongly affected.

Despite the fact that estimating direct genetic effects on traits is the principal goal of GWAS, other sources of genetic associations can be extremely informative for our understanding of the complexity of traits and their interplay. Knowledge of indirect genetic effects can be used to elucidate maternal effects, or the extent to which diseases are mediated by our environments [38, 40]. Thus, there is a need for family-based GWAS that also provide estimates of indirect genetic effects, along the unbiased direct ones.

[12] *Direct genetic effect*: direct effect of inheriting a genetic variant or a correlated variant on the expression of a specific trait

[13] *Indirect genetic effect*: effects of relative genotypes (via relative phenotypes and shared environment) on the individual's phenotype

1.4 Beyond association and correlation: Causation

One of the main goals of epidemiology is the discovery and study of risk factors that are causal for common complex diseases affecting our public health. This is commonly done through **observational analysis**, where you observe a certain risk factor and study its effect on a disease or outcome of interest. However, observational analyses are prone to bias as the observed association between the risk factor and the outcome can be caused by a **confounder**[14] of this relationship or by **reverse causation** from the outcome onto the risk factor, e.g. lower lipids seem to increase cardiovascular disease when in fact, cardiovascular medication (a consequence of the disease) lowers lipid levels.

One way to overcome this bias is by introducing randomisation, as is done in **Randomised Control Trials (RCTs)** [41]. RCTs are the gold standard to estimate causality between trait pairs as they randomly allocate trial participants to two or more groups and introduce a suspected risk factor or intervention to one (or more) group(s) while keeping the other as control. After a follow up time of exposure, any measured difference in the outcome between the groups is said to be solely caused by the exposure or risk factor.

For a thorough review on causal inference and MR, see Sanderson et al. [42]

A disadvantage of RCTs however, is that they are time-consuming by design, are often costly and could pose ethical challenges [43]. So we turn to an alternative that has emerged from genetic data, termed **Mendelian Randomisation (MR)** [42, 44]

MR parallels RCTs in the following way:

- Instead of a sample of trial participants, MR assumes that a population is undergoing the experiment
- The randomisation of trial participants is replaced by the random segregation of alleles during meiosis at conception
- The follow-up time of exposure in RCTs is represented by the traits in the population that associate to different alleles in individuals
- In MR, the measurement of the outcome difference between different groups is measured at any time point in the population instead of at the end of the trial
- Similarly however, the two methods do estimate the significant difference between the various exposure groups and their outcomes

To give an example of an MR study, we turn to the Aldehyde dehydrogenase 2 gene (*ALDH2*) located on chromosome 12. This

[14] *Confounder*: a variable that influences both the dependent variable and independent variable, causing a spurious association between them

gene encodes an enzyme that is of the major oxidative pathway of alcohol metabolism, and a mutation in its sequence leads to the inactivity of said enzyme. In turn, alcohol metabolism is disrupted, leading to consequences of varying severity such as facial flushing, nausea, and asthma bronchoconstriction. This tends to reduce the alcohol drinking in the population with the mutation, naturally creating a binary level of exposure: alcohol consumption.
At a given time point we can measure the overall cardiovascular health in the population (blood pressure, cholesterol levels) and study if there is a strong difference in the disease status between the two exposure groups, indicating a true causal effect of alcohol consumption on cardiovascular health. From several concordant studies of this example [45, 46], one MR study revealed a causal effect of alcohol on cardiovascular health, where a 1 SD increase in genetically predicted alcohol consumption was associated with 1.3-fold (95% CI [1.2-1.4]) higher risk of hypertension and 1.4-fold (95% CI [1.1-1.8]) higher risk of coronary artery disease [47].

MR is a statistical method that uses the genetic landscape of the population to discover causal relationships between modifiable traits and outcomes. It gets around the confounding and reverse causality biases seen in observational analysis by utilising genetic variants as **instrumental variables (IVs)** that robustly associate with the exposure of interest, thus any confounding occurring between the exposure and outcome trait (e.g. sex, age, environmental factors) is independent of the genetic variant, and there is no reverse causality from the outcome trait back to the variant in the human germ line [44].
There are several assumptions required for unbiased MR casual effect estimation [44], the principle three illustrated in Figure 1.5 are:

- The **relevance assumption**, where the IV is robustly associated to the exposure trait
- The **exchangeability assumption**, stating that there is no confounder of the IV and the outcome trait pair (examples include: population stratification, assortative mating)
- The **exclusion restriction assumption,** stating that there is no pathway or association between the IV and the outcome except through the exposure.

Other MR assumptions or considerations , principally used for accurate effect estimation, include: **effect homogeneity** which assumes either the association between the genetic IV and the exposure is homogeneous across the population or that the effect of the exposure on the outcome is; **gene–environment**

equivalence which states that perturbing the exposure genetically or environmentally, should produce the same downstream effect on an outcome; and **effect linearity**, where the effect of the exposure on the outcome is assumed to be linear.

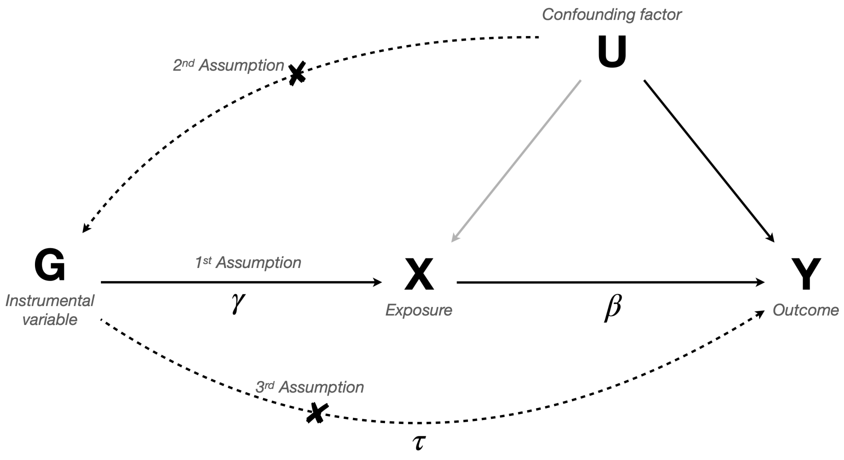


Figure 1.5: Core assumptions of Mendelian randomisation. Genetic marker G representing an IV has a direct effect on exposure/risk factor X denoted as γ. It also has an effect on outcome/disease Y through X. The measured effect of G on Y is denoted as τ. U represents potential confounding factors such as population stratification or sex. The dashed lines illustrate potentially violated MR assumptions. β represents the MR causal effect.
Adapted from Bowden, Davey Smith, and Burgess [48].

The simplest and most common way to estimate the causal effect of an exposure trait on an outcome, is to use the **Wald Ratio method** [49], given an IV that is robustly associated to the exposure trait, and is not associated to the outcome except through its association with the exposure.
The Wald ratio causality estimate (denoted by $\hat{\beta}$ below) is calculated by dividing the coefficient from regressing the outcome onto the IV i (effect size estimate from the association summary data, $\hat{\tau}$), by the coefficient of regressing the exposure onto the same IV ($\hat{\gamma}$).

$$\hat{\beta}_i = \frac{\hat{\tau}_i}{\hat{\gamma}_i} \quad , \quad \text{with } SE = \frac{SE(\hat{\tau}_i)}{\hat{\gamma}_i} \tag{1.5}$$

For multiple IVs, you can obtain the overall causal effect by performing an **inverse variance weighted meta-analysis** on their individual Wald ratio estimates, this method is known as IVW-MR [50]. It is important to ensure that the exposure and outcome data are harmonised to ensure that the $\hat{\gamma}$ and $\hat{\tau}$ association refer to the same alleles.

MR studies can be conducted either on individual level genetic data or summary data commonly obtained from GWAS. Analysis conducted on individual level data offers more estimate precision as the sample size increases, and is commonly known as **one-sample MR**, as the genetic and phenotypic data of both exposure and outcome come from a single cohort. Causal effect estimates in one-sample MR are obtained from a **two-stage least square regression (2SLS)**, where robust IVs are selected and used to obtain a prediction of the exposure in the first stage, followed by regressing the outcome onto the predicted exposure in the second stage [44].

Similarly MR analysis conducted on summary statistics data depend on the association accuracy between the IV and the traits of interest, and are known as **two-sample MR**. Here, GWAS summary statistics of the exposure and the outcome come from distinct cohorts, but are assumed to be from the same underlying population or similar populations. Furthermore, it is assumed that there is no sample overlap between the two cohorts to ensure lack of bias due to over-fitting [51, 52].

Over-fitting refers to exaggerated/inflated results occurring when MR studies are done within the same GWAS from which the genetic IVs are selected

For the analysis performed in my thesis, I used GWAS summary data as they are more widely available for public use, and thus will continue all MR explanation and analysis assuming summary data is used.

The first MR assumption can be easily verified thanks to the increasing number of GWAS studies and their ever-increasing sample size, where SNPs with **genome wide significant (GWS)** associations ($P \leq 5 \times 10^{-8}$) to traits of interest can be selected as IVs. However, if the sample used to select the GWS IVs is the same from which the association summaries to the exposure are obtained, this could exaggerate the true IV-exposure association in what is termed as **Winner's curse**, resulting in an underestimation of the causal effect in the case on non-overlapping samples [52]. Employing a three-sample genome wide design, where you would use a selection GWAS dataset for IV selection, separate from the exposure GWAS dataset, thereby circumventing the Winner's Curse issue is a valid approach. However, summary statistics from such additional datasets are rarely available.

When multiple SNPs are used as IVs, each with its own small effect on the exposure, this can lead to **weak instrument bias** [52], which increases in severity as the average variance of the exposure explained by the IVs decreases. The resulting underestimation of the causal effect, also in the case of non-overlapping samples, can be avoided through bias correction calculations, using a two-sample approach and applying a 2SLS regression, or by ensuring that we use proper IVS by measuring instrument strength using the **F-statistic** [53].

The F-statistic is related to the proportion of phenotypic variance explained by the genetic variants, sample size, and number of instruments. The bias can thus be reduced by increasing the sample size or by excluding instruments that are not contributing to the explanation of the phenotypic variance of the exposure, both of which increase the F-statistic. As a rule of thumb, an F-statistic of 10 or higher is needed to run an analysis with no weak instrument bias.

It is worth noting that both these types of biases are affected by the degree of sample overlap; when using overlapping samples, the causal effect estimate will be biased towards the observational correlation instead. Although these sources of bias have previously been studied independently, a recent study has proposed a new two-sample MR framework, termed MRlap, that simultaneously takes into account weak instrument bias and winner's curse, while accounting for potential sample overlap and its effect as a modifier of these biases, to obtain a corrected causal effect estimate [54].

Unfortunately, the second and third assumptions are difficult to verify, however, they can sometimes be disproved. Any attempt and subsequent failure to disprove them can be interpreted as validation of these conditions.

Although, as stated earlier, MR attempts to overcome possible confounding seen between trait-pairs in observational analysis by using genetic variants as instruments for the exposure trait they associate to, it can still be prone to confounding between the outcome and the IV instead. Sources of this confounding vary between population stratification, assortative mating, and dynastic effects and are difficult to correct for with the current MR methods.

Similarly, the third assumption can be violated by **pleiotropy**[15], where genetic instruments are associated to multiple traits [55]. This phenomenon can be seen in many GWAS, where a single variant is often associated with multiple phenotypes. However, in the context of MR, certain types of pleiotropy can violate the assumptions and introduce a correlation between the trait pair, that is often falsely interpreted as causation as seen in Figure 1.6. There are multiple MR methods and extensions that aim to overcome this source of bias, including two developed over the course of my studies.

An important assumption of IVW-MR is that the genetic instruments are independent of each other, which can be verified using LD based clumping of IVs. Second, IVW assumes balanced pleiotropy, meaning that it has zero mean and satisfies the **Instrument Strength Independent of Direct Effect (InSIDE)** assumption where the direct pleiotropic effects of the genetic variants on the outcome are distributed independently of the genetic associations with the exposure [48].

Numerous MR extensions, some outlined in Table 1.1, have been developed to address the diverse forms of pleiotropy, each grounded in distinct assumptions regarding the causal nature of this pleiotropy. These extensions mainly rely on three strategies:

[15] *Pleiotropy*: in its true biological definition, it is the association of a single variant to multiple traits. However other types exist as seen in Figure 1.6, such as:
Vertical pleiotropy: occurs when a genetic variant influences both the exposure of interest and the outcome being studied.
Horizontal/Correlated pleiotropy: arises when a genetic variant influences the outcome through pathways unrelated to the exposure, either through LD with an instrument directly affecting the outcome, or by association to a latent heritable variable or confounder of the exposure and outcome traits.
Mediated pleiotropy: occurs when a genetic variant affects a trait, which in turn influences another trait that affects the outcome.

excluding outliers, modifying outliers, and accommodating specific forms of pleiotropy.

Outlier removal estimation involves recognising and eliminating individual genetic variants whose sole causal effect estimate falls beyond the expected range based on estimates from other variants, thus mitigating their impact on the final total causal effect. Common methods that employ outlier removal include **weighted median** [56] (allows for balanced/sparse pleiotropy) and **weighted mode** [57] (allows for some directional pleiotropy).

Outlier adjustment methods identify outlier variants, and then make adjustments either to the effect estimate of that genetic variant or to the weighting of the estimate from that variant, reducing its influence on the overall estimation outcome. Such methods include **MR-RAPS** [58] (allows for balanced pleiotropy) and **MR-CAUSE** [59], **MR-PRESSO** [60] or **MR-TRYX** [61] (allow for directional pleiotropy).

The last category of pleiotropy-robust methods for summary-data MR estimation encompasses approaches that permit most or all of the genetic variants used in the estimation to exhibit pleiotropic effects on the outcome while imposing additional constraints on these pleiotropic effects. Such an extension that handles directional pleiotropy (when the mean of the pleiotropic distribution is non-zero) is called **MR-Egger** [48]. The MR-Egger regression model estimates both the causal effect and a measure of directional pleiotropy. Another example is **MultiVariable MR (MVMR)** [62, 63].

MVMR is an extension that simultaneously estimates the causal effect of several exposure traits on a single outcome conditional on each other. It accounts for horizontal pleiotropy by adjusting the causal effect estimate of one trait based on several other candidate pleiotropic traits, thus it is able to discern between risk factors, determining which ones are causal and which ones are merely correlated, mediating, or confounding factors. However, MVMR is limited by its assumption that all pleiotropic traits are known and are fitted in the model.

Furthermore, MVMR utilises the IVs of all fitted exposures together, and thus a conditional F-statistic for each exposure ought to be calculated to ensure that the IVs being used will not affect the results through weak instrument bias [53].

Other MR extensions include non-linear MR methods that estimate the non-linear relationship between exposure-outcome pairs (e.g. LACE [64] and polyMR [65]) and multiple-outcome methods (e.g. MR2 [66]), that are designed for multiple outcomes, in order to identify exposures that cause more than one

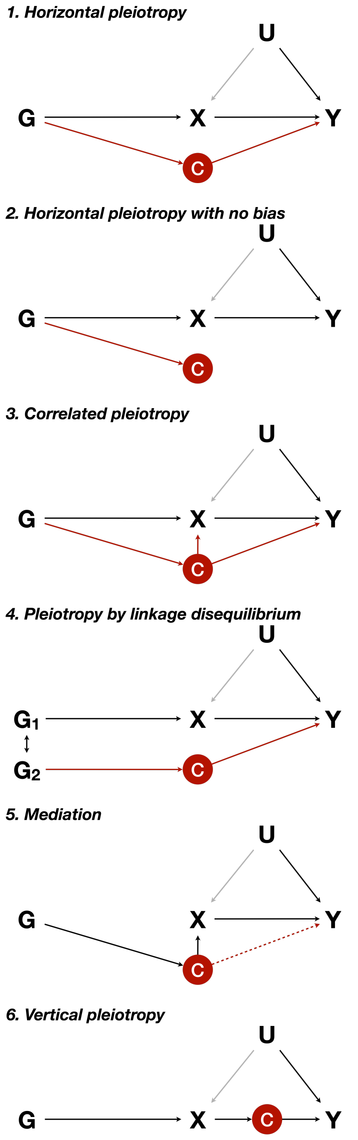


Figure 1.6: The different types of pleiotropy. G represents an instrumental variable with a direct effect on exposure/risk factor X. It can also have an effect on outcome/disease Y through C, a latent heritable trait shown in red. U represents potential confounding factors such as population stratification or sex. The dashed arrows illustrate potential relationships between traits, whereas red arrows illustrate relationships that will bias the causal effect estimation between X and Y.

outcome or exposures that have effects on distinct responses.

Table 1.1: Table of some MR methods and extensions that handle various assumption violations by relaxing the targeted assumption. PWC-MR and LHC-MR are discussed further in the next chapter.

Method	Action	Relaxed MR assumption
Wald ratio (both individual- and summary-level data), IVW, 2SLS regression (individual-level data)	Basic MR	None
MR-RAPS, NOME adjustment [67]	Weak instrument robust methods	1st: allows for weak instruments
Weighted median	Variant selection/Outlier removal	3rd: allows for balanced/sparse pleiotropy
Weighted mode, Steiger filtering [68], MR-LASSO [69], MR-Clust [70], **PWC-MR** [71]	Variant selection/Outlier removal	3rd: allows for some directional pleiotropy and effect heterogeneity
MR-TRYX, MR-RAPS, MR-CAUSE, **LHC-MR** [72], MR-PRESSO	Variant/Outlier adjustment	3rd: allows for some directional/balanced pleiotropy
MR-Egger, MVMR	Estimation adjustment	3rd: allows for some directional pleiotropy

As the field of MR continues to evolve, I hoped to contribute, with my thesis work shown in the following chapters, to the development of new methods that aid in utilising or overcoming MR assumption violations.

My main focuses were on *(i) violations of the third assumption*, whether in the form of horizontal/correlated pleiotropy that I tackled in **Chapter 2**, or in the form of causal effect heterogeneity as shown in **Chapter 3**, and *(ii) violations the second assumption* in the form of dynastic effects. Although, there is a dearth in family-based GWAS data, we attempted to better understand and study dynastic causal effects using indirect genetic effects estimated from first-degree relatives, explored in **Chapter 4**. These chapters along with **Chapter 5** also detail additional works and contributions that originated through either external or internal collaborations.

Chapter 2 2

2.1 Accounting for latent heritable confounding in Mendelian Randomisation

The article, *Simultaneous estimation of bi-directional causal effects and heritable confounding from GWAS summary statistics* (Darrous, Mounier, and Kutalik (2021) - see Appendix A), proposed a Latent Heritable Confounder MR (**LHC-MR**) method that can overcome some limitations of other MR extensions, including the under-exploitation of genome-wide markers, and sensitivity to the presence of a heritable confounder of the exposure-outcome relationship.

LHC-MR extends the typical MR model by accounting for the presence of a LHC with effects on the exposure and outcome traits. This is done through a **structural equation model (SEM)** that models the bi-directional causal effects between the two traits (X and Y), and the confounder (U) effect on each.
LHC-MR optimises the likelihood function associated with the SEM, taking as input observed genome-wide association summary statistics for X and Y, in order to **simultaneously estimate the bi-directional causal effect, the confounder effect, the heritability and polygenicity of each trait**, as well as several other trait characteristics. As the method uses all genome-wide markers instead of GWS instruments, LHC-MR is not affected by winner's curse nor weak instrument bias. LHC-MR also accounts for the LD structure amongst the variants as well as sample overlap, and can be viewed as the integration of LDSC and classical MR.

We compared the performance of LHC-MR against various standard and robust MR methods in multiple simulation scenarios where we violated standard MR assumptions, as well as our own (e.g. presence of multiple confounders despite modelling only one, normality assumption of SNP effects).
In the majority of scenarios, LHC-MR estimated causal effects with less bias and variance than other MR methods, even in the presence of a heritable confounder. Furthermore, LHC-MR was also immune to the presence of a reverse causal effect with an opposite effect sign, or the presence of more than one discordant or concordant confounders.

2.1 Accounting for latent heritable confounding in Mendelian Randomisation 21
2.2 R package: `lhcMR` . . . 23
2.3 Estimating the causal effect of stratified physical activity on cognitive functioning 24

nature COMMUNICATIONS

ARTICLE

Check for updates

https://doi.org/10.1038/s41467-021-26970-w OPEN

Simultaneous estimation of bi-directional causal effects and heritable confounding from GWAS summary statistics

Liza Darrous [1,2,4], Ninon Mounier [1,2,4] & Zoltán Kutalik [1,2,3✉]

Mendelian Randomisation (MR) is an increasingly popular approach that estimates the causal effect of risk factors on complex human traits. While it has seen several extensions that relax its basic assumptions, most suffer from two major limitations; their under-exploitation of genome-wide markers, and sensitivity to the presence of a heritable confounder of the exposure-outcome relationship. To overcome these limitations, we propose a Latent Heritable Confounder MR (LHC-MR) method applicable to association summary statistics, which estimates bi-directional causal effects, direct heritabilities, and confounder effects while accounting for sample overlap. We demonstrate that LHC-MR outperforms several existing MR methods in a wide range of simulation settings and apply it to summary statistics of 13 complex traits. Besides several concordant results with other MR methods, LHC-MR unravels new mechanisms (how disease diagnosis might lead to improved lifestyle) and reveals new causal effects (e.g. HDL cholesterol being protective against high systolic blood pressure), hidden from standard MR methods due to a heritable confounder of opposite effect direction.

[1] University Center for Primary Care and Public Health, University of Lausanne, Lausanne, Switzerland. [2] Swiss Institute of Bioinformatics, Lausanne, Switzerland. [3] Department of Computational Biology, University of Lausanne, Lausanne, Switzerland. [4] These authors contributed equally: Liza Darrous, Ninon Mounier. ✉email: zoltan.kutalik@unil.ch

NATURE COMMUNICATIONS | (2021)12:7274 | https://doi.org/10.1038/s41467-021-26970-w | www.nature.com/naturecommunications 1

Figure 2.1: The manuscript *Simultaneous estimation of bi-directional causal effects and heritable confounding from GWAS summary statistics* and its supplementary materials can be found on *Nature communications* here, or in Appendix A. This work was selected as one of SIB's remarkable outputs for 2021.

We then applied LHC-MR to 13 complex traits, estimating their pairwise bi-directional causal effects using summary statistics from the UK Biobank and other large consortia. Comparing our findings to those of other MR methods, we found a general agreement in causal effect estimates when both methods showed significant estimates.
Moreover, LHC-MR found additional significant estimates between traits pairs, as expected considering its use of genome-wide instruments. We also identified significant confounding effects between 16 trait pairs, including HDL cholesterol levels and systolic blood pressure (SBP). In this case, LHC-MR estimated a causal effect of HDL on SBP equal to -0.13 ($P = 5.38 \times 10^{-05}$) with a significant positive confounder acting on the two traits, concordant with observational studies [73, 74], whereas the standard MR methods showed a non-significant (attenuated) negative effect.
Lastly, LHC-MR decomposes the observed genetic correlation into bi-directional causal effect-driven and confounder-driven contributions. Our findings showed that the total genetic correlation estimates derived from LHC-MR were highly consistent with those obtained using LDSC, where most seem to be driven by bi-directional causal effects.

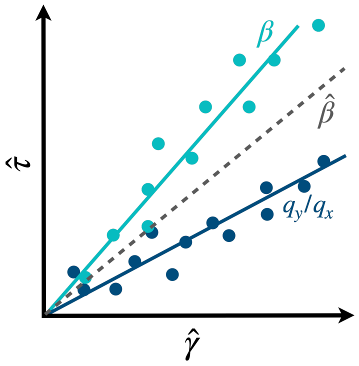


Figure 2.2: Illustration of causal effects estimated from SNP-outcome and SNP-exposure associations. The regression of the SNP-outcome association $\hat{\tau}$, onto the SNP-exposure association $\hat{\gamma}$ reveals in reality two separate SNP clusters, in the presence of a heritable confounder: those truly associated with the exposure with an effect on the outcome shown in light blue and having a causal effect β, and those that are primarily associated with the confounder having an effect equivalent to the ratio of the confounder's effect on outcome to exposure (q_y/q_x). Running standard IVW would result in an underestimated causal effect represented by the slope of the dashed grey line, taking into account all SNPs. LHC-MR aims to disentangle these separate slopes/effects.

As all methods do, LHC-MR has its limitations, some of which are: it provides biased causal effect estimates if the summary statistics used are affected by population stratification and dynastic effects (biases common to population-based GWAS). Also, LHC-MR's model is unidentifiable, meaning that the true causal slope (β) is sometimes indistinguishable from the confounder-associated slope (q_y/q_x), as explained in Figure 2.2. Thus two distinct sets of parameter estimates fit the input data equally well, especially if the alternate set of parameter estimates calculated fall within the parameter ranges specified. However, biological considerations and other pointers can aid in the choosing of the more likely set of estimates with the true causal effect.

LHC-MR is the fruit of two and a half years of labour, where its modelling has evolved several times over this course. This is in large thanks to reviewers' comments and the continuous discussions with peers and the greater scientific community. Our understanding of a plausible genetic structure, the varied aspects of pleiotropy, and our ever-evolving understanding of the interplay among these factors have significantly influenced the assumptions we make in our modelling.

This project was conceived and designed by Zoltán Kutalik (originally as a Master thesis project undertaken by Liza Darrous).

The mathematical and statistical derivations were performed by Zoltán Kutalik, and they were translated from Matlab to R by Liza Darrous. Liza Darrous and Ninon Mounier also contributed to the development of the approach, and to the implementation of the research. All three authors contributed to the analysis of the results and to the writing of the manuscript.

2.2 R package: `lhcMR`

As the methodology behind LHC-MR is not trait-specific, creating an R package that allows others to implement the method on any trait-pair they wish to investigate was a clear outcome of the project. The code for the R package `lhcMR` can be found on Github here.

As mentioned previously, `lhcMR` only requires the summary statistics of the traits being studied as input. We provide two additional files to be used as input; (i) the LD scores and regression weights of 4’650’107 common, high-quality SNPs, and (ii) the spike and slab distribution approximation of the local LD pattern using 2’500 SNPs left and right of each of the 4.65 million focal SNPs.
Since LHC-MR uses the R packages `TwoSampleMR` [75] to estimate standard MR causal effects, and `GenomicSEM` [76] to estimate trait heritabilities, these packages and their required files should also be installed.

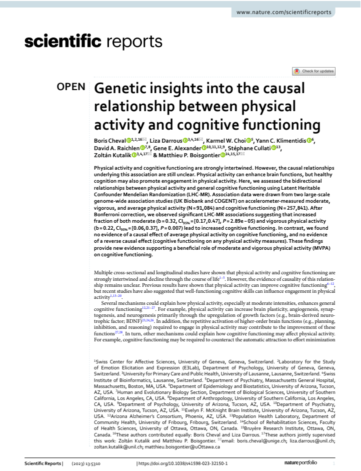
There are three main functions in `lhcMR`:

- `merge_sumstats()` reads in the summary statistics of the trait-pair and the LD score files, and merges the data into a single data frame with harmonised SNPs.
- `calculate_SP()` uses the previously generated data frame to smartly generate starting points using TwoSampleMR and GenomicSEM, for the parameter estimation in the trait-pair analysis done in the next step.
- `lhc_mr()` uses the input data frame and the stating points to optimise the likelihood function and estimate parameters such as the bidirectional causal effect, confounder effect and trait heritability, as well as their standard errors using block jackknife.

I am the main author of this R package with guidance and testing provided by Ninon Mounier and Zoltán Kutalik.

2.3 Estimating the causal effect of stratified physical activity on cognitive functioning

An **application of LHC-MR** to investigate the causal relationship between various levels of physical activity and cognitive functioning was the result of a collaboration between us, Boris Cheval from the University of Geneva and Matthieu Boisgontier from the University of Ottawa.
Since the relationship between these two traits is unclear - physical activity can enhance cognitive functions, but healthy cognition may also encourage engagement in physical activity - estimating the bidirectional causal effect while accounting for potential confounding using LHC-MR was undertaken in the manuscript *Genetic insights into the causal relationship between physical activity and cognitive functioning* (Cheval, Darrous et al. (2023) - see Appendix B).



www.nature.com/scientificreports

scientific reports

Check for updates

OPEN

Genetic insights into the causal relationship between physical activity and cognitive functioning

Boris Cheval[1,2,16], Liza Darrous[3,4,16], Karmel W. Choi[5], Yann C. Klimentidis[6], David A. Raichlen[7,8], Gene E. Alexander[10,11,12,9], Stéphane Cullati[13], Zoltán Kutalik[3,4,17] & Matthieu P. Boisgontier[14,15,17]

Physical activity and cognitive functioning are strongly intertwined. However, the causal relationships underlying this association are still unclear. Physical activity can enhance brain functions, but healthy cognition may also promote engagement in physical activity. Here, we assessed the bidirectional relationships between physical activity and general cognitive functioning using Latent Heritable Confounder Mendelian Randomization (LHC-MR). Association data were drawn from two large-scale genome-wide association studies (UK Biobank and COGENT) on accelerometer-measured moderate, vigorous, and average physical activity (N = 91,084) and cognitive functioning (N = 257,841). After Bonferroni correction, we observed significant LHC-MR associations suggesting that increased fraction of both moderate (b = 0.32, $CI_{95\%}$ = [0.17,0.47], *P* = 2.89e − 05) and vigorous physical activity (b = 0.22, $CI_{95\%}$ = [0.06,0.37], *P* = 0.007) lead to increased cognitive functioning. In contrast, we found no evidence of a causal effect of average physical activity on cognitive functioning, and no evidence of a reverse causal effect (cognitive functioning on any physical activity measures). These findings provide new evidence supporting a beneficial role of moderate and vigorous physical activity (MVPA) on cognitive functioning.

Multiple cross-sectional and longitudinal studies have shown that physical activity and cognitive functioning are strongly intertwined and decline through the course of life[1–5]. However, the evidence of causality of this relationship remains unclear. Previous results have shown that physical activity can improve cognitive functioning[6–12], but recent studies have also suggested that well-functioning cognitive skills can influence engagement in physical activity[1,13–20].

Several mechanisms could explain how physical activity, especially at moderate intensities, enhances general cognitive functioning[12,21–27]. For example, physical activity can increase brain plasticity, angiogenesis, synaptogenesis, and neurogenesis primarily through the upregulation of growth factors (e.g., brain-derived neurotrophic factor; BDNF)[23,24,26]. In addition, the repetitive activation of higher-order brain functions (e.g., planning, inhibition, and reasoning) required to engage in physical activity may contribute to the improvement of these functions[27,28]. In turn, other mechanisms could explain how cognitive functioning may affect physical activity. For example, cognitive functioning may be required to counteract the automatic attraction to effort minimization

[1]Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland. [2]Laboratory for the Study of Emotion Elicitation and Expression (E3Lab), Department of Psychology, University of Geneva, Geneva, Switzerland. [3]University for Primary Care and Public Health, University of Lausanne, Lausanne, Switzerland. [4]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [5]Department of Psychiatry, Massachusetts General Hospital, Massachusetts, Boston, MA, USA. [6]Department of Epidemiology and Biostatistics, University of Arizona, Tucson, AZ, USA. [7]Human and Evolutionary Biology Section, Department of Biological Sciences, University of Southern California, Los Angeles, CA, USA. [8]Department of Anthropology, University of Southern California, Los Angeles, CA, USA. [9]Department of Psychology, University of Arizona, Tucson, AZ, USA. [10]Department of Psychiatry, University of Arizona, Tucson, AZ, USA. [11]Evelyn F. McKnight Brain Institute, University of Arizona, Tucson, AZ, USA. [12]Arizona Alzheimer's Consortium, Phoenix, AZ, USA. [13]Population Health Laboratory, Department of Community Health, University of Fribourg, Fribourg, Switzerland. [14]School of Rehabilitation Sciences, Faculty of Health Sciences, University of Ottawa, Ottawa, ON, Canada. [15]Bruyère Research Institute, Ottawa, ON, Canada. [16]These authors contributed equally: Boris Cheval and Liza Darrous. [17]These authors jointly supervised this work: Zoltán Kutalik and Matthieu P. Boisgontier. ✉email: boris.cheval@unige.ch; liza.darrous@unil.ch; zoltan.kutalik@unil.ch; matthieu.boisgontier@uOttawa.ca

Scientific Reports | (2023) 13:5310 | https://doi.org/10.1038/s41598-023-32150-1 nature portfolio 1

Figure 2.3: The manuscript *Genetic insights into the causal relationship between physical activity and cognitive functioning* can be found on *Scientific Reports - Nature* here, or in Appendix B.

To run the analysis, we used association summary statistics of accelerometer-based average physical activity and cognitive functioning from two large consortia; COGENT [77] and UK Biobank [17]. We also ran our own GWAS to obtain association summary statistics of stratified measures of accelerometer based physical activity (moderate and vigorous) from individual level data of the UK Biobank.

LHC-MR findings suggested that moderate (0.32, $P = 2.89 \times 10^{-05}$) and vigorous physical activity (0.22, $P = 0.007$) lead to increased cognitive functioning, in line with previous findings of observational analysis [78–80].
However, LHC-MR found no evidence of a causal effect of average physical activity on cognitive functioning, and no evidence of a reverse causal effect (cognitive functioning on any physical activity measures). In comparison, standard MR methods found no significant causal effects between any trait pair in either direction, which may be primarily due to the fact that standard MR methods use only GWS SNPs as IVs, and that all physical activity measures had no GWS SNPs that could be used for the analysis (threshold was instead lowered to 6.33×10^{-5}) .
The findings highlight the essential role of engaging in moderate and vigorous physical activity to maintain or enhance overall cognitive function, which health policies and interventions can benefit from.

I contributed to this research by running the statistical analysis (GWAS, LHC-MR, standard MR methods), and writing of the manuscript (wherever relevant to causal inference, LHC-MR, or MR).

Chapter 3 | 3

3.1 PheWAS-based clustering of Mendelian Randomisation instruments 27
3.2 Application: clustering of obesity, a composite trait 29

3.1 PheWAS-based clustering of Mendelian Randomisation instruments

The article, *PheWAS-based clustering of Mendelian Randomisation instruments reveals distinct mechanism-specific causal effects between obesity and educational attainment* (Darrous, Hemani, Davey Smith, and Kutalik (2023) - see Appendix C), aims to investigate the mechanisms underlying heterogeneous causal effect estimates. The MR assumption of homogeneous causal effects can be violated when various underlying processes contribute to how a complex trait affects an outcome. Here, we seek to identify these mechanisms, whether they are different pleiotropic pathways or confounding effects, and estimate their contributions to the overall exposure effect.

Originally, we were motivated by the surprisingly large and negative **causal effect of body mass index (BMI) on educational attainment (EDU)** to investigate potential sources of bias between the trait-pair.

In addition to the improbability of BMI (a later-in-life measured trait) affecting EDU (an early-life trait), causal estimates from the Howe et al. within-sibship GWAS study [39] revealed a significantly attenuated causal effect of BMI on EDU compared to that obtained using population GWAS estimates of unrelated samples, which further motivated our investigation.

In order to discover the various possible pleiotropic pathways of BMI, we performed informative **K-means clustering** on the GWS BMI-associated SNPs using their association to ~400 other traits (from *PheWAS* data), in an approach we termed **PWC-MR**. This resulted in 6 clusters of BMI SNPs, each distinctly enriched for different traits and highlighting the complexity of BMI and its highly pleiotropic nature.

One cluster was strongly enriched for lean-mass traits such as 'Trunk predicted mass' and 'Arm fat-free mass', while another cluster was strongly enriched for socioeconomic position (SEP) related traits such as 'Age completed full time education' and 'Average total household income before tax'. Other clusters were enriched for food supplements, a mix of height-, blood-, and lung capacity measurement-related traits, or mixed traits (with

PheWAS-based clustering of Mendelian Randomisation instruments reveals distinct mechanism-specific causal effects between obesity and educational attainment

Liza Darrous[1,2,3,†], Gibran Hemani[4,5], George Davey Smith[4,5], and Zoltán Kutalik[1,2,3,†]

[1]University Center for Primary Care and Public Health, University of Lausanne, Switzerland
[2]Swiss Institute of Bioinformatics, Lausanne, Switzerland
[3]Department of Computational Biology, University of Lausanne, Lausanne, Switzerland
[4]Medical Research Council Integrative Epidemiology Unit, Population Health Sciences, University of Bristol, Bristol, United Kingdom
[5]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom
[†]*Correspondence should be addressed to darrous.liza@gmail.com or zoltan.kutalik@unil.ch*

Abstract

Mendelian Randomisation (MR) is a statistical method that estimates causal effects between risk factors and common complex diseases using genetic instruments. Heritable confounders, pleiotropy and heterogeneous causal effects violate MR assumptions and can lead to biases. To tackle these, we propose an approach employing a PheWAS-based clustering of the MR instruments (PWC-MR). We apply this method to revisit the surprisingly large apparent causal effect of body mass index (BMI) on educational attainment (EDU): $\widehat{\alpha}$ = -0.19 [-0.22, -0.16].

As a first step of PWC-MR, we clustered 324 BMI-associated genetic instruments based on their association profile across 407 traits in the UK Biobank, which yielded six distinct groups. The subsequent cluster-specific MR revealed heterogeneous causal effect estimates on EDU. A cluster strongly enriched for traits related to socio-economic position yielded the largest BMI-on-EDU causal effect estimate ($\widehat{\alpha}$ = -0.49 [-0.56, -0.42]) whereas a cluster enriched for primary impact on body-mass had the smallest estimate ($\widehat{\alpha}$ = -0.09 [-0.13, -0.05]). Several follow-up analyses confirmed these findings: (i) within-sibling MR results ($\widehat{\alpha}$ = -0.05 [-0.09, -0.01]); (ii) MR for childhood BMI on EDU ($\widehat{\alpha}$ = -0.03 [-0.06, -0.002]); (iii) step-wise multivariable MR (MVMR) ($\widehat{\alpha}$ = -0.06 [-0.09, -0.04]) where time spent watching television and past tobacco smoking (two proxies for potential confounders) were jointly modelled.

Through a detailed examination of the BMI-EDU causal relationship we demonstrated the utility of our PWC-MR approach in revealing distinct pleiotropic pathways and confounder mechanisms.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

1

Figure 3.1: The manuscript *PheWAS-based clustering of Mendelian Randomisation instruments reveals distinct mechanism-specific causal effects between obesity and educational attainment* and its supplementary materials can be found on *medRxiv* here. A revised version of the manuscript is found in Appendix C.

This manuscript has been submitted to *Nature Communications*, where it is under a second round of revision.

lower enrichment ratios).

Comparing the causal effect estimate of each cluster of BMI SNPs on EDU, to the overall causal effect obtained from all BMI SNPs revealed significantly heterogeneous estimates ranging from -0.5 to -0.09.
The results further revealed that the cluster with the most negative causal effect was the one enriched for SEP-related traits, whereas the cluster with the smallest causal effect was that enriched for lean-mass and body related traits. **We hypothesised that the SEP-related cluster was an example of correlated pleiotropy between BMI and EDU, and was thus biasing the true causal effect estimate towards it.**

We verified our findings by running several post-hoc analysis; the first was to re-run the clustering and subsequent causal inference on a trait that is proxying childhood BMI (under the assumption that adult SEP is less associated with childhood traits). We discovered 4 clusters in this case, none of which were strongly enriched for SEP-related traits, although one was enriched for body measurement-related traits. The causal effect estimates from the clusters were not significantly heterogeneous from the estimate obtained using all SNPs, which itself was significantly lower than that obtained using all adult BMI SNPs (-0.03, $P = 0.04$).
Secondly, using bi-directional MR, we used each of ~400 various traits as exposure and estimated its causal effect on BMI as the outcome once, and on EDU as the outcome another time (as seen in Figure 3.2). In doing so, we were able to find putative **confounder traits** by selecting those that had an effect on both BMI and EDU. Out of 19 such confounder traits, 3 survived stepwise-MVMR selection and were **arguably associated with SEP**: 'Time spent watching TV', 'Past tobacco smoking', and 'Muesli eating'. Their causal effect on EDU was simultaneously estimated with BMI using MVMR, revealing an attenuated conditional effect of BMI on EDU (-0.05, $P = 2.07\times10^{-5}$) matching those of our previous findings. We ran sensitivity analyses and compared our method against that of other clustering MR methods such as MR-Clust.

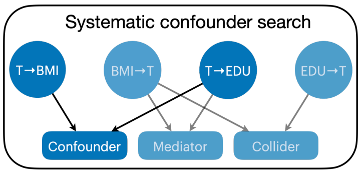


Figure 3.2: A simplified graph representation of a systematic confounder search. Around 400 traits were each used as an exposure (represented by T), and their bi-directional causal effect estimate on BMI and on EDU were separately estimated. Traits with significant causal effects on both BMI and EDU were labelled as putative confounder traits.

The advantages that PWC-MR offer include (i) not requiring within-family based association summary statistics which are scarcely available, (ii) not requiring association summary statistics of early traits which are also not widely available, and (iii) revealing heterogeneous causal effect estimates, some of which could be reflecting confounder effects.

This project originated from our investigation into discordant genetic correlation and causal effect estimates between traits.

We found this discordance larger when using population-based GWAS compared to within-sibship GWAS.
Specifically, BMI and EDU had a r_G value of -0.38 ($P = 6.3 \times 10^{-99}$) using population based GWAS, whereas using sib-ship GWAS, this correlation was no longer significant (0.14, $P = 0.38$). This prompted us to investigate the BMI-EDU relationship more closely, with respect to unbiased sib-ship association estimates and subsequent attenuated causal effects.
The project was conceived and designed by Liza Darrous and Zoltán Kutalik. Statistical analyses were suggested by Zoltán Kutalik and trials were carried out by Liza Darrous (SVD, K-means clustering) and Zoltán Kutalik (Fuzzy clustering). The manuscript was written by Liza Darrous with the help of Zoltán Kutalik, and valuable input for the project and the manuscript was provided by co-authors Gibran Hemani, and George Davey Smith.

3.2 Application: clustering of obesity, a composite trait

Following the suggestion of George Davey Smith, we ran additional analysis to investigate the findings of PWC-MR when using the components of a composite trait as exposure/trait of interest.
In Sulc et al. [81], they performed a PCA using 14 anthropometric traits from the UK Biobank to obtain distinct orthogonal components of obesity, each representing different features of the human body shape. They also showed that these body-shape related measures can be summarised by the first four principal components influencing body size, adiposity, abdominal fat deposition, and lean mass respectively.

Therefore, we re-ran PWC-MR on each of the first four obesity PCs. By selecting the GWS SNPs for each PC, we performed informative K-means clustering and enrichment analysis using PheWAS data of ~400 traits, the results of which can be seen in Figure 3.3.
We followed with causal inference analysis on EDU by running IVW on each of the SNP-clusters of the 4 PCs as well as on all their GWS SNPs respectively (see Figure 3.4).

Our findings revealed the following:

- PC1, affecting body size, clustered into 6 groups, 4 of which were strongly and distinctly enriched for body impedance, food substrates, lung capacity measures, and

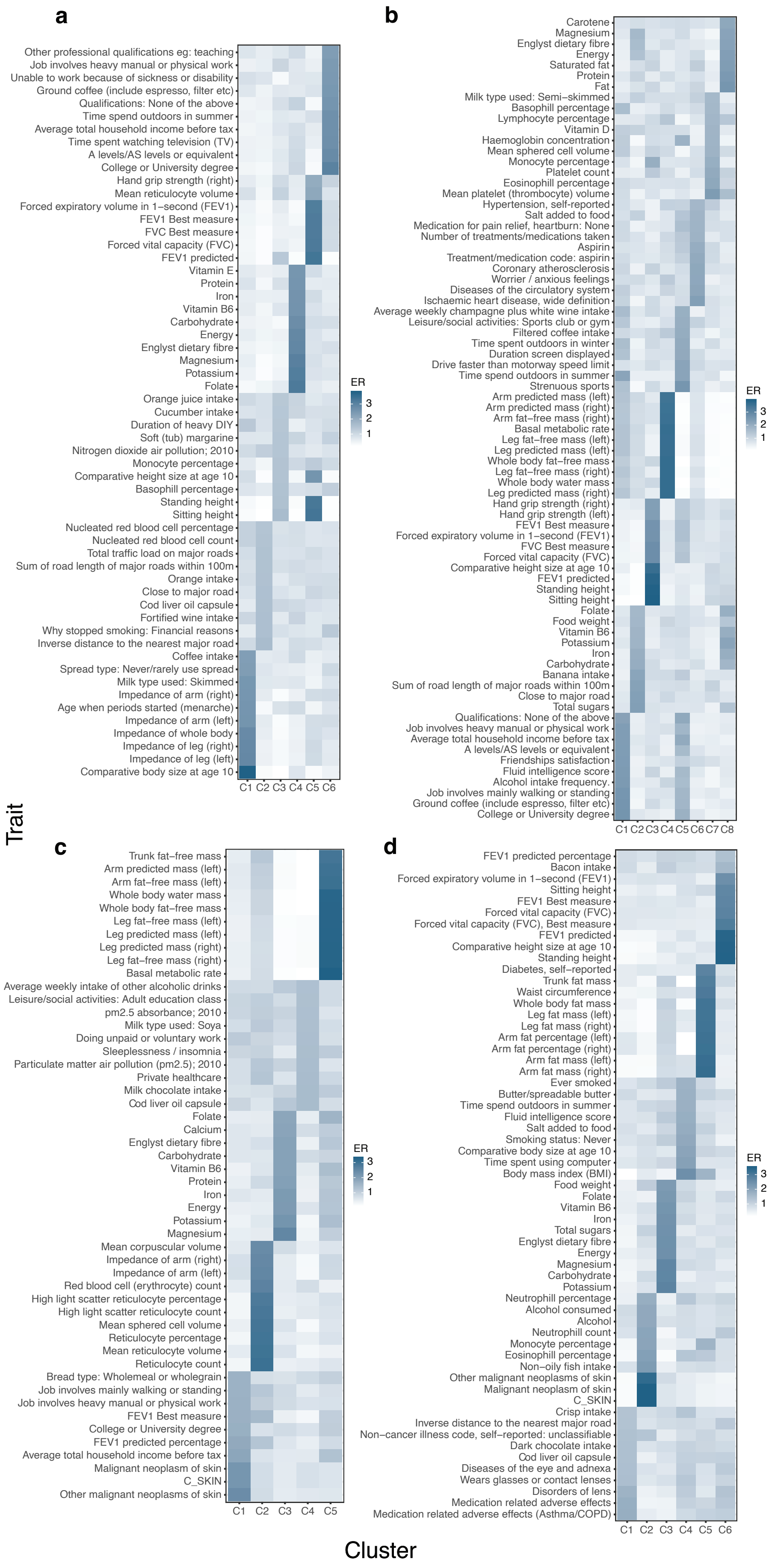


Figure 3.3: Heat map of the trait enrichment ratios in each cluster for PCs of the obesity composite trait. **a** PC1 representing body size clustered into 6 groups. **b** PC2 representing adiposity clustered into 8 groups. **c** PC3 representing abdominal fat clustered into 5 groups. **d** PC4 representing lean mass clustered into 6 groups. The darker the colour, the stronger the enrichment ratio (ER).

SEP-related traits like job type and qualifications respectively. The causal effect estimates of the clusters were significantly heterogeneous (Q-test of 62.50, $P = 1.40{\times}10^{-11}$), with the cluster that was strongly enriched for SEP-proxy traits having the largest negative causal effect on EDU (-0.32, $P = 2.01 \times 10^{-13}$). Cluster 1, which was enriched for body impedance traits had an attenuated causal effect on EDU of -0.06 ($P = 0.030$).

- PC2, influencing adiposity, clustered into 8 groups instead. However, most clusters were not very strongly enriched. Clusters 1 and 5 were enriched for SEP-proxy traits, whereas cluster 4 was enriched for fat-free mass-related traits. Unsurprisingly, the causal effect estimates for all clusters on EDU were heterogeneous (69.40, $P = 6.46 \times 10^{-12}$), and the clusters that were enriched for SEP-proxy traits had the largest negative causal effects (-0.46, $P = 2.47{\times}10^{-16}$, and -0.42, $P = 9.78{\times}1{-}^{-16}$ respectively).
- PC3 with its effect on abdominal fat clustered into 5 clusters, one of which was strongly enriched for fat-free and body mass related traits. There was little enrichment for SEP-proxy traits overall, but some was found in cluster 1. Consequently, the causal effect estimates of these clusters were not significantly heterogeneous (8.00, $P = 0.16$), and the overall causal effect on EDU was -0.03 ($P = 0.054$).
- Lastly, the clustering of PC4, which represented lean mass, resulted in 6 clusters. One of which was strongly enriched for body measurement and fat related traits, another was enriched for a mix of lung measurements, height, and blood traits. Similarly to PC3, there was very little enrichment for SEP-proxy traits, and the causal effects estimated were not heterogeneous (10.36, $P = 0.11$). The overall causal effect on EDU was non-significant (0.019, $P = 0.23$).

These findings support our main analysis and results [71] in that SEP-proxy traits, that are strongly enriched for in PC1 and PC2, seem to confound the BMI-EDU relationship, inducing an overestimated causal effect that may actually be much smaller, or even non-existent.

They also compliment the findings of Sulc et al. [81], where the decomposition of obesity into various components sheds light onto which are more likely to be disease-causing or associated with various life styles.

Their findings showed that PC1 (body size) and PC2 (adiposity), which explained 73.3% and 19.9% of the total variance respectively, increased the risk of many diseases, especially obesity-related ones. An increase of one standard deviation

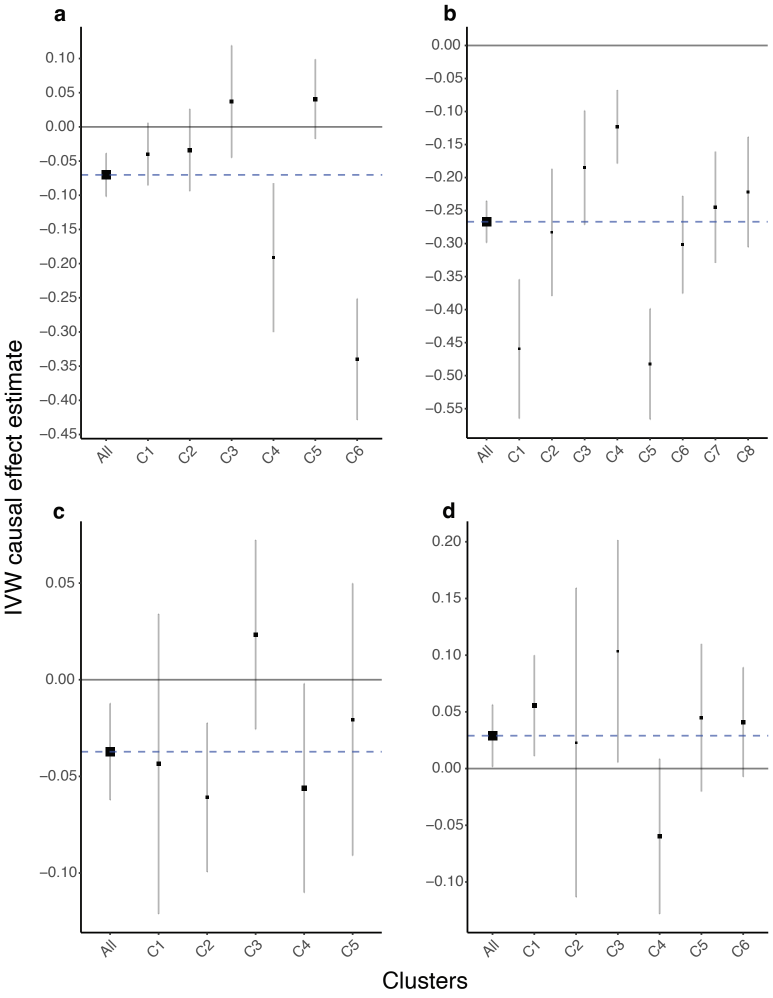


Figure 3.4: IVW causal effect estimates of all GWS SNPs as well as cluster-specific SNPs for each PC of the obesity composite trait. **a** Causal effect estimate of PC1 representing body size on EDU. **b** Causal effect estimate of PC2 representing adiposity on EDU. **c** Causal effect estimate of PC3 representing abdominal fat on EDU. **d** Causal effect estimate of PC4 representing lean mass on EDU.

(SD) of PC1 increased the absolute risk of diabetes by 1.7% (95% CI: 1.3–2.1), whereas a 1 SD increase of PC2 had a 2.1% risk increase. Similarly, these two also had a decreasing effect on lifestyle factors; PC1 slightly reduced SES, and PC2 had a similar yet more pronounced effect on SES, as well as links to decreased income, fluid intelligence score, education, and physical activity.

In contract, PC3 (predisposition to abdominal fat deposition) and PC4 (lean mass), which explained 6.2% of the variance combined, had no significant causal effects on traits such as SES, job-type nor education. PC3 however, did have a detrimental effect on diabetes (1 SD increase resulted in 1.6% increase of absolute risk).

By taking advantage of the orthogonality of PCs and comparing their effects, we can better understand and dissect the causes of obesity-related diseases and lifestyle consequences.

These specific components with their contrasting homogeneous and heterogeneous causal effects, as seen in our results, can help pinpoint mechanisms through which particular sub-types of obesity, rather than broad measures like BMI, can influence traits such as educational attainment.

Chapter 4 4

4.1 Ongoing: Estimating family-to-offspring causal effects from genetic data of first-degree relatives 35
4.2 Methods 36
4.3 Results 39
4.4 Discussion 42

4.1 Ongoing: Estimating family-to-offspring causal effects from genetic data of first-degree relatives

GWAS have recently shifted focus onto estimating direct genetic effects using family-based cohorts. This shift was necessitated by the recognition of bias inherent in population-based GWAS, such as the presence of confounding factors (population stratification, dynastic effect, or assortative mating) and cryptic relatedness which can introduce spurious associations between genetic variants and traits. These biases can lead to false-positive or inflated results, and hinder the accurate identification of true causal genetic variants.

By focusing on individuals within families, shared genetic and environmental factors that often confound population-based studies can be controlled for. Family-based designs inherently account for shared genetic ancestry and provide a more controlled setting to tease apart the direct effects of genetic variants from extraneous influences, thus reducing the risk of false positives and enhancing the reliability of GWAS findings.

However, while past studies investigated direct genetic effects of complex trait, we were interested in estimating the indirect effects, also known as parental/family effects, that arise when the genetics of an individual affect the trait of a family member. While unbiased direct genetic effect estimates are valuable for understanding complex traits, estimating indirect effects is equally important for comprehending the influence of the environment or rearing factors on these traits.

Indirect, *family*, *offspring*, and *untransmitted effects* will be used interchangeably in this section

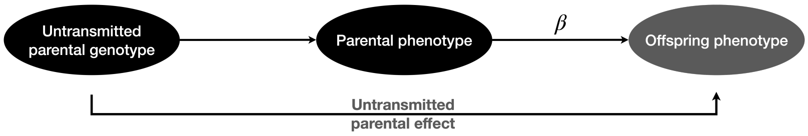

Untransmitted parental genotype
Parental phenotype
β
Offspring phenotype
Untransmitted parental effect

Figure 4.1: MR representation of estimating the causal effect of parental rearing environment on offspring phenotype (β), using unstransmitted parental/indirect effects.

Thus, to estimate the parental environment-to-offspring causal effects using MR, as seen in Figure 4.1, we first estimated these untransmitted effects by modelling both direct and indirect effects jointly using genetic data of first-degree relatives.

Although estimating indirect genetic effects of parents on offspring are of most interest, because they are likely to be the largest, indirect genetic effects of *siblings* or more distal relatives are also important, as these family relationships are more likely to be found in large cohorts.

4.2 Methods

In order to estimate both the direct and the family genetic effects (indirect effects), we used individual level data of first-degree relatives, specifically siblings, from the UK Biobank. We filtered our individuals to be of white British origin based on self-identified ethnicity, and still consenting to participate in UK Biobank research.

To select siblings, we filtered for ID pairs that satisfied the following criteria: 0.3534 ≥ Relatedness ≤ 0.707 and IBS0≥ 0.0012. For these subset of individuals, we extracted phenotypic, covariate and additional information such as their age, sex, first 40 PCs, and missing rate (for quality assurance). For our preliminary analysis, we chose to focus on two traits: BMI and EDU, where we re-coded EDU to include educational attainment of tertiary level (college or university degree).

Relatedness here is equivalent to kinship×2, where the kinship coefficient is a simple measure of relatedness, estimated from the GRM of the study

IBS0 stands for identical by state zero, and is a measure of lack of genetic similarity

For ease of analysis, we limited the individuals to unique sibling pairs with no duplicate IDs in either index or related individuals. This left us with ~17’300 unique sibling pairs.

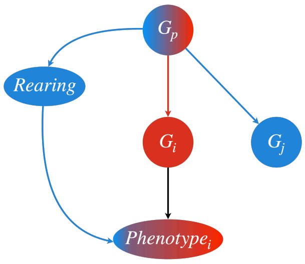


Figure 4.2: Graph representation of direct and indirect effects of parental genotype onto offspring genotype and subsequently, phenotype. G_p represents the parental genotype, where a single allele is inherited by the index individual’s genotype (G_i). In this scenario, the other allele is inherited by the sibling (G_j). This untransmitted effect, which we aim to estimate, is hypothesised to act via a rearing/environmental effect on the offspring phenotype.

Given G_i and G_j, representing the genotype of an *index individual i* and the genotype of their *sibling j* respectively, we first attempted to obtain the untransmitted effect of the index individual, denoted as G_{iu}, by regressing the genotype of the sibling onto the genotype of the index individual, and obtaining the regression residual from:

$$G_j \sim G_i$$

where the regression coefficient is 0.5, thus the residual is:

$$G_{iu} = G_j - 0.5 \times G_i$$

Visualising this process in Figure 4.3, we see that with respect to the effect allele T, the dosage among siblings could differ depending on parental allele transmission. We have shown this in the three separate scenarios, where the parents are heterozygous AT carriers, and the siblings vary from being homozygous AA or TT, to heterozygous AT. It is important to note that the parental allele dosage is unknown to us, and thus we can only **estimate** what **the untransmitted allele dosage** can be from sibling data, instead of having precise knowledge.

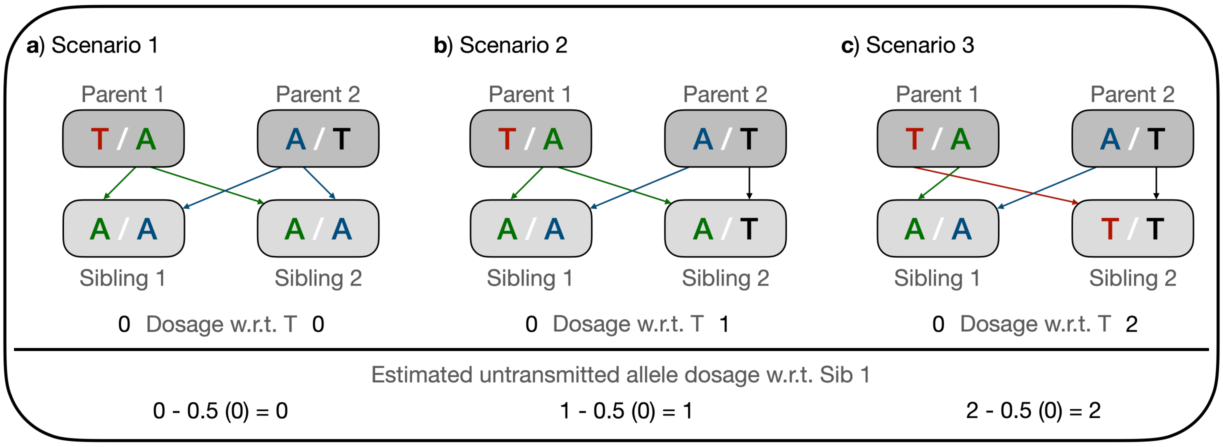


Figure 4.3: Illustration of allele transmission from parents to offspring Three scenarios are shown of differing allele transmission leading to different allele status for sibling 2. Consequently, the allele dosage differs, and so does the estimated untransmitted allele with respect to sibling 1.

When both siblings are homozygous AA, then the estimated untransmitted allele dosage with respect to sibling 1 is 0, despite the fact that the T allele is untransmitted from both parents. In scenario 2, sibling 2 is heterozygous, and so the estimated untransmitted allele dosage is 1. Lastly, when sibling 2 is homozygous T in the third scenario, then the dosage is 2.

Given this estimate of untransmitted allele, we then estimated both the direct and indirect effects by regressing the index individual's phenotype, Y_i, jointly onto its own genotype G_i, the untransmitted dosage (obtained as the regression residual G_{iu} in the first step), and various covariates:

$$Y_i \sim G_i + G_{iu} + age_i + age_i^2 + sex_i + PC1_i + ... + PC40_i$$

The regression was performed per SNV, across all sibling pairs. To speed up this process, the regression was parallelised over genome chunks that were created and stored in an initial step for the filtered subset of first-degree relatives.
Due to the bi-directional relationship between sibling pairs with respect to indirect effects inherited from the parents, the association effect can be estimated twice, once for each sibling in the pair acting as the index individual; as shown below for individual j in sibship $i - j$:

$$Y_j \sim G_j + G_{ju} + age_j{+}age_j^2 + sex_j + PC1_j + ... + PC40_j$$
$$\text{where } G_{ju} = G_i - 0.5 \times G_j$$

This resulted in a doubling of the sample size for sibling pairs, leaving us with a total sample size of ~34'600.

We then meta-analysed the estimates from sibling pairs (with each sibling acting as index individual once) by using a meta-analysis approach that accounts for relatedness of overlapping/correlated subjects in the two studies. This method uses the covariance structure between the two "studies" to adjust the weighting of the coefficients (see equations 3-6 in Lin and Sullivan [82]).

In order to verify our effect estimates, we performed quality control that included trait heritability estimation using LDSC, and GWAS visualisation (QQ-plot and Manhattan plot).
We then calculated the genetic correlation of the indirect effects of our two traits of interest across ~190 other UK Biobank traits. These traits were selected from ~1480 UK Biobank traits for being continuous or ordinal in nature, and filtered for right-specific traits when there was either left or right-side measurements for the same trait. We also contrasted our findings to genetic correlations of population-based GWAS association estimates, and within-sibship GWAS association estimates (representing unbiased direct effects) of our two focal traits to the same ~200 other traits.
Within-sibship GWAS effects for both BMI and EDU came from Howe et al. [39]. However, to reduce bias given that our indirect effects were estimated using individual level data from the UK Biobank data, we used external and large GWAS for both EDU (Okbay et al. [83]) and BMI (Yengo et al. [84]) for the genetic correlation estimation.

Lastly, in order to test our hypothesis of whether parental/-family traits act indirectly on offspring traits by proxying a rearing environment, we estimated the causal effect of over 150 parental traits on indirect offspring EDU and BMI. Those with a nominally significant causal effect were then used in a stepwise-MVMR in order to estimate their conditional causal effect on the two offspring traits.

4.3 Results

After running the GWAS twice for each sibling pair (N = ~34’600), we meta-analysed the direct and indirect effect estimates for both BMI and EDU across ~8.6 million SNPs. The heritability of the indirect effects for EDU and BMI were modest; 0.06 (SE = 0.02) and 0.01 (SE = 0.02) respectively. As seen in Supplementary Figure 0.1, there were no GWS hits for either trait.

Our comparisons of the genetic correlation for both BMI and EDU - coming from three different sources of GWAS: population-based GWAS, within-sibship GWAS, sibling meta-analysed indirect effects - across ~190 other traits are shown in Supplementary Figure 0.2.
We notice that the genetic correlation of population-based and within-sibship BMI or EDU GWAS across various traits such as BMI, impedance, and alcohol intake frequency is nearly identical. Only sixteen traits show a significant genetic correlation with the indirect BMI effect estimates, while 43 traits exhibit a significant genetic correlation with the indirect EDU effect estimates..

Some patterns we observe (highlighted in Figure 4.4) include the lack of genetic correlation between parental (population-based) and offspring (sibling meta-analysed indirect effect) obesity. The r_G of both parental BMI and body fat percentage with indirect BMI effects are non-significant. Conversely, healthy parental dietary habits (increased dried fruit/cereal intake, reduced beef/pork/poultry intake) is inversely correlated with offspring BMI.

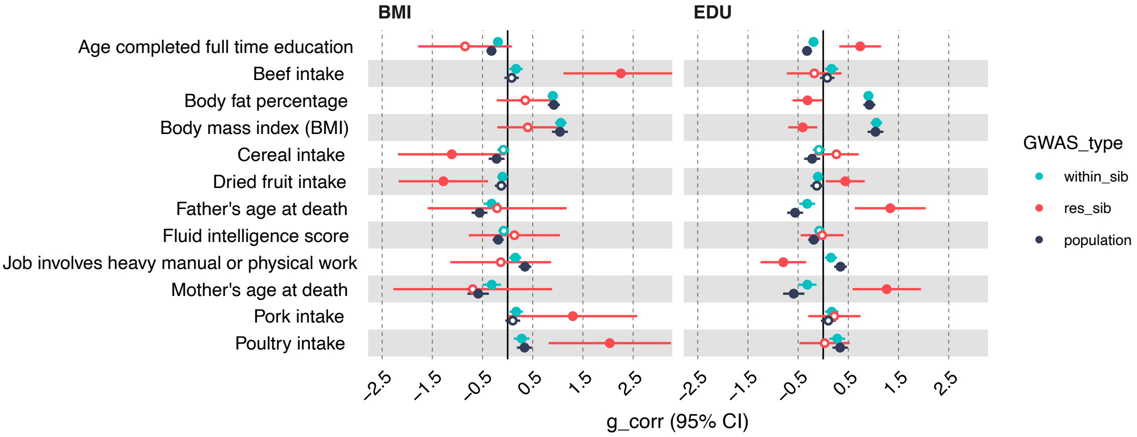


Figure 4.4: Forest plot of the genetic correlation of both BMI and EDU across 12 other selected traits. Effects of BMI and EDU come from 3 different sources of GWAS: population-based, within-sibship, sibling meta-analysed indirect effects (res_sib). The effects of the other traits all come from population-based GWAS. 95% confidence intervals are shown as error bars. Points that are not filled indicate a non-significant genetic correlation estimate.

In contrast to BMI, parental and offspring education share extensive genetic basis ($r_G = 0.74, P = 5.25 \times 10^{-4}$). However, this is not the case for fluid intelligence; EDU indirect effects had no significant genetic correlation ($r_G = -0.02, P = 0.92$). While parental jobs involving physical labour are anticorrelated to offspring EDU, parental longevity is strongly positively correlated to offspring EDU: 'Father's age at death' had a $r_G = 1.33$ ($P = 2.16 \times 10{-}4$), 'Mother's age at death' had a $r_G = 1.26$ ($P = 2.74 \times 10{-}4$).

For our second analysis, we ran univariate MR using over ~150 traits representing parental effects as exposures and measured their causal effect on the untransmitted BMI/EDU offspring effects.

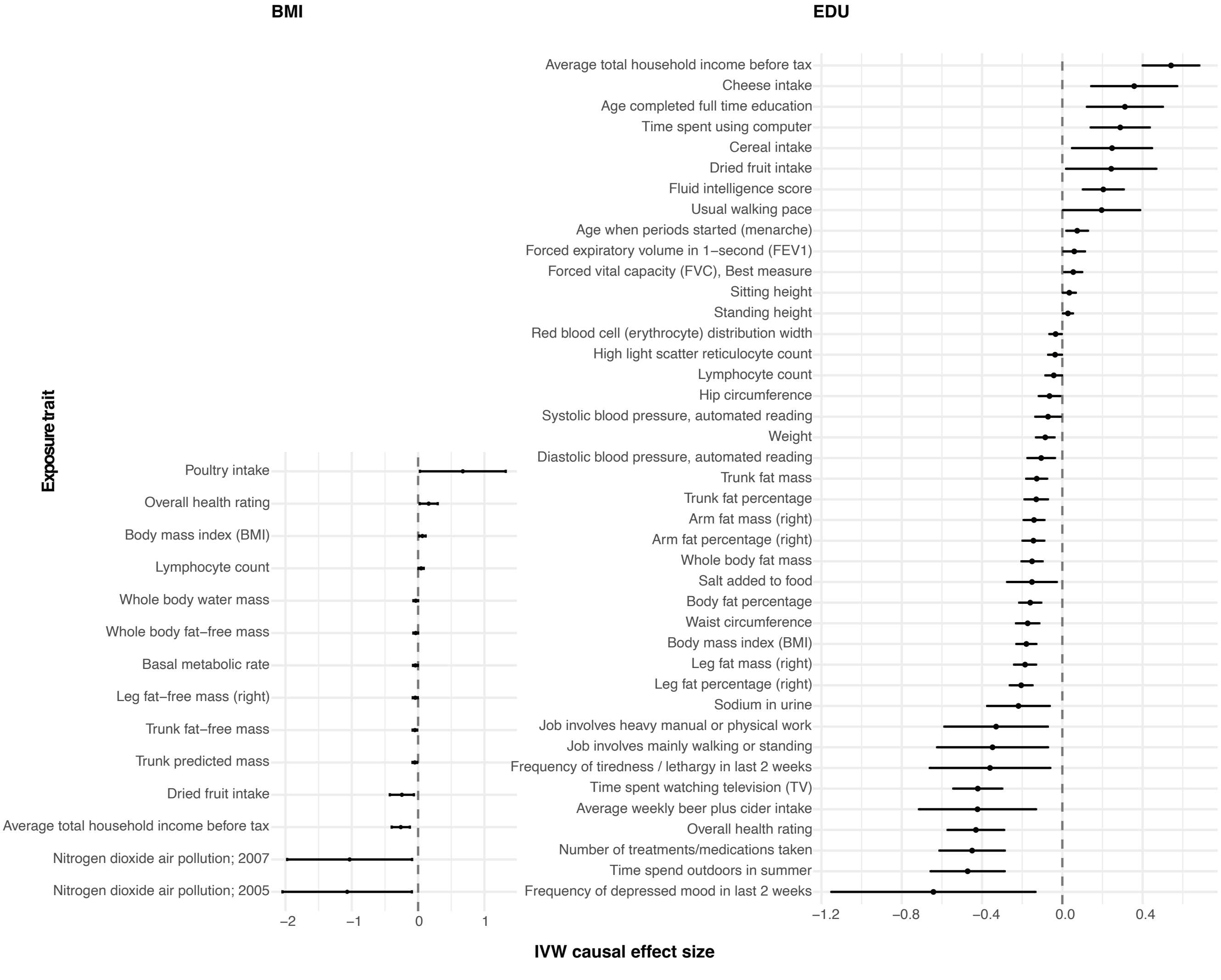


Figure 4.5: Nominally significant IVW causal effect estimates of various traits on both untransmitted BMI and EDU (meta-analysed sibling effects).

As seen in Figure 4.5, there were 14 traits with nominally significant causal effect on offspring BMI, which ranged from dietary

intake, body measurements and SEP-proxy traits. Poultry intake had the largest positive causal effect, compared to the arguably healthier dried fruit intake which had a negative causal effect on offspring BMI.
MR revealed that there is no parent-to-offspring transmission of BMI (causal effect of parental BMI is negligible), rather parental dietary habits and SEP-related traits are the most likely drivers of offspring obesity.
This was further supported when stepwise-MVMR was run using all these 14 traits as exposures and offspring BMI as outcome. The results in Table 4.1 show that 'Average total household income before tax' (SEP-proxy trait) has a negative causal effect on offspring obesity: $-0.10, P = 7.87 \times 10{-}4$.

Table 4.1: Stepwise-MVMR causal effect estimates on offspring BMI. One parental trait that survives stepwise-MVMR has its causal effect on untransmitted BMI estimated using MVMR.

Exposure	F-statistic	$\hat{\beta}$	SE	P
Average total household income before tax	41.24	-0.0997	0.0226	7.87E-04

Similarly, 41 traits ranging from dietary intake, body measurements, SEP-proxy traits and a mix of lung and blood measurements had nominally significant causal effects on offspring EDU. Running stepwise-MVMR on these traits as exposures and offspring EDU as outcome (see Table 4.2), revealed that both parental BMI ($-0.04, P = 2.5\times10{-}6$) and fluid intelligence score ($0.19, P = 6.24 \times 10{-}5$) have a significant causal effect on offspring EDU.

Table 4.2: Stepwise MVMR causal effect estimates on offspring EDU. Two parental traits that survive stepwise-MVMR have their conditional causal effect on untransmitted EDU estimated using MVMR.

Exposure	Conditional F-statistic	$\hat{\beta}$	SE	P
Body mass index (BMI)	16.9	-0.0424	0.0121	6.34E-04
Fluid intelligence score	5.97	0.1875	0.0452	6.24E-05

4.4 Discussion

In this study, we estimated the untransmitted genetic effects of two traits, BMI and EDU, using genetic data from first-degree relatives in order to investigate potential family-to-offspring causal effects.
Our motivation stemmed from the findings of PWC-MR, which suggested a potential confounder influencing both BMI and EDU, possibly of parental origin. Consequently, our primary interest was in estimating parental effects on offspring, and to do so, we needed to first estimate these untransmitted effects.

Initially, we explored methods to distinguish direct and indirect effects, with a focus on estimating the indirect effects as a measure of rearing/environmental factors in the family. We first tried GWAS-by-subtraction using GenomicSEM [76], given that the data available to us was population based (encompassing both direct and indirect effects) and within-sibship GWAS (indirect effects). However, the input traits (population-based and within-sibship-based GWAS summary statistics of the same trait) were too genetically correlated to reveal a latent trait representing the indirect effect. Instead, we took inspiration from Howe et al. [39], and decided to model our own GWAS using sibling data from the UK Biobank to obtain estimates of **indirect effects** instead.
In order to ensure that the family component of the GWAS model is not correlated to the index individual's genotype, we instead used a two-step GWAS model, where we i) subtract the genotype of the index sibling from the other sibling, ii) use the residual as a variable in our GWAS model to represent untransmitted/indirect genetic effects from the parent to the index individual (offspring).
Our estimates showed low heritability for our two focal traits, as expected due to their hypothesised role as primarily environmentally determined rearing behaviours.

These indirect effect estimates, which were fed into genetic correlation and causal inference analyses, shed light onto i) whether traits such as BMI and EDU are entirely genetically inherited, and if not ii) how the rearing environment plays a role in shaping an offspring's BMI and educational attainment. Our results highlight that a high socioeconomic environment and healthy parental diet have a favourable effect on offspring BMI, whereas sedentary habits, such as excessive TV watching, tend to decrease the offspring's educational attainment.

Our work is preliminary in nature, and despite its valuable insight into the significance of environmental and rearing factors in understanding "heritable" traits, like other scientific methods, it has its limitations:

- Our sample size of near 35 thousand pairs, is relatively small. However, this limitation arises from the scarcity of available sibling or family-based cohorts with easy open access. For future analyses, we aim to incorporate family-specific data from sources such as the Estonian biobank and the MoBa study.
- We can attempt to expand our sample beyond just sibling pairs, by incorporating parents, cousins, uncles and aunts as first-, second- and third-degree relatives.
  However, the more relatives we add, the noisier our estimates may become. This is not surprising, given the small sample sizes of these relatives, and that the further away from the index individual that you get, the less likely it is that you are truly estimating the untransmitted effect between the parent of the index individual and themselves. For example, an aunt's genotype has a 50% chance of being shared with the parent (its sibling), which means there is only a 50% probability of being truly untransmitted between the parent and the index offspring. A possible way to account for this would be to weigh these estimates in relation to their (transmission) distance from the index individual.
- Although we attempt to estimate the untransmitted effect by comparing and subtracting the genotype of two siblings, we cannot definitively confirm whether the estimated indirect effect accurately represents the truly untransmitted allele.
  For a more accurate estimation, we would need to employ knowledge gained by haplotypes and their phasing, in order to precisely identify parental alleles that were untransmitted and estimate their indirect effects on offsprings.
- In our step-wise MVMR analysis, the conditional F-statistic for the most likely environmental traits through which parental rearing acts, is less than the typically accepted value of 10. However, this lower value may represent a compromise between two sources of biases: weak instrument bias *vs.* bias due to omitting relevant confounders.

This ongoing project was originally conceived and designed by Zoltán Kutalik and Gibran Hemani. Statistical analyses were proposed by Zoltán Kutalik, Gibran Hemani and Liza Darrous and carried out by Liza Darrous.

Chapter 5 | 5

5.1 Minor contributions to other publications . . . 45

5.1 Minor contributions to other publications

Ojavee et al. [85] investigated how genetic effects of age-at-menopause can change across time by running a marginal Cox age-specific mixed proportional hazards (CAMP) model. Their results show that 74% of 245 associations show a form of age-specificity, and they were able to replicate their 19 novel findings in an independent cohort. To test whether these stratified age-at-menopause effects have a causal effect on various traits such as BMI, cholesterol, stroke and educational attainment, I ran bi-directional standard MR methods.

Additionally, I made contributions to the following manuscripts, Sulc et al. [81] and Porcu et al. [86], by providing input on the conceptual framework and analyses, and conducting thorough proofreading.

Discussion 6

6.1 The golden thread: confounding 47
6.2 Moving forward: MVMR and its caveats 50
6.3 Future works: Using haplotype data for indirect genetic effect estimation 54
6.4 Conclusion 56

In the introduction of this work, I provided an overview of human genetics and its recent advancements, focusing on genome-wide association studies of various traits and of exceedingly larger cohorts.
GWAS have led to the discovery of thousands of genetic variants associated with various diseases, including common complex diseases like diabetes, heart disease, and certain cancers. These associations could point to specific genes or biological pathways implicated in disease susceptibility. Knowledge of these associations helps researchers understand the underlying mechanisms of diseases, which in turn, can lead to the development of targeted therapies.
Beyond diseases, GWAS have also shed light on the genetic underpinnings of various complex traits, including traits related to behaviour, cognition, and physical characteristics. This has enhanced our understanding of the genetic basis of human phenotypic diversity, and further allowed us to investigate how complex traits are linked to each other.
During my research, I aimed to investigate causal relationships that risk factors may have with common complex diseases by using a statistical method called Mendelian Randomisation (MR). MR uses genetic variants as instrumental variables for the exposure that they associate with, and leverages the principles of Mendelian inheritance to estimate causal effects on an outcome of interest, in the presence of unobserved confounding.

6.1 The golden thread: confounding

The common theme in my research has been the study of model violations biasing causal inference, and how to account for them. Sources of model violations with respect to MR can take many forms such as reverse causality, non-linear relationship, over fitting, and population stratification. However, my work focused on two specific types of violation: **correlated pleiotropy** and **effect heterogeneity**.
I first tackled correlated pleiotropy in the form of heritable confounding, by accounting for its presence in a typical MR framework through LHC-MR. We developed a structural equation (mixed effect) model that accounted for the presence of

a latent heritable confounder of an exposure-outcome relationship, in order to estimate unbiased bi-directional causal effects between the two traits. LHC-MR had an advantage over standard MR methods, in that it used whole-genome SNPs instead of GWS SNPs only. This allowed it to have more power to detect causal relationships between traits otherwise missed by standard MR methods, as well as detect the presence of latent heritable confounders of trait pairs.

LHC-MR however, was limited by certain assumptions about the genetic architecture of traits (two-component Gaussian mixture of effect sizes), and that of a single general confounder of the exposure-outcome relationship. Indeed, for some trait pairs, we found a significant effect of the confounder on either the exposure or the outcome alone, hinting at a more complex genetic architecture for that trait than a two-component Gaussian mixture of effects. Any potential latent confounder might have been missed in this case, if it had a small effect on the trait pair.

Furthermore, although we attempted to identify potential traits that fit as confounders for some of our trait pairs, we could not accurately distinguish if there was a single confounder trait or multiple, with either concordant or discordant effects on the trait pair. While simulations we conducted revealed more accurate causal effect estimation between trait pairs with two confounders (either discordant or concordant) when compared to standard MR, LHC-MR itself could not identify the specific potential confounders.

Lastly, we also assumed that the correlation (across markers) between the direct effect of a genetic variant on the exposure, outcome and latent confounder is zero, i.e. the effects on each are independent. This assumption caused LHC-MR to be incapable of detecting parental/dynastic effects as potential confounders of possible trait pairs, as dynastic effects are correlated/share genetic markers with the exposure or outcome trait.

Given the above-mentioned limitations, we then attempted to extended LHC-MR's concept of classifying SNPs into those with a direct effect on an exposure and those that act through a confounder, to classifying SNPs into multiple different groups based on their association profile across several other traits.

Generally, MR has presented bias stemming from heterogeneous causal effects through various distinct pathways, and bias due to confounding of the instrument-outcome association as distinct mechanisms. In this study, we aimed at softening the homogeneous causal effect assumption of MR, by utilising an approach based on pheWAS-based clustering which can categorise instruments into distinct groups based on their association profile across several other traits, independently of their causal effect on any single trait. Some of these groups may represent different exposure subtypes or mechanisms through which the exposure exerts its effect, while others can include IVs primarily associated with confounding factors.

Using PWC-MR, we investigated BMI as an exposure and grouped its GWS SNPs based on their association to ~400 UK Biobank traits using K-means clustering.
This revealed 6 SNP clusters, some of which were enriched for distinct traits that highlighted the mechanisms through which BMI can be modulated; body-related measurement traits, food supplements/nutrients, and SEP-related traits. Estimating the individual causal effects of each cluster on EDU revealed significantly heterogeneous causal effect estimates. This variation in estimates reinforced our suspicion that the MR causal estimate of BMI on EDU tends to be overestimated when using population-based estimates of SNP effect sizes, primarily because of the presence of confounding factors.

Our findings have two significant implications: 1) The cluster of IVs related to lean mass suggests that the causal effect of BMI on EDU is nearly negligible, 2) we have also uncovered that IVs related to SEP indicate a substantial negative impact of BMI on EDU. One likely explanation for the observed bias is dynastic effect via parental SEP traits acting as confounders on both offspring adult EDU and adult BMI.
This hypothesis is supported by the findings of Howe et al. [39], where assortative mating, dynastic effects and population stratification were all accounted for in their sib-design, and their within-sibship GWAS effects revealed a non-significant MR causal effect of BMI on EDU: -0.05 (95% CI: -0.09, -0.01).

Despite our best efforts to be impartial when it came to pheWAS based clustering, by obtaining BMI SNP associations with as many traits as possible, while still filtering out traits that were strongly correlated with the exposure BMI to avoid redundancy and self causation, our method still has its limitations.
Our ability to create informative clusters of IVs is constrained by the availability of traits that have PheWAS data. This limitation could result in an inability to identify key pathways, potentially causing us to overlook clusters that represent significant subgroups related to mediators, sub-phenotypes, or confounding factors.

Another potential limitation pertains to our use of only GWS SNPs as exposure-IVS for clustering. This differs from the approach employed in LHC-MR, where genome-wide SNPs

are utilised as input. It would be interesting to test the possible clustering difference when leveraging information from more IVs, by decrementing the threshold used to filter for exposure-associated IVs. We would still need to ensure that the SNPs are primarily exposure-associated, by performing a trait-wide variant of Steiger-filtering.
However, we would also need to find a balance between the number of IVs used for clustering, and the potential to increase noise when IVs with smaller associations are used, or those with false positive exposure-associations.

PWC-MR was a logical continuation of LHC-MR, whereby we attempted to biologically interpret the pleiotropic effects observed between pairs of traits by considering the influence of potential confounding traits.
However, moving forward, we could benefit immensely from the secondary analysis carried out in PWC-MR; where we systematically searched for several candidate confounder traits and measured their causal effect on EDU conditionally on each other and on our focal exposure trait, BMI, using **MultiVariable MR**.

6.2 Moving forward: MVMR and its caveats

In an ideal scenario, we would preselect known confounder traits and have sufficiently large sample sizes for selecting strongly associated IVs to be used in a MVMR, enabling us to properly disentangle confounding from the causal effects between our focal trait pair. There would be no need for additional

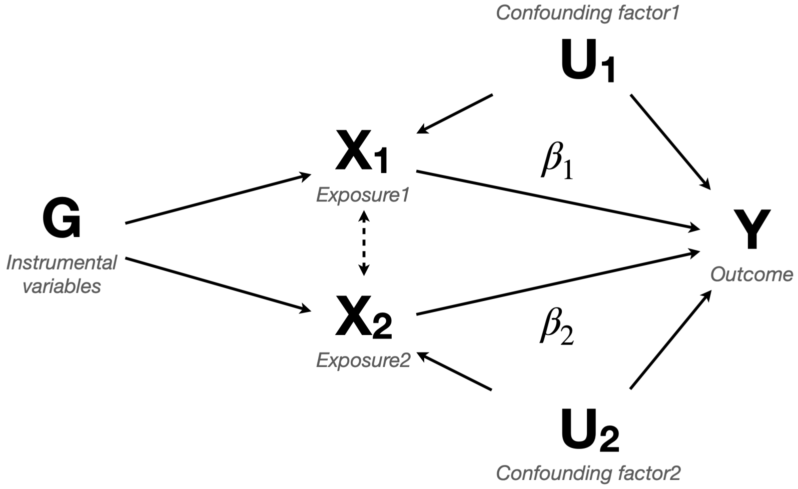


Figure 6.1: A simplified multivariable Mendelian randomisation (MVMR) model with two exposures. G represents a group of IVs, each associated with at least one of the two exposures. The line between the first exposure X_1, and the second X_2 is bidirectional and dashed as no assumptions are made about this relationship in the estimation of their respective causal effects, β_1 and β_2, on the outcome (Y). Confounders U_1 and U_2 are assumed to be unknown.

MR analysis such as outlier removal, instrument clustering, or modelling of latent heritable confounder(s).

MVMR is a powerful method with the potential to provide valuable insights into causal relationships. It allows us to account for confounding and mediating variables, while assessing the causal effects of multiple exposures simultaneously and conditionally, as seen in Figure 6.1.
Burgess and Thompson [62] show how MVMR can be implemented as an extension of IVW, using GWAS summary estimates of the association between SNP j and the outcome $\hat{\tau}_j$, the first exposure $\hat{\gamma_{1,j}}$, and the second exposure $\hat{\gamma_{2,j}}$. This is done by regressing the effect of each SNP on the outcome ($\hat{\tau}$), on the effect of each SNP on each exposure ($\hat{\gamma}$), i.e. by fitting the following model:

$$\hat{\tau}_j = \beta_1 \hat{\gamma_{1,j}} + \beta_2 \hat{\gamma_{2,j}} + \epsilon_j$$

Weighted by the inverse variance of $\hat{\tau}_j$, and where ϵ_j is a random error term for each SNP.

Whether X_2's relationship towards X_1 was that of a confounder or a collider (in this case, both X_1 and Y would have an effect on X_2), the direct effect of X_1 on Y estimated by MVMR is equal to the total effect of X_1 on Y estimated through standard MR using IVs that are strictly associated to X_1.
However, if X_2 was a mediator of X_1, a complication arises, whereby the direct effect estimated by MVMR is not equal to the total effect of X_1 on Y ($\beta_1 + \alpha\beta_2$, where α is the mediation effect of X_1 on X_2) estimated by standard MR using X_1-associated IVs. In this scenario, the mediation effect could in theory be estimated as the difference between the total and the direct effect of X_1 on Y; given that the exclusion restriction assumption is valid, and there are no IVs acting on the outcome except through X_1.

The causal effect of an exposure on an outcome, including any effect through potential mediators, is known as the *total effect*. This can be decomposed into *direct effect* of the exposure on the outcome, and *indirect effect* of the exposure on the outcome operating through the mediators included in the model

The benefit of MVMR has been demonstrated in several studies, where it helped in :

- investigating the causal relationships between multiple lipid traits (e.g., LDL cholesterol, HDL cholesterol, triglycerides) and cardiovascular disease outcomes. MVMR studies [87, 88] have helped clarify the role of different lipid components in heart disease risk, contrasting results from observational analysis.
- exploring the causal effects of obesity-related traits (e.g., body mass index, waist-to-hip ratio) on circulating lipoprotein, lipid, and metabolite levels. Here, MVMR [89] showed that excess adiposity likely raises atherogenic

lipid and metabolite levels exclusively via adiposity stored centrally.
- separating the effects of early- and later-life adiposity on disease risk. MVMR studies [90–92] suggested that childhood body size does not directly influence outcomes such as coronary artery disease and type 2 diabetes, but rather only has an effect via adulthood body size.

However, we have yet to reach an ideal scenario and thus MVMR has its own limitations and assumptions. These include extensions of the three core assumptions of standard MR (accounting for multiple exposures), and its success in revealing true direct causal effects can be limited by factors such as pleiotropy, measurement error, sample size, the presence of unmeasured confounders, and others as detailed below:

- Assumption of no pleiotropy: MVMR analysis assumes that the genetic variants used as instruments are not pleiotropic, meaning that they do not affect the outcome through pathways other than the exposure of interest. In other words, all traits that are potentially involved with the outcome should be used as exposures.
- Limited statistical power: MVMR analysis requires a large sample size to achieve sufficient statistical power to detect multiple causal effects. However, the number of genetic variants available as instruments for each exposure may be limited, reducing the statistical power of the analysis.
- Limited ability to control for confounding: MVMR analysis can only control for measured confounders that are included in the analysis. Unmeasured confounders may still bias the estimates of the causal effect if they are also associated with the IVs.
- Directionality and interactions: MVMR assumes clear directionality and linear relationships between exposures and outcomes. If the relationships are bidirectional, nonlinear, or involve feedback loops, MVMR may not provide valid estimates.
- Assumption of no measurement error: MVMR analysis assumes that the genetic variants used as instruments are not subject to measurement error. However, this assumption may not hold true in some cases, leading to regression dilution bias.

To perform an MVMR analysis, it is essential to have a minimum number of genetic instruments equal to the number of exposures to be instrumented in the model. It is also beneficial, as per the first assumption, to include as many (pleiotropic) traits as possible.

However, a delicate balance between selecting strongly associated IVs and relevant exposures needs to be achieved. The more exposures we include, the more IVs are selected. However, if these exposures are not all related, there is a possibility that some IVs will have strong(er) effects on select exposures, and weak effects on the unrelated exposures. To reduce the noise that arises from using IVs with small effects on secondary exposures, we could implement instrument shrinkage, where we shrink the effects of IVs on traits that fall below a certain threshold to zero. Another challenge faced is the clumping method used to choose between correlated IVs that are considered primary SNPs for different exposures.

Furthermore, we can test the association strength of the instruments of each exposure, conditionally, in the presence of other exposures in the model by calculating the conditional F-statistic [53]. This statistic provides an indicator of the strength of an exposure's IVs relative to the sample size and the number of IVs. A common threshold is an F-statistic greater than 10, which is considered indicative of strong instruments (average bias of the MVMR estimates is 10%).
Researchers can use the conditional F-statistic to guide their selection of IVs that are strong instruments for a particular exposure, allowing for more robust MVMR analyses.

However, the conditional F-statistic can also be used for trait selection when having a focal exposure trait and multiple other secondary pleiotropic traits. The conditional F-statistic of the focal trait calculated based on different inclusion-combinations of secondary traits in the model will vary; generally, the more traits included in the model, the lower the value for the focal trait is, due to added noise. When this value falls below a selected threshold, it indicates that the secondary traits included, which could be either mediators or even the true focal exposure, are tagging the selected focal exposure equally well. Thus, including too many such traits, could cause the MVMR to underestimate the true and direct causal effect of our focal trait on the outcome.

All of these considerations were taken into account when conducting our MVMR analysis in PWC-MR. We even implemented an additional trait filtering step by running step-wise MVMR inspired by the bGWAS package [93].
This was done by first running univariate MR between all the candidate exposure traits and the outcome. Then, the exposure traits with significant causal effects were added to the MVMR model in a stepwise manner, ensuring that each addition was significant with an estimate P-value below a pre-specified thresh-

old. Finally an MVMR model was run with all the surviving traits included, and the focal exposure trait among them.

6.3 Future works: Using haplotype data for indirect genetic effect estimation

As discussed in Section 4.4, one of the limitations we face when estimating indirect effects using individual level genetic data of siblings, was the inability to verify that the indirect effect estimated is that of the truly untransmitted allele or not.
We see in Figure 6.2, two examples of this limitation, where the true inheritance of alleles between parents and two siblings, is shown. Both scenarios demonstrate how the estimated untransmitted allele dosage obtained from the sibling genotypes is not accurate when it comes to the true dosage of untransmitted parental alleles. In our previous analysis, we were arbitrarily associating the phenotype of each sibling with an average of the possible untransmitted allele dosage given the different possible transmission scenarios, without knowing the true underlying one.
However, a more accurate estimation can be obtained if we leverage the information gained from haplotype phasing to precisely identify the unstransmitted allele from parent to offspring, and correctly associate it with the phenotype.

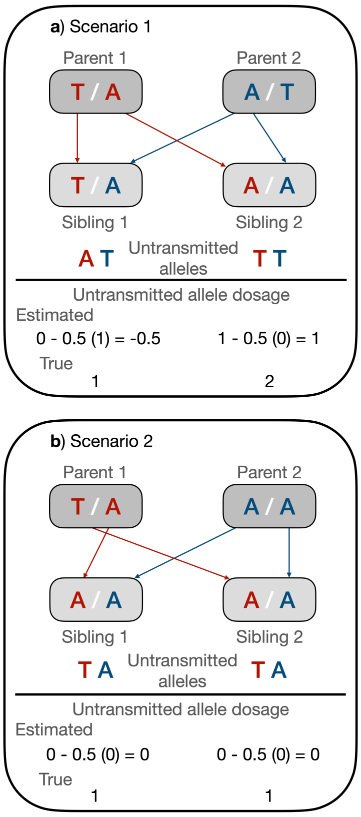


Figure 6.2: Graph representation of transmitted and untransmitted parental alleles onto siblings. **a** shows scenario 1, where the T allele of sibling 1 is inherited from parent 1, leaving the A allele to be untransmitted. **b** shows scenario 2 where the A allele of parent 1 is inherited by sibling 1, leaving the T allele to be untransmitted.

Haplotype phasing is the process of determining the specific combination of alleles that are inherited together on a single chromosome from one parent. With pedigree phasing, we infer the haplotypes of individuals within a family based on their familial relationships and genetic data (see the top panel of Figure 6.3). When a trio (parents and offspring) is available, the genetic data from parents and their offspring can be used to directly infer the haplotypes of each parent, and thereby phase the offspring's haplotypes.
However, other family members and their genetic relationships can also contribute to phasing; siblings, cousins, and other extended family members can provide information about shared haplotypes and genetic linkages. Interestingly, the process of pedigree phasing can be flipped, in what is known as **reverse-pedigree phasing**, to infer the set of transmitted and *untransmitted alleles* (bottom panel of Figure 6.3).
Reverse-pedigree phasing uses the genotype of the offspring as a reference to phase the genotype of the parents and then extract the set of untransmitted alleles of each.

However, in the case where parental genomes are not available, we can use sibling pairs to identify untransmitted alleles by mapping IBD segments. Shared IBD segments are transmitted from either of the parents, however segments that are not shared between the siblings indicate untransmitted alleles from one parent, with respect to a specific sibling, as seen in Figure 6.4. In this scenario, with sibling 1 as reference, we can determine which alleles were untransmitted by observing the regions that are not shared between the siblings in sibling 2 (shown in grey). These regions could be diploid for some alleles, meaning that neither alleles on either chromosomes are shared with sibling 1. This makes it difficult to predict the untransmitted allele for sibling 1.
The regions could be haploid, meaning one allele in that locus is known and shared with sibling 1. In this case, the missing allele can be predicted using population allele frequency and regarded as the untransmitted allele of sibling 1.

The same technique of IBD mapping followed by allele inference can be used for second degree relatives (aunt/uncle, niece/nephew etc...) to determine the untransmitted alleles of individuals. However, given the one degree of separation with respect to the index individual's parent (aunt/uncle to parent), the estimates ought to be weighed by half, since there is a 50% chance that the suspected untransmitted allele is shared between the parent and the relative.
Similarly, for third degree relatives, i.e. first cousins. An additional step in this scenario would be to determine the chromosome that belongs to the same family as that of the index individual, in order to then identify the haplotype inherited from common ancestor by IBD mapping. This is then followed by the same weighting scheme.

Once the untransmitted alleles of the index individual (G_{iu}) are identified, we can repeat our previous analysis of estimating the indirect effects on an individuals phenotype (Y_i).

$$Y_i \sim G_i + \boldsymbol{G_{iu}} + age_i + age_i^2 + sex_i + PC1_i + ... + PC40_i$$

Future work can also benefit from obtaining a larger sample size of relatives, by incorporating data from biobanks such as the Estonian biobank, and the MoBa study in addition to the UK Biobank.
Furthermore, although we began the analysis using two focal traits; BMI and EDU, we can estimate the indirect effects of other traits which are suspected of having an indirect or environmental component, such as age at first birth.
Our investigation into the causal relationship that parental/-

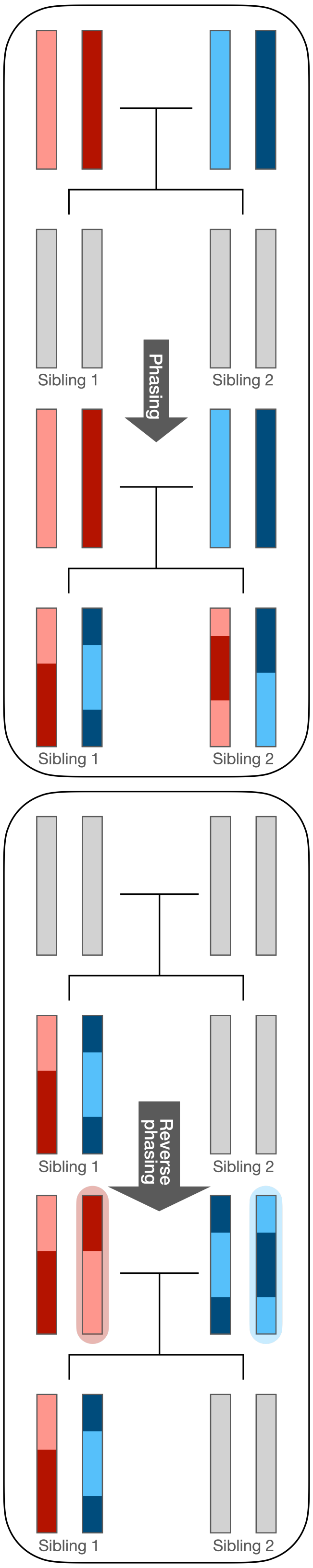


Figure 6.3: Graphical representation of pedigree haplotype phasing and reverse pedigree haplotype phasing. Top panel shows haplotype phasing of offspring genomes using parental genomes. Bottom panel shows reverse haplotype phasing using offspring genome to extract untransmitted alleles from the parents (highlighted in red and blue). Adapted from Robin Hofmeister.

family effects have on offspring traits (estimated untransmitted effects) can also benefit from curating more traits to be used as exposures in MR. This can include food-liking traits[94], or more specific SEP-proxy and social traits.

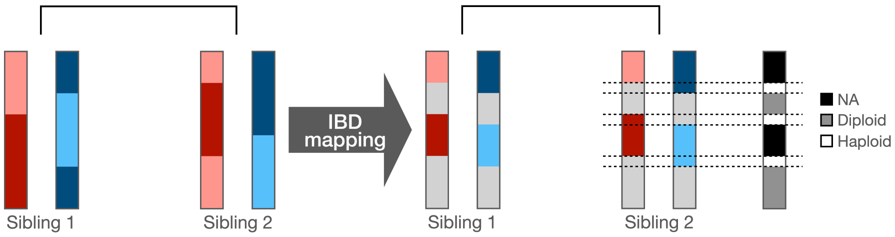


Figure 6.4: Graphical representation of IBD mapping between sibling pairs. Chromosomal areas in grey are not shared between either sibling. Taking sibling 1 as reference, grey areas on either chromosome for sibling 2 that are not shared could be either diploid for unknown alleles or haploid. Adapted from Robin Hofmeister.

6.4 Conclusion

In summary, the field of statistical genetics has undergone significant changes in recent years. There has been a notable increase in the number of GWAS studies, accompanied by much larger sample sizes. This shift has fundamentally improved our capacity to understand the genetic foundations of complex traits. It has also enhanced the accuracy of our genetic analyses, allowing us to gain a deeper understanding of the genetic factors contributing to various traits, and the interplay amongst these traits.

Moreover, Mendelian Randomisation (MR) has emerged as a powerful tool for causal inference, surpassing the limitations of traditional observational studies and randomised control trials. MR has not only provided a framework for causal analysis but has also evolved through multiple innovative extensions tackling its various assumption violations, including two such methods presented in this thesis, LHC-MR and PWC-MR, which enabled us to explore pleiotropic effects and potential confounders while improving the estimation of causal effects between traits.
While MR can be applied to virtually any pair of traits, the true value of the results emerges when we consider the broader biological context. It's imperative to ask questions about the relationship between these traits: Are there potential mediators that link them? Could unaccounted confounding factors distort our causal estimates? Is there any potential collider bias emerging due to the data collection or pre-processing methods used?

To better understand the complex web of cause-and-effect relationships, including what influences them and what might introduce biases, we can use Multivariable MR (MVMR). By extending beyond exposure-outcome pairs and incorporating multiple traits or trait components as exposures, MVMR allows us to disentangle true causal effects from the influence of confounders or mediators within a network of traits, providing us with a more holistic view on complex biological relationships.

In recent years, there has been a growing recognition within the scientific community regarding the advantages of incorporating family-based designs into GWAS. One of the primary advantages is the ability to obtain direct genetic effects, which can produce unbiased estimations of heritability and causal effects. Furthermore, family-based designs provide a unique opportunity to investigate indirect genetic effects. These effects capture the influence of one individual's genetics on the traits of another individual, often within a family context. For instance, researchers can explore how parental genetic factors play a role in shaping the traits of their offspring. This extends beyond the direct transmission of genetic information to encompass the complex interplay of genetic, environmental, and behavioural factors within families.

In conclusion, the convergence of several key factors – the rapid expansion of GWAS, the adaptable nature of MR and its innovative extensions, and the rigorous scrutiny of confounders and trait relationships in genetic investigations – has ushered the field of statistical genetics into an era of unparalleled exploration. These strides in genetic research provide a pathway towards a more precise and comprehensive understanding of the genetic (and environmental) underpinnings of complex traits and their causal associations.

Bibliography

Here are the references in citation order.

[1] D. S. Falconer and Trudy F. C. Mackay. *Introduction to quantitative genetics*. Longman, 1996 (cited on page 2).

[2] Nicholas J. Timpson et al. 'Genetic architecture: the shape of the genetic contribution to human traits and disease'. In: *Nature Reviews Genetics* 19.2 (2018), pp. 110–124 (cited on page 2).

[3] Eric S. Lander et al. 'Initial sequencing and analysis of the human genome'. In: *Nature* 409.6822 (2001), pp. 860–921 (cited on page 2).

[4] International Human Genome Sequencing Consortium. 'Finishing the euchromatic sequence of the human genome'. In: *Nature* 431.7011 (2004), pp. 931–945 (cited on page 2).

[5] David Altshuler, Peter Donnelly, and The International HapMap Consortium. 'A haplotype map of the human genome'. In: *Nature* 437.7063 (2005), pp. 1299–1320 (cited on page 3).

[6] Richard M. Durbin et al. 'A map of human genome variation from population-scale sequencing'. In: *Nature* 467.7319 (2010), pp. 1061–1073 (cited on pages 3, 4).

[7] Adam Auton et al. 'A global reference for human genetic variation'. In: *Nature* 526.7571 (2015), pp. 68–74 (cited on page 3).

[8] Monkol Lek et al. 'Analysis of protein-coding genetic variation in 60,706 humans'. In: *Nature* 536.7616 (2016), pp. 285–291 (cited on page 3).

[9] Konrad J. Karczewski et al. 'The ExAC browser: displaying reference data information from over 60 000 exomes'. In: *Nucleic Acids Research* 45.D1 (Nov. 2016), pp. D840–D845 (cited on page 3).

[10] Nicola Whiffin et al. 'Characterising the loss-of-function impact of 5'untranslated region variants in 15,708 individuals'. In: *Nature Communications* 11.1 (2020), p. 2523 (cited on page 3).

[11] Sergey Nurk et al. 'The complete sequence of a human genome'. In: *Science* 376.6588 (2022), pp. 44–53 (cited on page 3).

[12] Victor Guryev et al. 'Haplotype Block Structure Is Conserved across Mammals'. In: *PLOS Genetics* 2.7 (July 2006), pp. 1–8 (cited on page 4).

[13] Gustavo Glusman, Hannah C. Cox, and Jared C. Roach. 'Whole-genome haplotyping approaches and genomic medicine'. In: *Genome Medicine* 6.9 (2014), p. 73 (cited on page 4).

[14] Jonathan K. Pritchard and Molly Przeworski. 'Linkage Disequilibrium in Humans: Models and Data'. In: *The American Journal of Human Genetics* 69.1 (2001), pp. 1–14 (cited on page 4).

[15] John A Sved and William G Hill. 'One Hundred Years of Linkage Disequilibrium.' In: *Genetics* 209.3 (July 2018), pp. 629–636 (cited on page 4).

[16] Emil Uffelmann et al. ‘Genome-wide association studies’. In: *Nature Reviews Methods Primers* 1.1 (2021), p. 59 (cited on page 5).

[17] Cathie Sudlow et al. ‘UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age’. In: *PLOS Medicine* 12.3 (Mar. 2015), pp. 1–10 (cited on pages 7, 24).

[18] Neale Lab. *UK BioBank - round 2*. http://www.nealelab.is/uk-biobank/. 2018 (cited on page 7).

[19] Hakon Hakonarson, Jeffrey R Gulcher, and Kari Stefansson. ‘deCODE genetics, Inc.’ In: *Pharmacogenomics* 4.2 (Mar. 2003), pp. 209–215 (cited on page 7).

[20] Mitja I. Kurki et al. ‘FinnGen provides genetic insights from a well-phenotyped isolated population’. In: *Nature* 613.7944 (2023), pp. 508–518 (cited on page 7).

[21] The Psychiatric GWAS Consortium Steering Committee. ‘A framework for interpreting genome-wide association studies of psychiatric disorders’. In: *Molecular Psychiatry* 14.1 (2009), pp. 10–17 (cited on page 7).

[22] Cristen J Willer et al. ‘Six new loci associated with body mass index highlight a neuronal influence on body weight regulation’. In: *Nature Genetics* 41.1 (2009), pp. 25–34 (cited on page 7).

[23] Cristen J Willer et al. ‘Discovery and refinement of loci associated with lipid levels’. In: *Nature Genetics* 45.11 (2013), pp. 1274–1283 (cited on page 7).

[24] William S. Bush and Jason H. Moore. ‘Chapter 11: Genome-Wide Association Studies’. In: *PLOS Computational Biology* 8.12 (Dec. 2012), pp. 1–11 (cited on page 8).

[25] Abdel Abdellaoui et al. ‘15 years of GWAS discovery: Realizing the promise’. In: *The American Journal of Human Genetics* 110.2 (2023), pp. 179–194 (cited on page 7).

[26] Peter M. Visscher, William G. Hill, and Naomi R. Wray. ‘Heritability in the genomics era —concepts and misconceptions’. In: *Nature Reviews Genetics* 9.4 (2008), pp. 255–266 (cited on page 9).

[27] Jian Yang et al. ‘GCTA: a tool for genome-wide complex trait analysis.’ In: *Am J Hum Genet* 88.1 (Jan. 2011), pp. 76–82 (cited on page 9).

[28] Doug Speed et al. ‘Improved Heritability Estimation from Genome-wide SNPs’. In: *The American Journal of Human Genetics* 91.6 (Dec. 2012), pp. 1011–1021 (cited on page 9).

[29] Doug Speed, John Holmes, and David J. Balding. ‘Evaluating and improving heritability models using summary statistics’. In: *Nature Genetics* 52.4 (2020), pp. 458–462 (cited on page 9).

[30] Brendan Bulik-Sullivan et al. ‘An atlas of genetic correlations across human diseases and traits’. In: *Nature Genetics* 47.11 (2015), pp. 1236–1241. DOI: 10.1038/ng.3406 (cited on pages 9, 10).

[31] Vivian Tam et al. ‘Benefits and limitations of genome-wide association studies’. In: *Nature Reviews Genetics* 20.8 (2019), pp. 467–484 (cited on page 11).

[32] Ben Brumpton et al. ‘Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses’. In: *Nature Communications* 11.1 (2020), p. 3519 (cited on page 11).

[33] Matthew R Robinson et al. 'Population genetic differentiation of height and body mass index across Europe.' In: *Nat Genet* 47.11 (Nov. 2015), pp. 1357–1362 (cited on page 11).

[34] Richard Border et al. 'Assortative mating biases marker-based heritability estimators'. In: *Nature Communications* 13.1 (2022), p. 660 (cited on page 11).

[35] Beben Benyamin, Peter M Visscher, and Allan F McRae. 'Family-based genome-wide association studies.' In: *Pharmacogenomics* 10.2 (Feb. 2009), pp. 181–190 (cited on pages 11, 12).

[36] N Risch and J Teng. 'The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases I. DNA pooling.' In: *Genome Res* 8.12 (Dec. 1998), pp. 1273–1288 (cited on page 12).

[37] Jurg Ott, Yoichiro Kamatani, and Mark Lathrop. 'Family-based designs for genome-wide association studies'. In: *Nature Reviews Genetics* 12.7 (2011), pp. 465–474 (cited on page 12).

[38] Nicole M Warrington et al. 'Using structural equation modelling to jointly estimate maternal and fetal effects on birthweight in the UK Biobank'. In: *International Journal of Epidemiology* 47.4 (Feb. 2018), pp. 1229–1241 (cited on pages 12, 13).

[39] Laurence J. Howe et al. 'Within-sibship genome-wide association analyses decrease bias in estimates of direct genetic effects'. In: *Nature Genetics* 54.5 (2022), pp. 581–592 (cited on pages 12, 27, 38, 42, 49).

[40] Augustine Kong et al. 'The nature of nurture: Effects of parental genotypes'. In: *Science* 359.6374 (2018), pp. 424–428 (cited on page 13).

[41] A B HILL. 'The clinical trial.' In: *N Engl J Med* 247.4 (July 1952), pp. 113–119 (cited on page 14).

[42] Eleanor Sanderson et al. 'Mendelian randomization'. In: *Nature Reviews Methods Primers* 2.1 (2022), p. 6 (cited on page 14).

[43] Robert William Sanson-Fisher et al. 'Limitations of the Randomized Controlled Trial in Evaluating Population-Based Health Interventions'. In: *American Journal of Preventive Medicine* 33.2 (2007), pp. 155–161 (cited on page 14).

[44] Debbie A Lawlor et al. 'Mendelian randomization: using genes as instruments for making causal inferences in epidemiology.' In: *Stat Med* 27.8 (Apr. 2008), pp. 1133–1163 (cited on pages 14, 15, 17).

[45] Min Jeong Shin, Yoonsu Cho, and George Davey Smith. 'Alcohol Consumption, Aldehyde Dehydrogenase 2 Gene Polymorphisms, and Cardiovascular Health in Korea.' In: *Yonsei Med J* 58.4 (July 2017), pp. 689–696 (cited on page 15).

[46] Susanna C. Larsson et al. 'Alcohol Consumption and Cardiovascular Disease'. In: *Circulation: Genomic and Precision Medicine* 13.3 (2020), e002814 (cited on page 15).

[47] Kiran J. Biddinger et al. 'Association of Habitual Alcohol Intake With Risk of Cardiovascular Disease'. In: *JAMA Network Open* 5.3 (Mar. 2022), e223849–e223849 (cited on page 15).

[48] Jack Bowden, George Davey Smith, and Stephen Burgess. 'Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression.' In: *Int J Epidemiol* 44.2 (Apr. 2015), pp. 512–525 (cited on pages 16, 18, 19).

[49] Abraham Wald. 'The Fitting of Straight Lines if Both Variables are Subject to Error'. In: *The Annals of Mathematical Statistics* 11.3 (1940), pp. 284–300. (Visited on 08/15/2023) (cited on page 16).

[50] Stephen Burgess, Adam Butterworth, and Simon G Thompson. 'Mendelian randomization analysis with multiple genetic variants using summarized data.' In: *Genet Epidemiol* 37.7 (Nov. 2013), pp. 658–665 (cited on page 16).

[51] Stephen Burgess, Neil M. Davies, and Simon G. Thompson. 'Bias due to participant overlap in two-sample Mendelian randomization'. In: *Genetic Epidemiology* 40.7 (2016), pp. 597–608 (cited on page 17).

[52] Debbie A Lawlor. 'Commentary: Two-sample Mendelian randomization: opportunities and challenges'. In: *International Journal of Epidemiology* 45.3 (July 2016), pp. 908–915 (cited on page 17).

[53] Eleanor Sanderson, Wes Spiller, and Jack Bowden. 'Testing and correcting for weak and pleiotropic instruments in two-sample multivariable Mendelian randomization'. In: *Statistics in Medicine* 40.25 (2021), pp. 5434–5452 (cited on pages 17, 19, 53).

[54] Ninon Mounier and Zoltán Kutalik. 'Bias correction for inverse variance weighting Mendelian randomization'. In: *bioRxiv* (2022) (cited on page 18).

[55] Kyoko Watanabe et al. 'A global overview of pleiotropy and genetic architecture in complex traits'. In: *Nature Genetics* 51.9 (2019), pp. 1339–1348 (cited on page 18).

[56] Jack Bowden et al. 'Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator.' In: *Genet Epidemiol* 40.4 (May 2016), pp. 304–314 (cited on page 19).

[57] Fernando Pires Hartwig, George Davey Smith, and Jack Bowden. 'Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption'. In: *International Journal of Epidemiology* 46.6 (July 2017), pp. 1985–1998 (cited on page 19).

[58] Qingyuan Zhao et al. 'Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score'. English. In: *Annals of Statistics* 48.3 (July 2020) (cited on page 19).

[59] Jean Morrison et al. 'Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics.' In: *Nat Genet* 52.7 (July 2020), pp. 740–747 (cited on page 19).

[60] Marie Verbanck et al. 'Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases'. In: *Nature Genetics* 50.5 (2018), pp. 693–698 (cited on page 19).

[61] Yoonsu Cho et al. 'Exploiting horizontal pleiotropy to search for causal pathways within a Mendelian randomization framework'. In: *Nature Communications* 11.1 (2020), p. 1010 (cited on page 19).

[62] Stephen Burgess and Simon G. Thompson. 'Multivariable Mendelian Randomization: The Use of Pleiotropic Genetic Variants to Estimate Causal Effects'. In: *American Journal of Epidemiology* 181.4 (Jan. 2015), pp. 251–260 (cited on pages 19, 51).

[63] Eleanor Sanderson et al. 'An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings.' In: *Int J Epidemiol* 48.3 (June 2019), pp. 713–727 (cited on page 19).

[64] James R Staley and Stephen Burgess. 'Semiparametric methods for estimation of a nonlinear exposure-outcome relationship using instrumental variables with application to Mendelian randomization.' In: *Genet Epidemiol* 41.4 (May 2017), pp. 341–352 (cited on page 19).

[65] Jonathan Sulc, Jennifer Sjaarda, and Zoltán Kutalik. 'Polynomial Mendelian randomization reveals non-linear causal effects for obesity-related traits.' In: *HGG Adv* 3.3 (July 2022), p. 100124 (cited on page 19).

[66] Verena Zuber et al. 'Multi-response Mendelian randomization: Identification of shared and distinct exposures for multimorbidity and multiple related disease outcomes'. In: *The American Journal of Human Genetics* 110.7 (2023), pp. 1177–1199 (cited on page 19).

[67] Jack Bowden et al. 'Improving the accuracy of two-sample summary-data Mendelian randomization: moving beyond the NOME assumption'. In: *International Journal of Epidemiology* 48.3 (Dec. 2018), pp. 728–742 (cited on page 20).

[68] Gibran Hemani, Kate Tilling, and George Davey Smith. 'Orienting the causal relationship between imprecisely measured traits using GWAS summary data'. In: *PLOS Genetics* 13.11 (Nov. 2017), pp. 1–22 (cited on page 20).

[69] Jessica M. B. Rees et al. 'Robust methods in Mendelian randomization via penalization of heterogeneous causal estimates'. In: *PLOS ONE* 14.9 (Sept. 2019), pp. 1–24 (cited on page 20).

[70] Christopher N Foley et al. 'MR-Clust: clustering of genetic variants in Mendelian randomization with similar causal estimates'. In: *Bioinformatics* 37.4 (Sept. 2020), pp. 531–541 (cited on page 20).

[71] Liza Darrous et al. 'PheWAS-based clustering of Mendelian Randomisation instruments reveals distinct mechanism-specific causal effects between obesity and educational attainment'. In: *medRxiv* (2023) (cited on pages 20, 31).

[72] Liza Darrous, Ninon Mounier, and Zoltán Kutalik. 'Simultaneous estimation of bi-directional causal effects and heritable confounding from GWAS summary statistics'. In: *Nature Communications* 12.1 (2021), p. 7274 (cited on page 20).

[73] Kyung-Hyun Cho, Hye-Jeong Park, and Jae-Ryong Kim. 'Decrease in Serum HDL-C Level Is Associated with Elevation of Blood Pressure: Correlation Analysis from the Korean National Health and Nutrition Examination Survey 2017'. In: *International Journal of Environmental Research and Public Health* 17.3 (2020) (cited on page 22).

[74] David E. Laaksonen et al. 'Dyslipidaemia as a predictor of hypertension in middle-aged men'. In: *European Heart Journal* 29.20 (Feb. 2008), pp. 2561–2568 (cited on page 22).

[75] G. Hemani et al. 'The MR-Base platform supports systematic causal inference across the human phenome'. In: *eLife* 7 (2018), e34408. doi: 10.7554/eLife.34408 (cited on page 23).

[76] Andrew D. Grotzinger et al. 'Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits'. In: *Nature Human Behaviour* 3.5 (2019), pp. 513–525 (cited on pages 23, 42).

[77] James J Lee et al. 'Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals.' In: *Nat Genet* 50.8 (July 2018), pp. 1112–1121 (cited on page 24).

[78] F. Sofi et al. ‘Physical activity and risk of cognitive decline: a meta-analysis of prospective studies’. In: *Journal of Internal Medicine* 269.1 (2011), pp. 107–117 (cited on page 24).

[79] David A. Raichlen and Gene E. Alexander. ‘Adaptive Capacity: An Evolutionary Neuroscience Model Linking Exercise, Cognition, and Brain Health’. In: *Trends in Neurosciences* 40.7 (2017), pp. 408–421 (cited on page 24).

[80] Sarah J. Blondell, Rachel Hammersley-Mather, and J. Lennert Veerman. ‘Does physical activity prevent cognitive decline and dementia?: A systematic review and meta-analysis of longitudinal studies’. In: *BMC Public Health* 14.1 (2014), p. 510 (cited on page 24).

[81] Jonathan Sulc et al. ‘Composite trait Mendelian randomization reveals distinct metabolic and lifestyle consequences of differences in body shape’. In: *Communications Biology* 4.1 (2021), p. 1064 (cited on pages 29, 31, 45).

[82] Dan-Yu Lin and Patrick F Sullivan. ‘Meta-analysis of genome-wide association studies with overlapping subjects.’ In: *Am J Hum Genet* 85.6 (Dec. 2009), pp. 862–872 (cited on page 38).

[83] Aysu Okbay et al. ‘Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals’. In: *Nature Genetics* 54.4 (2022), pp. 437–449 (cited on page 38).

[84] Loic Yengo et al. ‘Meta-analysis of genome-wide association studies for height and body mass index in 700000 individuals of European ancestry.’ In: *Hum Mol Genet* 27.20 (Oct. 2018), pp. 3641–3649 (cited on page 38).

[85] Sven E. Ojavee et al. ‘Genetic insights into the age-specific biological mechanisms governing human ovarian aging’. In: *The American Journal of Human Genetics* (2023) (cited on page 45).

[86] Eleonora Porcu et al. ‘Triangulating evidence from longitudinal and Mendelian randomization studies of metabolomic biomarkers for type 2 diabetes’. In: *Scientific Reports* 11.1 (2021), p. 6197 (cited on page 45).

[87] Tom G Richardson et al. ‘Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: A multivariable Mendelian randomisation analysis.’ In: *PLoS Med* 17.3 (Mar. 2020), e1003062 (cited on page 51).

[88] Yuan-De Tan, Peng Xiao, and Chittibabu Guda. ‘In-depth Mendelian randomization analysis of causal factors for coronary artery disease’. In: *Scientific Reports* 10.1 (2020), p. 9208 (cited on page 51).

[89] Joshua A. Bell et al. ‘Effects of general and central adiposity on circulating lipoprotein, lipid, and metabolite levels in UK Biobank: A multivariable Mendelian randomization study’. In: *The Lancet Regional Health - Europe* 21 (2022), p. 100457 (cited on page 51).

[90] Tom G Richardson et al. ‘Use of genetic variation to separate the effects of early and later life adiposity on disease risk: mendelian randomisation study’. In: *BMJ* 369 (2020) (cited on page 52).

[91] Grace M. Power et al. ‘Mendelian randomization analyses suggest childhood body size indirectly influences end points from across the cardiovascular disease spectrum through adult body size’. In: *Journal of the American Heart Association* 10.17 (2021) (cited on page 52).

[92] Tom G Richardson et al. 'Evaluating the direct effects of childhood adiposity on adult systemic metabolism: a multivariable Mendelian randomization analysis'. In: *International Journal of Epidemiology* 50.5 (Mar. 2021), pp. 1580–1592 (cited on page 52).

[93] Ninon Mounier and Zoltán Kutalik. 'bGWAS: an R package to perform Bayesian genome wide association studies'. In: *Bioinformatics* 36.15 (May 2020), pp. 4374–4376 (cited on page 53).

[94] Sebastian May-Wilson et al. 'Large-scale GWAS of food liking reveals genetic determinants and genetic correlations with distinct neurophysiological traits'. In: *Nature Communications* 13.1 (2022), p. 2743 (cited on page 56).

Appendix

Supplementary Figures

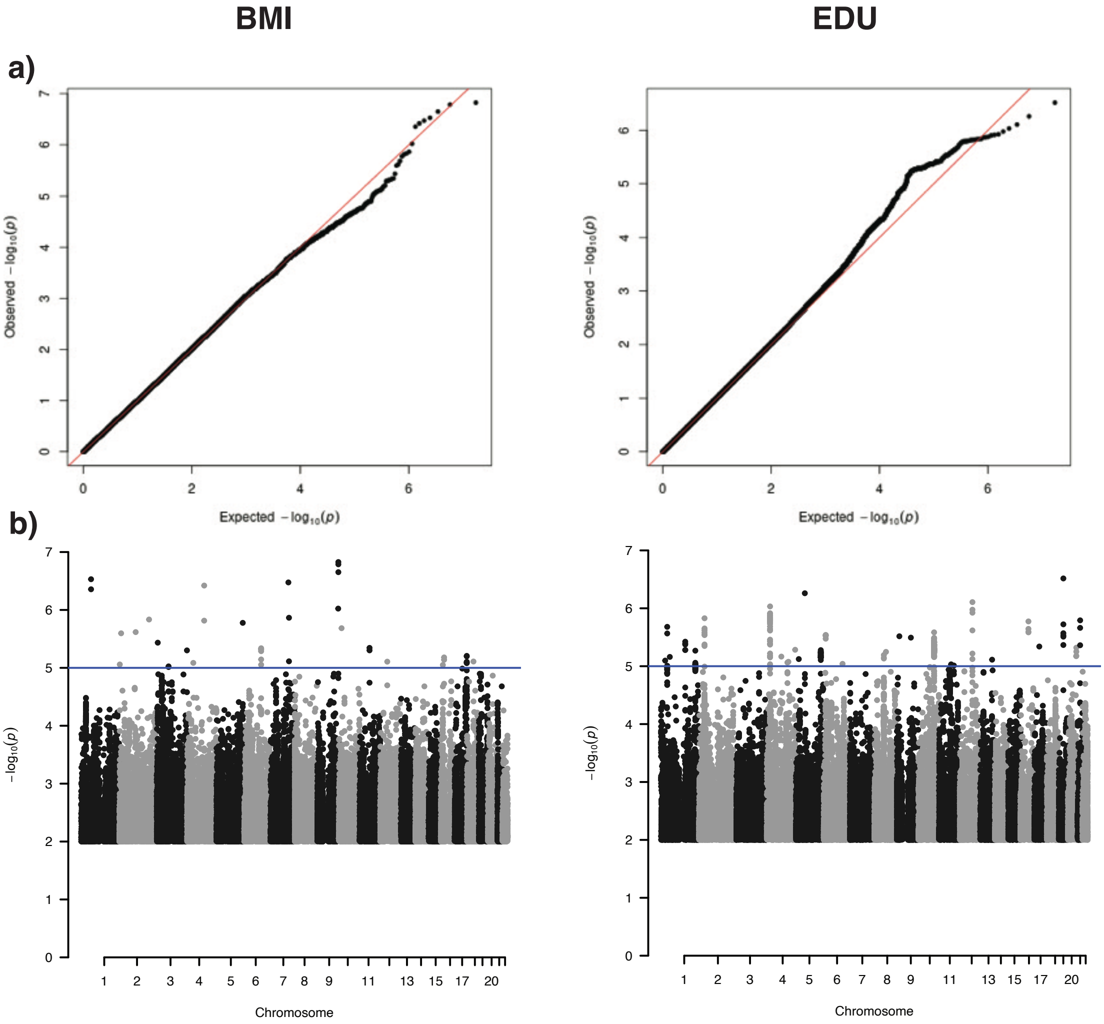
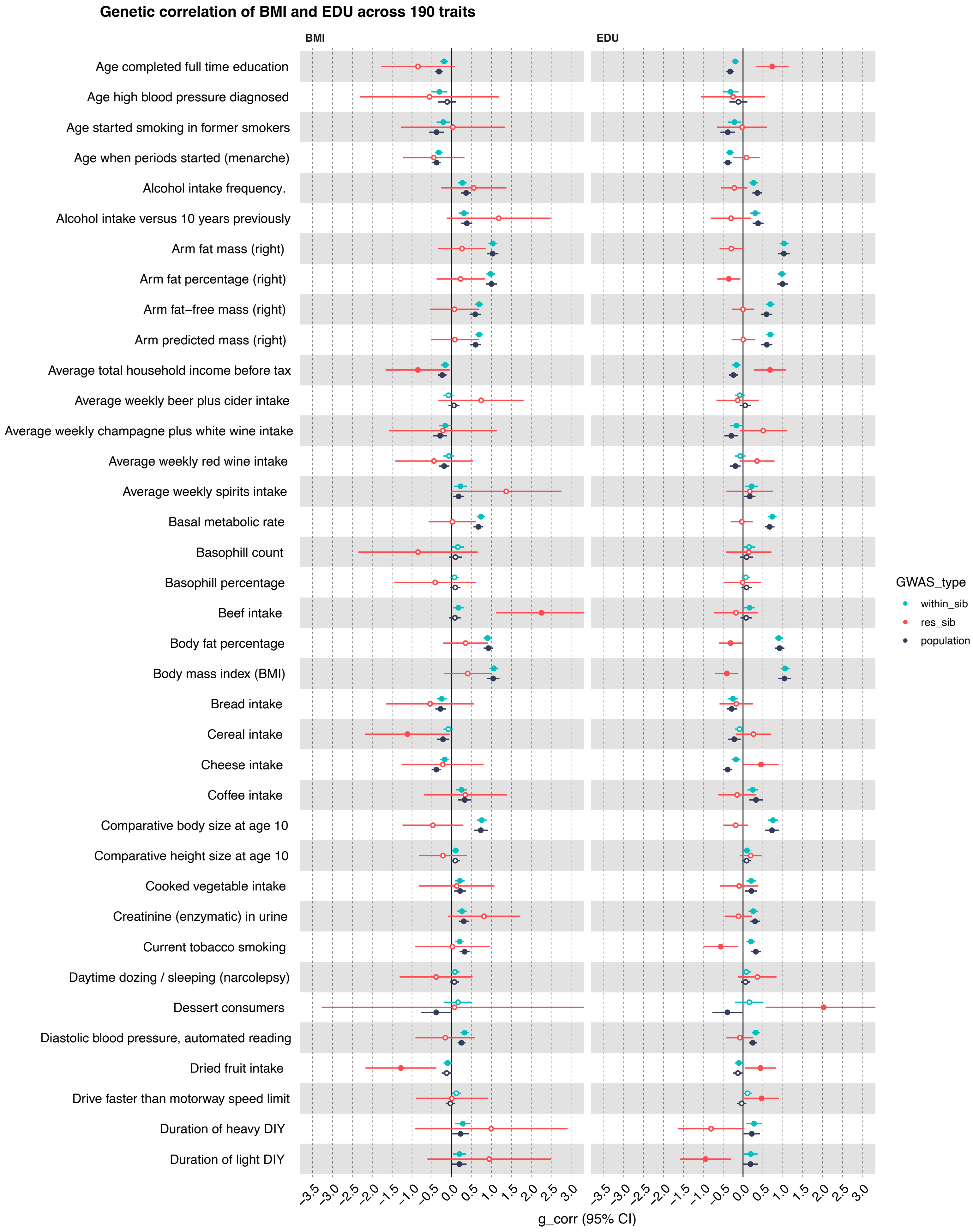
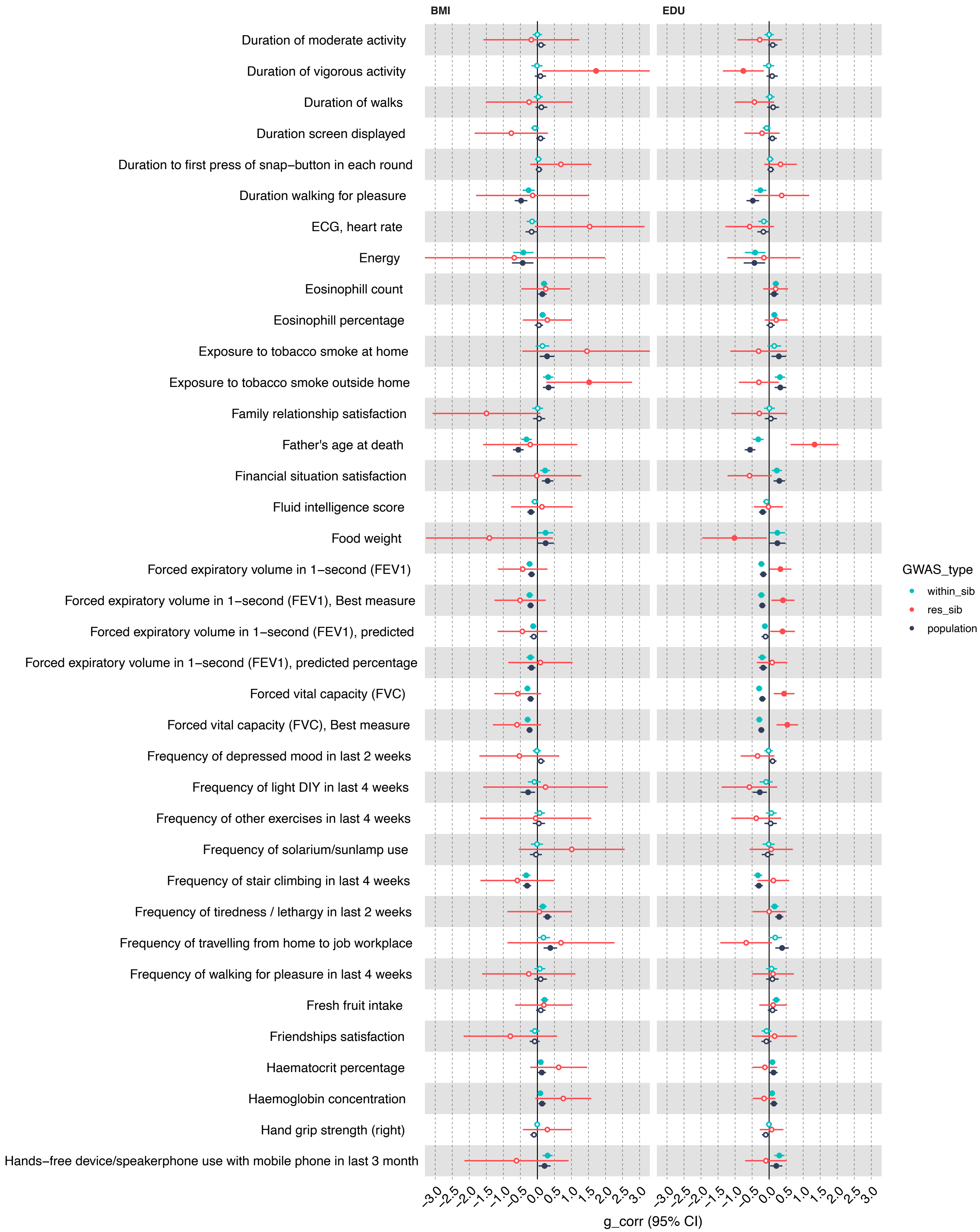
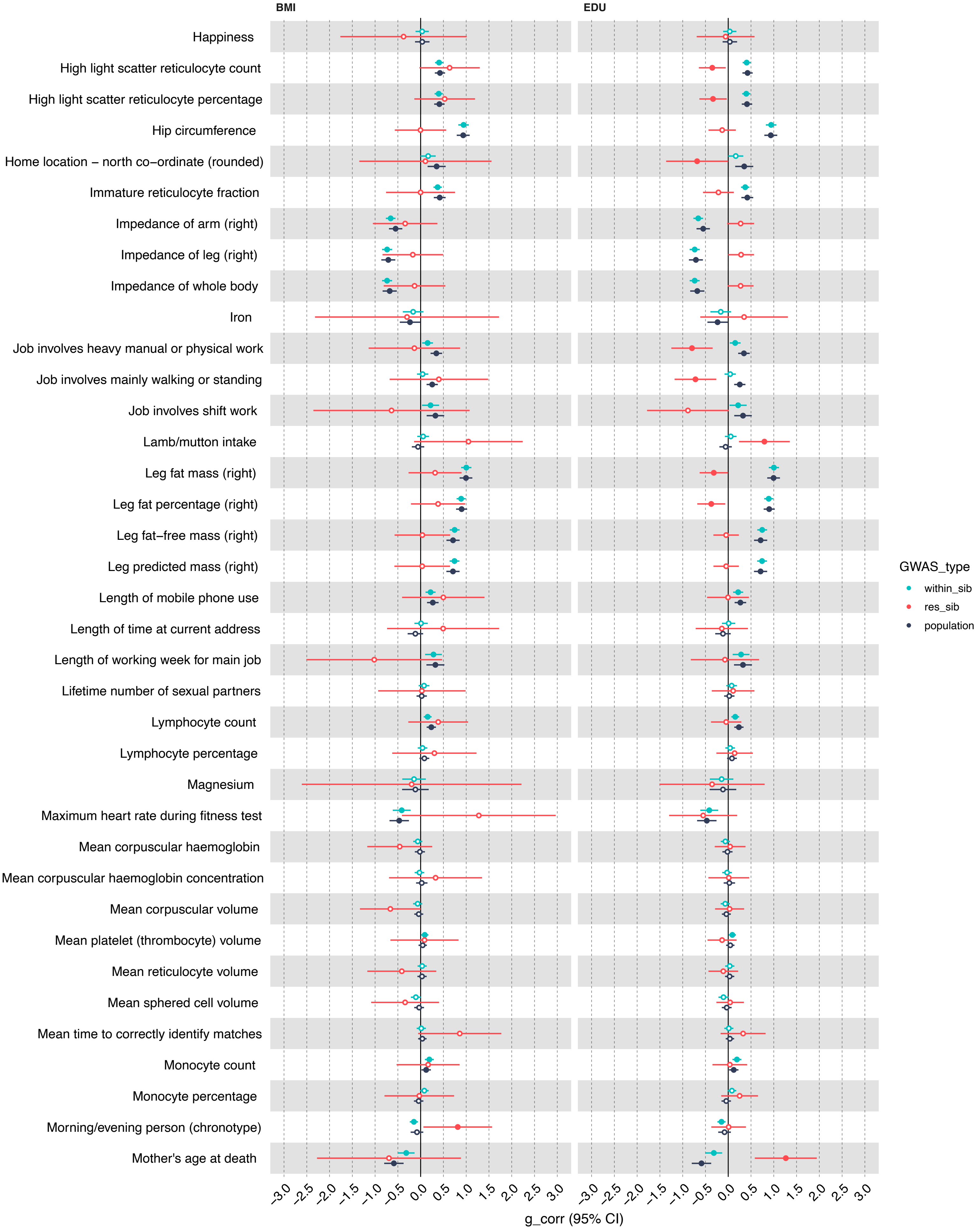


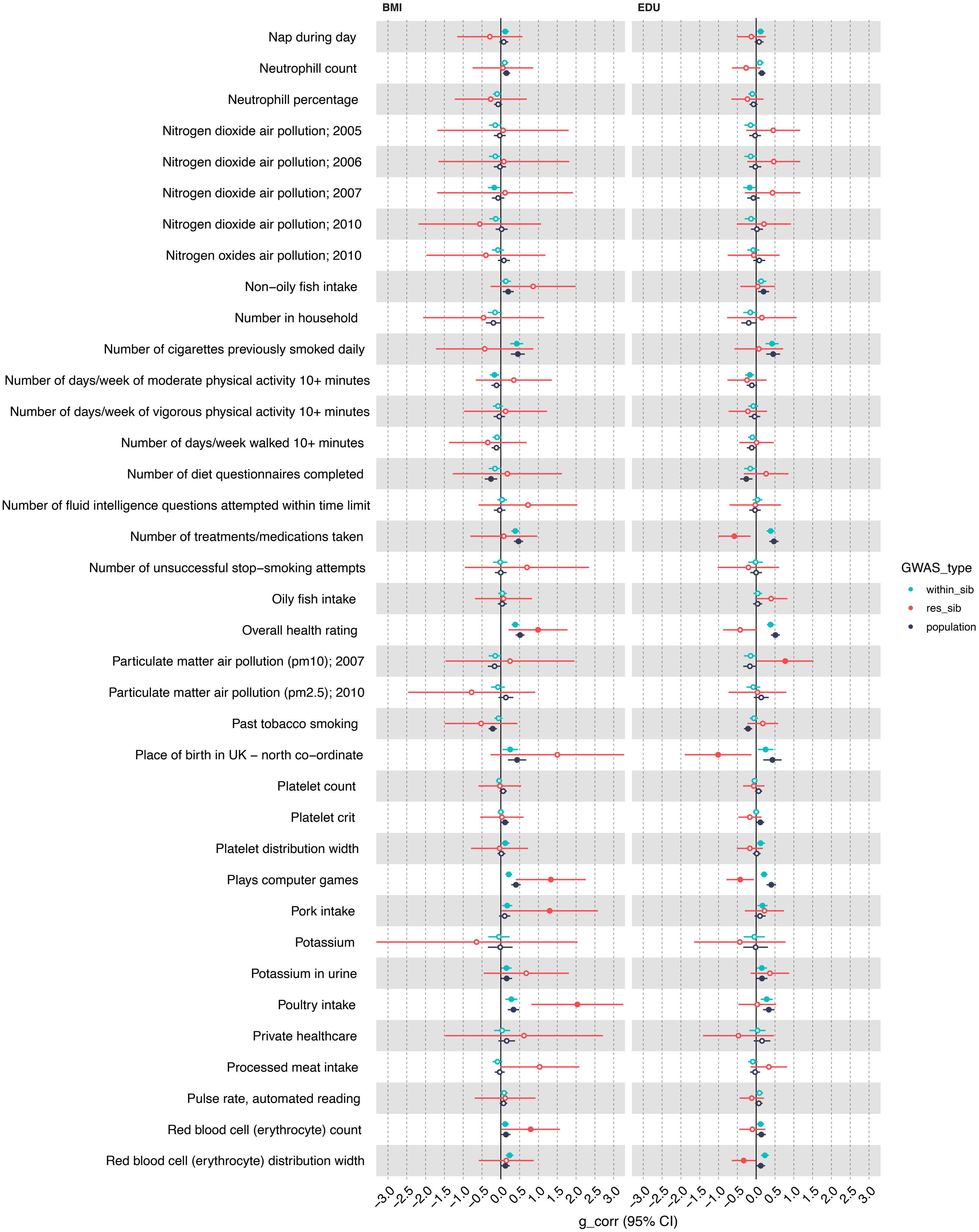
Figure 0.1: QQ-plot and Manhattan plot for untransmitted BMI and EDU effects (meta-analysed sibling effects). **a** QQ-plot of GWAS P-values. **b** Manhattan plot of selected SNPs with $-log_{10}(P) \geq 2$.



Genetic correlation of BMI and EDU across 190 traits
BMI
EDU
Age completed full time education
Age high blood pressure diagnosed
Age started smoking in former smokers
Age when periods started (menarche)
Alcohol intake frequency.
Alcohol intake versus 10 years previously
Arm fat mass (right)
Arm fat percentage (right)
Arm fat–free mass (right)
Arm predicted mass (right)
Average total household income before tax
Average weekly beer plus cider intake
Average weekly champagne plus white wine intake
Average weekly red wine intake
Average weekly spirits intake
Basal metabolic rate
Basophill count
Basophill percentage
Beef intake
Body fat percentage
Body mass index (BMI)
Bread intake
Cereal intake
Cheese intake
Coffee intake
Comparative body size at age 10
Comparative height size at age 10
Cooked vegetable intake
Creatinine (enzymatic) in urine
Current tobacco smoking
Daytime dozing / sleeping (narcolepsy)
Dessert consumers
Diastolic blood pressure, automated reading
Dried fruit intake
Drive faster than motorway speed limit
Duration of heavy DIY
Duration of light DIY
-3.5 -3.0 -2.5 -2.0 -1.5 -1.0 -0.5 0.0 0.5 1.0 1.5 2.0 2.5 3.0
-3.5 -3.0 -2.5 -2.0 -1.5 -1.0 -0.5 0.0 0.5 1.0 1.5 2.0 2.5 3.0
g_corr (95% CI)
GWAS_type
within_sib
res_sib
population


BMI
EDU
Duration of moderate activity
Duration of vigorous activity
Duration of walks
Duration screen displayed
Duration to first press of snap−button in each round
Duration walking for pleasure
ECG, heart rate
Energy
Eosinophill count
Eosinophill percentage
Exposure to tobacco smoke at home
Exposure to tobacco smoke outside home
Family relationship satisfaction
Father's age at death
Financial situation satisfaction
Fluid intelligence score
Food weight
Forced expiratory volume in 1−second (FEV1)
Forced expiratory volume in 1−second (FEV1), Best measure
Forced expiratory volume in 1−second (FEV1), predicted
Forced expiratory volume in 1−second (FEV1), predicted percentage
Forced vital capacity (FVC)
Forced vital capacity (FVC), Best measure
Frequency of depressed mood in last 2 weeks
Frequency of light DIY in last 4 weeks
Frequency of other exercises in last 4 weeks
Frequency of solarium/sunlamp use
Frequency of stair climbing in last 4 weeks
Frequency of tiredness / lethargy in last 2 weeks
Frequency of travelling from home to job workplace
Frequency of walking for pleasure in last 4 weeks
Fresh fruit intake
Friendships satisfaction
Haematocrit percentage
Haemoglobin concentration
Hand grip strength (right)
Hands−free device/speakerphone use with mobile phone in last 3 month
−3.0 −2.5 −2.0 −1.5 −1.0 −0.5 0.0 0.5 1.0 1.5 2.0 2.5 3.0
−3.0 −2.5 −2.0 −1.5 −1.0 −0.5 0.0 0.5 1.0 1.5 2.0 2.5 3.0
g_corr (95% CI)
GWAS_type
within_sib
res_sib
population


BMI
EDU
Happiness
High light scatter reticulocyte count
High light scatter reticulocyte percentage
Hip circumference
Home location – north co–ordinate (rounded)
Immature reticulocyte fraction
Impedance of arm (right)
Impedance of leg (right)
Impedance of whole body
Iron
Job involves heavy manual or physical work
Job involves mainly walking or standing
Job involves shift work
Lamb/mutton intake
Leg fat mass (right)
Leg fat percentage (right)
Leg fat–free mass (right)
Leg predicted mass (right)
Length of mobile phone use
Length of time at current address
Length of working week for main job
Lifetime number of sexual partners
Lymphocyte count
Lymphocyte percentage
Magnesium
Maximum heart rate during fitness test
Mean corpuscular haemoglobin
Mean corpuscular haemoglobin concentration
Mean corpuscular volume
Mean platelet (thrombocyte) volume
Mean reticulocyte volume
Mean sphered cell volume
Mean time to correctly identify matches
Monocyte count
Monocyte percentage
Morning/evening person (chronotype)
Mother's age at death
-3.0
-2.5
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
g_corr (95% CI)
GWAS_type
within_sib
res_sib
population


BMI
EDU
Nap during day
Neutrophill count
Neutrophill percentage
Nitrogen dioxide air pollution; 2005
Nitrogen dioxide air pollution; 2006
Nitrogen dioxide air pollution; 2007
Nitrogen dioxide air pollution; 2010
Nitrogen oxides air pollution; 2010
Non−oily fish intake
Number in household
Number of cigarettes previously smoked daily
Number of days/week of moderate physical activity 10+ minutes
Number of days/week of vigorous physical activity 10+ minutes
Number of days/week walked 10+ minutes
Number of diet questionnaires completed
Number of fluid intelligence questions attempted within time limit
Number of treatments/medications taken
Number of unsuccessful stop−smoking attempts
Oily fish intake
Overall health rating
Particulate matter air pollution (pm10); 2007
Particulate matter air pollution (pm2.5); 2010
Past tobacco smoking
Place of birth in UK – north co−ordinate
Platelet count
Platelet crit
Platelet distribution width
Plays computer games
Pork intake
Potassium
Potassium in urine
Poultry intake
Private healthcare
Processed meat intake
Pulse rate, automated reading
Red blood cell (erythrocyte) count
Red blood cell (erythrocyte) distribution width
−3.0 −2.5 −2.0 −1.5 −1.0 −0.5 0.0 0.5 1.0 1.5 2.0 2.5 3.0
−3.0 −2.5 −2.0 −1.5 −1.0 −0.5 0.0 0.5 1.0 1.5 2.0 2.5 3.0
g_corr (95% CI)
GWAS_type
● within_sib
● res_sib
● population

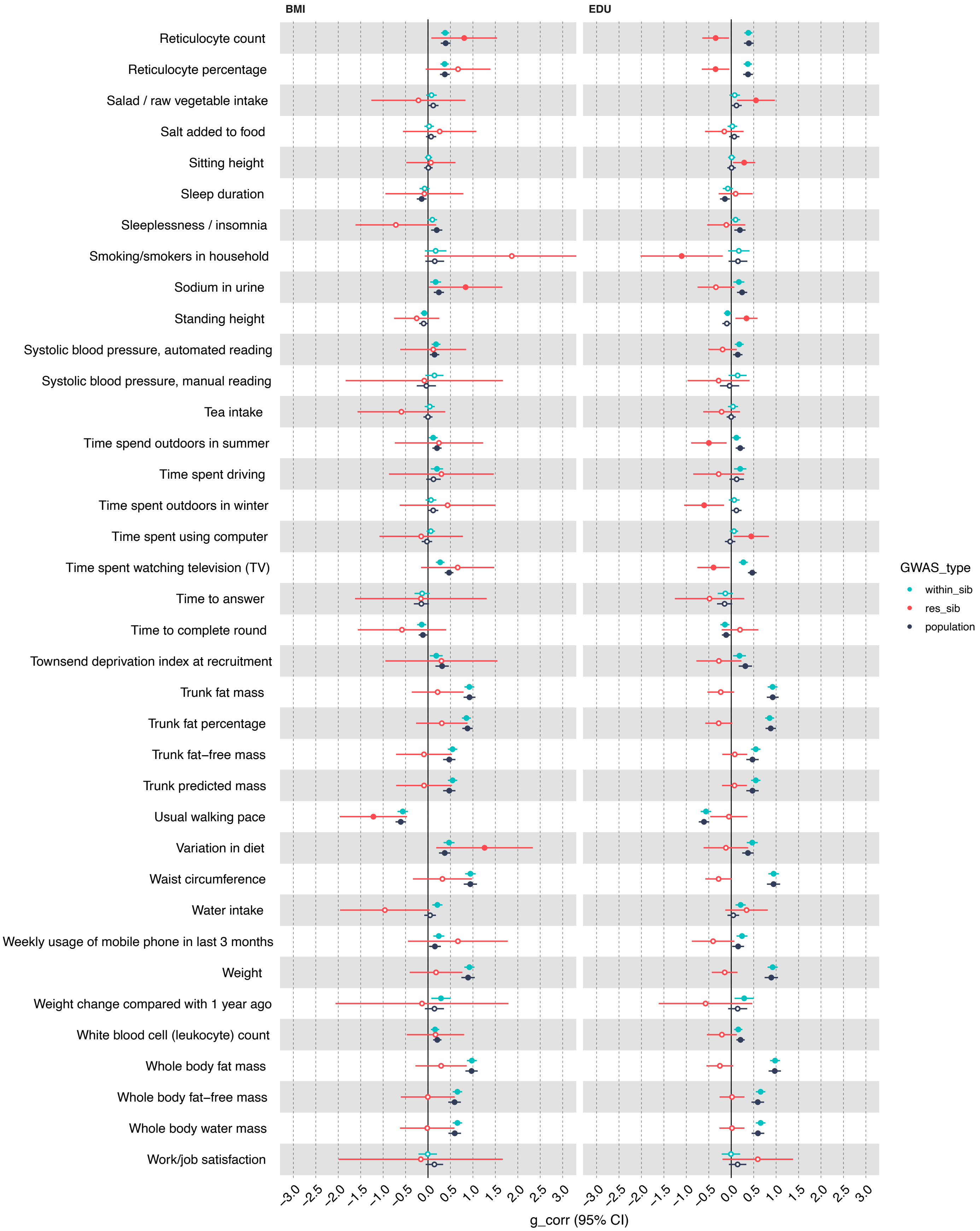


Figure 0.2: Forest plot of the genetic correlation of both BMI and EDU across ∼190 **other traits.** The effects of both BMI and EDU come from 3 different sources of GWAS: population-based, within-sibship, sibling meta-analysed indirect effects. The effects of other traits all come from population-based GWAS. 95% confidence intervals are shown as error bars. Points that are not filled indicate a non-significant genetic correlation estimate.

Appendix A: Simultaneous estimation of bi-directional causal effects and heritable confounding from GWAS summary statistics

This article is presented in chapter 2.1.

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-021-26970-w OPEN

Simultaneous estimation of bi-directional causal effects and heritable confounding from GWAS summary statistics

Liza Darrous[1,2,4], Ninon Mounier[1,2,4] & Zoltán Kutalik[1,2,3]✉

Mendelian Randomisation (MR) is an increasingly popular approach that estimates the causal effect of risk factors on complex human traits. While it has seen several extensions that relax its basic assumptions, most suffer from two major limitations; their under-exploitation of genome-wide markers, and sensitivity to the presence of a heritable confounder of the exposure-outcome relationship. To overcome these limitations, we propose a Latent Heritable Confounder MR (LHC-MR) method applicable to association summary statistics, which estimates bi-directional causal effects, direct heritabilities, and confounder effects while accounting for sample overlap. We demonstrate that LHC-MR outperforms several existing MR methods in a wide range of simulation settings and apply it to summary statistics of 13 complex traits. Besides several concordant results with other MR methods, LHC-MR unravels new mechanisms (how disease diagnosis might lead to improved lifestyle) and reveals new causal effects (e.g. HDL cholesterol being protective against high systolic blood pressure), hidden from standard MR methods due to a heritable confounder of opposite effect direction.

[1] University Center for Primary Care and Public Health, University of Lausanne, Lausanne, Switzerland. [2] Swiss Institute of Bioinformatics, Lausanne, Switzerland. [3] Department of Computational Biology, University of Lausanne, Lausanne, Switzerland. [4] These authors contributed equally: Liza Darrous, Ninon Mounier. ✉email: zoltan.kutalik@unil.ch

The identification of frequent risk factors and the quantification of their impact on common diseases is a principal quest for public health policy makers. Epidemiological studies aim to address this issue, but they are most often based on observational data due to their abundance over the years. Despite major methodological advances, a large majority of such studies have inherent limitations and suffer from confounding and reverse causation[1,2]. For these reasons, many of the reported associations found in classical epidemiological studies are mere correlates of disease risk, rather than causal factors directly involved in disease progression. Due to this, additional evidence is required before developing public health interventions in a bid to reduce the future burden of diseases. While well-designed and carefully conducted randomised control trials (RCTs) remain the gold standard for causal inference, they are exceedingly expensive, time-consuming, may not be feasible for ethical reasons, and have high failure rates[3,4].

Mendelian randomisation (MR), a natural genetic counterpart to RCTs, is an instrumental variable (IV) technique used to infer the strength of a causal relationship between a risk factor (X) and an outcome (Y)[5]. To do so, it uses genetic variants (G) as instruments and relies on three major assumptions (see Supplementary Fig. 1): (1) Relevance—G is robustly associated with the exposure. (2) Exchangeability—G is not associated with any confounder of the exposure-outcome relationship. (3) Exclusion restriction—G is independent of the outcome conditional on the exposure and all confounders of the exposure-outcome relationship (i.e. the only path between the instrument and the outcome is via the exposure).

The advantage of the MR approach is that for most heritable exposures, dozens (if not hundreds) of genetic instruments are known to date thanks to well-powered genome-wide association studies (GWASs). Each instrument can provide a causal effect estimate, which can be combined with others, by using an inverse variance-weighting (IVW) scheme (e.g. Burgess et al.[6]). However, the last assumption is particularly problematic, because genetic variants tend to be pleiotropic, i.e. exert effect on multiple traits independently. Still, it can be shown that if the instrument strength is independent of the direct effect on the outcome (InSIDE assumption) and the direct effects are on average zero, IVW-based methods will still yield consistent estimates. Methods, such as MR-Egger[7], produce consistent estimates even if direct effects are allowed to have a non-zero offset. The third assumption can be further reduced to assuming that >50% of the instruments (or in terms of their weight) are valid (median-based estimators[8]) or that zero-pleiotropy instruments are the most frequent (mode-based estimators[9]).

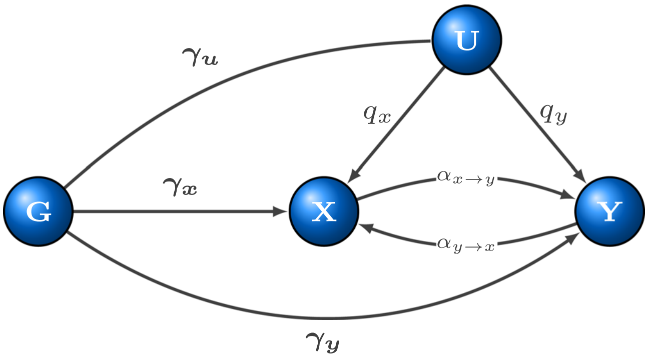


Fig. 1 Schematic representation of the extended structural equation model (SEM). X and Y are two complex traits under scrutiny with a latent (heritable) confounder U with causal effects q_x and q_y on them. G represents genetic variants, with effects $\boldsymbol{\gamma_x}$, $\boldsymbol{\gamma_y}$ and $\boldsymbol{\gamma_u}$, respectively. Traits X and Y have causal effects on each other, which are denoted by $\alpha_{x\to y}$ and $\alpha_{y\to x}$.

The InSIDE assumption (i.e. horizontal pleiotropic effects ($G \to Y$) are independent of the direct effect ($G \to X$)) is reasonable if the pleiotropic path $G \to Y$ does not branch off to X. However, if there is such a branching off, the variable representing the split is a confounder of the $X - Y$ relationship and we fall back on the violation of the second assumption (exchangeability), making it the most problematic. Therefore, in this paper, we extend the standard MR model to incorporate the presence of a latent (i.e. unmeasured) heritable confounder (U) and estimate its contribution to traits X and Y, while simultaneously estimating the bi-directional causal effect between the two traits. Standard MR methods are vulnerable to such heritable confounders, since any genetic marker directly associated with the confounder may be selected as an instrument for the exposure. However, such instruments will have a direct effect on the outcome that is correlated to their instrument strength, violating the InSIDE assumption and biasing the causal effect estimate.

In this paper, we first introduce the extended MR model and derive the likelihood function for the observed genome-wide summary statistics (for X and Y). We then test and compare the method against conventional and more advanced (such as CAUSE[10] and MR-RAPS[11]) MR approaches through extensive simulation settings, including several violations of the model assumptions. Finally, the approach is applied to association summary statistics (based on the UK Biobank and meta-analysis studies) of 13 complex traits to re-assess all pairwise bi-directional causal relationships between them.

Results

Overview of the method. We set up a structural equation model (SEM) (Fig. 1) and derived how its parameters are linked to genome-wide association summary statistics of two studied complex traits. We then maximised the resulting likelihood function in order to estimate bi-directional causal effects between them (for details see Methods), in addition to inferring direct heritabilities for X and Y, confounder effects, cross-trait and individual trait LD-score intercepts and the polygenicity for X and Y. All SNPs associated with the heritable confounder (U) are indirectly associated with X and Y with effects that are proportional (ratio q_y/q_x). SNPs that are directly associated with X (and not with U) are also associated with Y with proportional effects (ratio $1/\alpha_{x\to y}$). Finally, SNPs that are directly Y-associated are also X-associated with a proportionality ratio of $1/\alpha_{y\to x}$. These three groups of SNPs are illustrated on the β_x-vs-β_y scatter plot (Supplementary Fig. 2). In simple terms, the aim of our method is to identify the different clusters, estimate the slopes and distinguish which corresponds to the causal- and confounder effects. In this paper, we focus on the properties of the maximum likelihood estimates (MLEs) (and their variances) for the bi-directional causal effects arising from our SEM.

Simulation results. We started off with a realistic simulation setting of 234,000 SNPs on chromosome 10 (LD patterns used from the UK10K panel) and 50,000 samples for both traits. Traits X, Y and confounder U had average polygenicity ($\pi_x = 5 \times 10^{-3}$, $\pi_y = 1 \times 10^{-2}$, $\pi_u = 5 \times 10^{-2}$), with substantial direct heritability for X and Y ($h_x^2 = 0.25, h_y^2 = 0.2$), mild confounding on X and Y ($t_x = 0.16, t_y = 0.11$, where $t_x = \sqrt{h_u^2 \cdot q_x^2}$ and $t_y = \sqrt{h_u^2 \cdot q_y^2}$), and a causal effect between X and Y ($\alpha_{x\to y} = 0.3, \alpha_{y\to x} = 0$). Note that with these settings, SNPs associated with U would violate the InSIDE assumption but might still be used by conventional MR methods. Under this standard setting, there were no genome-wide significant SNPs for standard MR methods, and estimates derived using SNPs with a p-value $< 5 \times 10^{-6}$ showed a downward bias for all MR methods (Fig. 2a). MR-RAPS using filtered SNPs (p-value $< 5 \times 10^{-4}$) was similarly downward biased whereas MR-RAPS using the entire set of SNPs was upward biased with the least amount of variance compared to all methods including LHC-MR. LHC-MR in this scenario slightly overestimated the causal effect in comparison but had the smallest RMSE after MR-RAPS (0.13 vs 0.06, Supplementary Data 1).

We ran all our simulation scenarios with a smaller and a larger sample size (50,000 and 500,000) and observed that the relative performance of the methods were in some cases sample size specific. Smaller sample sizes often meant that standard MR methods had little to no IVs reaching genome-wide (GW) significance and hence we were forced to use IVs from less stringent thresholds ($< 5 \times 10^{-4}$ and $< 5 \times 10^{-6}$). Therefore, the causal effects were estimated with a substantial downward bias due to weak instrument bias (and winner's curse). LHC-MR in these cases was able to estimate the causal effect with less bias but with a larger variance compared to most standard MR methods—still outperforming them in terms of RMSE in most settings. In the larger sample size setting, standard MR methods had IVs for every threshold cutoff. However, a pattern also observed with smaller sample sizes—but to a lesser extent—emerged, where the causal estimates of some methods changed (either in mean or in variance, most noticeably observed in weighted median and IVW) as the threshold became more stringent. This is of particular concern and highlights that while in this simulation setting the 5×10^{-8} threshold may have optimally cancelled out the different biases for IVW (downward bias due to winner's curse and weak instrument bias, upward bias due to genetic confounding), its estimate remains strongly setting-dependent. LHC-MR performed reasonably well, exhibiting lower RMSE than most other methods, except for IVW and MR-RAPS for the 5×10^{-4} threshold (Supplementary Fig. 4a). However, we observed that the performance of MR-RAPs is particularly setting and threshold dependent.

Furthermore, unequal sample sizes for the two traits showed an underestimation of the causal effects for almost all MR methods, while LHC-MR remained the most accurate in the case where n_x (50,000) was smaller than n_y (500,000). However, the performances in the reverse scenario, where n_x was larger in size, were akin to the large sample size standard setting, where only IVW and filtered MR-RAPS ($<5 \times 10^{-4}$) showed superior performance to LHC-MR both in terms of bias and variance (see Supplementary Fig. 5).

When testing scenarios in the absence of a causal or a confounder effect (imitating the classical MR assumptions), with a smaller causal effect ($\alpha_{x \to y} = 0.1$), or with both forward- and reverse causal effects, we note that LHC-MR outperforms the standard MR methods as well as MR-RAPS in all these scenarios.

When there was no causal effect ($\alpha_{x \to y} = 0$), LHC-MR had the smallest bias out of all the methods in both sample sizes (0.004 in both, Supplementary Fig. 6a and Supplementary Fig. 7a). The variance of the LHC-MR estimates in the larger sample size was much lower (0.0001 vs 0.01), similarly the other methods had a smaller variance in the larger sample size and had more clearly seen upward biased estimates. The increased upward bias of standard MR methods is due to the fact that confounder-associated SNPs could only be detected in the larger sample size and those lead to positive bias (due to the concordant effect of the confounder on the two traits). Note that the variances of standard MR methods are low simply because, in these settings, we were forced to lower the instrument selection threshold, hence artificially included many (potentially invalid) instruments, which lowers the estimator variance while increasing bias. MR-RAPS greatly overestimates the causal effects when the sample size is larger.

In the absence of a confounder effect, there is not much of a difference between the two sample sizes; standard MR methods have a large variance and are downward biased, LHC-MR is less biased compared to them but MR-RAPS performs best with the least bias and variance when all the SNPs are used as instruments (Supplementary Fig. 6b and Supplementary Fig. 7b). Trying a smaller causal effect led to an upward bias for all MR methods including both filterings of MR-RAPS in the larger sample size. Alternately, when $n_x = n_y = 50,000$, the MR methods are downward biased (Supplementary Figs. 6c and 7c). Lastly, when a (negative) reverse causal effect is introduced, all MR methods and MR-RAPS are negatively biased in their estimation of the causal effect (see Fig. 2b). LHC-MR has a much smaller bias for the forward causal effect estimate in this case, and a generally small bias for the reverse causal effect in both sample sizes (0.05 for $n = 50,000$ and 0.03 for $n = 500,000$, Supplementary Fig. 4b).

Increasing the indirect genetic effects, by intensifying the contribution of the confounder to X and Y ($t_x = 0.41, t_y = 0.27$), led to a general overestimation of the causal effects by all methods including LHC-MR, though more drastically seen in standard MR methods and MR-RAPS in the larger sample size, when there is sufficient power to pick up these confounder-associated SNPs. The causal effect estimates of standard MR methods in the smaller sample size were much less affected by the presence of a strong confounder compared to LHC-MR and MR-RAPS (Supplementary Fig. 8). The reason for this is that the confounder-associated SNPs remain undetectable at lower sample size and hence instruments will not violate the classical MR assumptions.

Further testing the effects of the confounder trait on the causal estimation, we tested the impact of confounders with opposite effects on X and Y. We observe a major underestimation of the causal effects for standard MR methods as well as MR-RAPS, whereas LHC-MR performs better for both sample sizes (RMSE = 0.01 and 0.1 for larger and smaller n respectively), see Fig. 2c and Supplementary Fig. 4c.

Our LHC-MR method is influenced by the unlikely scenario of extreme polygenicity for traits X, Y and U, and it suffers from increased bias and variance regardless of sample size (see Supplementary Fig. 9). Standard MR methods as well as filtered MR-RAPS underestimated the causal effect when $n = 50,000$. Some also underestimated $\alpha_{x \to y}$ when $n = 500,000$, with the exception of IVW, Mode and filtered MR-RAPS, that outperformed the rest. Decreasing the proportion of confounder-associated SNPs to 1% only, does not seem to affect our method and shows similar results to the standard setting (Supplementary Fig. 10).

Furthermore, we simulated summary statistics, where (contrary to our modelling assumptions) the $X - Y$ relationship has two confounders, U_1 and U_2. When the ratio of the causal effects of these two confounders on X and Y ($q_y^{(1)}/q_x^{(1)}$ and $q_y^{(2)}/q_x^{(2)}$, respectively) agreed in sign, the corresponding causal effects of standard MR methods were overestimated in the larger sample size and, conversely, underestimated in the smaller sample size (Supplementary Figs. 11a and 12a). LHC-MR and weighted median performed better however in the larger sample size and had a bias of 0.03 and 0.07, respectively. However, when the signs were opposite ($q_x^{(1)} = 0.3, q_y^{(1)} = 0.2$ for U_1 and $q_x^{(2)} = 0.3, q_y^{(2)} = -0.2$ for U_2), conventional MR methods and MR-RAPS in this case almost all underestimated the causal effect regardless of sample size. LHC-MR outperformed them both in the larger sample size (bias of 0.007) and in the smaller sample size (bias of −0.003), see Supplementary Figs. 11b and 12b.

Finally, we explored how sensitive our method is to different violations of our modelling assumptions. First, we simulated

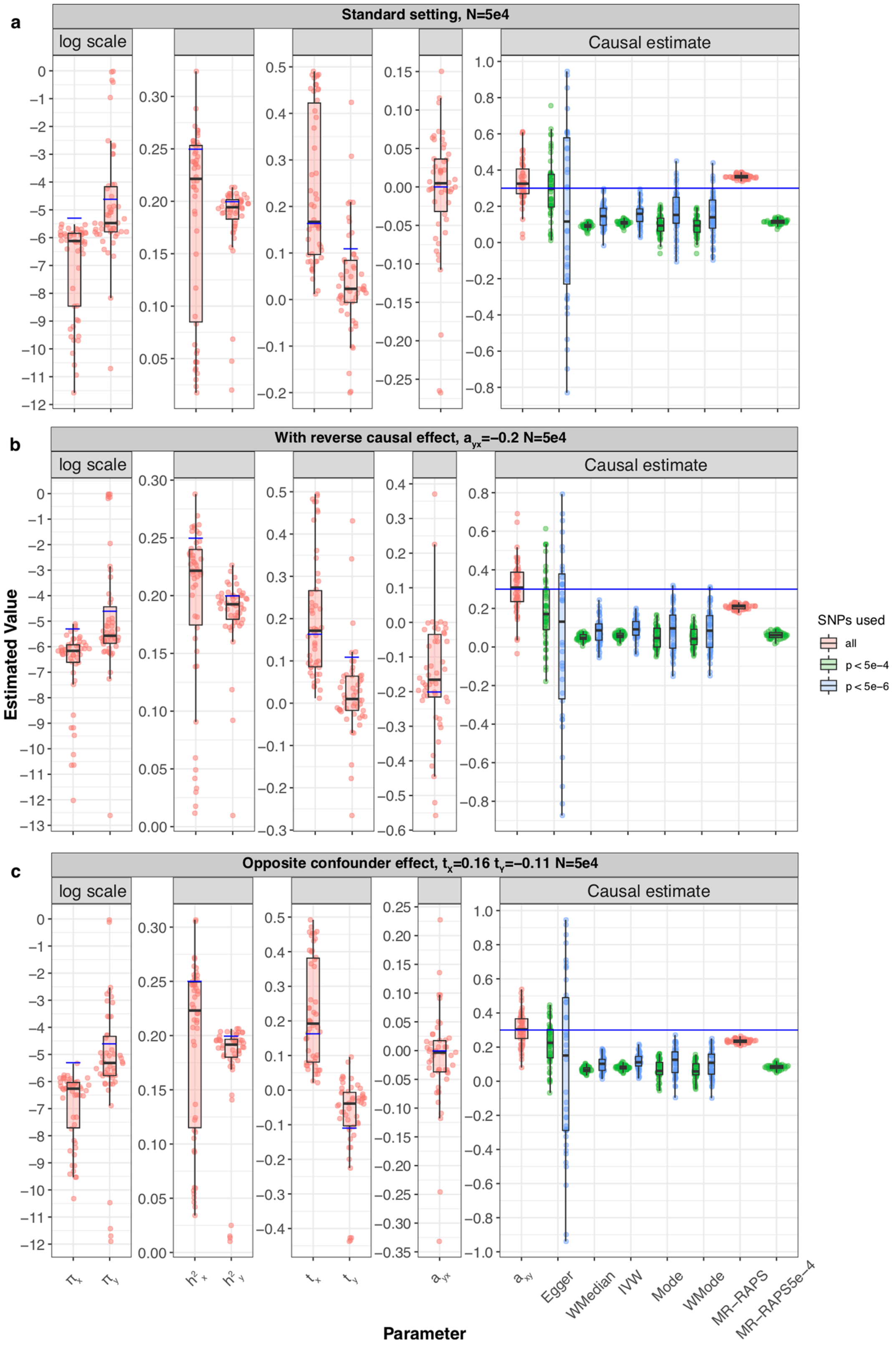


Fig. 2 Simulation results under various scenarios. These modified Sina-boxplots represent the distribution of parameter estimates from 50 different data generations under various conditions. For each generation, standard MR methods as well as our LHC-MR were used to estimate a causal effect. In the boxplots, the lower and upper hinges correspond to the first and third quartiles, the middle bar corresponds to the median, whereas the upper whisker is the largest dataset estimate smaller than 1.5×inter-quartile range above the third quartile. The lower whisker is defined analogously. The true values of the parameters used in the data generations are represented by the blue dots/lines. **a** Estimation under standard settings ($\pi_x = 5\times10^{-3}, \pi_y = 1\times10^{-2}$, $\pi_u = 5\times10^{-2}, h_x^2 = 0.25, h_y^2 = 0.2, h_u^2 = 0.3, t_x = 0.16, t_y = 0.11$). **b** Addition of a reverse causal effect $\alpha_{y\to x} = -0.2$. **c** Confounder with opposite causal effects on X and Y ($t_x = 0.16, t_y = -0.11$).

summary statistics when the underlying non-zero effects come from a non-Gaussian distribution. Interestingly, we observed that, for the smaller sample size, the variance of the causal effect estimate was dependent on the kurtosis for most MR methods. LHC-MR estimations yielded slightly more pronounced upward bias than IVW, while still exhibiting the lowest RMSE among all methods (Fig. 3a). Similar results are seen in the larger sample size with smaller variance for all methods under all degrees of kurtosis except for IVW, which showed a better performance than LHC-MR (Supplementary Fig. 13a). Second, we simulated effect sizes coming from a three-component Gaussian mixture distribution (null/small/large effects), instead of the classical spike-and-slab assumption of our model. The smaller sample size estimates mirror those of the standard setting with n also equal to 50,000 (see Fig. 3b). However, in the larger sample size, LHC-MR overestimates the causal effect. This bias could be due to the merging of true effect estimates with confounder effect leading to an overestimation of $\alpha_{x\to y}$ (Supplementary Fig. 13b). MR-Egger, IVW and filtered MR-RAPS have the smallest RMSE in this case.

Comparing CAUSE and LHC-MR. When running CAUSE on data simulated using the LHC-MR model framework in order to estimate a causal effect (γ in their notation), we investigated three different scenarios, each with multiple data generations: one where the underlying model has a shared factor/confounder with effect on both exposure and outcome only, another where the underlying model has a causal effect of 0.3 only, a third where the underlying model has both a causal effect and a shared factor. The data generated using the LHC-MR model was done under the standard settings ($\pi_x = 5\times10^{-3}, \pi_y = 1\times10^{-2}, \pi_u = 5\times10^{-2}, h_x^2 = 0.25, h_y^2 = 0.2, h_u^2 = 0.3, t_x = 0.16, t_y = 0.11, \alpha_{x\to y} = 0.3, \alpha_{y\to x} = 0, m = 234,000, n_x = n_y = 50,000$). For each setting, 50 different replications were investigated.

In the case of an underlying shared effect only, CAUSE preferred the sharing model 100% of the time, and thus there was no causal estimation, however it underestimated both η and q. When there was an underlying causal effect only, CAUSE preferred the causal model only 4% of the times, where it slightly underestimated the causal effect ($\widehat{\gamma} = 0.241$). Although the true values of η and q are null in this scenario, the sharing model returned estimates for these two parameters overestimating them both (probably driven by their priors), as seen in Supplementary Fig. 14. In the third case, and in the presence of both, CAUSE preferred the sharing model in 48 of the 50 simulations, yet it underestimated η (corresponding to t_y/t_x for our model) but overestimated q ($t_x^2/(t_x^2 + h_x^2)$ in our model) (mean of 0.566 and 0.222, respectively, where the true values are 0.667 and 0.097) showing a similar estimation pattern to the second case. Interestingly, for the larger sample size, CAUSE selects the correct model 100% of the time, but still underestimates γ, as shown in Supplementary Fig. 15.

In the reverse situation, where data was generated using the CAUSE framework (with parameters $h_1 = h_2 = 0.25, m = 97,450, N1 = N2 = 50,000$) and LHC-MR was used to estimate the causal effect, we saw the following results (see Supplementary Fig. 16). First, when we generated data in the absence of causal effect ($\gamma = 0, \eta = \sqrt{0.05}, q = 0.1$), CAUSE does extremely well in estimating a null causal effect 100% of the time. Standard MR methods yield a slight overestimation of the (null) causal effect with varying degrees of variance, whereas LHC-MR shows both a greater variance and an upward bias—still leading to a causal effect compatible with zero. Second, in the absence of a confounder combined with non-zero causal effect ($\gamma = \sqrt{0.05} = 0.22, \eta = 0, q = 0$), CAUSE underestimates the causal effect ($\widehat{\gamma} = 0.18$) compared to LHC-MR which overestimates the causal effect: the mean of the estimates was 0.38 (over the 50 runs). Finally, in the presence of both a confounder and a causal effect ($\gamma = \sqrt{0.05}, \eta = \sqrt{0.05}, q = 0.1$), CAUSE slightly underestimates the causal effect ($\widehat{\gamma} = 0.20$), whereas LHC-MR overestimates the effects and shows estimates reaching the boundaries 11 out of 50 times (mean of the converged $\widehat{\gamma} = 0.39$ over the 39 data simulations, see Supplementary Fig. 16c)— indicating that this setting of the CAUSE model is not compatible with the LHC-MR model framework. Interestingly, classical MR methods outperform CAUSE in this case. Note that in the interest of run time we used less SNPs (than usual) for parameter estimations. The analysis of the three separate scenarios was repeated for a larger sample size of 500,000 (Supplementary Fig. 17), with more favourable results for LHC-MR. In the absence of a causal effect, we had similar results to the smaller sample size, whereas in the absence of a shared effect, LHC-MR estimates the causal effect accurately with a mean of 0.22, CAUSE underestimates it and the rest of the MR methods are less biased. In the presence of both causal and shared factor, CAUSE recovers the causal effect. IVW, unlike the other MR methods and CAUSE, is more affected by the presence of the confounder, while LHC-MR exhibits upward bias with a mean estimate of 0.27.

Application to association summary statistics of complex traits. We applied our LHC-MR and other MR methods to estimate all pairwise causal effects between 13 complex traits (156 causal relationships in both directions). Our results are presented as a heatmap in Fig. 4 (and are detailed in Supplementary Data 2). Further, we calculated the alternate set of estimated parameters that naturally results from our model (for reference see Section The observed association summary statistics and Supplementary Methods 1.4). Among trait pairs for which the exposure had sufficient heritability (>2.5%), the alternate parameters of 102 trait pairs were within the possible ranges mentioned in methods (i.e. the confounder and the exposure are interchangeable). However, for all of these pairs, the alternative parameter optima lead to lower direct- than indirect heritability, which we deem unrealistic. Therefore, we report only the primary set of estimated optimal parameters in the main results and provide the alternative parameters in the Supplementary Data 3. The comparison of the results obtained by LHC-MR and standard MR methods is detailed below and more extensively in Supplementary Data 4–5. In summary, LHC-MR provided reliable causal effect estimates for 132 out of 156 exposure traits (i.e. those exposures had an estimated total heritability greater than 2.5%). These estimates were compared to five different MR methods. Seventy-four causal relationships were deemed significant by LHC-MR. Furthermore, for 117 out of those 132 comparable causal relationships, our LHC-MR causal effect estimates were concordant (not significantly different) with at least two out of five standard MR methods' estimates.

By simply comparing the significance status and the direction of the causal effects between the methods, we see that LHC-MR agrees in sign and significance (or the lack there of) with at least 3 MR methods 77 times. For 31 relationships, LHC-MR results lead to different conclusions than those of standard MR methods. For 28 of those, LHC-MR identified a causal effect missed by all standard MR methods. For the other three, we observed a disagreement in sign: LDL has a negative effect on BMI according to weighted mode and weighted median, whereas we show a positive effect, HDL and LDL show a negative bi-directional causal effect for weighted mode but a positive bi-directional effect with LHC-MR. Despite the conflicting evidence for the causal relationship of LDL on BMI, studies have shown that the relationship between them is non-linear[12], possibly explaining the discrepancy between the results.

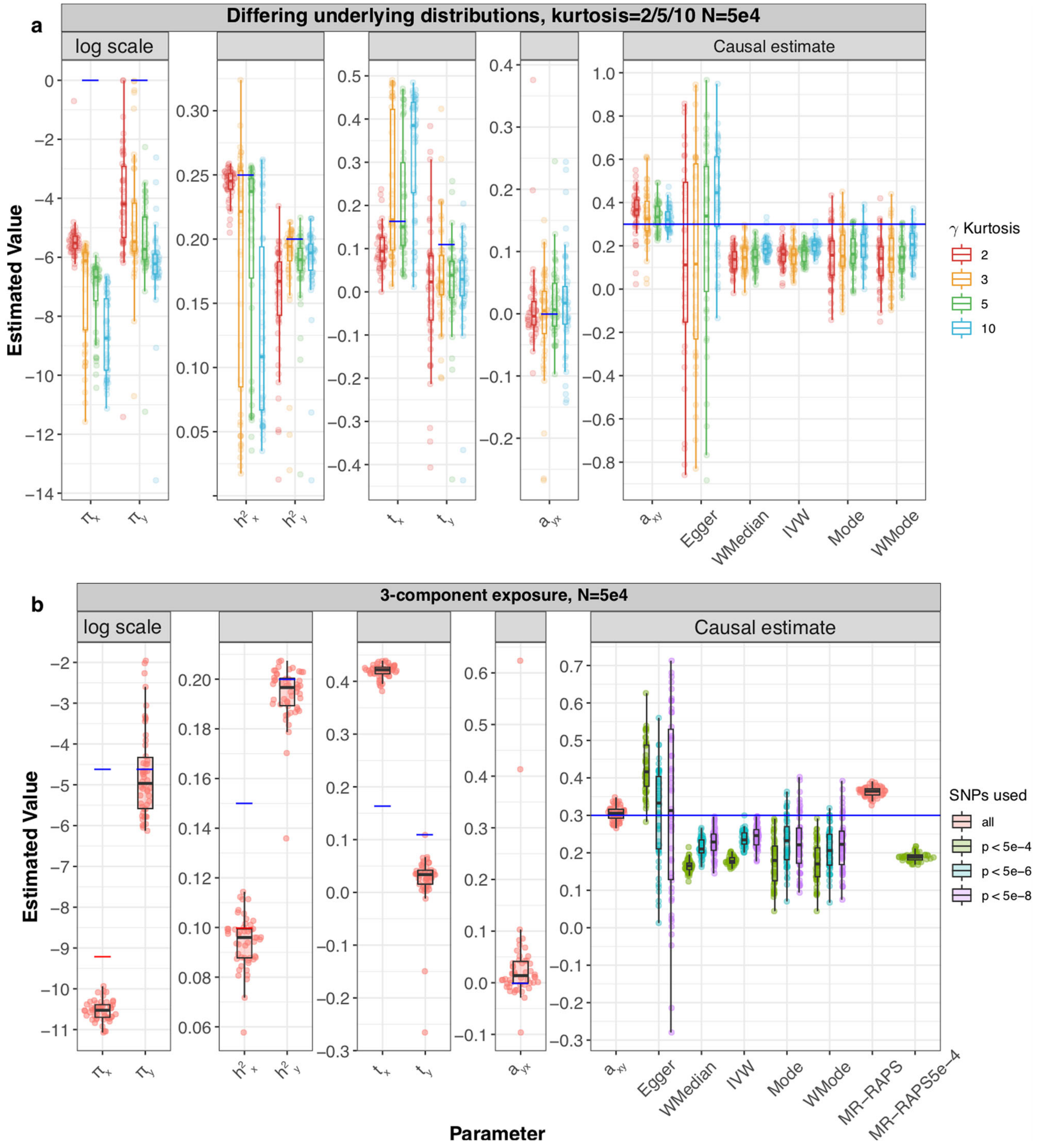


Fig. 3 Simulation results under various scenarios. These modified Sina-boxlots represent the distribution of parameter estimates from 50 different data generations under various conditions. For each generation, standard MR methods as well as our LHC-MR were used to estimate a causal effect. In the boxplots, the lower and upper hinges correspond to the first and third quartiles, the middle bar corresponds to the median, whereas the upper whisker is the largest dataset estimate smaller than 1.5×inter-quartile range above the third quartile. The lower whisker is defined analogously. The true values of the parameters used in the data generations are represented by the blue dots/lines. **a** The different coloured boxplots represent the underlying non-normal distribution used in the simulation of the three $\boldsymbol{\gamma_x}, \boldsymbol{\gamma_y}, \boldsymbol{\gamma_u}$ vectors associated to their respective traits. The Pearson distributions had the same zero mean and skewness, however their kurtosis ranged between 2 and 10, including the kurtosis of 3, which corresponds to a normal distribution assumed by our model. The standard MR results reported had IVs selected with a *p*-value threshold of 5×10^{-6}. **b** Addition of a third component for exposure *X*, while decreasing the strength of *U*. True parameter values are in colour, blue and red for each component ($\pi_{x1} = 1 \times 10^{-4}, \pi_{x2} = 1 \times 10^{-2}, h^2_{x1} = 0.15, h^2_{x2} = 0.1$).

LHC-MR agreed with most MR estimates and confirmed many previous findings, such as increased BMI leading to elevated blood pressure[13,14], diabetes mellitus[15,16] (DM), myocardial infarction[17] (MI) and coronary artery disease[18] (CAD). Furthermore, we confirmed previous results[19] that diabetes increases SBP ($\hat{\alpha}_{x\to y} = 0.39 - P = 1.70 \times 10^{-9}$).

Interestingly, it revealed that higher BMI increases smoking intensity, concordant with other studies[20,21]. It also showed the

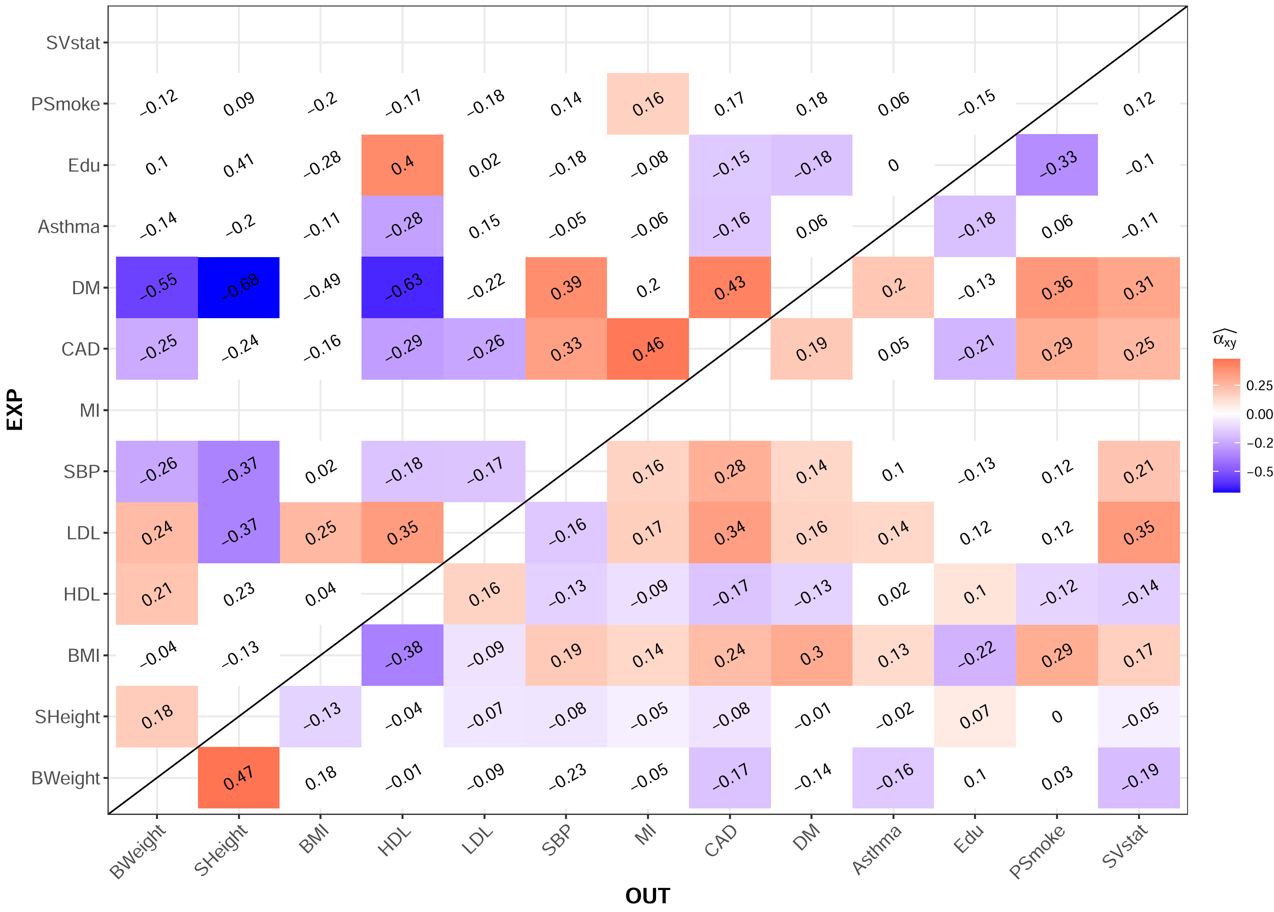


Fig. 4 Heatmap representing the bi-directional causal relationship between the 13 UK Biobank traits. The causal effect estimates in coloured tiles all have a significant *p*-value surviving Bonferroni multiple testing correction with a threshold of 3.2×10^{-4}. We did not report an estimated causal effect for exposures with an estimated total heritability less than 2.5%. White tiles show an absence of a significant causal effect estimate. BMI: Body Mass Index, BWeight: Birth Weight, CAD: Coronary Artery Disease, DM: Diabetes Mellitus, Edu: Years of Education, HDL: High-Density Lipoprotein, LDL: Low-Density Lipoprotein, MI: Myocardial Infarction, PSmoke: # of Cigarettes Previously Smoked, SBP: Systolic Blood Pressure, SHeight: Standing Height, SVstat: Medication-Simvastatin.

protective effect of education against a range of diseases (e.g. CAD and diabetes[22,23]) and risk factors such as smoking[24,25], in agreement with previous observational and MR studies. Probably reflecting lifestyle change recommendations by medical doctors upon disease diagnosis, statin use is greatly increased when being diagnosed with CAD, (systolic) hypertension, dislipidemia and diabetes as is shown by both LHC-MR and standard MR methods.

Furthermore, causal effects of height on CAD, DM and SBP have been previously examined in large MR studies[26,27]. LHC-MR, agreeing with these claims, did not find significant evidence to support the effect of height on DM, but did find a significant protective effect on CAD and SBP. However, unlike the first two, the relationship between height and SBP also revealed the existence of a confounder with causal effects 0.14 ($P = 9.2 \times 10^{-11}$) and 0.11 ($P = 3.39 \times 10^{-8}$) on height and SBP respectively. Another example of a trait pair for which LHC-MR found an opposite sign confounder effect is HDL and its protective effect on SBP. The confounder had a positive effect ratio of $t_y/t_x = 0.84$, opposing the negative causal effect of $\hat{\alpha}_{x\to y} = -0.13$ supported by observational studies[28]. This causal effect was not found by any other MR method.

It is important to note that while the effects of parental exposures on offspring outcomes can be seen as genetic confounding, LHC-MR would not be able to distinguish parental and offspring causal effects, because the LHC-MR model assumes that there is no correlation between the genetic effects on the exposure and the genetic effects on the confounder (which is not the case for parental vs offspring traits). Thus, LHC-MR causal effect estimates are just as likely to reflect parental effects as any other MR method[29]. This may be the case, for example, for the detrimental effect of increased (parental) BMI on education (supported by longitudinal studies[30]), the positive effect of (parental) height on birth weight[31], or on education[32]. There are also some associations identified only by LHC-MR that might reflect parental effects: the negative causal effect of CAD on education or on birth weight, the positive impact of HDL on birth weight, or DM reducing height. All these pair associations uniquely found by LHC-MR are examples of LHC-MR's use of whole-genome SNPs instead of GW-significant SNPs only, as our estimates are of larger magnitude than those found by standard MR. Interestingly, for the CAD → birth weight relationship, LHC-MR revealed a confounder of opposite causal effects, which could have masked/mitigated the causal effect of standard MR methods.

A systematic comparison between IVW and LHC-MR has shown generally good agreement between the two methods, which is illustrated in Fig. 5. To identify discrepancies between our causal estimates and those of the standard MR results, we grouped the estimates into several categories, either non-significant *p*-value for both or either, significant with an agreeing sign for the causal estimate, or significant with a disagreeing sign. The diagonal (seen in Fig. 5) representing the agreement in

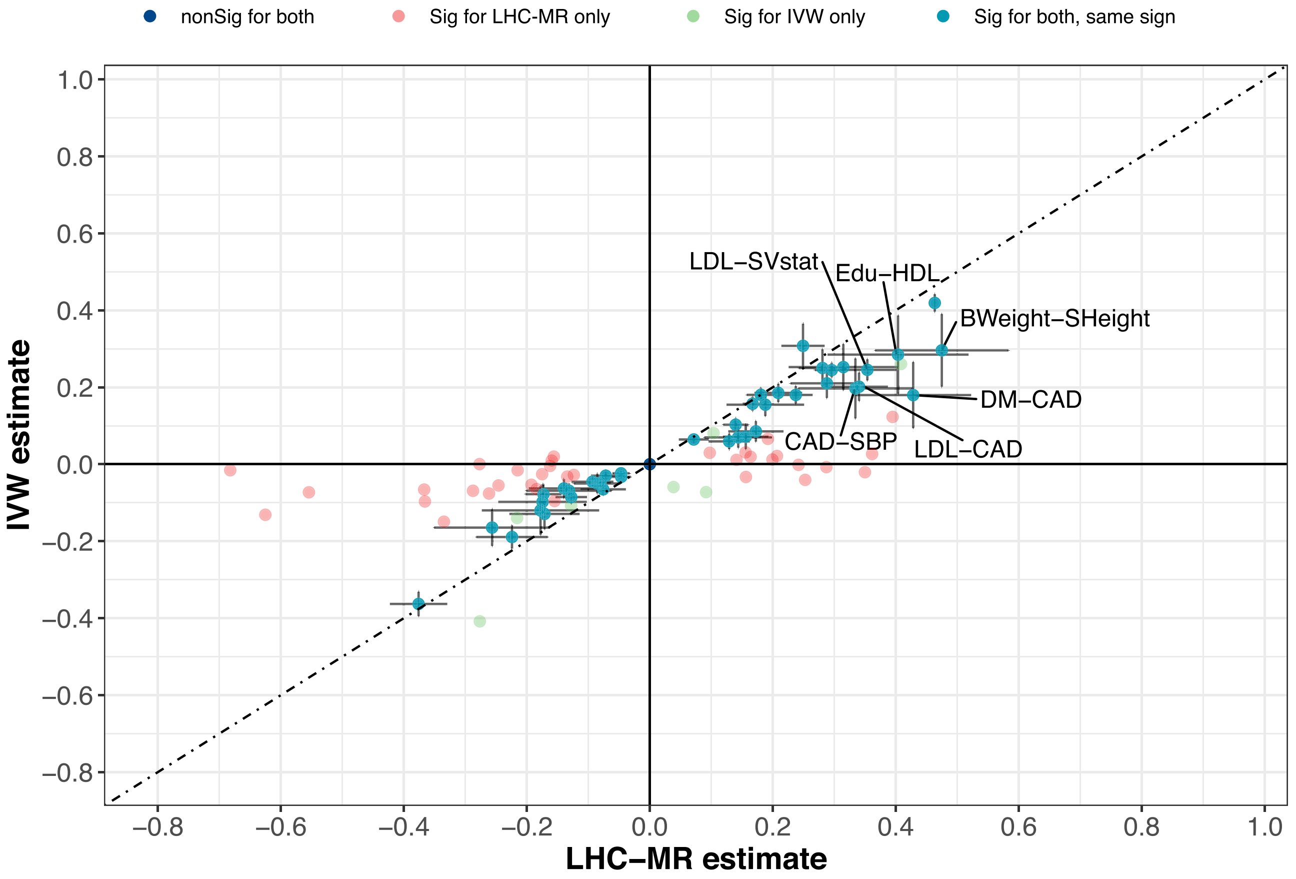


Fig. 5 A scatter plot of the causal effect estimates between LHC-MR and IVW. To improve visibility, non-significant estimates by both methods are placed at the origin, while significant causal estimates by both methods appear on the diagonal with 95% CI error bars. Pairs with an absolute value difference > 0.1 are labelled.

significance status and sign between the two methods, is heavily populated. On the other hand, 34 pairs have causal links that are significantly non-zero according to LHC-MR, but are non-significant for IVW, while the opposite is true for seven pairs. We believe that many of these seven pairs may be false positives, since four of them are picked up by no other MR method, two are confirmed by only one other method and the last one by two methods. Further comparisons of significance between LHC-MR estimates and the remaining standard MR methods can be found in Supplementary Table 2.

LHC-MR identified a confounder for 16 trait pairs out of the possible 78. In order to support these findings, we used EpiGraphDB[33,34] to systematically identify those potential confounders. EpiGraphDB could identify reliable confounders for ten out of the 16 trait pairs. Notably, for the birth weight–diabetes pair, the average epigraph confounder-effect ratio (r_3/r_1) clearly agreed in sign with our t_y/t_x ratio, indicating that the characteristics of the confounder(s) evidenced by LHC-MR agree with those found in an exhaustive confounder search, and are mainly obesity-related traits (Supplementary Fig. 18a). Six other trait pairs showed mixed signs of different confounders, indicating the possibility of having heterogeneous confounders (Supplementary Fig. 18b-e). Finally, three trait pairs showed a disagreement between our estimated confounder-effect ratio and the bulk of those found by epighraphDB as seen in Supplementary Fig. 18f-j. However, at least one of the top ten potential confounders showed effects that are in agreement with our ratio for each of these pairs. Note that since the reported causal effects of the confounders on X and Y reported in EpiGraphDB are not necessarily on the same scale, we do not expect the magnitudes to agree.

As described in the methods (Eq. (32)), genetic correlation can be computed from our estimated model parameters. To verify that the fitted LHC-MR model leads to a genetic correlation similar to the one obtained from LD-score regression[35] (LDSC), we compared whether the two approaches produce similar genetic correlation estimates. We did this by taking the estimated parameters obtained from the 200 block jackknife to estimate the genetic correlations between traits (and their standard errors), and plotted them against LD-score regression values as seen in Fig. 6. As expected, we observe an overall good agreement between the estimates of the two methods, with only six trait pairs differing in sign. Of these six, only 2 were nominally significantly different between the two methods (LDL → Asthma and LDL → DM). Further decomposition of the genetic covariance into heritable confounder-led or causal effect-led covariance revealed that most of the genetic covariance between traits can be attributed to bi-directional causal effects. A reason for this could be that confounders would need to have very strong effects to substantially contribute to the genetic correlation ($\approx t_x \cdot t_y$) compared to the bi-directional causal effects ($\approx \alpha^2_{x\to y} \cdot h^2_x + \alpha^2_{y\to x} \cdot h^2_y$).

As for the comparison of LHC-MR against CAUSE for real trait pairs, we ran CAUSE on all 156 trait pairs (bi-directional), and extracted the parameter estimates that corresponded to the methods winning model. The p-value threshold was corrected for multiple testing and was equivalent to 0.05/156. Based on that threshold, the p-value that compared between the causal and the sharing model of CAUSE was used to choose one of the two. Then the parameters estimated from the winning model, γ (only

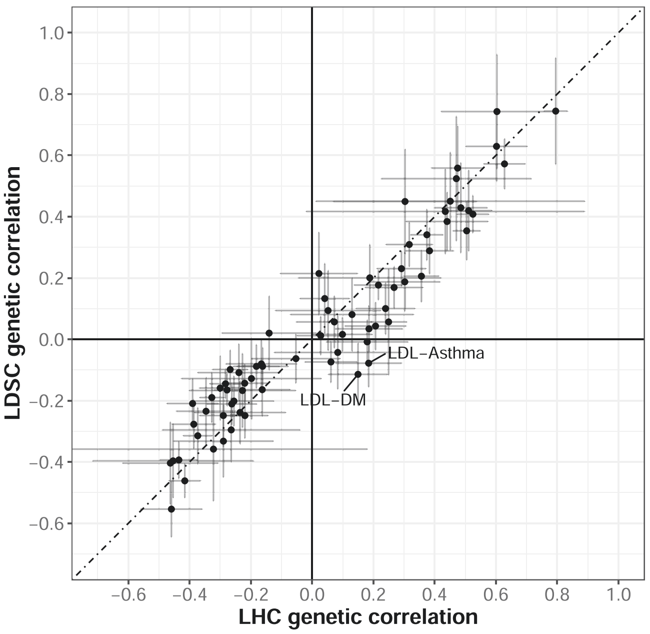


Fig. 6 Scatter plot comparing the genetic correlation for each trait obtained from LDSC against the value calculated using parameter estimates from the LHC-MR model. LHC-MR calculated genetic correlations from 200 parameter estimates generated during the block-jackknife procedure, where the mean values of these 200 estimates are shown here. A 95% CI for both method-calculation is shown for each point, and pairs with an absolute value difference > 0.2 are labelled. Values from both methods are reported in Supplementary Data 6.

for causal model), η and q, were compared to their counterparts in LHC-MR. A visual comparison of LHC-MR's causal estimates and those of CAUSE can be seen in Supplementary Fig. 19.

Whenever the causal effect estimates were significant both for CAUSE and LHC-MR (30 causal relationships), they always agreed in sign (Supplementary Table 3) with a high Pearson correlation of 0.592. Calculating the correlation for their estimates regardless of significance yielded a smaller value of 0.377. When compared to the causal effect estimate from IVW, LHC-MR was strongly correlated (0.585), whereas CAUSE had a slightly weaker correlation (0.471) using all estimates.

Similarly, the significant confounder-effect ratio of LHC-MR (t_y/t_x) can be compared to the significant confounder-effect estimate of CAUSE (η) when a sharing model is chosen. These 12 confounding quantities by CAUSE and LHC-MR disagreed in sign for all but one trait pair (Height → MI), with a Pearson correlation compatible with zero (−0.357 (95% CI [−0.77, 0.27])).

Discussion

We have developed a structural equation (mixed-effects) model to account for a latent heritable confounder (U) of an exposure (X)–outcome (Y) relationship in order to estimate bi-directional causal effects between the two traits (X and Y). The method, termed LHC-MR, fits this model to association summary statistics of genome-wide genetic markers to estimate various global characteristics of these traits, including bi-directional causal effects, confounder effects, direct heritabilities, polygenicities and population stratification.

We first demonstrated through simulations that in most scenarios, the method produces causal effect estimates with substantially less bias and variance (in the larger sample size) than other MR tools. The direction and magnitude of the bias of classical MR approaches varied across scenarios and sample sizes. This bias was mainly influenced by two often opposite forces: downward bias resulting from winner's curse and weak instruments, and upward bias due to a positive confounder of the $X-Y$ relationship, evident in the larger sample size. In the scenario lacking a confounder (thus respecting all MR assumptions), MR methods were distinctly underestimating the causal effect, except for LHC-MR and to a better extent MR-RAPS. However, under standard settings with an added small heritable confounder and no reverse causality present, all classical MR methods still slightly underestimated the causal effect in the smaller sample size, except for the MR-RAPS estimate which was now overestimated. For the same standard setting scenario but in a larger sample size where confounder effects were more detectable, IVW had an estimation that was close to the true causal value chosen ($\alpha_{x\to y}=0.3$) due to the opposite biases cancelling out. However, when the causal effect was set to be smaller ($\alpha_{x\to y}=0.1$), the estimates of IVW became biased. More substantial violations of classical MR assumptions, such as the presence of negative-effect confounder or a negative reverse causal effect, led to more substantial biases that impacted all methods (including MR-RAPS) except LHC-MR.

Interestingly, in the smaller sample size, standard MR methods showed a slight decreasing trend in the variance of the causal effect estimate as the kurtosis of the underlying effect size distribution went up from 2 to 10. On the other hand, LHC-MR did not show a similar trend with growing kurtosis, and estimated the causal effect with a smaller bias. As confounder causal effects (q_x, q_y) increased, classical MR methods (except weighted ones) were prone to produce overestimated causal effects with at least twice the bias than that of LHC-MR, especially in the large sample size where the confounder-associated SNPs make it to the set of GW-significant instruments for all methods. Furthermore, mode-based estimators were robust to the presence of two concordant confounders, yet their bias was still 10-fold higher than LHC-MR's, and they did not perform as well in the presence of discordant confounders. In summary, LHC-MR was robust to a wide range of violations of the classical MR assumptions and was less impacted than standard MR methods. Thus it outperformed all MR methods in virtually all tested scenarios, many of which violated even its own modelling assumptions.

We then applied our method to summary statistics of 13 complex traits from large studies, including the UK Biobank. We observed a general trend in our results (in agreement with epidemiological studies) that higher BMI and LDL are risk factors for most diseases such as diabetes and CAD. We also note the protective effect HDL has on these same diseases. Moreover, we observe many disease traits increasing the intake of lipid-lowering medication (simvastatin), reflecting the recommendation/treatment of medical personnel following the diagnosis.

LHC-MR can have discordant results compared to other MR methods for many possible reasons. The positive causal effect of smoking on MI, diabetes on asthma, the protective impact of higher birth weight on asthma, or higher education on smoking intensity, all of which were missed by standard MR could reflect the increased power of LHC-MR with its use of full-genome SNPs as opposed to genome-wide significant SNPs of classical MR approaches.

Estimates from classical MR methods could also be impacted by sample overlap between the exposure and outcome datasets, whereas LHC-MR takes this into account. However, when using large sample sizes, the bias due to sample overlap is expected to be very small, and therefore not sufficient to explain any discrepancy in the results[36]. Another possible reason for the discrepancy between our findings and those of standard MR methods is the presence of a significant heritable confounder found by LHC-MR with opposite effect to the estimated causal effect between the pair. These two opposite forces lead to association summary statistics that may be compatible with reduced (or even null) causal effect when the confounder is ignored. Possible examples of this scenario can be observed when (parental) traits, e.g. diabetes and CAD, act on birth weight. These pairs have a confounder of opposite effects, possibly related to (parental) obesity. Similarly, standard MR methods show little evidence for a causal effect of SBP on height, while our LHC-MR estimate is -0.37 ($P=4.81\times10^{-8}$) which most probably reflects parental (maternal) effects as seen in previous studies[37,38]. The protective effect of HDL on SBP is another example where a confounder of opposite sign to that of the causal effect allows it to be uniquely found by LHC-MR. LHC-MR assumes no genetic correlation between the confounder and the direct effects on the exposure, which may be violated when the confounder is the same trait as the exposure, but in the parent. Such parental effects can mislead most MR methods[39], including ours, and hence we may observe biased results for traits such as BMI → education and HDL → birth weight.

Sixteen trait pairs showed a strong confounder effect, in the form of significant t_x and t_y estimates. These pairs were investigated for the presence of confounders using EpiGraphDB, and 10 of them returned possible confounders. The bulk of such pairs returned confounders with both agreeing and disagreeing effect directions on X and Y, making it difficult to pinpoint a group of concordant and dominant confounders. However, for the birth weight-DM pair, where LHC-MR identifies a negative reverse causal effect and a confounder with effects $t_x=0.10(P=6.77\times10^{-8})$ and $t_y=0.15$ ($P=3.13\times10^{-7}$) on birth weight and DM respectively, EpigraphDB confirmed several confounders related to body fat distribution and weight that matched in sign with our estimated confounder effect (Supplementary Fig. 18a). Note that EpiGraphDB causal estimates are not necessarily on the scale of SD outcome difference upon 1 SD exposure change scale, hence they are not directly comparable with the t_y/t_x ratio, but are rather indicative of the sign of the causal effect ratio of the confounder. Furthermore, if EpigraphDB does not find a causal relationship between the trait pair in either directions, then it does not return any possible confounders of the two, a reason why only 10 out of 16 confounder-associated trait pairs returned any hits.

Lastly, our comparison of the genetic correlations calculated from our estimated parameters against those calculated from LD-score regression showed good concordance, confirming that the detailed genetic architecture proposed by our model is compatible with the observed genetic covariance. The major difference between the genetic correlation obtained by LD-score regression *vs* LHC-MR is that our model approximates all existing confounders by a single latent variable, which may be inaccurate when multiple ones exist with highly variable t_y/t_x ratios. Furthermore, LHC-MR decomposed the observed genetic correlation into confounder and bi-directional causality driven components, revealing that most genetic correlations are primarily driven by bi-directional causal effects. Note that we have much higher statistical power to detect situations when the confounder effects are of opposite sign compared to the causal effects, because opposing genetic components are more distinct.

To our knowledge only two recent papers use similar models and genome-wide summary statistics. The LCV approach[40] is a special case of our model, where the causal effects are not included in the model, but they estimate the confounder effect mixed with the causal effect to estimate a quantity of genetic causality proportion (GCP). In agreement with others[10,41], we would not interpret non-zero GCP as evidence for causal effect. Moreover, in other simulation settings, LCV has shown very low power to detect causal effects (by rejecting GCP = 0) (Fig S15 in Howey et al.[42]). Another very recent approach, CAUSE[10], proposes a structural equation mixed effect model similar to ours. However, there are several differences between LHC-MR and CAUSE: (a) we allow for bi-directional causal effects and model them simultaneously, while CAUSE is fitted twice for each direction of causal effect; (b) they first use an adaptive shrinkage method to integrate out the multivariable SNP effects and then go on to estimate other model parameters, while we fit all parameters at once; (c) CAUSE estimates the correlation parameter empirically; (d) we assume that direct effects come from a two-component Gaussian mixture, while they allow for larger number of components; (e) their likelihood function does not explicitly model the shift between univariate vs multivariate effects (i.e. the LD); (f) CAUSE adds a prior distribution for the causal/confounder effects and the proportion π_u, while LHC-MR does not; (g) to calculate the significance of the causal effect they estimate the difference in the expected log point-wise posterior density and its variance through importance sampling, whereas we use a simple block-jackknife method. Because of point (a), the CAUSE model can be viewed as a special case of ours when there is no reverse causal effect. We have the advantage of fitting all parameters simultaneously, while they only approximate this procedure. Although they allow for more than a two-component Gaussian mixture, for most traits with realistic sample sizes we do not have enough power to distinguish whether two or more components fit the data better. Therefore, we believe that a two-component Gaussian is a reasonable simplification. Due to the more complicated approach described in points (e-g), CAUSE is computationally more intense than LHC-MR, taking up to 1.25 CPU-hours in contrast to our 2.5 CPU-minute run time for a single starting point optimisation (which is massively parallelisable).

When we compared the performance of CAUSE and LHC-MR, we found that for large sample sizes both LHC-MR and CAUSE performed well not only when applied to data simulated by their own model, but also by the model of the other method. For smaller sample sizes, both methods performed poorly when applied to data generated by the other model. However, LHC-MR was less biased when applied to data generated by its own model than CAUSE was on data simulated based on its own model, where it provided rather conservative estimates. This is somewhat expected, since the primary aim of CAUSE is model selection and it is less geared towards parameter estimation, especially for settings where both sharing and causal effects are present (leading to very broad estimates). Also, CAUSE parameter estimates have shown to be somewhat sensitive to the choice of the prior.

Finally, when applying both LHC-MR and CAUSE to 156 complex trait pairs, we observed that the causal effects are reasonably well correlated (0.38 for all estimates, 0.59 for significant estimates) and agree in sign for trait pairs deemed significantly causal by either or both methods. In addition, LHC-MR causal estimates were more similar to those of IVW than the estimates provided by CAUSE. Surprisingly, when a confounding factor was identified by both methods, the confounder effects (LHC-MR t_y/t_x ratio and CAUSE η parameter) were uncorrelated. There are two possible explanations for this: (i) CAUSE may confuse/merge the confounder with the reverse causal effect, since it does not explicitly model the latter one. (ii) The two models assume different marginal effect size distributions, hence when multiple heterogeneous confounders exist, one method may detect one of the confounders, while the other method picks up the other confounder, depending on which has more similar genetic architecture to the assumed one.

Our approach has its own limitations, which we list below. Like any MR method, LHC-MR provides biased causal effect estimates if the input summary statistics are flawed (e.g. not corrected for complex population stratification, parental/dynasty effects). As mentioned in the Methods section, our model is strictly-speaking unidentifiable and two distinct sets of parameters fit the data equally well, if the alternate set of parameters fall within the parameter ranges. As opposed to classical MR methods that give a single (biased) causal effect estimate, ours can detect and calculate the competing model. Due to biological considerations, from these competing models, we chose the one which yielded larger direct heritability than confounder-driven (indirect) heritability. Additional pointers to decide which parameter optimum we choose can be to pick the one with smaller magnitude of causal effects (large causal effects are unrealistic) or pick the one that includes causal effects that agree better with those of other MR methods.

LHC-MR is not an optimal solution for traits whose genetic architecture substantially deviates from a two-component Gaussian mixture of effect sizes. Also, for traits with low heritability (<2.5%), it is particularly important to compare the causal effect estimates to those from standard MR methods as results from LHC-MR may be less robust. In addition, trait pairs with multiple confounders with heterogeneous effect ratios can violate the single confounder assumption of the LHC model and can lead to biased causal effect estimates. Finally, LHC-MR, like other methods, is not immune to parental effects that are correlated with offspring effects. In such cases, the parental effect is grouped with the exposure (due to their strong genetic correlation) and not viewed as a confounder of the exposure-outcome relationship.

Methods

The underlying structural equation model. Let X and Y denote continuous random variables representing two complex traits. Let us assume (for simplicity) that there is one heritable confounder U of these traits. To simplify notation we assume that $E(X) = E(Y) = E(U) = 0$ and $Var(X) = Var(Y) = Var(U) = 1$. The genome-wide sequence data for M sequence variants is denoted by $G = (G_1, G_2, \dots, G_M)$. The aim of our work is to dissect the effects of the heritable confounding factor U from the bi-directional causal effects of these two traits (X and Y). For this we consider a model (see Fig. 1) defined by the following equations:

$$X = q_x \cdot U + \alpha_{y\to x} Y + G \cdot \boldsymbol{\gamma_x} + \boldsymbol{e_x} \qquad \text{with} \qquad \boldsymbol{e_x} \sim \mathcal{N}(0, \nu_x^2) \tag{1}$$

$$Y = q_y \cdot U + \alpha_{x\to y} X + G \cdot \boldsymbol{\gamma_y} + \boldsymbol{e_y} \qquad \text{with} \qquad \boldsymbol{e_y} \sim \mathcal{N}(0, \nu_y^2) \tag{2}$$

$$U = G \cdot \boldsymbol{\gamma_u} + \boldsymbol{e_u} \qquad \text{with} \qquad \boldsymbol{e_u} \sim \mathcal{N}(0, \nu_u^2) \tag{3}$$

where $\boldsymbol{\gamma_x}, \boldsymbol{\gamma_y}, \boldsymbol{\gamma_u} \in \mathcal{R}^M$ denote the (true multivariable) direct effect of all M genetic variants on X, Y and U, respectively. All error terms ($\boldsymbol{e_x}$, $\boldsymbol{e_y}$ and $\boldsymbol{e_u}$) are assumed to be independent of each other and normally distributed with variances ν_x^2, ν_y^2 and ν_u^2, respectively.

Note that we do not include in the model reverse causal effects on the confounder ($X \to U$ and $Y \to U$). The reason for this is the following: Let s_x and s_y denote those causal effect of X and Y on U. We can see that by reparameterising the original model to $\alpha' := \alpha_{x\to y} + s_x \cdot q_y$, $\alpha' := \alpha_{y\to x} + s_y \cdot q_x$ and $q_x' := q_x/(1 - q_x \cdot s_x)$, $q_y' := q_y/(1 - q_y \cdot s_y)$, the genetic effects produced by the extended model with reverse causal effects on U and the simpler model (Fig. 1) with the updated parameters are indistinguishable. Thus these extra parameters are not identifiable and the reparameterisation means that $\alpha_{x\to y}$ and $\alpha_{y\to x}$ in our model represent the total causal effects, some of which may be mediated by U.

Note that the model cannot be represented by classical directed acyclic graphs, as the bi-directional causal effects could form a cycle. However, the equations can be reorganised to avoid recursive formulation as follows:

$$X = q_x \cdot U + \alpha_{y\to x} \cdot \left(q_y \cdot U + \alpha_{x\to y} X + G \cdot \boldsymbol{\gamma_y} + \boldsymbol{e_y} \right) + G \cdot \boldsymbol{\gamma_x} + \boldsymbol{e_x} \tag{4}$$

$$Y = q_y \cdot U + \alpha_{x\to y} \cdot \left(q_x \cdot U + \alpha_{y\to x} Y + G \cdot \boldsymbol{\gamma_x} + \boldsymbol{e_x} \right) + G \cdot \boldsymbol{\gamma_y} + \boldsymbol{e_y} \tag{5}$$

$$U = G \cdot \boldsymbol{\gamma_u} + \boldsymbol{e_u} \tag{6}$$

Regrouping the terms gives

$$(1 - \alpha_{y\to x}\alpha_{x\to y}) \cdot X = (q_x + \alpha_{y\to x} \cdot q_y) \cdot U + \alpha_{y\to x}(G \cdot \boldsymbol{\gamma_y}) + G \cdot \boldsymbol{\gamma_x} + (\boldsymbol{e_x} + \alpha_{y\to x} \cdot \boldsymbol{e_y}) \tag{7}$$

$$(1 - \alpha_{x\to y}\alpha_{y\to x}) \cdot Y = (q_y + \alpha_{x\to y} \cdot q_x) \cdot U + \alpha_{x\to y}(G \cdot \boldsymbol{\gamma_x}) + G \cdot \boldsymbol{\gamma_y} + (\boldsymbol{e_y} + \alpha_{x\to y} \cdot \boldsymbol{e_x}) \tag{8}$$

$$U = G \cdot \boldsymbol{\gamma_u} + \boldsymbol{e_u} \tag{9}$$

Substituting U into the first two equations yields

$$X = \frac{q_x + \alpha_{y\to x} \cdot q_y}{1 - \alpha_{y\to x}\alpha_{x\to y}} \cdot (G \cdot \boldsymbol{\gamma_u}) + \frac{\alpha_{y\to x}}{1 - \alpha_{y\to x}\alpha_{x\to y}} (G \cdot \boldsymbol{\gamma_y}) + \frac{1}{1 - \alpha_{y\to x}\alpha_{x\to y}} (G \cdot \boldsymbol{\gamma_x}) + \epsilon_x \tag{10}$$

$$Y = \frac{q_y + \alpha_{x\to y} \cdot q_x}{1 - \alpha_{x\to y}\alpha_{y\to x}} \cdot (G \cdot \boldsymbol{\gamma_u}) + \frac{\alpha_{x\to y}}{1 - \alpha_{x\to y}\alpha_{y\to x}} (G \cdot \boldsymbol{\gamma_x}) + \frac{1}{1 - \alpha_{x\to y}\alpha_{y\to x}} (G \cdot \boldsymbol{\gamma_y}) + \epsilon_y \tag{11}$$

with

$$\epsilon_x := \frac{\boldsymbol{e_x} + \alpha_{y\to x} \cdot \boldsymbol{e_y} + (q_x + \alpha_{y\to x} \cdot q_y) \cdot \boldsymbol{e_u}}{1 - \alpha_{y\to x}\alpha_{x\to y}} \sim \mathcal{N}(0, i_x) \tag{12}$$

$$\epsilon_y := \frac{\boldsymbol{e_y} + \alpha_{x\to y} \cdot \boldsymbol{e_x} + (q_y + \alpha_{x\to y} \cdot q_x) \cdot \boldsymbol{e_u}}{1 - \alpha_{x\to y}\alpha_{y\to x}} \sim \mathcal{N}(0, i_y) \tag{13}$$

where $i_x := (\nu_x^2 + \alpha_{y\to x}^2 \nu_y^2 + (q_x + \alpha_{y\to x} q_y)\nu_u^2)/(1 - \alpha_{y\to x}\alpha_{x\to y})^2$ and $i_y := (\nu_y^2 + \alpha_{x\to y}^2 \nu_x^2 + (q_y + \alpha_{x\to y} q_x)\nu_u^2)/(1 - \alpha_{x\to y}\alpha_{y\to x})^2$. Note that i_x is equivalent to the LD-score regression intercept[43].

We model the genetic architecture of these direct effects with a spike-and-slab distribution, assuming that only $0 \le \pi_x, \pi_y, \pi_u \le 1$ proportion of the genome have a direct effect on X, Y, U, respectively, and these direct effects come from a Gaussian distribution. Namely,

$$\boldsymbol{\gamma_x} = \boldsymbol{\zeta_x} \odot \boldsymbol{\kappa_x} \qquad \text{with} \qquad \boldsymbol{\kappa_x} \sim \mathcal{N}(0, \sigma_x^2 \cdot I) \qquad \text{and} \qquad \boldsymbol{\zeta_x} \sim \mathcal{B}_m(1, \pi_x) \tag{14}$$

$$\boldsymbol{\gamma_y} = \boldsymbol{\zeta_y} \odot \boldsymbol{\kappa_y} \qquad \text{with} \qquad \boldsymbol{\kappa_y} \sim \mathcal{N}(0, \sigma_y^2 \cdot I) \qquad \text{and} \qquad \boldsymbol{\zeta_y} \sim \mathcal{B}_m(1, \pi_y) \tag{15}$$

$$\boldsymbol{\gamma_u} = \boldsymbol{\zeta_u} \odot \boldsymbol{\kappa_u} \qquad \text{with} \qquad \boldsymbol{\kappa_u} \sim \mathcal{N}(0, \sigma_u^2 \cdot I) \qquad \text{and} \qquad \boldsymbol{\zeta_u} \sim \mathcal{B}_m(1, \pi_u) \tag{16}$$

Here, $\odot$ denotes element-wise multiplication and $\mathcal{B}_m(1, q)$ the m dimensional independent Bernoulli distribution. Further, we assume that all $\boldsymbol{\kappa_x}, \boldsymbol{\kappa_y}, \boldsymbol{\kappa_u}$s are independent of each other and so are all $\boldsymbol{\zeta_x}, \boldsymbol{\zeta_y}, \boldsymbol{\zeta_u}$s. We can refer to $h_x^2 := M \cdot \pi_x \cdot \sigma_x^2$ as the direct heritability of X, i.e. independent of the genetic basis of U and Y. Similar notation is adapted for U ($h_u^2 := M \cdot \pi_u \cdot \sigma_u^2$) and Y ($h_y^2 := M \cdot \pi_y \cdot \sigma_y^2$). Note that when $q_x = 0$ and $q_y \neq 0$ (or vice versa), this means that there is no confounder U present, but the genetic architecture of Y (or X) can be better described by a three-component Gaussian mixture distribution.

We assume that the correlation (across markers) between the direct effects of a genetic variant on X, Y and U is zero, i.e. $cov(\boldsymbol{\gamma_x}, \boldsymbol{\gamma_y}) = cov(\boldsymbol{\gamma_x}, \boldsymbol{\gamma_u}) = cov(\boldsymbol{\gamma_u}, \boldsymbol{\gamma_y}) = 0$. Note that this assumption still allows for a potential correlation between the total effect of G on X and its horizontal pleiotropic effect on Y, but only due to the confounder U and through the reverse causal effect $Y \to X$. As we argued above, this is a reasonable assumption, since the most plausible reason (apart from outcome-dependent sampling, which is out of the scope of this paper) for the violation of the InSIDE assumption may be one or more heritable confounder(s).

For simplicity, we also assume that the set of genetic variants with direct effects on each trait overlap only randomly, i.e. the fraction of the genome directly associated with both X and Y is $\pi_x \cdot \pi_y$, etc. This assumption is in line with recent observation that the bulk of observed pleiotropy can be explained by extreme polygenicity with random overlap between trait loci[44]. Note that uncorrelated effects (e.g. $cov(\boldsymbol{\gamma_x}, \boldsymbol{\gamma_y}) = 0$) do not ensure that the active variant sets overlap randomly, this is a slightly stronger assumption.

The observed association summary statistics. Let us now assume that we observe univariable association summary statistics for these two traits from two (potentially overlapping) finite samples N_x and N_y of size n_x, n_y, respectively. In the following, we will derive observed summary statistics in sample N_x and then we will repeat the analogous exercise for sample N_y. Let the realisations of X, Y and U be denoted by $\boldsymbol{x}, \boldsymbol{y}$ and $\boldsymbol{u} \in \mathcal{R}^{n_x}$. The genome-wide genetic data are represented by $G_x \in \mathcal{R}^{n_x \times M}$ and the genetic data for a single nucleotide polymorphism (SNP) k tested for association is $\boldsymbol{g_k} \in \mathcal{R}^{n_x}$. Note the distinction between the k-th column of G_x, which is the k-th sequence variant, in contrast to $\boldsymbol{g_k}$, which is the k-th SNP tested for association in the GWAS. We assume that all SNP genotypes have been standardised to have zero mean and unit variance. The marginal effect size estimate for SNP k of trait X can then be written as $\widehat{\beta}_k^x = \boldsymbol{g}_k' \cdot \boldsymbol{x}/\boldsymbol{n_x}$, which is a special case of univariable standard normal linear regression when both the outcome and the predictor is standardised to have zero mean and unit variance[43]. Note that $\boldsymbol{x}'$ denotes the transpose of the column vector $\boldsymbol{x}$. This can be further transformed as

$$\begin{aligned} \widehat{\beta}_k^x &= \boldsymbol{g}_k' \cdot \boldsymbol{x}/\boldsymbol{n_x} \\ &= \frac{q_x + \alpha_{y\to x} \cdot q_y}{1 - \alpha_{y\to x}\alpha_{x\to y}} \cdot \boldsymbol{g}_k' \cdot G_x \cdot \boldsymbol{\gamma_u}/n_x + \frac{\alpha_{y\to x}}{1 - \alpha_{y\to x}\alpha_{x\to y}} \cdot \boldsymbol{g}_k' \cdot G_x \cdot \boldsymbol{\gamma_y}/n_x \\ &\quad + \frac{1}{1 - \alpha_{y\to x}\alpha_{x\to y}} \cdot \boldsymbol{g}_k' \cdot G_x \cdot \boldsymbol{\gamma_x}/n_x + \boldsymbol{g}_k' \cdot \epsilon_x/n_x \end{aligned} \tag{17}$$

By denoting the linkage disequilibrium (LD) between variant k and all markers in the genome with $\boldsymbol{\rho_k} = G'_x \cdot \boldsymbol{g}_k/n_x$ we get

$$\widehat{\beta}_k^x = \frac{q_x + \alpha_{y\to x} \cdot q_y}{1 - \alpha_{y\to x}\alpha_{x\to y}} \cdot \boldsymbol{\rho'_k} \cdot \boldsymbol{\gamma_u} + \frac{\alpha_{y\to x}}{1 - \alpha_{y\to x}\alpha_{x\to y}} \cdot \boldsymbol{\rho'_k} \cdot \boldsymbol{\gamma_y} + \frac{1}{1 - \alpha_{y\to x}\alpha_{x\to y}} \cdot \boldsymbol{\rho'_k} \cdot \boldsymbol{\gamma_x} + \eta_k^x \tag{18}$$

with $\eta_k^x := \boldsymbol{g}'_k \cdot \epsilon_x/n_x \sim \mathcal{N}(0, i_x/n_x)$. Given the above-defined genetic effect size distribution the equation becomes

$$\begin{aligned}\widehat{\beta}_k^x = {} & \frac{q_x + \alpha_{y\to x} \cdot q_y}{1 - \alpha_{y\to x}\alpha_{y\to x}} \cdot \underbrace{\boldsymbol{\rho'_k} \cdot (\boldsymbol{\zeta_u} \odot \boldsymbol{\kappa_u})}_{z_k^{(u)}} + \frac{\alpha_{y\to x}}{1 - \alpha_{y\to x}\alpha_{x\to y}} \cdot \underbrace{\boldsymbol{\rho'_k} \cdot (\boldsymbol{\zeta_y} \odot \boldsymbol{\kappa_y})}_{z_k^{(y)}} \\ & + \frac{1}{1 - \alpha_{y\to x}\alpha_{x\to y}} \cdot \underbrace{\boldsymbol{\rho'_k} \cdot (\boldsymbol{\zeta_x} \odot \boldsymbol{\kappa_x})}_{z_k^{(x)}} + \eta_k^x \qquad (19) \\ = {} & \frac{q_x + \alpha_{y\to x} \cdot q_y}{1 - \alpha_{y\to x}\alpha_{x\to y}} \cdot z_k^{(u)} + \frac{\alpha_{y\to x}}{1 - \alpha_{y\to x}\alpha_{x\to y}} \cdot z_k^{(y)} + \frac{1}{1 - \alpha_{y\to x}\alpha_{x\to y}} z_k^{(x)} + \eta_k^x\end{aligned}$$

Similarly, assuming that the LD structures ($\boldsymbol{\rho_k}$) in the two samples are comparable, for $\widehat{\beta}_k^y$ estimated in the other sample (N_y) we obtain

$$\widehat{\beta}_k^y = \frac{\alpha_{x\to y} \cdot q_x + q_y}{1 - \alpha_{x\to y}\alpha_{y\to x}} \cdot z_k^{(u)} + \frac{\alpha_{x\to y}}{1 - \alpha_{x\to y}\alpha_{y\to x}} \cdot z_k^{(x)} + \frac{1}{1 - \alpha_{x\to y}\alpha_{y\to x}} z_k^{(y)} + \eta_k^y \tag{20}$$

with $\eta_k^y \sim \mathcal{N}(0, i_y/n_y)$.

Therefore, the joint effect size estimates can be written as

$$\begin{pmatrix} \widehat{\beta}_k^x \\ \widehat{\beta}_k^y \end{pmatrix} = \frac{1}{1 - \alpha_{x\to y}\alpha_{y\to x}} \left(\begin{pmatrix} (\alpha_{y\to x} \cdot q_y + q_x) \\ (\alpha_{x\to y} \cdot q_x + q_y) \end{pmatrix} z_k^{(u)} + \begin{pmatrix} 1 \\ \alpha_{x\to y} \end{pmatrix} z_k^{(x)} + \begin{pmatrix} \alpha_{y\to x} \\ 1 \end{pmatrix} z_k^{(y)} \right) + \begin{pmatrix} \eta_k^x \\ \eta_k^y \end{pmatrix} \tag{21}$$

Following the same rational as the cross-trait LD-score regression[45], the noise term distribution is readily obtained

$$\begin{pmatrix} \eta_k^x \\ \eta_k^y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} i_x/n_x & \frac{n_{x\cap y}}{n_x \cdot n_y} \cdot r_{x,y} \\ \frac{n_{x\cap y}}{n_x \cdot n_y} \cdot r_{x,y} & i_y/n_y \end{pmatrix} \right) \tag{22}$$

where $r_{x,y}$ is the observational correlation between variables X and Y and $n_{x\cap y}$ is the size of the overlapping samples for X and Y. Since both $n_{x\cap y}$ and $r_{x,y}$ cannot be estimated, we simply denote $i_{x,y} := r_{x,y} \cdot \frac{n_{x\cap y}}{\sqrt{n_x \cdot n_y}}$ as the only estimated parameter and parameterise the covariance term as $\frac{i_{x,y}}{\sqrt{n_x \cdot n_y}}$. Note that $i_{x,y}$ is the cross-trait LD-score regression intercept.

While the bivariate probability density function (PDF) of these summary statistics cannot be obtained analytically, we could derive its characteristic function (see Supplementary Methods 1.1), which is the product of some transformed version of the characteristic functions of $z_k^{(x)}, z_k^{(u)}, z_k^{(y)}$ and (η_k^x, η_k^y), yielding

$$\begin{aligned}\varphi_{\left(\widehat{\beta}_k^x, \widehat{\beta}_k^y\right)}(v, w) &= E\left[\exp\left(i \cdot \left(v \cdot \widehat{\beta}_k^x + w \cdot \widehat{\beta}_k^y\right)\right)\right] \\ &= \varphi_{z_k^{(u)}}\left(\frac{v \cdot (\alpha_{y\to x} \cdot q_y + q_x) + w \cdot (\alpha_{x\to y} \cdot q_x + q_y)}{1 - \alpha_{x\to y}\alpha_{y\to x}}\right) \\ &\quad \times \varphi_{z_k^{(x)}}\left(\frac{v + \alpha_{x\to y} \cdot w}{1 - \alpha_{x\to y}\alpha_{y\to x}}\right) \cdot \varphi_{z_k^{(y)}}\left(\frac{w + \alpha_{y\to x} \cdot v}{1 - \alpha_{x\to y}\alpha_{y\to x}}\right) \cdot \varphi_{(\eta_k^x, \eta_k^y)}(v, w)\end{aligned} \tag{23}$$

Approximating the local correlations of SNP k (ρ_k) by a spike and slab distribution, parameterised by the fraction of non-zero correlations (π_k) and the variance of the non-zero correlations (σ_k^2), allows the derivation of a closed form expressions of $\varphi_{z_k^{(u)}}$, $\varphi_{z_k^{(x)}}$ and $\varphi_{z_k^{(y)}}$.

Derivation of the likelihood function. Given that the characteristic function can be analytically derived, we used the inversion theorem (for characteristic functions) to obtain the joint distribution of $\left(\widehat{\beta}_k^x, \widehat{\beta}_k^y\right)$ as

$$f_{\left(\widehat{\beta}_k^x, \widehat{\beta}_k^y\right)}(x, y) = \left(\frac{1}{2\pi}\right)^2 \cdot \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(-i \cdot (x \cdot v + y \cdot w)) \cdot \varphi_{\left(\widehat{\beta}_k^x, \widehat{\beta}_k^y\right)}(v, w)\, dv\, dw \tag{24}$$

This integral can be efficiently computed by the Fast Fourier Transformation (FFT, see ref. [46] and references within. Detailed derivation is found in Supplementary Methods 1.2). To speed up computation, we bin SNPs according to their π_k and σ_k values which characterise the local LD distribution for each SNP k (10×10 bins with equidistant centres - see Supplementary Methods 1.3) and for SNPs in the same bin the PDF function is evaluated over a fine grid ($2^7 \times 2^7$ combinations) using the FFT.

To reduce the number of parameters we define $t_x := \sigma_u \cdot q_x$ and $t_y := \sigma_u \cdot q_y$ since σ_u and q_x are separately not identifiable, but only their product is. Extensive simulations have shown that π_u is unidentifiable, and hence is set to an arbitrary value of 0.1. For improved interpretability, we slightly reparameterise the likelihood function by using $h_x^2 := \pi_x \cdot M \cdot \sigma_x^2, h_y^2 := \pi_y \cdot M \cdot \sigma_y^2$. Since different SNPs are correlated we have to estimate the over-counting of each SNP. We choose the same strategy as LD-score regression[43] and weigh each SNP by the inverse of its restricted LD score, i.e. $w_k = 1/\sum_{j=1}^{m_0} r_{jk}^2$, where r_{jk} is the correlation between GWAS SNPs k and j. The log-likelihood function is, thus, of the form

$$\log\left(\mathcal{L}\left(\boldsymbol{\theta}| \begin{pmatrix} \widehat{\boldsymbol{\beta}}^x \\ \widehat{\boldsymbol{\beta}}^y \end{pmatrix}\right)\right) \propto \sum_{k=1}^{K} w_k \cdot f_k\left(\widehat{\beta}_k^x, \widehat{\beta}_k^y\right) \tag{25}$$

where $f_k\left(\widehat{\beta}_k^x, \widehat{\beta}_k^y\right)$ is the log-likelihood function value for SNP k. Parameters $\{n_x, n_y, m, \sigma_{k=1,\ldots,K}, \pi_{k=1,\ldots,K}\}$ are known and the other 11 parameters

$$\boldsymbol{\theta} = \{\pi_x, \pi_y, h_x^2, h_y^2, t_x, t_y, \alpha_{x\to y}, \alpha_{y\to x}, i_x, i_y, i_{x,y}\}$$

are to be estimated from the observed association summary statistics. In order to further speed up computation, we estimate the 11 parameters in two separate steps: we first estimate for each trait the parameters π_x, i_x and π_y, i_y (SNP polygenicity and LD-score intercept) and the total heritability (unlike the direct heritability obtained by the full-model of LHC-MR) by using a simplified model with only the trait of interest, without a second trait or confounder, e.g. we fit only π_x, h_x^2 and i_x using $\widehat{\boldsymbol{\beta}}^x$ and assume that π_x and i_x do not change when two traits are taken into account. Note that π_x may change slightly (decreasing from the total to direct polygenicity), but its value has little impact on the likelihood function. The estimates from the first step are then fixed for the parameter estimation of trait pairs in the second step. Since only π_x, i_x and π_y, i_y are fixed, the remaining parameters to estimate are now:

$$\boldsymbol{\theta} = \{h_x^2, h_y^2, t_x, t_y, \alpha_{x\to y}, \alpha_{y\to x}, i_{x,y}\}$$

It is key to note that our approach does not aim to estimate individual (direct or indirect) SNP effects, as these are handled as random effects. By replacing U with $-U$ we swap the signs of both t_x and t_y, therefore these parameters are unique only if the sign of one of them is fixed. Thus, we will have the following restrictions on the parameter ranges: h_x^2, h_y^2, t_x are in $[0, 1]$, t_y, $\alpha_{x\to y}$, $\alpha_{y\to x}$, $i_{x,y}$ are in $[-1, 1]$.

Likelihood maximisation and standard error calculation. Our method, termed Latent Heritable Confounder Mendelian Randomisation (LHC-MR), maximises this likelihood function to obtain the MLE. Due to the complexity of the likelihood surface, we initialise the maximisation using 50 different starting points, where they come from a uniform distribution within the parameter-specific ranges mentioned above. We then choose parameter estimates corresponding to the highest likelihood of the 50 runs. Run time depends on the number of iterations during the maximisation procedure, and is linear with respect to the number of SNPs. It takes ~0.25 CPU-minute to fit the complete model to 50,000 SNPs with a single starting point.

Given the particular nature of the underlying directed graph, two different sets of parameters lead to an identical fit of the data, resulting in two global optima. The reason for this is the difficulty in distinguishing the ratio of the confounder effects (t_y/t_x) from the causal effect ($\alpha_{x\to y}$), as illustrated in Supplementary Fig. 2 by the slopes belonging to different SNP-clusters. More rigorously, it can be show that if $\{h_x, h_y, \alpha_{x\to y}, \alpha_{y\to x}, t_x, t_y\}$ is an optimum, then so will be $\{h'_x, h'_y, \alpha'_{x\to y}, \alpha'_{y\to x}, t'_x, t'_y\}$, where

$$h'_x = t_x + t_y \cdot \alpha_{y\to x} \tag{26}$$

$$h'_y = h_y \tag{27}$$

$$\alpha' = \frac{\alpha_{x\to y} + w}{1 + \alpha_{y\to x} \cdot w} \tag{28}$$

$$\alpha' = \alpha_{y\to x} \tag{29}$$

$$t'_x = h_x \cdot (1 + \alpha_{y\to x} \cdot w) \tag{30}$$

$$t'_y = -h_x \cdot w \tag{31}$$

with $w = t_y/t_x$ (for further derivations, see Supplementary Methods 1.4). This allows us to directly obtain both optima, even if the optimisation only revealed one of them. It happens very often that one of these parameter sets are outside of the allowed ranges and hence can be automatically excluded. If not, we keep track of both parameter estimates maximising the likelihood function. Note that, we call the one for which the direct heritability is larger than the indirect one, i.e. $h_x^2 > t_x^2$, the primary solution. We show that for real data application this solution is far more plausible than the alternative optimum. Finally, note that such bimodality can be observed at different levels: (i) For one given data generation, using multiple starting points leads to different optima; (ii) LHC-MR applied to multiple different data generations for a fixed parameter setting can yield different optima. Both of these situations are signs of the same underlying phenomenon and most often co-occur.

We implemented the block-jackknife procedure that is also used by LD-score regression to calculate the standard errors. For this we split the genome into 200 jackknife blocks and compute MLE in a leave-one-block-out fashion yielding $\widehat{\boldsymbol{\theta}}^{(-i)}, i = 1, \ldots, 200$ estimates. The variance of the full SNP MLE is then defined as $Var(\widehat{\theta}) := \frac{m - m\cdot(1/200)}{m\cdot(1/200)} \cdot \frac{1}{200-1} \sum_{i=1}^{200} (\widehat{\boldsymbol{\theta}}^{(-i)} - \widehat{\boldsymbol{\theta}})^2 = \sum_{i=1}^{200} (\widehat{\boldsymbol{\theta}}^{(-i)} - \widehat{\boldsymbol{\theta}})^2$.

Decomposition of genetic correlation. Given the starting equations for X and Y (Eqs. (2)–(3)) we can calculate their genetic correlation as the ratio between their genetic covariance and variance (calculated from their heritabilities) as such:

$$corr(\delta_x, \delta_y) = \frac{(t_x + \alpha_{y\to x} t_y)(t_y + \alpha_{x\to y} t_x) + \alpha_{y\to x} h_y^2 + \alpha_{x\to y} h_x^2}{\sqrt{\left((t_x + \alpha_{y\to x} t_y)^2 + \alpha_{y\to x}^2 h_y^2 + h_x^2\right)\left((t_y + \alpha_{x\to y} t_x)^2 + \alpha_{x\to y}^2 h_x^2 + h_y^2\right)}} \tag{32}$$

The full details of the derivation is found in Supplementary Methods 1.5. Using our estimated parameters, we first calculate the correlation based on Eq. (32) and then compare them to those obtained by LD-score regression.

Simulation settings. First, we tested LHC-MR using realistic parameter settings with a mild violation of the classical MR assumptions. These standard parameter settings consisted of simulating $m = 234{,}000$ SNPs for two non-overlapping cohorts of equal size (for simplicity) of $n_x = n_y = 50{,}000$ for each trait. X, Y and U were simulated with moderate polygenicity ($\pi_x = 5\times10^{-3}, \pi_y = 1\times10^{-2}, \pi_u = 5\times10^{-2}$), and considerable direct heritability ($h_x^2 = 0.25, h_y^2 = 0.2, h_u^2 = 0.3$). U had a confounding effect on the two traits as such, $q_x = 0.3, q_y = 0.2$ (resulting in $t_x = 0.16$, $t_y = 0.11$), and X had a direct causal effect on Y ($\alpha_{x\to y} = 0.3$), while the reverse causal effect from Y to X was set to null. Note that in this setting the total heritability of each of these traits is principally driven by direct effects and less than 10% of the total heritability is through a confounder and in case of Y less than an additional 8% of its total heritability is through X.It is important to note that for each tested parameter setting, we generated 50 different datasets, and each data generation underwent a likelihood maximisation of Eq. (25) using 50 starting points, and produced estimated parameters corresponding to the highest likelihood (simplified schema in Supplementary Fig. 3).

In the following simulations, we changed various parameters of these standard settings to test the robustness of the method. We explored how increased sample size ($n_x = n_y = 500{,}000$) or differences in sample sizes ($(n_x, n_y) = (50{,}000, 500{,}000)$ and $(n_x, n_y) = (500{,}000, 50{,}000)$) influence causal effect estimates of LHC-MR and other MR methods. We also simulated data with no causal effect (or with no confounder) and then examined how LHC-MR estimates those parameters. Next, we varied our causal effects between the two traits by lowering $\alpha_{x\to y}$ to 0.1, and in another setting by introducing a reverse causal effect ($\alpha_{y\to x} = -0.1$). In addition, we tried to create extremely unfavourable conditions for all MR analyses by varying the confounding effects. We did this in several ways: (i) increasing q_x and q_y ($q_x = 0.75, q_y = 0.50$), (ii) having a confounder with causal effects of opposite signs on X and Y ($q_x = 0.3, q_y = -0.2$). We also drastically increased the proportion of SNPs with non-zero effect on traits X, Y and U (π_x, π_y and $\pi_u = 0.1, 0.15, 0.2$ respectively). We also simulated data whereby the confounder has lower ($\pi_u = 0.01$) polygenicity than the two focal traits.

Finally, we explored various violations of the assumptions of our model (see Methods Section). First, we introduced two confounders in the simulated data, once with causal effects on X and Y that were concordant ($t_x^{(1)} = 0.16, t_y^{(1)} = 0.11, t_x^{(2)} = 0.22, t_y^{(2)} = 0.16$) in sign, and another with discordant effects ($t_x^{(1)} = 0.16, t_y^{(1)} = 0.11, t_x^{(2)} = 0.22, t_y^{(2)} = -0.16$), while still fitting the model with only one U. Second, we breached the assumption that the non-zero effects come from a Gaussian distribution. By design, the first three moments of the direct effects are fixed: they have zero mean, their variance is defined by the direct heritabilities and they must have zero skewness because the effect size distribution has to be symmetrical. Therefore, to violate the normality assumption, we varied the kurtosis (2, 3, 5 and 10) of the distribution drawn from the Pearson's distribution family. Third, we tested the assumption of the direct effects on our traits coming from a two-component Gaussian mixture by introducing a third component and observing how the estimates were effected. In this simulation scenario we introduced a large effect third component for X while decreasing the polygenicity of U ($\pi_{x1} = 1\times10^{-4}, \pi_{x2} = 1\times10^{-2}, h_{x1}^2 = 0.15$, $h_{x2}^2 = 0.1, \pi_u = 1\times10^{-2}$).

Application to real summary statistics. Once we demonstrated favourable performance of our method on simulated data, we went on to apply LHC-MR to summary statics obtained from the UK Biobank and other meta-analytic studies (Supplementary Table 1) in order to estimate pairwise bi-directional causal effect between 13 complex traits. The traits varied between conventional risk factors (such as low education, high body mass index (BMI), dislipidemia) and diseases (including diabetes and coronary artery disease among others). SNPs with imputation quality greater than 0.99, and minor allele frequency (MAF) greater than 0.5% were selected. Moreover, SNPs found within the human leukocyte antigen (HLA) region on chromosome 6 were removed due to the abundance of SNPs associated with autoimmune and infectious diseases as well as the complicated LD structure present in that region. For traits with total heritability below 2.5%, the outgoing causal effect estimates were ignored since instrumenting such barely heritable traits is questionable.

In order to perform LHC-MR between trait pairs, a set of overlapping SNPs was used as input for each pair. The effects of these overlapping SNPs were then aligned to the same effect allele in both traits. To decrease computation time further (while only minimally reducing power), we selected every 10th QC-filtered SNP as input for the analysis. We calculated regression weights using the UK10K panel, which may be sub-optimal for summary statistics not coming from the UK Biobank, but we have previously shown[47] that estimating LD in a ten-times larger dataset (UK10K) outweighs the benefit of using smaller, but possibly better-matched European panel (1000 Genomes[48]).

We also ran IVW for each trait pair in both directions to estimate bi-directional causal effects as well as LD-score regression to get the cross-trait intercept term. We then added uniformly distributed ($\sim U(-0.1, 0.1)$) noise to these pre-estimated parameters to generate starting points for the second step of the likelihood optimisation. These closer-to-target starting points did not change the optimisation results, simply sped up the likelihood maximisation and increased the chances to converge to the same (primary) optimum. The LHC-MR procedure was run for each pair of traits 100 times, each using a different set of randomly generated starting points within the ranges of their respective parameters. For the optimisation of the likelihood function (Eq. (25)), we used the R function 'optim' from the 'stats' R package[49]. Once we fitted this *complete* model estimating 11 parameters in two steps $\{i_x, i_y, \pi_x, \pi_y, h_x^2, h_y^2, t_x, t_y, \alpha_{x\to y}, \alpha_{y\to x}, i_{xy}\}$, we then ran block jackknife to obtain the SE of the parameters estimated in the second step: $\{h_x^2, h_y^2, t_x, t_y, \alpha_{x\to y}, \alpha_{y\to x}, i_{xy}\}$.

To support the existence of the confounders identified by LHC-MR, we used EpiGraphDB[33,34] to systematically identify those potential confounders. The database provided for each potential confounder of a causal relationship, a causal effect on trait X and Y ($r1$, and $r3$ in their notation), the sign of the ratio of which ($sign(r_3/r_1)$) was compared to the sign of the LHC-MR estimated t_y/t_x values representing the strength of the confounder acting on the two traits. We restricted our comparison to the sign only, since the $r1, r3$ values reported in EpiGraphDB are not necessarily on the same scale.

Comparison against conventional MR methods and CAUSE. We compared the causal parameter estimates of the LHC-MR method to those of five conventional MR approaches (MR-Egger, weighted median, IVW, mode MR and weighted mode MR) using a Z-test[50]. The 'TwoSampleMR' R package[51] was used to get the causal estimates for all the pairwise traits as well as their standard errors from the above-mentioned MR methods. The same set of genome-wide SNPs that were used by LHC-MR, were used as input for the package. SNPs associated with the exposure were selected to various degrees (for simulation we selected SNPs over a range of thresholds: absolute p-value $< 5\times10^{-4}$ to $< 5\times10^{-8}$), and SNPs more strongly associated with the outcome than with the exposure (p-value < 0.05 in one-sided t-test) were removed. The default package settings for the clumping of SNPs ($r^2 = 0.001$) were used and the analysis was run with no further changes. We tested the agreement between the significance and direction of our estimates and that of standard MR methods, with the focus being on finding differences in statistical conclusions regarding causal effect sizes.

We compared our causal estimates from all our simulation settings to the causal estimates obtained by running MR-RAPS[11] also using the 'TwoSampleMR' R package, once by using the entire set of SNPs, and another by filtering for SNPs with a significance threshold of $<5\times10^{-4}$. We also compared both our simulation as well as real data results against those of CAUSE[10]. We first generated simulated data under the LHC model and used them as input to estimate the causal effect using CAUSE. We then generated simulated data using the CAUSE framework and inputted them into LHC-MR (as well as standard MR methods) to estimate the causal parameters. Lastly, we compared causal estimates obtained for the 78 trait pairs (156 bi-directional causal effects) from LHC-MR to those obtained when running CAUSE.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The origin of the summary statistics data used is referenced in Supplementary Table 1. The UK Biobank summary statistics data used in this study came from the Neale Lab[52], and can be downloaded from http://www.nealelab.is/uk-biobank. Data on coronary artery disease[53] have been contributed by the CARDIoGRAMplusC4D and UK Biobank CardioMetabolic Consortium CHD working group who used the UK Biobank Resource (application number 9922). Data have been downloaded from http://www.cardiogramplusc4d.org/data-downloads/. The computed local LD scores described in Supplementary Methods 1.3 can be downloaded from https://wp.unil.ch/sgg/lhc-mr/. We also used EpiGraphDB, an analytical platform and database to support data mining in epidemiology, to preform Phenome-wide MR search. Access to EpigraphDB is free and may be done through their web application https://epigraphdb.org or their R package https://github.com/MRCIEU/epigraphdb-r.

Code availability

The source code[54] for this work can be found on https://github.com/LizaDarrous/LHC-MR_v2/(https://doi.org/10.5281/zenodo.5534639), it has also been developed into an R package that can be downloaded from https://github.com/LizaDarrous/lhcMR.

Received: 27 August 2020; Accepted: 28 October 2021;
Published online: 14 December 2021

References

1. Fewell, Z., Davey Smith, G. & Sterne, J. A. C. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. *Am. J. Epidemiol.* **166**, 646–655 (2007).
2. Pingault, J.-B. et al. Using genetic data to strengthen causal inference in observational research. *Nat. Rev. Genet.* **19**, 566–580 (2018).
3. Barter, P. J. et al. Effects of torcetrapib in patients at high risk for coronary events. *N. Engl. J. Med.* **357**, 2109–2122 (2007).
4. Fordyce, C. B. et al. Cardiovascular drug development: is it dead or just hibernating? *J. Am. Coll. Cardiol.* **65**, 1567–1582 (2015).
5. Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N. & Davey Smith, G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* **27**, 1133–1163 (2008).
6. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665 (2013).
7. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).
8. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).
9. Hartwig, F. P., Davey Smith, G. & Bowden, J. Robust inference in summary data mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* **46**, 1985–1998 (2017).
10. Morrison, J., Knoblauch, N., Marcus, J. H., Stephens, M. & He, X. Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. *Nat. Genet.* **52**, 740–747 (2020).
11. Zhao, Q., Wang, J., Hemani, G., Bowden, J. & Small, D. S. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *Ann. Stat.* **48**, 1742 – 1769 (2020).
12. Laclaustra, M. et al. Ldl cholesterol rises with bmi only in lean individuals: Cross-sectional u.s. and spanish representative data. *Diabetes Care* **41**, 2195–2201 (2018).
13. Drøyvold, W. B., Midthjell, K., Nilsen, T. I. L. & Holmen, J. Change in body mass index and its impact on blood pressure: a prospective population study. *Int. J. Obes.* **29**, 650–655 (2005).
14. Lee, M.-R., Lim, Y.-H. & Hong, Y.-C. Causal association of body mass index with hypertension using a mendelian randomization design. *Medicine (Baltimore)* **97**, e11252 (2018).
15. Corbin, L. J. et al. Bmi as a modifiable risk factor for type 2 diabetes: refining and understanding causal estimates using mendelian randomization. *Diabetes* **65**, 3002–3007 (2016).
16. Narayan, K., Boyle, J. P., Thompson, T. J., Gregg, E. W. & Williamson, D. F. Effect of bmi on lifetime risk for diabetes in the u.s. *Diabetes Care* **30**, 1562–1566 (2007).
17. Yusuf, S. et al. Obesity and the risk of myocardial infarction in 27,000 participants from 52 countries: a case-control study. *Lancet* **366**, 1640–1649 (2005).
18. Riaz, H. et al. Association between obesity and cardiovascular outcomes: a systematic review and meta-analysis of Mendelian Randomization Studies. *JAMA Netw. Open* **1**, e183788–e183788 (2018).
19. Sun, D. et al. Type 2 diabetes and hypertension. *Circulation Res.* **124**, 930–937 (2019).
20. Tomeo, C. A., Field, A. E., Berkey, C. S., Colditz, G. A. & Frazier, A. L. Weight concerns, weight control behaviors, and smoking initiation. *Pediatrics* **104**, 918–924 (1999).
21. Cawley, J., Markowitz, S. & Tauras, J. Lighting up and slimming down: the effects of body weight and cigarette prices on adolescent smoking initiation. *J. Health Econ.* **23**, 293–311 (2004).
22. Cao, M. & Cui, B. Association of educational attainment with adiposity, type 2 diabetes, and coronary artery diseases: a Mendelian Randomization Study. *Front. Public Health* **8**, 112 (2020).
23. Loucks, E. B. et al. Education and coronary heart disease risk associations may be affected by early-life common prior causes: a propensity matching analysis. *Ann. Epidemiol.* **22**, 221–232 (2012).
24. Gage, S. H., Bowden, J., Davey Smith, G. & Munafò, M. R. Investigating causality in associations between education and smoking: a two-sample Mendelian randomization study. *Int. J. Epidemiol.* **47**, 1131–1140 (2018).
25. Sanderson, E., Davey Smith, G., Bowden, J. & Munafò, M. R. Mendelian randomisation analysis of the effect of educational attainment and cognitive ability on smoking behaviour. *Nat. Commun.* **10**, 2949 (2019).
26. Marouli, E. et al. Mendelian randomisation analyses find pulmonary factors mediate the effect of height on coronary artery disease. *Commun. Biol.* **2**, 119 (2019).
27. Tan, L. E., Llano, A., Aman, A., Dominiczak, A. F. & Padmanabhan, S. A18709 mendelian randomization study of causal relationship of height on blood pressure and arterial stiffness. *J. Hypertens.* **36**, e91–e92 (2018).
28. Laaksonen, D. E. et al. Dyslipidaemia as a predictor of hypertension in middle-aged men. *Eur. Heart J.* **29**, 2561–2568 (2008).
29. Davies, N. M. et al. Within family Mendelian randomization studies. *Hum. Mol. Genet.* **28**, R170–R179 (2019).
30. Benson, R., von Hippel, P. T. & Lynch, J. L. Does more education cause lower bmi, or do lower-bmi individuals become more educated? evidence from the national longitudinal survey of youth 1979. *Soc. Sci. Med.* **211**, 370–377 (2018).
31. Witter, F. R. & Luke, B. The effect of maternal height on birth weight and birth length. *Early Hum. Dev.* **25**, 181–186 (1991).
32. Tyrrell, J. et al. Height, body mass index, and socioeconomic status: mendelian randomisation study in uk biobank. *BMJ* **352**, i582 (2016).
33. MRC IEU. *EpiGraphDB*. http://epigraphdb.org/ (2019).
34. Liu, Y. et al. Epigraphdb: a database and data mining platform for health data science. *Bioinformatics* **37**, 1304–1311 (2021).
35. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
36. Mounier, N. & Kutalik, Z. Correction for sample overlap, winner's curse and weak instrument bias in two-sample mendelian randomization. *bioRxiv* https://www.biorxiv.org/content/10.1101/2021.03.26.437168v1?rss=1 (2021).
37. Thomas, D., Strauss, J. & Henriques, M.-H. How does mother's education affect child height? *J. Hum. Resour.* **26**, 183–211 (1991).
38. Warrington, N. M. et al. Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nat. Genet.* **51**, 804–814 (2019).
39. Brumpton, B. et al. Within-family studies for mendelian randomization: avoiding dynastic, assortative mating, and population stratification biases. *Nat. Commun.* **11**, 3519 (2020).
40. O'Connor, L. J. & Price, A. L. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat. Genet.* **50**, 1728–1734 (2018).
41. Brown, B. C. & Knowles, D. A. Phenome-scale causal network discovery with bidirectional mediated mendelian randomization. *bioRxiv* https://www.biorxiv.org/content/10.1101/2020.06.18.160176v2.full (2020).
42. Howey, R., Shin, S.-Y., Relton, C., Smith, G. D. & Cordell, H. J. Bayesian network analysis incorporating genetic anchors complements conventional mendelian randomization approaches for exploratory analysis of causal relationships in complex data. *PLoS Genet* **16**, e1008198 (2020).
43. Bulik-Sullivan, B. K. et al. Ld score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
44. Jordan, D. M., Verbanck, M. & Do, R. HOPS: a quantitative score reveals pervasive horizontal pleiotropy in human genetic variation is driven by extreme polygenicity of human traits and diseases. *Genome Biol.* **20**, 222 (2019).
45. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
46. Heideman, M. T., Johnson, D. H. & Burrus, C. S. Gauss and the history of the fast fourier transform. *Arch. Hist. Exact. Sci.* **34**, 265–277 (1985).
47. Rüeger, S., McDaid, A. & Kutalik, Z. Evaluation and application of summary statistic imputation to discover new height-associated loci. *PLoS Genet.* **14**, e1007371 (2018).
48. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
49. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2019).
50. Clogg, C. C., Petkova, E. & Haritou, A. Statistical methods for comparing regression coefficients between models. *Am. J. Sociol.* **100**, 1261–1293 (1995).
51. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7**, e34408 (2018).
52. Neale Lab. *UK BioBank*. http://www.nealelab.is/uk-biobank/ (2018).
53. van der Harst, P. & Verweij, N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. *Circ. Res.* **122**, 433–443 (2018).
54. Darrous, L. Simultaneous estimation of bi-directional causal effects and heritable confounding from GWAS summary statistics. *Software* https://doi.org/10.5281/zenodo.5534639 (2021).

Acknowledgements

This research has been conducted using the UK Biobank Resource under Application Number 16389. L.D. scores were calculated based on the UK10K data resource (EGAD00001000740, EGAD00001000741). Z.K. was funded by the Swiss National Science Foundation (31003A_169929, 310030_189147 and 32003B_173092). For computations we used the CHUV HPC cluster. We would like to thank Jack Bowden, Valentin Rousson, Matthew Robinson, George Davey Smith, Thomas Richardson, Eleonora Porcu and Sven Erik Ojavee for their valuable feedback and comments on this manuscript.

Author contributions

Z.K. devised and directed the project. Z.K., N.M. and L.D. contributed to the mathematical derivations, design and implementation of the research, the analysis of the results and to the writing of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-26970-w.

Correspondence and requests for materials should be addressed to Zoltán. Kutalik.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information

Supplementary Methods

1.1 Characteristic functions of $z_k^{(u)}$, $z_k^{(x)}$, $z_k^{(y)}$ and (η_k^x, η_k^y)

The bivariate probability density function (PDF) of these summary statistics cannot be obtained analytically, but in the following we demonstrate that the characteristic function can be derived. Let us first compute the characteristic function of this two-dimensional random variable, knowing that $z_k^{(x)}, z_k^{(u)}, z_k^{(y)}$ and (η_k^x, η_k^y) are independent, hence the characteristic function can be factorised:

$$\begin{aligned}
\varphi_{\left(\widehat{\beta}_k^x, \widehat{\beta}_k^y\right)}(v,w) &= E\left[\exp\left(i \cdot \left(v \cdot \widehat{\beta}_k^x + w \cdot \widehat{\beta}_k^y\right)\right)\right] \\
&= E\left[\exp\left(i \cdot \left(v \cdot \left(\frac{z_k^{(x)} + (\alpha_{y\to x} \cdot q_y + q_x) \cdot z_k^{(u)} + \alpha_{y\to x} \cdot z_k^{(y)}}{1 - \alpha_{x\to y}\alpha_{y\to x}} + \eta_k^x\right) + \right.\right.\right. \\
&\left.\left.\left. + \ w \cdot \left(\frac{z_k^{(y)} + (\alpha_{x\to y} \cdot q_x + q_y) \cdot z_k^{(u)} + \alpha_{x\to y} \cdot z_k^{(x)}}{1 - \alpha_{x\to y}\alpha_{y\to x}} + \eta_k^y\right)\right)\right)\right] \\
&= E\left[\exp\left(i \cdot z_k^{(u)} \cdot \frac{v \cdot (\alpha_{y\to x} \cdot q_y + q_x) + w \cdot (\alpha_{x\to y} \cdot q_x + q_y)}{1 - \alpha_{x\to y}\alpha_{y\to x}}\right)\right] \\
&\times E\left[\exp\left(i \cdot z_k^{(x)} \cdot \frac{v + \alpha_{x\to y} \cdot w}{1 - \alpha_{x\to y}\alpha_{y\to x}}\right)\right] \cdot E\left[\exp\left(i \cdot z_k^{(y)} \cdot \frac{w + \alpha_{y\to x} \cdot v}{1 - \alpha_{x\to y}\alpha_{y\to x}}\right)\right] \\
&\times E\left[\exp\left(i \cdot \left(v \cdot \eta_k^x + w \cdot \eta_k^y\right)\right)\right] \\
&= \varphi_{z_k^{(u)}}\left(\frac{v \cdot (\alpha_{y\to x} \cdot q_y + q_x) + w \cdot (\alpha_{x\to y} \cdot q_x + q_y)}{1 - \alpha_{x\to y}\alpha_{y\to x}}\right) \\
&\times \varphi_{z_k^{(x)}}\left(\frac{v + \alpha_{x\to y} \cdot w}{1 - \alpha_{x\to y}\alpha_{y\to x}}\right) \cdot \varphi_{z_k^{(y)}}\left(\frac{w + \alpha_{y\to x} \cdot v}{1 - \alpha_{x\to y}\alpha_{y\to x}}\right) \cdot \varphi_{(\eta_k^x, \eta_k^y)}(v,w)
\end{aligned} \tag{1}$$

In the following we will work out each of the characteristic functions on the right hand side.

It is reasonable to assume that linkage disequilibrium (LD) fades off beyond 1Mb distance. Thus, without loss of generality we can assume that non-zero LD does not extend beyond m_0 markers around the focal variant. Hence we can assume that the length of $\boldsymbol{\rho_k}$ is m_0 and only consider $\boldsymbol{\gamma_x}, \boldsymbol{\gamma_y}$ and $\boldsymbol{\gamma_u}$ to be of length m_0 instead of m. Let us first approximate the distribution of $\boldsymbol{\rho}_k$ values following a spike and slab Gaussian mixture, i.e. proportion π_k of the m_0 SNPs have non-zero LD, coming from a Gaussian distribution $\mathcal{N}(0, \sigma_k^2)$ and the remaining $(1-\pi_k)$ fraction of the LD values is zero. In mathematical notation

$$\boldsymbol{\rho}_k = \boldsymbol{r}_k \odot \boldsymbol{\kappa}_k \quad \text{with} \quad \boldsymbol{r_k} \sim \mathcal{N}(0, \sigma_k^2 \cdot I) \quad \text{and} \quad \boldsymbol{\kappa}_k \sim \mathcal{B}_{m_0}(1, \pi_k)$$

Therefore $z_k^{(u)}$ can be written of the form

$$\begin{aligned}
z_k^{(u)} &= \boldsymbol{\rho_k'} \cdot (\boldsymbol{\zeta_u} \odot \boldsymbol{\kappa_u}) = (\boldsymbol{r}_k \odot \boldsymbol{\kappa}_k)' \cdot (\boldsymbol{\zeta_u} \odot \boldsymbol{\kappa_u}) = (\boldsymbol{r}_k \odot \boldsymbol{\zeta_u})' \cdot (\underbrace{\boldsymbol{\kappa}_k \odot \boldsymbol{\kappa_u}}_{\boldsymbol{\kappa_{k,u}} \sim \mathcal{B}(1, \pi_k \cdot \pi_u)}) \\
&= \sum_{j=1}^{m_0} (\boldsymbol{r}_k \odot \boldsymbol{\zeta_u})_j \cdot (\boldsymbol{\kappa_{k,u}})_j
\end{aligned} \tag{2}$$

The PDF of the product of two zero-mean Gaussians ($\boldsymbol{r}_k$ and $\boldsymbol{\zeta_u}$) is a modified Bessel function of the second kind of order zero ($K_0(\omega)$) [1], more precisely

$$f_{(\boldsymbol{r}_k \odot \boldsymbol{\zeta_u})_j}(t) = \frac{1/\pi}{\sigma_u \cdot \sigma_k} \cdot K_0\left(\frac{|t|}{\sigma_u \cdot \sigma_k}\right) \tag{3}$$

and its characteristic function [2,3] is

$$\varphi_0(t) = E(\exp(i \cdot t \cdot (\boldsymbol{r}_k \odot \boldsymbol{\zeta_u})_j)) = \frac{1}{\sqrt{\sigma_u^2 \cdot \sigma_k^2 \cdot t^2 + 1}} \tag{4}$$

Next, the characteristic function of the product of $(\boldsymbol{r}_k \odot \boldsymbol{\zeta_u})_j$ and a Bernoulli distributed $(\boldsymbol{\kappa_{k,u}})_j$ is

$$\begin{aligned}\varphi_1(t) &= E\left(\exp\left(i \cdot t \cdot (\boldsymbol{r}_k \odot \boldsymbol{\zeta_u})_j\right) \cdot (\boldsymbol{\kappa_{k,u}})_j)\right) \\ &= \pi_k \cdot \pi_u \cdot E(\exp(i \cdot t \cdot (\boldsymbol{r}_k \odot \boldsymbol{\zeta_u})_j)) + (1 - \pi_k \cdot \pi_u) \cdot E(\exp(i \cdot t \cdot 0)) \\ &= \pi_k \cdot \pi_u \cdot \varphi_0(t) + (1 - \pi_k \cdot \pi_u) \\ &= \frac{\pi_k \cdot \pi_u}{\sqrt{\sigma_u^2 \cdot \sigma_k^2 \cdot t^2 + 1}} + (1 - \pi_k \cdot \pi_u) \end{aligned} \tag{5}$$

Hence the characteristic function of the sum of m_0 independent random variables is the product of them, we have

$$\varphi_{z_k^{(u)}}(t) = \left(\frac{\pi_k \cdot \pi_u}{\sqrt{\sigma_u^2 \cdot \sigma_k^2 \cdot t^2 + 1}} + (1 - \pi_k \cdot \pi_u)\right)^{m_0} \tag{6}$$

Finally, we apply a first order Taylor series approximation (around 1) of the log of the characteristic function in order to speed up computation and improve numerical accuracy

$$\begin{aligned}\log(\varphi_{z_k^{(u)}}(t)) &= m_0 \cdot \log\left(\frac{\pi_k \cdot \pi_u}{\sqrt{\sigma_u^2 \cdot \sigma_k^2 \cdot t^2 + 1}} + (1 - \pi_k \cdot \pi_u)\right) \\ &= m_0 \cdot \log\left(1 - \pi_k \cdot \pi_u \cdot \left(1 - \frac{1}{\sqrt{\sigma_u^2 \cdot \sigma_k^2 \cdot t^2 + 1}}\right)\right) \\ &\approx -m_0 \cdot \pi_k \cdot \pi_u \cdot \left(1 - \frac{1}{\sqrt{\sigma_u^2 \cdot \sigma_k^2 \cdot t^2 + 1}}\right) \end{aligned} \tag{7}$$

Analogously, the approximation of the logarithm of the characteristic functions of $z_k^{(x)}$ and $z_k^{(y)}$ is

$$\log(\varphi_{z_k^{(x)}}(t)) \approx -m_0 \cdot \pi_k \cdot \pi_x \cdot \left(1 - \frac{1}{\sqrt{\sigma_x^2 \cdot \sigma_k^2 \cdot t^2 + 1}}\right) \tag{8}$$

$$\log(\varphi_{z_k^{(y)}}(t)) \approx -m_0 \cdot \pi_k \cdot \pi_y \cdot \left(1 - \frac{1}{\sqrt{\sigma_y^2 \cdot \sigma_k^2 \cdot t^2 + 1}}\right) \tag{9}$$

Since the characteristic function of a centred multivariate Gaussian with variance-covariance matrix Σ is $\exp(-(1/2) \cdot \boldsymbol{t'} \cdot \Sigma \cdot \boldsymbol{t})$ we have

$$\log\left(\varphi_{(\eta_k^x, \eta_k^y)}(v, w)\right) = -\frac{1}{2} \cdot \left(\frac{i_x}{n_x} \cdot v^2 + 2 \cdot \frac{i_{x,y}}{\sqrt{n_x \cdot n_y}} \cdot v \cdot w + \frac{i_y}{n_y} \cdot w^2\right) \tag{10}$$

1.2 From characteristic function to probability density function

The final form of the logarithm of the joint characteristic function of the transformed summary statistics is

$$
\begin{aligned}
\log\left(\varphi_{\left(\widehat{\beta}_k^x,\widehat{\beta}_k^y\right)}(v,w)\right) &= \log\left(\varphi_{z_k^{(x)}}\left(\frac{v+\alpha_{x\to y}w}{1-\alpha_{x\to y}\alpha_{y\to x}}\right)\right)+\log\left(\varphi_{z_k^{(y)}}\left(\frac{w+\alpha_{y\to x}v}{1-\alpha_{x\to y}\alpha_{y\to x}}\right)\right)\\
&+\log\left(\varphi_{z_k^{(u)}}\left(\frac{v\cdot(\alpha_{y\to x}\cdot q_y+q_x)+w\cdot(\alpha_{x\to y}\cdot q_x+q_y}{1-\alpha_{x\to y}\alpha_{y\to x}})\right)\right)\\
&+\log\left(\varphi_{\left(\eta_k^x,\eta_k^y\right)}(v,w)\right)\\
&\approx -m_0\cdot\pi_k\cdot\pi_x\cdot\left(1-\frac{1}{\sqrt{\frac{\sigma_x^2\cdot\sigma_k^2\cdot(v+\alpha_{x\to y}w)^2}{(1-\alpha_{x\to y}\alpha_{y\to x})^2}+1}}\right)\\
&-m_0\cdot\pi_k\cdot\pi_y\cdot\left(1-\frac{1}{\sqrt{\frac{\sigma_y^2\cdot\sigma_k^2\cdot(w+\alpha_{y\to x}v)^2}{(1-\alpha_{x\to y}\alpha_{y\to x})^2}+1}}\right)\\
&-m_0\cdot\pi_k\cdot\pi_u\cdot\left(1-\frac{1}{\sqrt{\frac{\sigma_u^2\cdot\sigma_k^2\cdot(v\cdot(\alpha_{y\to x}\cdot q_y+q_x)+w\cdot(\alpha_{x\to y}\cdot q_x+q_y))^2}{(1-\alpha_{x\to y}\alpha_{y\to x})^2}+1}}\right)\\
&-\frac{1}{2}\cdot\left(\frac{i_x}{n_x}\cdot v^2+2\cdot\frac{i_{x,y}}{\sqrt{n_x\cdot n_y}}\cdot v\cdot w+\frac{i_y}{n_y}\cdot w^2\right)
\end{aligned}
\tag{11}
$$

Using the inversion theorem for characteristic functions we can express the joint distribution of $\left(\widehat{\beta}_k^x,\widehat{\beta}_k^y\right)$ as

$$
f_{\left(\widehat{\beta}_k^x,\widehat{\beta}_k^y\right)}(x,y)=\left(\frac{1}{2\pi}\right)^2\cdot\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp(-i\cdot(x\cdot v+y\cdot w))\cdot\varphi_{\left(\widehat{\beta}_k^x,\widehat{\beta}_k^y\right)}(v,w)\;dv\;dw \tag{12}
$$

This integral can be efficiently computed by Fast Fourier Transformation (FFT, see [4] and references within). To speed up computation, we bin SNPs according to their π_k and σ_k values (10×10 bins with equidistant centres) and for SNPs in the same bin the PDF function is evaluated over a fine grid ($2^7\times 2^7$ combinations) using the FFT.

Note that any derivative of the likelihood function can be readily calculated as a FFT of the derivative of the characteristic function, i.e.

$$
\frac{\partial}{\partial\theta}f_{\left(\widehat{\beta}_k^x,\widehat{\beta}_k^y\right)}(x,y)=\left(\frac{1}{2\pi}\right)^2\cdot\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp(-i\cdot(x\cdot v+y\cdot w))\cdot\frac{\partial}{\partial\theta}\varphi_{\left(\widehat{\beta}_k^x,\widehat{\beta}_k^y\right)}(v,w)\;dv\;dw \tag{13}
$$

1.3 Computation of the LD scores

We first took 4,773,627 SNPs with info (imputation certainty measure) ≥ 0.99 present in the association summary files from the second round of GWAS by the Neale lab[5]. This set was restricted to 4,650,107 common, high-quality SNPs, defined as being present in both UK10K and UK Biobank, having MAF $> 1\%$ in both data sets, non-significant ($P_{diff} > 0.05$) allele frequency difference between UK Biobank and UK10K and residing outside the HLA region (chr6:28.5-33.5Mb). For these SNPs, LD scores and regression weights were computed based on

3,781 individuals from the UK10K study[6]. To estimate the local LD distribution for each SNP (k), characterised by π_k, σ_k^2, we fitted a two-component Gaussian mixture distribution to the observed local correlations (focal SNP $+/-$ 2'500 markers with MAF$\geq$ 0.5% in the UK10K): (1) one Gaussian component corresponding to zero correlations, reflecting only measurement noise (whose variance is proportional to the inverse of the reference panel size) and (2) a second component with zero mean and a larger variance than the first component (encompassing measurement noise plus non-zero LD).

1.4 Likelihood function identifiability

The likelihood function is symmetric around U, but for simplicity we will consider the general case where the variables of U and X are flipped, although the same can be said for the variables of U and Y. The likelihood function is partially identifiable such that there exists for any given model parameters, another model with different parameters but with the exact same likelihood function.

Proof: given that the SNPs effects between trait X and the confounder U are flipped, the new parameters follow the following structure:

$$\begin{aligned}
h'_x &= t_x + t_y \cdot \alpha_{y\to x} && (14)\\
h'_y &= h_y && (15)\\
\alpha'_{y\to x} &= \alpha_{y\to x} && (16)\\
\alpha'_{x\to y} &= \frac{q_x \cdot \alpha_{x\to y} + q_y}{q_x + q_y \cdot \alpha_{y\to x}}\\
&= \frac{q_x(\alpha_{x\to y} + \frac{q_y}{q_x})}{q_x(1 + \frac{q_y}{q_x} \cdot \alpha_{y\to x})}\\
&= \frac{\alpha_{x\to y} + \frac{q_y}{q_x}}{1 + \frac{q_y}{q_x} \cdot \alpha_{y\to x}} && (17)
\end{aligned}$$

through inverse transformation,

$$\alpha_{x\to y} = \frac{\alpha'_{x\to y} + \frac{q'_y}{q'_x}}{1 + \frac{q'_y}{q'_x} \cdot \alpha_{y\to x}} \qquad (18)$$

Plugging in $\alpha'_{x\to y}$ in the above equation, and simplifying $\frac{ty}{tx}$ by w and $\frac{ty'}{tx'}$ by w' to get the confounding ratio:

$$\begin{aligned}
\alpha_{x\to y} &= \frac{\alpha'_{x\to y} + w'}{1 + w' \cdot \alpha_{y\to x}}\\
\alpha_{x\to y} + \alpha_{x\to y} \cdot w' \cdot \alpha_{y\to x} &= \alpha'_{x\to y} + w'\\
\alpha_{x\to y} - \alpha'_{x\to y} &= w' - \alpha_{x\to y} \cdot w' \cdot \alpha_{y\to x}\\
\alpha_{x\to y} - \alpha'_{x\to y} &= w'(1 - \alpha_{x\to y} \cdot \alpha_{y\to x})\\
w' &= \frac{\alpha_{x\to y} - \alpha'_{x\to y}}{1 - \alpha_{x\to y} \cdot \alpha_{y\to x}} && (19)
\end{aligned}$$

inserting the complete form of $\alpha'_{x\rightarrow y}$,

$$\begin{aligned}
w' &= \frac{\alpha_{x\rightarrow y} - \frac{\alpha_{x\rightarrow y}+\frac{q_y}{q_x}}{1+\frac{q_y}{q_x}\cdot\alpha_{y\rightarrow x}}}{1-\alpha_{x\rightarrow y}\cdot\alpha_{y\rightarrow x}} \\
&= \frac{\alpha_{x\rightarrow y}(1+w\cdot\alpha_{y\rightarrow x}) - w - \alpha_{x\rightarrow y}}{(1-\alpha_{x\rightarrow y}\cdot\alpha_{y\rightarrow x})(1+w\cdot\alpha_{y\rightarrow x})} \\
&= \frac{\alpha_{x\rightarrow y}\cdot w\cdot\alpha_{y\rightarrow x} - w}{(1-\alpha_{x\rightarrow y}\cdot\alpha_{y\rightarrow x})(1+w\cdot\alpha_{y\rightarrow x})} \\
&= \frac{w(\alpha_{x\rightarrow y}\cdot\alpha_{y\rightarrow x} - 1)}{(1-\alpha_{x\rightarrow y}\cdot\alpha_{y\rightarrow x})(1+w\cdot\alpha_{y\rightarrow x})} \\
&= \frac{-w}{1+w\cdot\alpha_{y\rightarrow x}} \qquad (20)
\end{aligned}$$

In order to obtain t'_y and t'_x, we use the equations of h'_x, $\alpha'_{x\rightarrow y}$ and by using the inverse transformation of $\alpha'_{y\rightarrow x} = \alpha_{y\rightarrow x}$, $\alpha_{x\rightarrow y}$ as well as w' as follows:

$$\begin{aligned}
t'_y &= \frac{-t'_x\cdot w}{1+w\cdot\alpha_{y\rightarrow x}} \qquad (21) \\
h_x &= t'_x + t'_y\cdot\alpha_{y\rightarrow x} \\
&= t'_x + \frac{-t'_x\cdot w}{1+w\cdot\alpha_{y\rightarrow x}}\cdot\alpha_{y\rightarrow x} \\
&= \frac{t'_x + t'_x\cdot w\cdot\alpha_{y\rightarrow x} - t'_x\cdot w\cdot\alpha_{y\rightarrow x}}{1+w\cdot\alpha_{y\rightarrow x}} \\
&= \frac{t'_x}{1+w\cdot\alpha_{y\rightarrow x}} \qquad (22) \\
t'_x &= h_x(1+w\cdot\alpha_{y\rightarrow x}) \qquad (23)
\end{aligned}$$

Replacing t'_x in h_x to get t'_y:

$$t'_y = h_x\cdot w \qquad (24)$$

Under these two models with equal likelihood, there are three slopes obtained from the observed data: two are the correlation of effect sizes ($\alpha_{x\rightarrow y}$ and $1/\alpha_{y\rightarrow x}$), where one of them is greater than, and the other is within the parameter bounds. The third is the correlation of the confounder $\frac{\alpha_{x\rightarrow y}+\frac{q_y}{q_x}}{1+\frac{q_y}{q_x}\cdot\alpha_{y\rightarrow x}}$.

More often than not, only one slope is recovered within the boundaries of the parameters set for LHC-MR. However, given the now known re-parameterisation, the second (and if found, third) slope can be simply calculated if not found by the likelihood function minimisation. It is reasonable to assume that the direct heritability of each trait is larger than the indirect heritability, hence we report parameter sets where $h^2_x > t^2_x$ or $h^2_y > t^2_y$.

1.5 Decomposition of genetic correlation

Given the starting equations for X and Y we can calculate their genetic correlation. Denoting the total (multivariate) genetic effect for X and Y as $\boldsymbol{\delta_x}$ and $\boldsymbol{\delta_y}$, we can express them as follows

$$\begin{align}\boldsymbol{\delta_x} &= q_x \cdot \boldsymbol{\gamma_u} + \alpha_{y\to x}\boldsymbol{\delta_y} + \boldsymbol{\gamma_x} \tag{25}\\ \boldsymbol{\delta_y} &= q_y \cdot \boldsymbol{\gamma_u} + \alpha_{x\to y}\boldsymbol{\delta_x} + \boldsymbol{\gamma_y} \tag{26}\end{align}$$

Substituting the second equation to the first yields

$$\begin{aligned}\boldsymbol{\delta_x} &= q_x \cdot \boldsymbol{\gamma_u} + \alpha_{y\to x}(q_y \cdot \boldsymbol{\gamma_u} + \alpha_{x\to y}\boldsymbol{\delta_x} + \boldsymbol{\gamma_y}) + \boldsymbol{\gamma_x}\\ &= (q_x + \alpha_{y\to x}q_y) \cdot \boldsymbol{\gamma_u} + (\alpha_{y\to x}\alpha_{x\to y})\boldsymbol{\delta_x} + \alpha_{y\to x}\boldsymbol{\gamma_y} + \boldsymbol{\gamma_x}\\ &= \left((q_x + \alpha_{y\to x}q_y) \cdot \boldsymbol{\gamma_u} + \alpha_{y\to x}\boldsymbol{\gamma_y} + \boldsymbol{\gamma_x}\right)/(1-\alpha_{y\to x}\alpha_{x\to y})\end{aligned} \tag{27}$$

Similarly,

$$\boldsymbol{\delta_y} = \left((q_y + \alpha_{x\to y}q_x) \cdot \boldsymbol{\gamma_u} + \alpha_{x\to y}\boldsymbol{\gamma_x} + \boldsymbol{\gamma_y}\right)/(1-\alpha_{y\to x}\alpha_{x\to y}) \tag{28}$$

Thus the genetic covariance is

$$\begin{aligned}E[\boldsymbol{\delta_x}\cdot\boldsymbol{\delta_y}] &= \left((q_x + \alpha_{y\to x}q_y) \cdot \boldsymbol{\gamma_u} + \alpha_{y\to x}\boldsymbol{\gamma_y} + \boldsymbol{\gamma_x}\right)\left((q_y + \alpha_{x\to y}q_x) \cdot \boldsymbol{\gamma_u} + \alpha_{x\to y}\boldsymbol{\gamma_x} + \boldsymbol{\gamma_y}\right)/(1-\alpha_{y\to x}\alpha_{x\to y})^2\\ &= \left((q_x + \alpha_{y\to x}q_y)(q_y + \alpha_{x\to y}q_x)h_u^2 + \alpha_{y\to x}h_y^2 + \alpha_{x\to y}h_x^2\right)/(1-\alpha_{y\to x}\alpha_{x\to y})^2\\ &= \left((t_x + \alpha_{y\to x}t_y)(t_y + \alpha_{x\to y}t_x) + \alpha_{y\to x}h_y^2 + \alpha_{x\to y}h_x^2\right)/(1-\alpha_{y\to x}\alpha_{x\to y})^2\end{aligned} \tag{29}$$

and the heritabilities are

$$\begin{align}E[\boldsymbol{\delta_x^2}] &= \left((t_x + \alpha_{y\to x}t_y)^2 + \alpha_{y\to x}^2h_y^2 + h_x^2\right)/(1-\alpha_{y\to x}\alpha_{x\to y})^2 \tag{30}\\ E[\boldsymbol{\delta_y^2}] &= \left((t_y + \alpha_{x\to y}t_x)^2 + \alpha_{x\to y}^2h_x^2 + h_y^2\right)/(1-\alpha_{y\to x}\alpha_{x\to y})^2 \tag{31}\end{align}$$

Therefore the genetic correlation takes the form

$$corr(\boldsymbol{\delta_x},\boldsymbol{\delta_y}) = \frac{(t_x + \alpha_{y\to x}t_y)(t_y + \alpha_{x\to y}t_x) + \alpha_{y\to x}h_y^2 + \alpha_{x\to y}h_x^2}{\sqrt{\left((t_x + \alpha_{y\to x}t_y)^2 + \alpha_{y\to x}^2h_y^2 + h_x^2\right)\left((t_y + \alpha_{x\to y}t_x)^2 + \alpha_{x\to y}^2h_x^2 + h_y^2\right)}} \tag{32}$$

These values can be compared to those obtained by LD score regression.

References

[1] Nadarajah, S. and Pogány, T. K. (2016). On the distribution of the product of correlated normal random variables. Comptes Rendus Mathematique *354*, 201–204.

[2] McNolty, F. (1973). Some probability density functions and their characteristic functions. Mathematics of computation 27, 495-504.

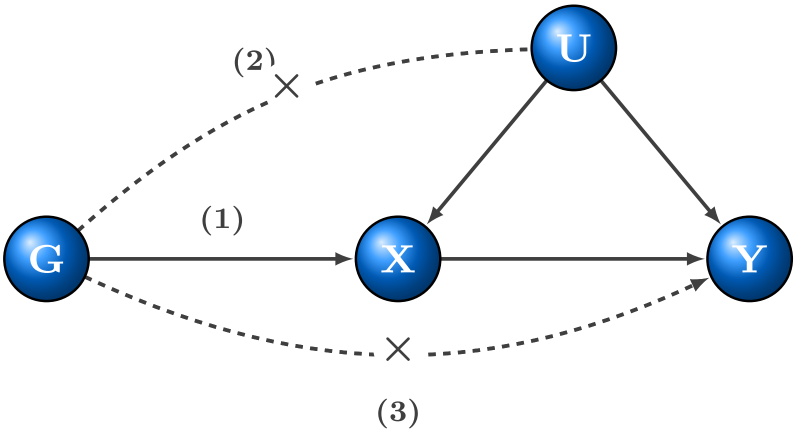
[3] Bateman, H. (1953). Higher Transcendental Functions, Volume I. `https://authors.library.caltech.edu/43491/`.

[4] Heideman, M. T., Johnson, D. H., and Burrus, C. S. (1985). Gauss and the history of the fast fourier transform. Archive for History of Exact Sciences *34*, 265-277.

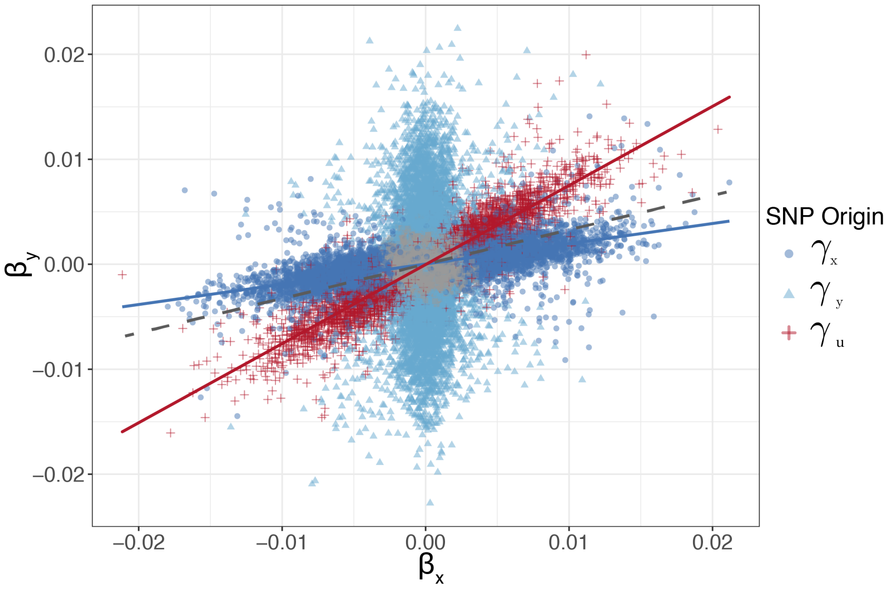
[5] Neale Lab (2018). UK BioBank. `http://www.nealelab.is/uk-biobank/`.

[6] Walter, K., Min, J. L., Huang, J., Crooks, L., Memari, Y., McCarthy, S., Perry, J. R. B., Xu, C., Futema, M., Lawson, D., et al. (2015). The uk10k project identifies rare variants in health and disease. Nature *526*, 82–90.

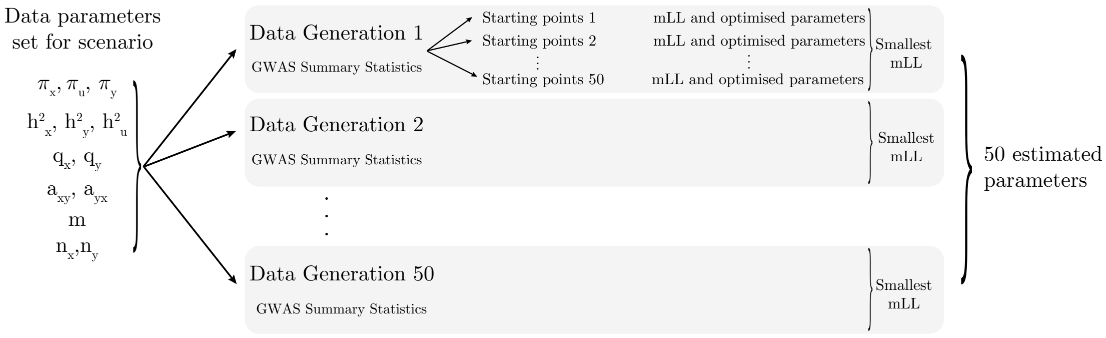
Supplementary Figures



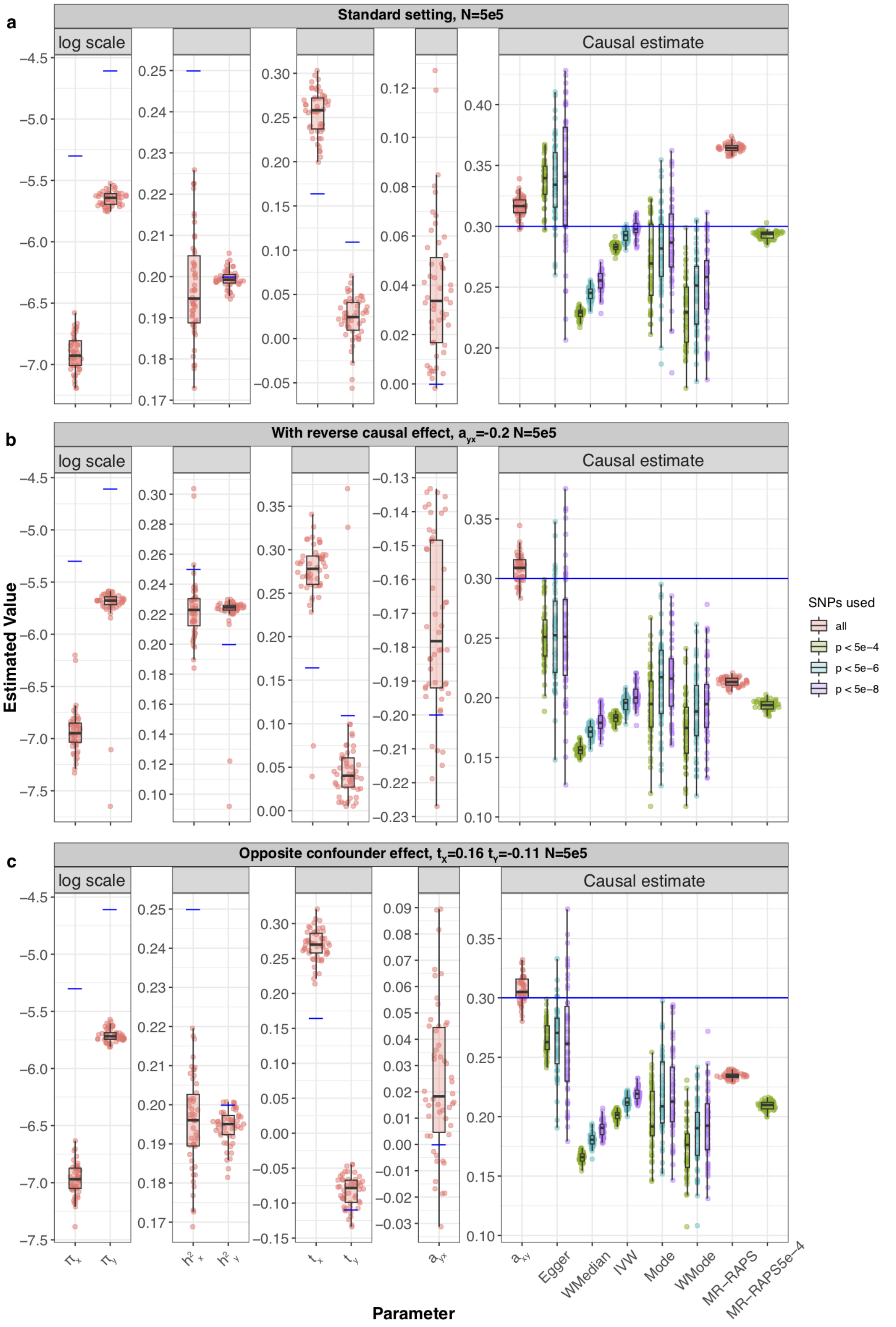
Supplementary Figure 1: **Basic assumptions of Mendelian randomisation**. (1) Relevance – genetic data, denoted by G, is robustly associated with the exposure. (2) Exchangeability – G is not associated with any confounder of the exposure-outcome relationship. (3) Exclusion restriction – G is independent of the outcome conditional on the exposure and all confounders of the exposure-outcome relationship (i.e. the only path between the instrument and the outcome is via the exposure).



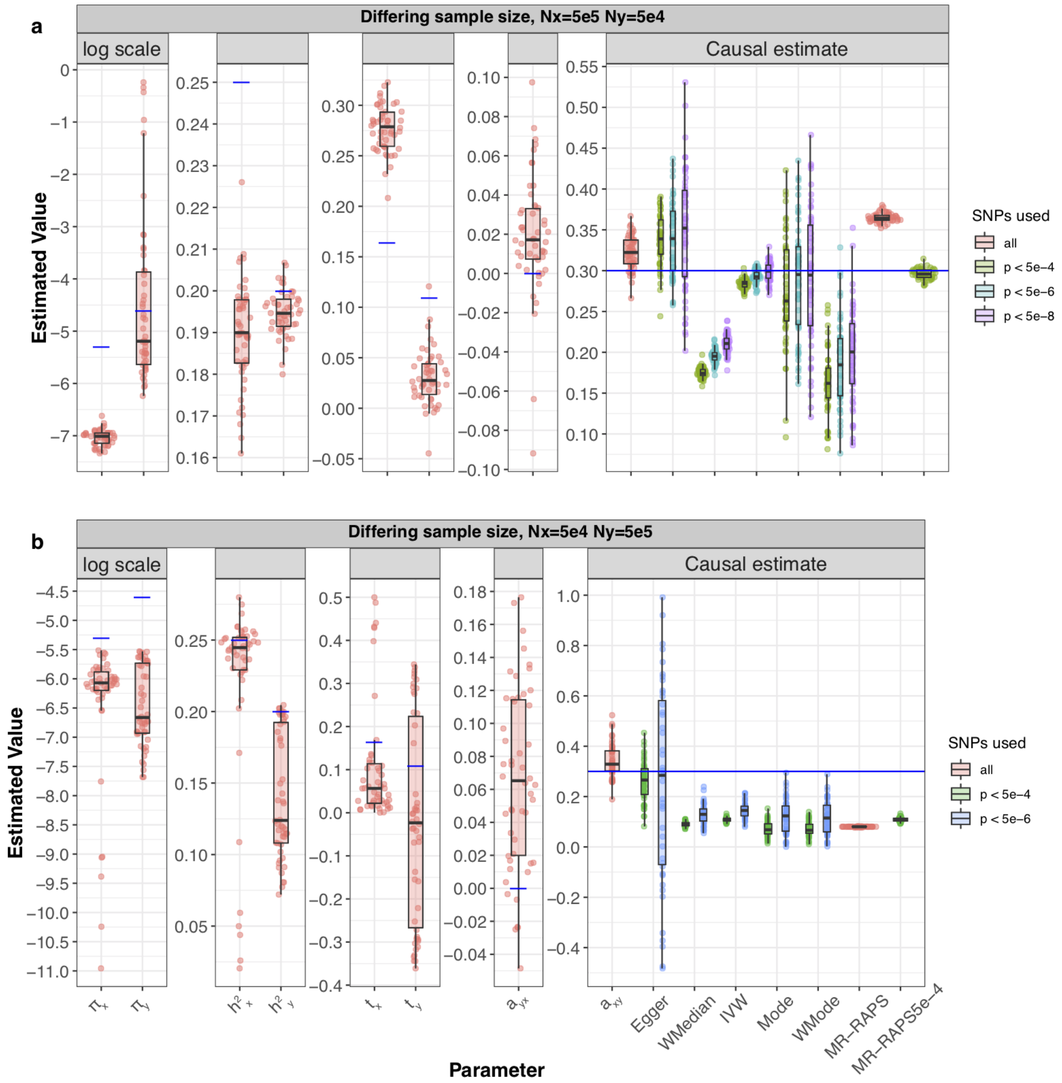
Supplementary Figure 2: An illustration of a scatter plot showing simulated observed SNP effects on traits X and Y, coloured by the strongest effect between the three vectors $\gamma_x, \gamma_y, \gamma_u$. SNPs in grey are those with no effect on any of the traits. This illustration shows the distinct clusters that could arise in the presence of a confounder. The dark blue cluster of SNPs represents those that are not in violation of any of the MR assumption, and hence its slope reflects the true causal effect of X on Y, while the red cluster of SNPs are those associated with the confounder. The steeper slope of the red cluster of SNPs causes a typical regression line - shown in grey - that represents the causal effect (estimated using conventional MR methods) to be overestimated.



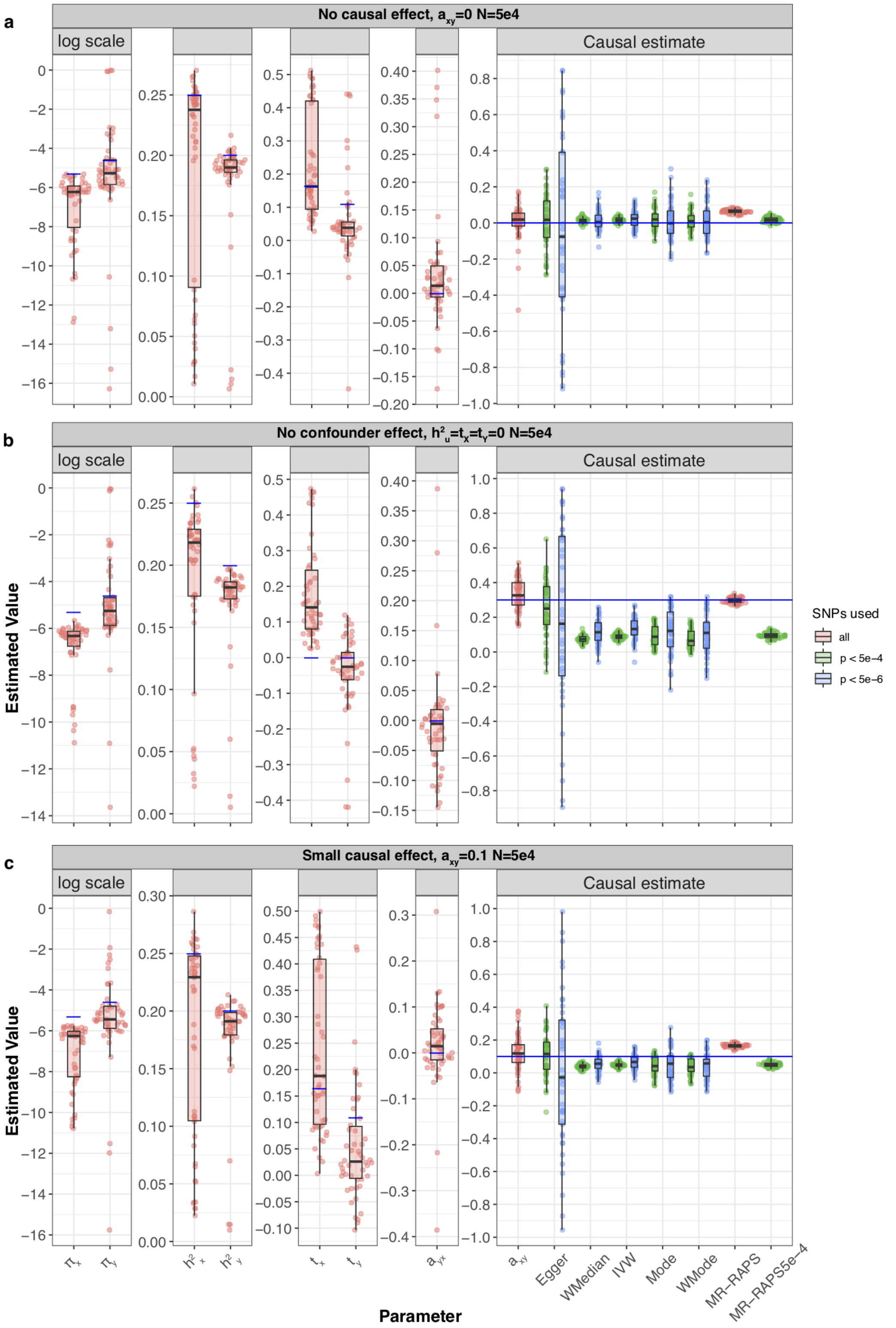
Supplementary Figure 3: **A schema showing the workflow of the simulation results**. For a single set of parameter settings, 50 different data generations of GWAS summary statistics are created for trait X and Y. The summary statistics of a single data generation, as well as the sample size, SNP number and SNP-based LD structure are used in the likelihood optimisation function that is run with 100 different random starting points in order to explore the likelihood surface. A single maximum likelihood and its corresponding estimated parameters are selected to represent the estimates of that data generation. And this is repeated for the other generations. The results for several data generation are often represented in boxplots throughout the paper.



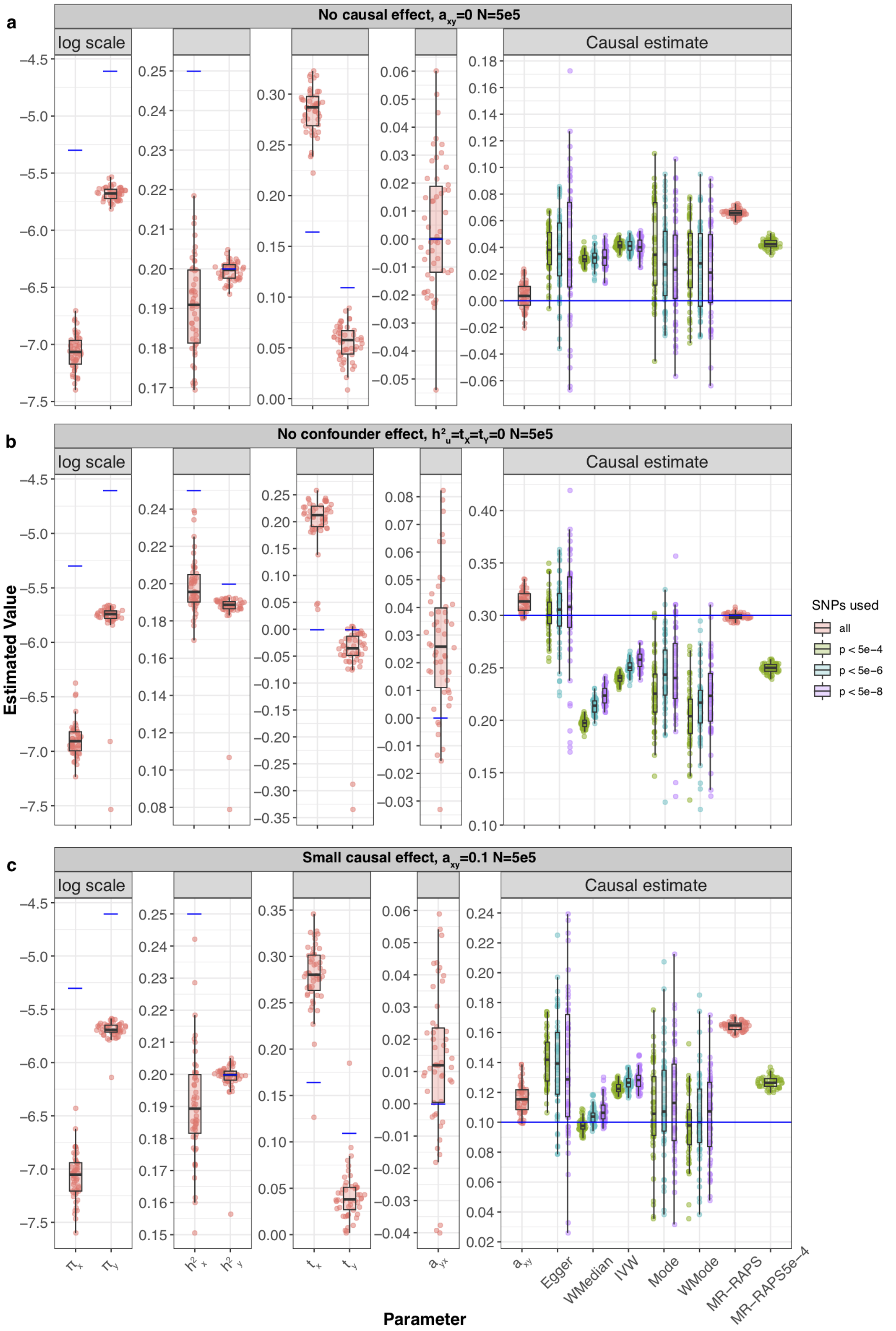
Supplementary Figure 4: Simulation results under various scenarios. These modified Sina-boxplot represent the distribution of parameter estimates from 50 different data generations under various conditions. For each generation, standard MR methods as well as our LHC-MR were used to estimate a causal effect. In the boxplots, the lower and upper hinges correspond to the first and third quartiles, the middle bar corresponds to the median, whereas the upper whisker is the largest dataset estimate smaller than 1.5∗inter-quartile range above the third quartile. The lower whisker is defined analogously. The true values of the parameters used in the data generations are represented by the blue dots/lines. **a** Estimation under standard settings ($\pi_x = 5 \times 10^{-3}, \pi_y = 1 \times 10^{-2}, \pi_u = 5 \times 10^{-2}, h^2_x = 0.25, h^2_y = 0.2, h^2_u = 0.3, t_x = 0.16, t_y = 0.11$). **b** Addition of a reverse causal effect $\alpha_{y \to x} = -0.2$. **c** Confounder with opposite causal effects on X and Y ($t_x = 0.16, t_y = -0.11$).



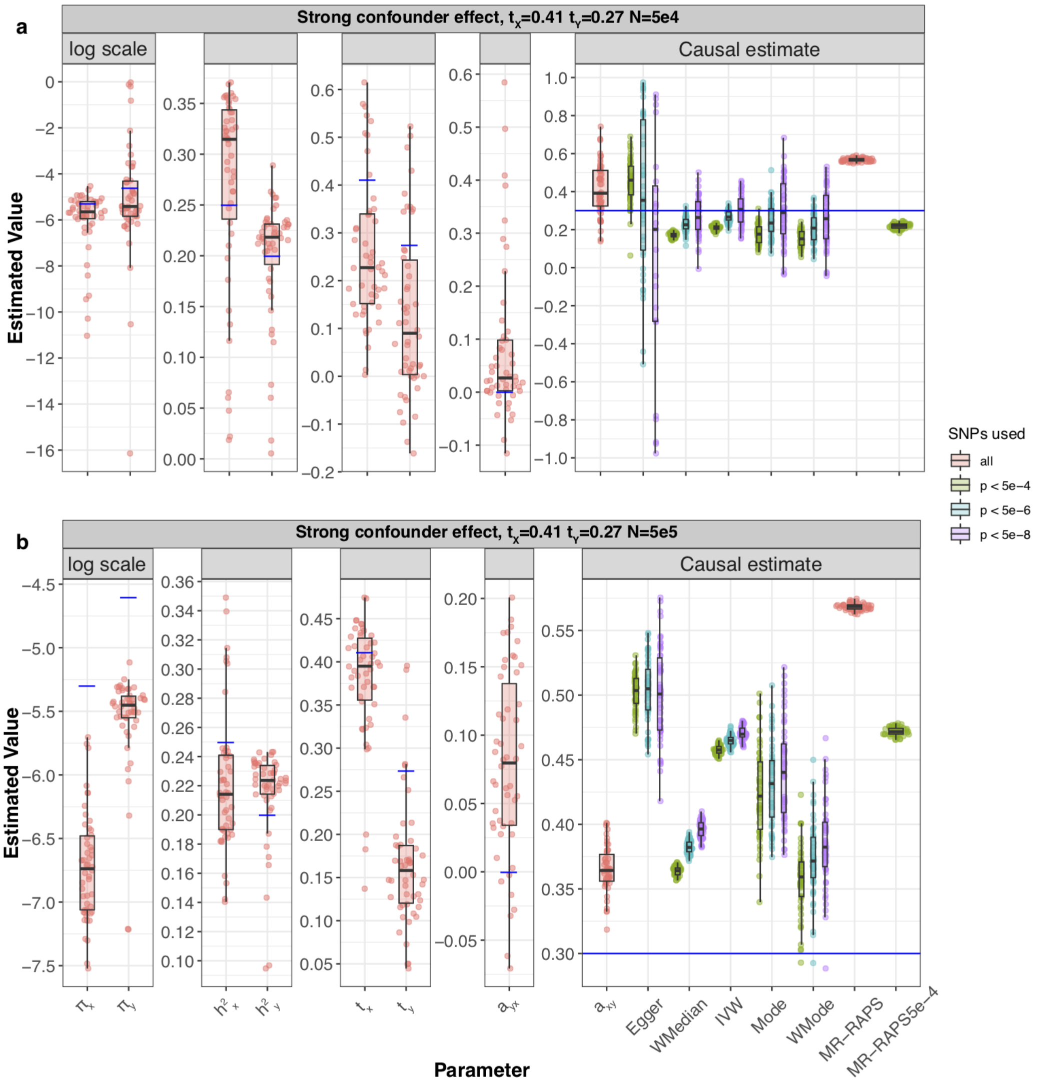
Supplementary Figure 5: Simulation results showing varying sample sizes for the two exposure and outcome samples. Modified Sina-boxplots representing the distribution of parameter estimates from 50 different data generations. For each generation, standard MR methods as well as our LHC-MR were used to estimate a causal effect. In the boxplots, the lower and upper hinges correspond to the first and third quartiles, the middle bar corresponds to the median, whereas the upper whisker is the largest dataset estimate smaller than 1.5∗inter-quartile range above the third quartile. The lower whisker is defined analogously. The true values of the parameters used in the data generations are represented by the blue dots/lines. In this figure, samples sizes for the two traits differ as such n_x = 500,000 and n_y = 50,000 for **a**, and n_x = 50,000 and n_y = 500,000 for **b**.



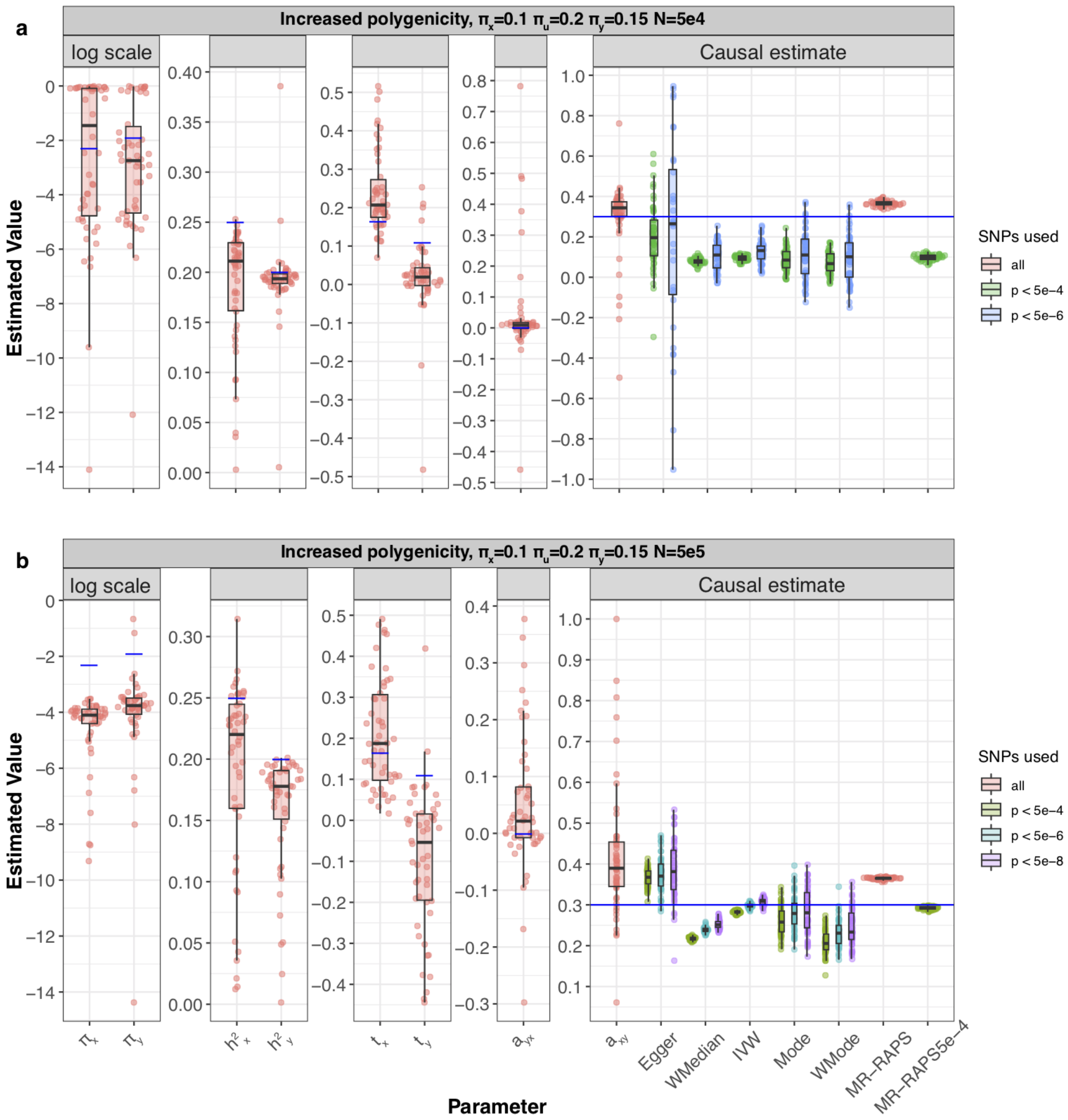
Supplementary Figure 6: Simulation results under various scenarios. These modified Sina-boxplots represent the distribution of parameter estimates from 50 different data generations under various conditions. For each generation, standard MR methods as well as our LHC-MR were used to estimate a causal effect. In the boxplots, the lower and upper hinges correspond to the first and third quartiles, the middle bar corresponds to the median, whereas the upper whisker is the largest dataset estimate smaller than 1.5∗inter-quartile range above the third quartile. The lower whisker is defined analogously. The true values of the parameters used in the data generations are represented by the blue dots/lines. **a** The data simulated had no causal effect in either direction. **b** The data simulated had no confounder effect with π_u, t_x, and $t_y = 0$. **c** This model had a small causal effect of $\alpha_{x\rightarrow y} = 0.1$.



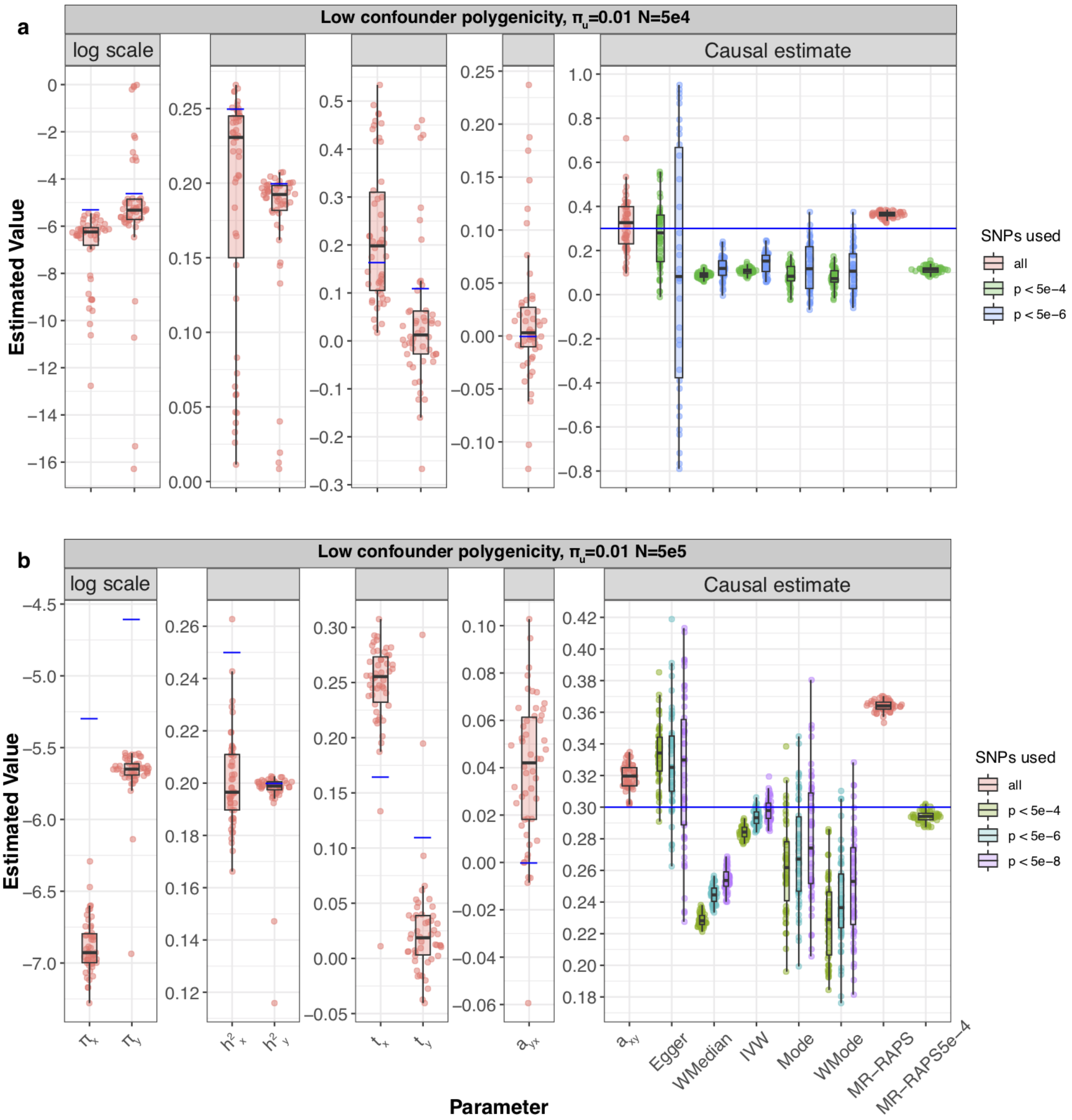
Supplementary Figure 7: Simulation results under various scenarios. These modified Sina-boxplots represent the distribution of parameter estimates from 50 different data generations under various conditions. For each generation, standard MR methods as well as our LHC-MR were used to estimate a causal effect. In the boxplots, the lower and upper hinges correspond to the first and third quartiles, the middle bar corresponds to the median, whereas the upper whisker is the largest dataset estimate smaller than 1.5∗inter-quartile range above the third quartile. The lower whisker is defined analogously. The true values of the parameters used in the data generations are represented by the blue dots/lines. **a** The data simulated had no causal effect in either direction. **b** The data simulated had no confounder effect with π_u, t_x, and $t_y = 0$. **c** This model had a small causal effect of $\alpha_{x\rightarrow y} = 0.1$.



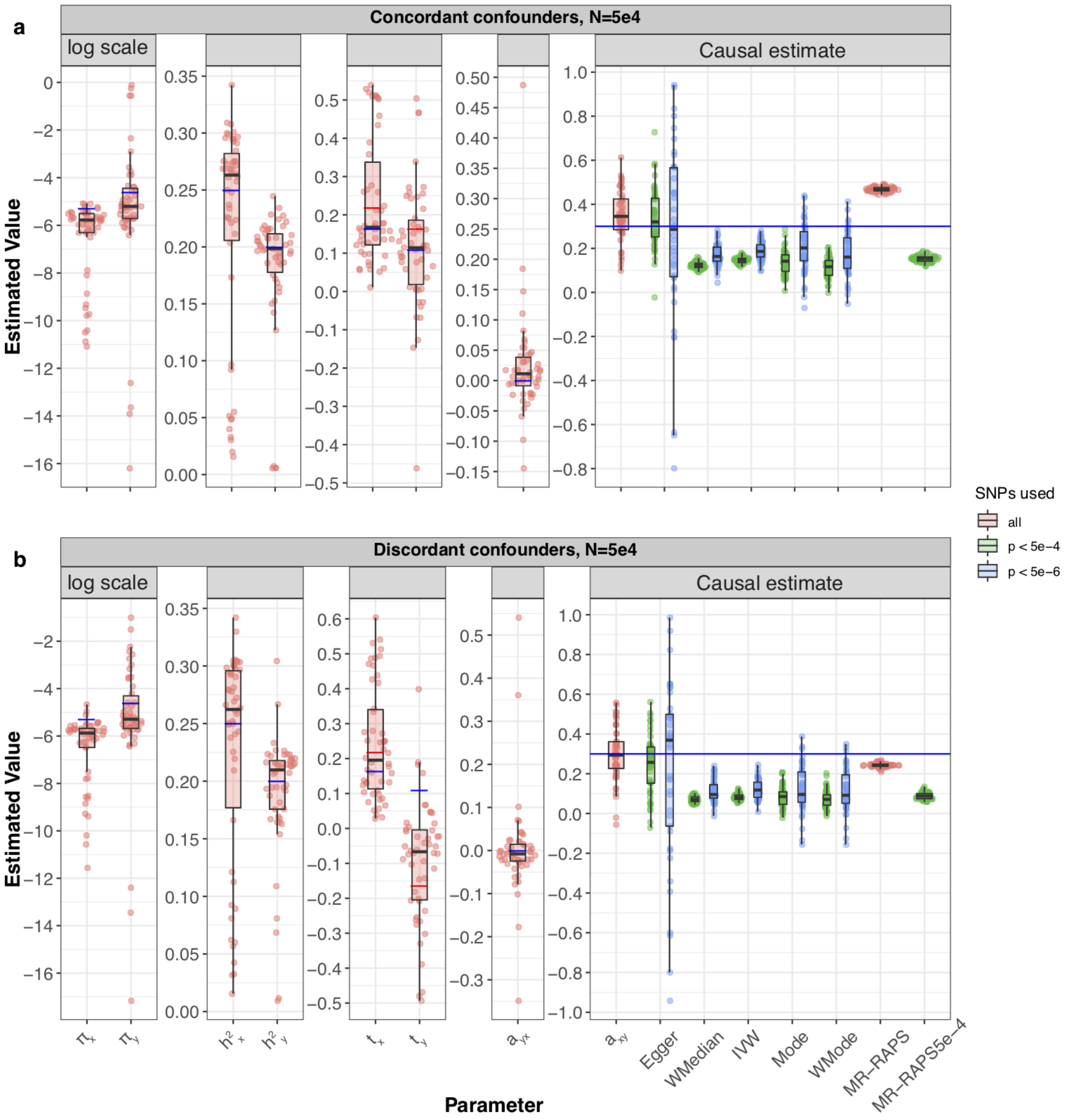
Supplementary Figure 8: **Simulation results under various scenarios.** These modified Sina-boxplots represent the distribution of parameter estimates from 50 different data generations under various conditions. For each generation, standard MR methods as well as our LHC-MR were used to estimate a causal effect. In the boxplots, the lower and upper hinges correspond to the first and third quartiles, the middle bar corresponds to the median, whereas the upper whisker is the largest dataset estimate smaller than 1.5∗inter-quartile range above the third quartile. The lower whisker is defined analogously. The true values of the parameters used in the data generations are represented by the blue dots/lines. **a** The data simulated shows the increased effect of U on X and Y through $t_x = 0.41, t_y = 0.27$ instead of the standard setting $t_x = 0.16, t_y = 0.11$. **b** This panel show the same thing but with a larger sample size of $n_x = n_y = 500,000$



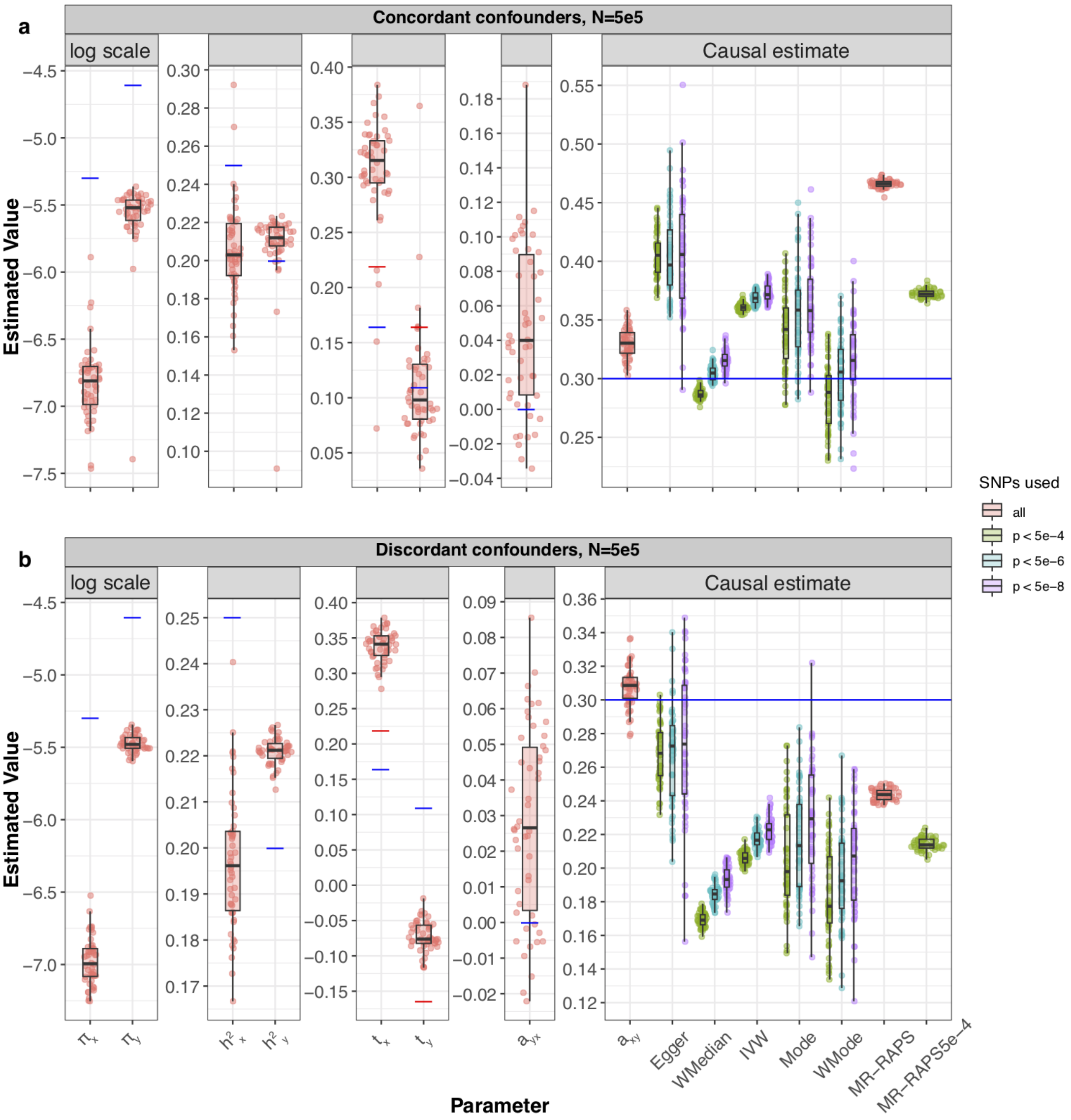
Supplementary Figure 9: Simulation results where there is an increased polygenicity for all traits. Modified Sina-boxplots representing the distribution of parameter estimates from 50 different data generations. For each generation, standard MR methods as well as our LHC-MR were used to estimate a causal effect. In the boxplots, the lower and upper hinges correspond to the first and third quartiles, the middle bar corresponds to the median, whereas the upper whisker is the largest dataset estimate smaller than 1.5∗inter-quartile range above the third quartile. The lower whisker is defined analogously. The true values of the parameters used in the data generations are represented by the blue dots/lines. The proportion of effective SNPs that make up the spike-and-slab distributions of the γ vectors in this setting is $10\%, 15\%, and 20\%$ for traits X, Y and U respectively. **a** Results for smaller sample size of $n_x = n_y = 50,000$. **b** Results for larger sample size of $n_x = n_y = 500,000$.



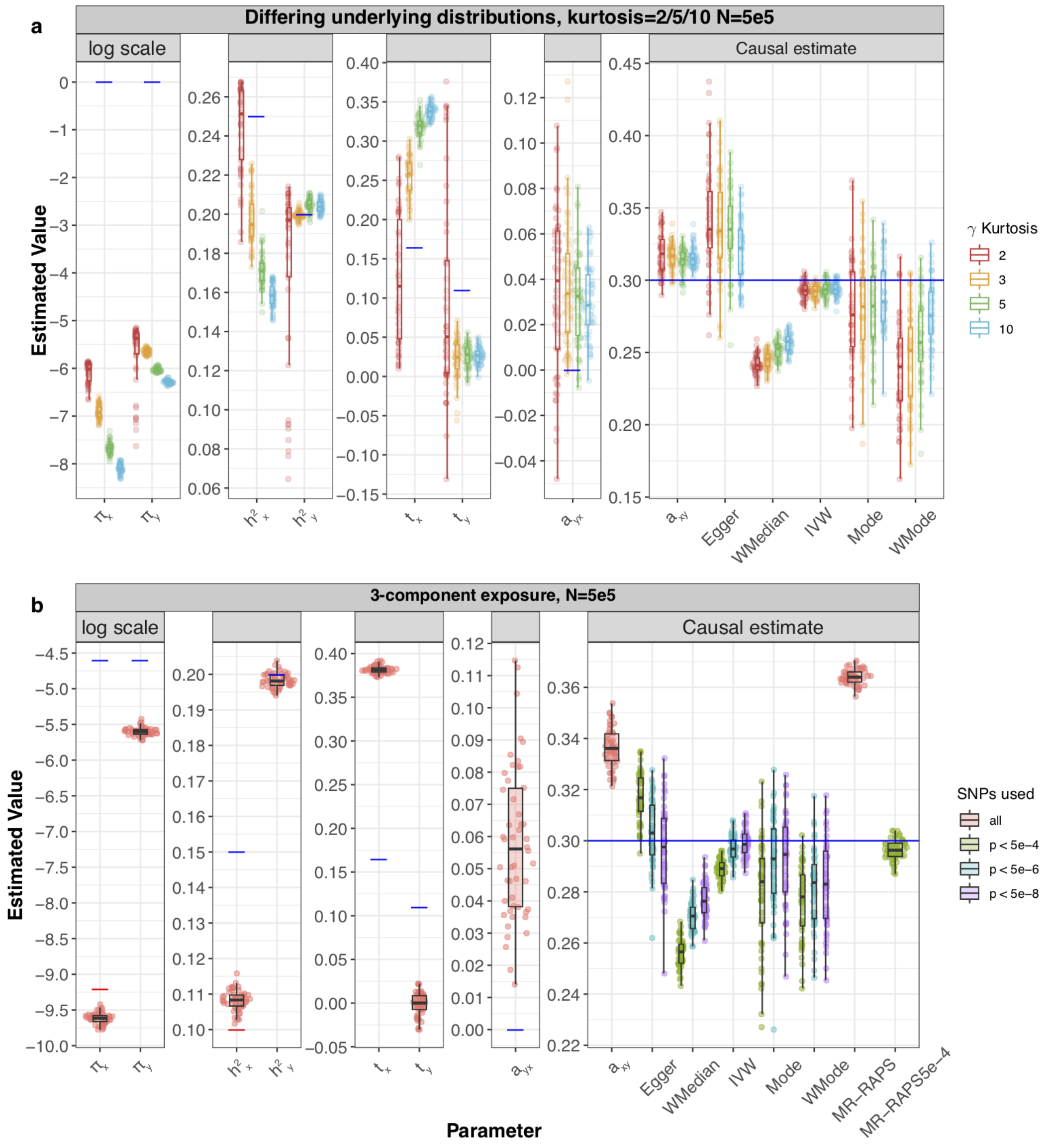
Supplementary Figure 10: Simulation results where the polygenicity of the confounder is reduced. Modified Sina-boxplots representing the distribution of parameter estimates from 50 different data generations. For each generation, standard MR methods as well as our LHC-MR were used to estimate a causal effect. In the boxplots, the lower and upper hinges correspond to the first and third quartiles, the middle bar corresponds to the median, whereas the upper whisker is the largest dataset estimate smaller than 1.5∗inter-quartile range above the third quartile. The lower whisker is defined analogously. The true values of the parameters used in the data generations are represented by the blue dots/lines. In this figure, the polygenicity for U is decreased in the form of lower $\pi_u = 0.01$. **a** Results for smaller sample size of $n_x = n_y = 50,000$. **b** Results for larger sample size of $n_x = n_y = 500,000$.



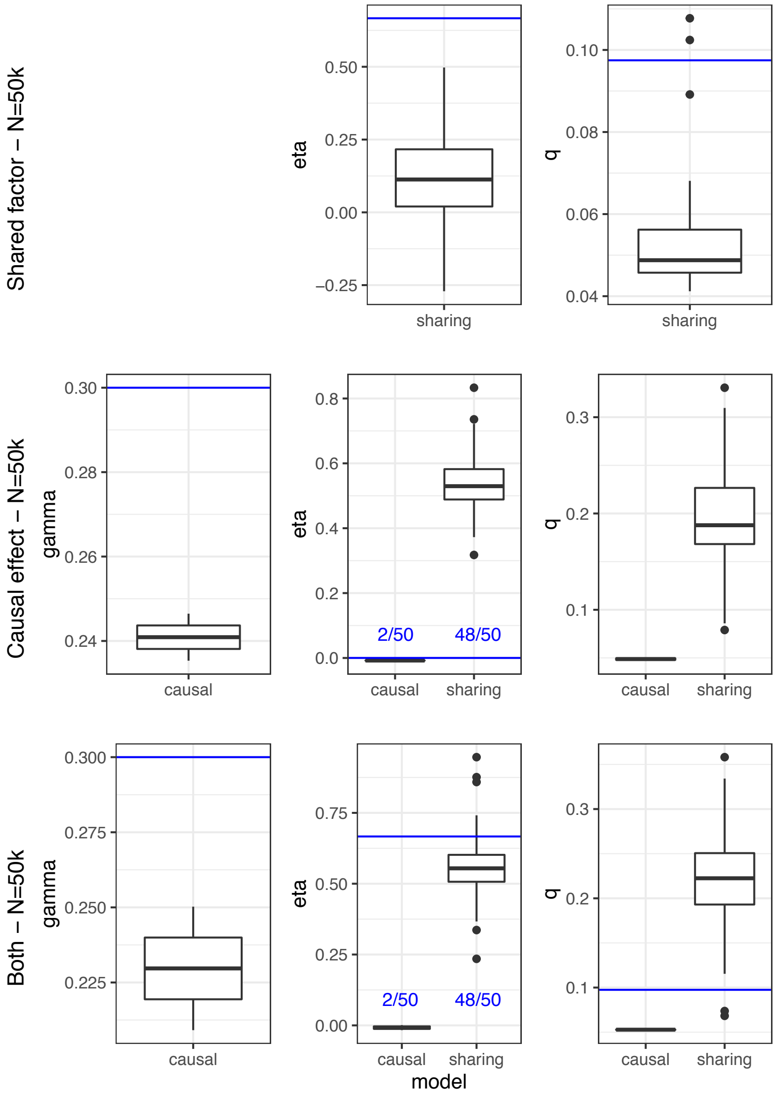
Supplementary Figure 11: Simulation results where there are two underlying confounders, once with concordant and another with discordant effects on the exposure-outcome pair. Modified Sina-boxplots representing the distribution of parameter estimates from 50 different data generations. For each generation, standard MR methods as well as our LHC-MR were used to estimate a causal effect. In the boxplots, the lower and upper hinges correspond to the first and third quartiles, the middle bar corresponds to the median, whereas the upper whisker is the largest dataset estimate smaller than 1.5∗inter-quartile range above the third quartile. The lower whisker is defined analogously. The true values of the parameters used in the data generations are represented by the blue dots/lines. **a** The underlying data generations have two concordant heritable confounders U_1 and U_2 with positive effects on traits X and Y. **b** The data generations have two discordant heritable confounders with $t_x^{(1)} = 0.16, t_y^{(1)} = 0.11$ shown as blue dots and $t_x^{(2)} = 0.22, t_y^{(2)} = -0.16$ shown as red dots.



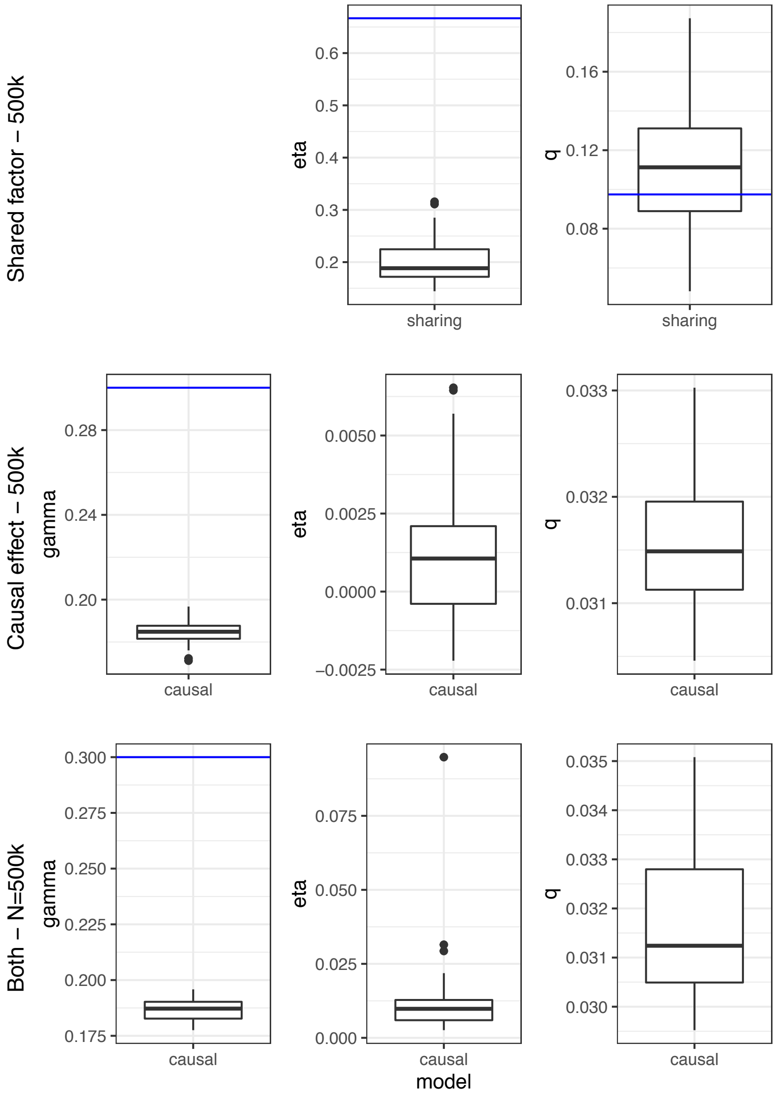
Supplementary Figure 12: Simulation results where there are two underlying confounders, once with concordant and another with discordant effects on the exposure-outcome pair. Modified Sina-boxplots representing the distribution of parameter estimates from 50 different data generations. For each generation, standard MR methods as well as our LHC-MR were used to estimate a causal effect. In the boxplots, the lower and upper hinges correspond to the first and third quartiles, the middle bar corresponds to the median, whereas the upper whisker is the largest dataset estimate smaller than 1.5∗inter-quartile range above the third quartile. The lower whisker is defined analogously. The true values of the parameters used in the data generations are represented by the blue dots/lines. **a** The underlying data generations have two concordant heritable confounders U_1 and U_2 with positive effects on traits X and Y. **b** The data generations have two discordant heritable confounders with $t_x^{(1)} = 0.16, t_y^{(1)} = 0.11$ shown as blue dots and $t_x^{(2)} = 0.22, t_y^{(2)} = -0.16$ shown as red dots.



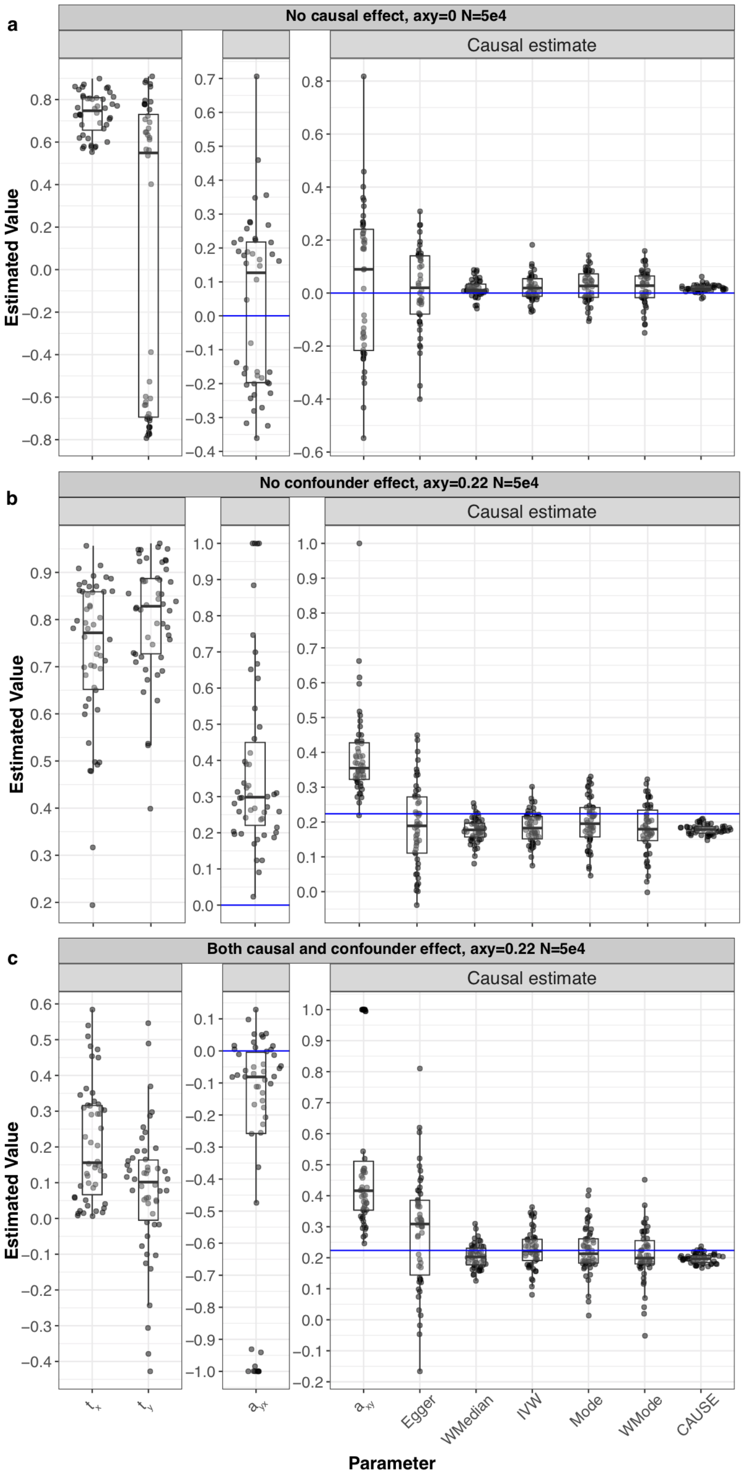
Supplementary Figure 13: Simulation results under various scenarios. These modified Sina-boxplots represent the distribution of parameter estimates from 50 different data generations under various conditions. For each generation, standard MR methods as well as our LHC-MR were used to estimate a causal effect. In the boxplots, the lower and upper hinges correspond to the first and third quartiles, the middle bar corresponds to the median, whereas the upper whisker is the largest dataset estimate smaller than 1.5∗inter-quartile range above the third quartile. The lower whisker is defined analogously. The true values of the parameters used in the data generations are represented by the blue dots/lines. **a** The different coloured boxplots represent the underlying non-normal distribution used in the simulation of the three $\gamma_x, \gamma_x, \gamma_u$ vectors associated to their respective traits. The Pearson distributions had the same 0 mean and skewness, however their kurtosis ranged between 2 and 10, including the kurtosis of 3, which corresponds to a normal distribution assumed by our model. The standard MR results reported had IVs selected with a p-value threshold of 5×10^{-6}. **b** Addition of a third component for exposure X, while decreasing the strength of U. True parameter values are in colour, blue and red for each component ($\pi_{x1} = 1 \times 10^{-4}, \pi_{x2} = 1 \times 10^{-2}, h^2_{x1} = 0.15, h^2_{x2} = 0.1$).



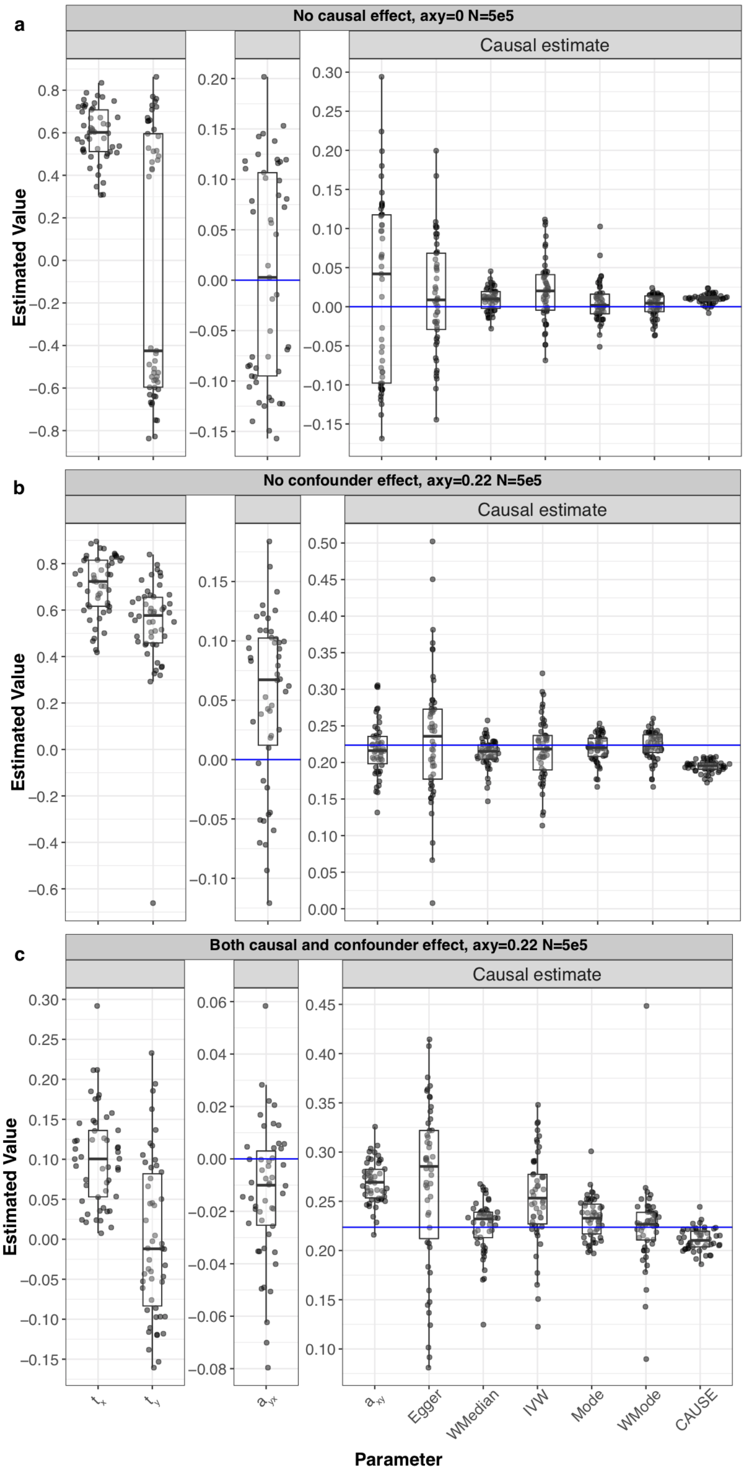
Supplementary Figure 14: Running CAUSE on LHC-MR simulated data under the standard settings. Boxplots of the parameter estimation of CAUSE on LHC-simulated data ($n_x = n_y = 50,000$), with 50 different data generations under three different scenarios: presence of a shared factor only, presence of a causal effect only, presence of both. In the boxplots, the lower and upper hinges correspond to the first and third quartiles, the middle bar corresponds to the median, whereas the upper whisker is the largest dataset estimate smaller than 1.5∗inter-quartile range above the third quartile. The lower whisker is defined analogously. CAUSE returns two possible models with a respective p-value, the sharing and the causal model, where the causal mode is the significant of the two. When only an underlying shared factor was present in the simulated data, CAUSE had no significant causal estimates. With a true underlying causal effect, or when both an underlying causal effect and a shared factor was present, the causal model was significant only 4% of the simulations.



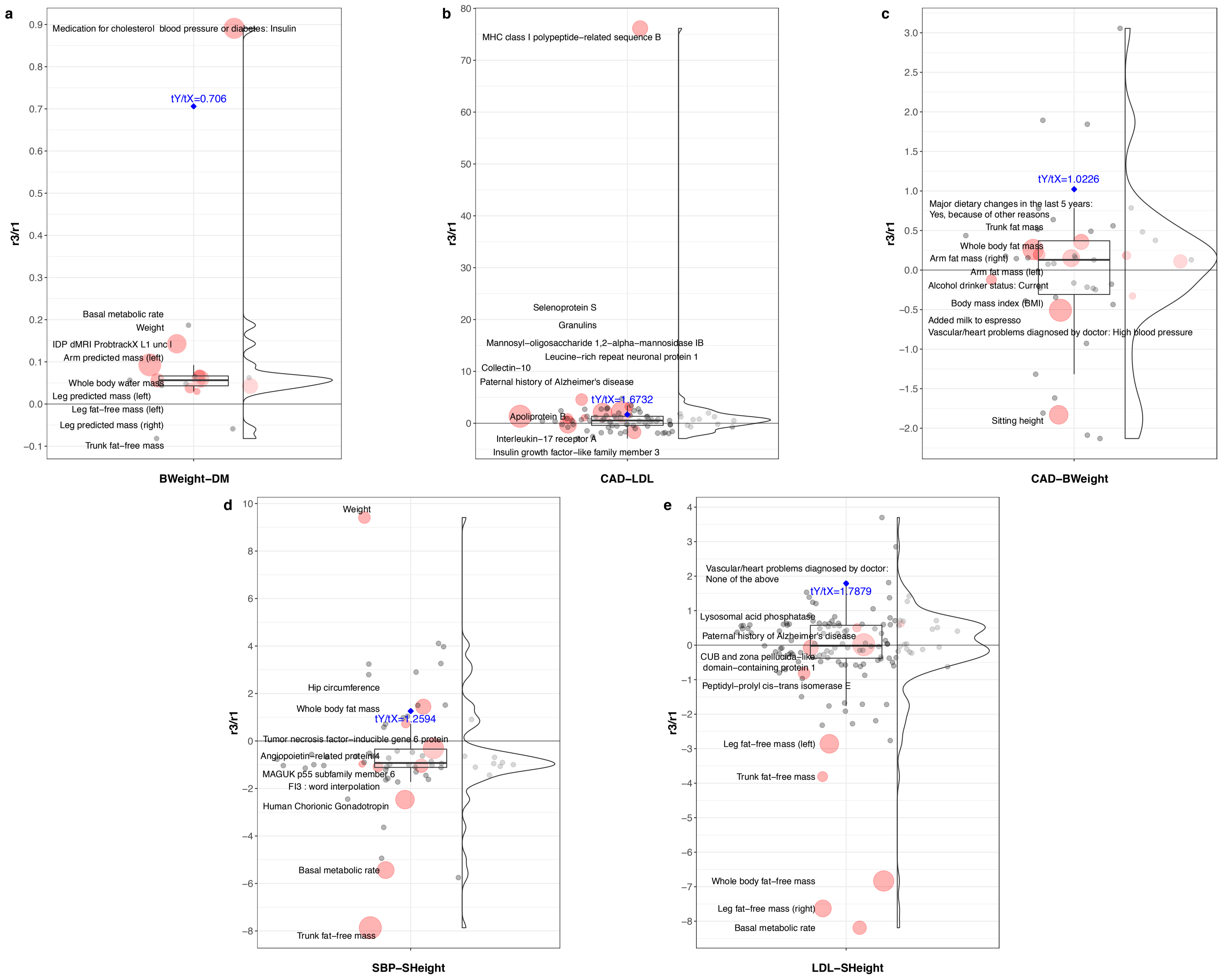
Supplementary Figure 15: Running CAUSE on LHC-MR simulated data under the standard settings. Boxplots of the parameter estimation of CAUSE on LHC-simulated data ($n_x = n_y = 500,000$), with 50 different data generations under three different scenarios: presence of a shared factor only, presence of a causal effect only, presence of both. In the boxplots, the lower and upper hinges correspond to the first and third quartiles, the middle bar corresponds to the median, whereas the upper whisker is the largest dataset estimate smaller than 1.5∗inter-quartile range above the third quartile. The lower whisker is defined analogously. CAUSE returns two possible models with a respective p-value, the sharing and the causal model, where the causal mode is the significant of the two. When only an underlying shared factor was present in the simulated data, CAUSE had no significant causal estimates. With a true underlying causal effect, or when both an underlying causal effect and a shared factor was present, the causal model was significant 100% of the simulations.

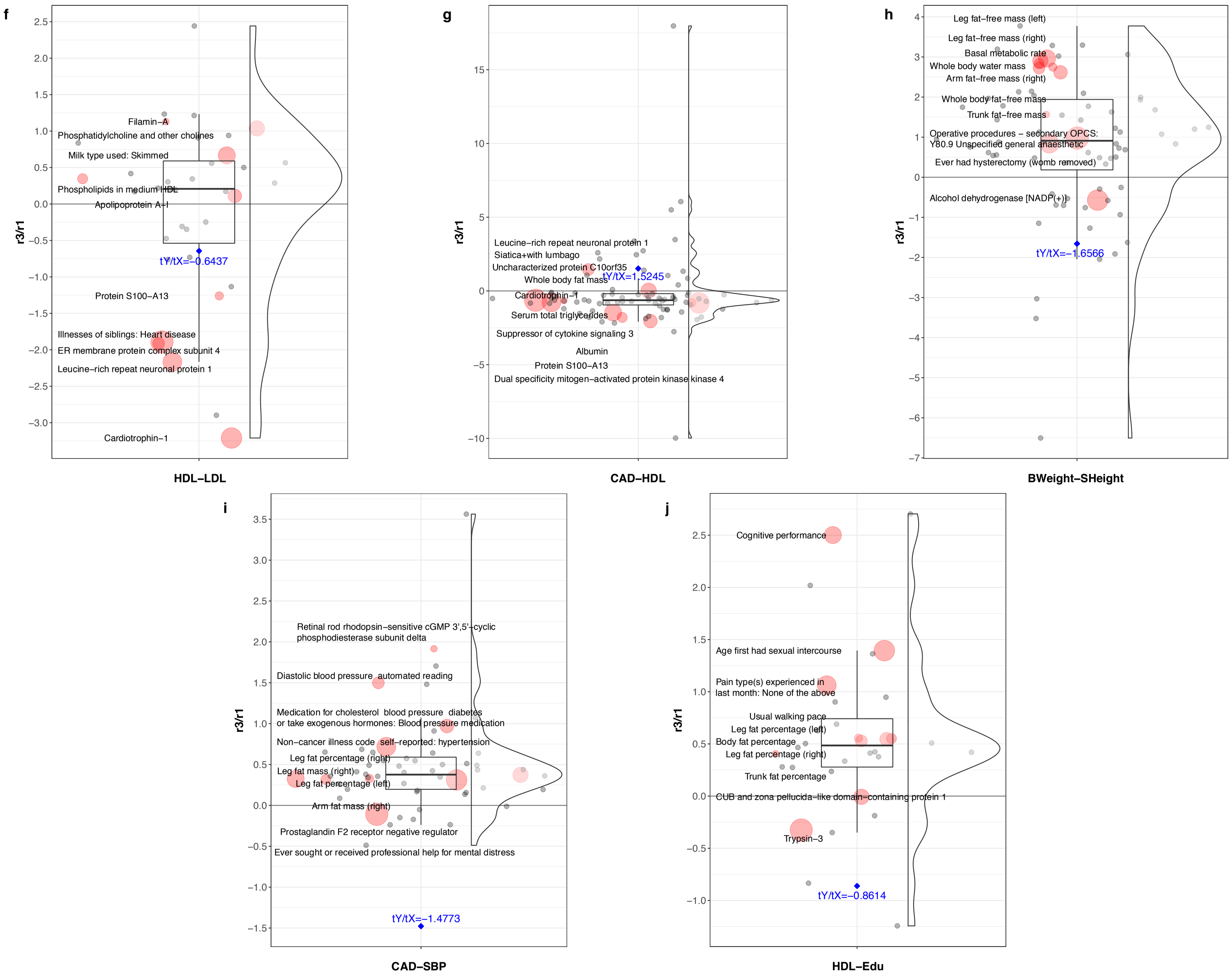


Supplementary Figure 16: Running LHC-MR on CAUSE simulated data under various scenarios. Modified Sina-boxplots representing the distribution of parameter estimates from LHC-MR of 50 different data generations using the CAUSE framework. For each generation, standard MR methods, CAUSE as well as our LHC-MR were used to estimate a causal effect. In the boxplots, the lower and upper hinges correspond to the first and third quartiles, the middle bar corresponds to the median, whereas the upper whisker is the largest dataset estimate smaller than 1.5*inter-quartile range above the third quartile. The lower whisker is defined analogously. The true values of the parameters used in the data generations are represented by the blue dots/lines. **a** CAUSE data was generated with no causal effect but with a shared factor with an η value of ~ 0.22. CAUSE chooses a sharing model 100% of the time with no estimate for a causal effect. **b** CAUSE is simulated with causal effect but with no shared factor. **c** CAUSE is simulated with both a causal effect and a shared factor.

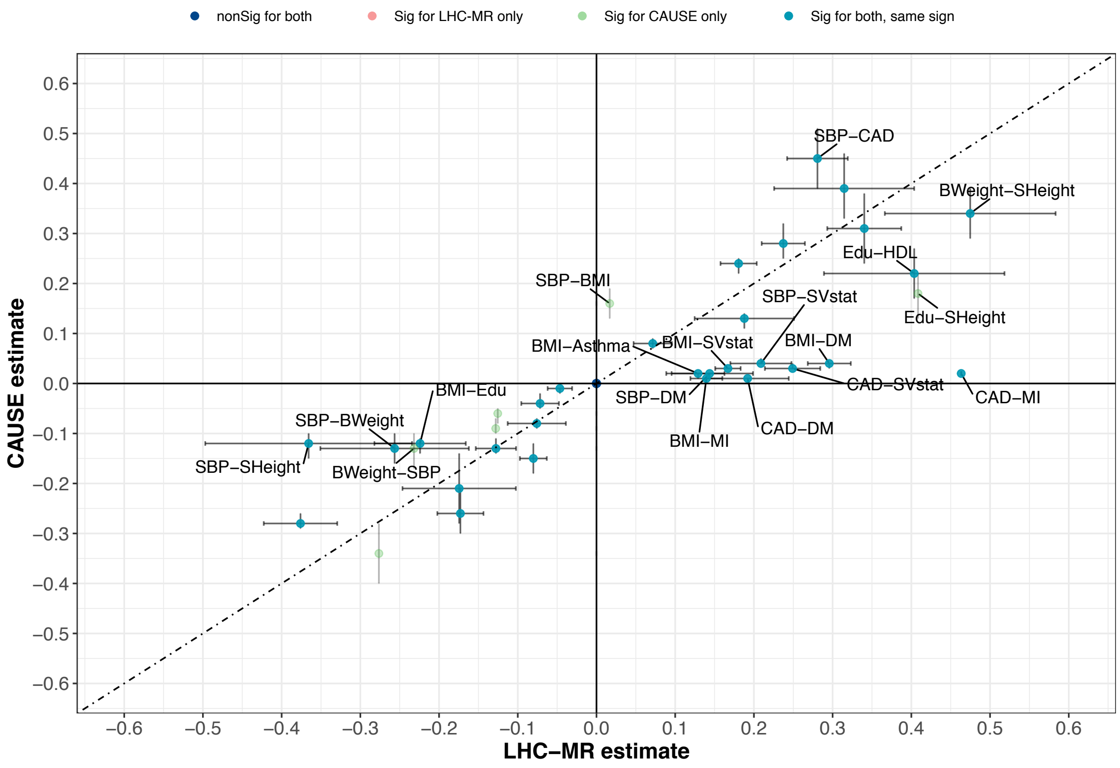


Supplementary Figure 17: Running LHC-MR on CAUSE simulated data under various scenarios. Modified Sina-boxplots representing the distribution of parameter estimates from LHC-MR of 50 different data generations using the CAUSE framework. For each generation, standard MR methods, CAUSE as well as our LHC-MR were used to estimate a causal effect. In the boxplots, the lower and upper hinges correspond to the first and third quartiles, the middle bar corresponds to the median, whereas the upper whisker is the largest dataset estimate smaller than 1.5∗inter-quartile range above the third quartile. The lower whisker is defined analogously. The true values of the parameters used in the data generations are represented by the blue dots/lines. **a** CAUSE data was generated with no causal effect but with a shared factor with an η value of ~ 0.22. **b** CAUSE is simulated with causal effect but with no shared factor. **c** CAUSE is simulated with both a causal effect and a shared factor. LHC-MR seems to exhibit a bimodal effect at first glance, but the two peaks are not connected.


a
r3/r1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
Medication for cholesterol blood pressure or diabetes: Insulin
tY/tX=0.706
Basal metabolic rate
Weight
IDP dMRI ProbtrackX L1 unc l
Arm predicted mass (left)
Whole body water mass
Leg predicted mass (left)
Leg fat−free mass (left)
Leg predicted mass (right)
Trunk fat−free mass
BWeight−DM
b
r3/r1
80
70
60
50
40
30
20
10
0
MHC class I polypeptide−related sequence B
Selenoprotein S
Granulins
Mannosyl−oligosaccharide 1,2−alpha−mannosidase IB
Leucine−rich repeat neuronal protein 1
Collectin−10
Paternal history of Alzheimer's disease
tY/tX=1.6732
Apolipoprotein B
Interleukin−17 receptor A
Insulin growth factor−like family member 3
CAD−LDL
c
r3/r1
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
tY/tX=1.0226
Major dietary changes in the last 5 years:
Yes, because of other reasons
Trunk fat mass
Whole body fat mass
Arm fat mass (right)
Arm fat mass (left)
Alcohol drinker status: Current
Body mass index (BMI)
Added milk to espresso
Vascular/heart problems diagnosed by doctor: High blood pressure
Sitting height
CAD−BWeight
d
r3/r1
10
8
6
4
2
0
−2
−4
−6
−8
Weight
Hip circumference
Whole body fat mass
tY/tX=1.2594
Tumor necrosis factor−inducible gene 6 protein
Angiopoietin−related protein 4
MAGUK p55 subfamily member 6
FI3 : word interpolation
Human Chorionic Gonadotropin
Basal metabolic rate
Trunk fat−free mass
SBP−SHeight
e
r3/r1
4
3
2
1
0
−1
−2
−3
−4
−5
−6
−7
−8
Vascular/heart problems diagnosed by doctor:
None of the above
tY/tX=1.7879
Lysosomal acid phosphatase
Paternal history of Alzheimer's disease
CUB and zona pellucida−like
domain−containing protein 1
Peptidyl−prolyl cis−trans isomerase E
Leg fat−free mass (left)
Trunk fat−free mass
Whole body fat−free mass
Leg fat−free mass (right)
Basal metabolic rate
LDL−SHeight



Supplementary Figure 18: Confounder effects obtained from EpiGraphDB plotted as a modified Sina-boxplot with the r_3/r_1 ratio on the y-axis. The blue diamonds represent the t_y/t_x ratio derived from the LHC model for that trait pair, also reported in blue text. Labelled dots in red and their varying size show the ten largest confounder traits in terms of their absolute effect product on the two traits, whereas grey dots represent the rest of the confounder traits found by EpigraphDB. In the boxplots, the lower and upper hinges correspond to the first and third quartiles, the middle bar corresponds to the median, whereas the upper whisker is the largest dataset point smaller than 1.5∗inter-quartile range above the third quartile. The lower whisker is defined analogously. Confounder effects of trait pairs: **a** Birth Weight - Diabetes Mellitus, **b** Coronary Artery Disease - Low-density Lipoprotein, **c** Coronary Artery Disease - Birth Weight, **d** Systolic Blood Pressure - Standing Height, **e** Low-density Lipoprotein - Standing Height, **f** High-density Lipoprotein - Low-density Lipoprotein, **g** Coronary Artery Disease - High-density Lipoprotein, **h** Birth Weight - Standing Height, **i** Coronary Artery Disease - Systolic Blood Pressure, **j** High-density Lipoprotein - Years of Education



Supplementary Figure 19: A scatter plot of the causal effect estimates between LHC-MR and CAUSE. To improve visibility, non-significant estimates by both methods are placed at the origin, while significant estimates by both methods appear on the diagonal with 95% CI error bars for LHC-MR causal estimates, and 95% credible interval error bars for CAUSE estimates. Labelled pairs are those with an estimate difference greater than 0.1.

Supplementary Tables

UKBB ID / Data Origin	Trait Name	Abbreviation	Sample Size	PMID
845	Age completed full time education	Edu	240,547	25826379
21001_irnt	Body mass index (BMI)	BMI	359,983	25826379
2443	Diabetes diagnosed by doctor	DM	360,192	25826379
20002_1075	Non-cancer illness code, self-reported: heart attack/myocardial infarction	MI	361,141	25826379
20002_1111	Non-cancer illness code, self-reported: asthma	Asthma	361,141	25826379
2887	Number of cigarettes previously smoked daily	PSmoke	84,456	25826379
20022_irnt	Birth weight	BWeight	205,475	25826379
50_irnt	Standing height	SHeight	360,388	25826379
4080	Systolic blood pressure, automated reading	SBP	340,159	25826379
20003_1140861958	Treatment/medication code: simvastatin	SVstat	361,141	25826379
30780_irnt	LDL Cholesterol	LDL	343,621	25826379
30760_irnt	HDL Cholesterol	HDL	315,133	25826379
UKBB + CARDIoGRAMplusC4D	Coronary Artery Disease	CAD	380,831	29212778

Supplementary Table 1: Details of the origin study of each trait, its abbreviation used in this paper, the sample size of the study for that trait, as well as the PubMed article ID.

a

MR-Egger

LHC-MR	Sig+	Sig−	nonSig
Sig+	10	0	27
Sig−	0	2	35
nonSig	0	2	56

b

WMedian

LHC-MR	Sig+	Sig−	nonSig
Sig+	21	1	15
Sig−	0	12	25
nonSig	1	6	51

c

IVW

LHC-MR	Sig+	Sig−	nonSig
Sig+	23	0	14
Sig−	0	17	20
nonSig	2	5	51

d

Mode

LHC-MR	Sig+	Sig−	nonSig
Sig+	12	0	25
Sig−	0	2	35
nonSig	0	0	58

e

WMode

LHC-MR	Sig+	Sig−	nonSig
Sig+	13	3	21
Sig−	0	7	30
nonSig	0	2	56

f

WMode

MR-Egger	Sig+	Sig−	nonSig
Sig+	10	0	0
Sig−	0	1	3
nonSig	3	11	104

Supplementary Table 2: **Cross tables between LHC-MR and various standard MR methods comparing the significance and sign of each respective causal estimate**. **f** shows a cross table between the two-least correlated MR methods in terms of their estimates.

Pair	$\alpha_{x\rightarrow y}$	**p-value**	γ	**IVW** $\alpha_{x\rightarrow y}$	**p-value**
BMI-Asthma	0.1290	4.99E-14	0.02 (0.01, 0.02)	0.0593	1.00E-08
BMI-DM	0.2958	1.07E-99	0.04 (0.03, 0.04)	0.2447	2.25E-140
BMI-SBP	0.1878	5.55E-09	0.13 (0.11, 0.14)	0.1547	1.11E-24
BMI-SVstat	0.1670	2.08E-91	0.03 (0.03, 0.03)	0.1570	4.26E-63
BMI-MI	0.1396	1.67E-41	0.01 (0.01, 0.01)	0.1027	9.16E-32
BWeight-SHeight	0.4748	9.60E-18	0.34 (0.29, 0.39)	0.2959	8.01E-10
SHeight-BWeight	0.1806	1.93E-53	0.24 (0.22, 0.25)	0.1803	7.21E-86
SBP-DM	0.1437	3.17E-07	0.02 (0.01, 0.02)	0.0697	3.69E-07
DM-SVstat	0.3147	4.11E-12	0.39 (0.33, 0.46)	0.2524	1.28E-16
SHeight-Edu	0.0715	8.42E-09	0.08 (0.07, 0.09)	0.0643	2.28E-21
SBP-SVstat	0.2089	4.84E-26	0.04 (0.04, 0.05)	0.1853	1.46E-52
Edu-HDL	0.4037	5.25E-12	0.22 (0.17, 0.27)	0.2848	4.06E-08
BMI-CAD	0.2373	2.37E-64	0.28 (0.25, 0.32)	0.1800	2.42E-53
CAD-DM	0.1920	5.92E-13	0.01 (0.01, 0.01)	0.0659	0.002455431
DM-CAD	0.4283	5.60E-19	1.95 (1.26, 2.64)	0.1796	4.15E-05
SBP-CAD	0.2807	2.86E-46	0.45 (0.39, 0.51)	0.2500	9.77E-24
CAD-SVstat	0.2491	8.82E-44	0.03 (0.03, 0.04)	0.3077	1.15E-25
CAD-MI	0.4634	0	0.02 (0.02, 0.02)	0.4191	3.07E-285
LDL-CAD	0.3402	1.17E-45	0.31 (0.24, 0.38)	0.2014	8.56E-27
BMI-Edu	-0.2241	3.74E-14	-0.12 (-0.14, -0.11)	-0.1892	6.15E-35
SHeight-BMI	-0.1278	1.40E-22	-0.13 (-0.14, -0.11)	-0.0854	9.01E-23
SBP-BWeight	-0.2565	9.85E-08	-0.13 (-0.16, -0.1)	-0.1646	1.20E-11
SBP-SHeight	-0.3657	4.81E-08	-0.12 (-0.15, -0.1)	-0.0967	0.004422636
SHeight-SBP	-0.0759	5.74E-05	-0.08 (-0.09, -0.07)	-0.0652	1.25E-15
SHeight-SVstat	-0.0465	4.76E-09	-0.01 (-0.02, -0.01)	-0.0328	6.78E-12
BMI-HDL	-0.3760	3.54E-56	-0.28 (-0.29, -0.26)	-0.3630	3.17E-111
SHeight-LDL	-0.0716	4.26E-09	-0.04 (-0.05, -0.02)	-0.0298	5.07E-06
BWeight-CAD	-0.1745	2.05E-06	-0.21 (-0.28, -0.14)	-0.0978	2.83E-05
SHeight-CAD	-0.0802	3.72E-20	-0.15 (-0.18, -0.12)	-0.0482	2.18E-12
HDL-CAD	-0.1729	7.00E-31	-0.26 (-0.3, -0.21)	-0.0778	5.45E-10

Supplementary Table 3: Table comparing the causal estimates of LHC-MR, CAUSE, and IVW for trait pairs that had a significant causal effect in LHC-MR and CAUSE. The column showing the gamma (causal effect) estimate of the CAUSE method also reports its 95% credible intervals. A complete table for all the studied pairs is found in the Supplementary Data 4.

Appendix B: Genetic insights into the causal relationship between physical activity and cognitive functioning

This article is presented in chapter 2.3.

scientific reports

OPEN

Genetic insights into the causal relationship between physical activity and cognitive functioning

Boris Cheval[1,2,16], Liza Darrous[3,4,16], Karmel W. Choi[5], Yann C. Klimentidis[6], David A. Raichlen[7,8], Gene E. Alexander[10,11,12,9], Stéphane Cullati[13], Zoltán Kutalik[3,4,17] & Matthieu P. Boisgontier[14,15,17]

Physical activity and cognitive functioning are strongly intertwined. However, the causal relationships underlying this association are still unclear. Physical activity can enhance brain functions, but healthy cognition may also promote engagement in physical activity. Here, we assessed the bidirectional relationships between physical activity and general cognitive functioning using Latent Heritable Confounder Mendelian Randomization (LHC-MR). Association data were drawn from two large-scale genome-wide association studies (UK Biobank and COGENT) on accelerometer-measured moderate, vigorous, and average physical activity (N = 91,084) and cognitive functioning (N = 257,841). After Bonferroni correction, we observed significant LHC-MR associations suggesting that increased fraction of both moderate (b = 0.32, CI$_{95\%}$ = [0.17,0.47], *P* = 2.89e − 05) and vigorous physical activity (b = 0.22, CI$_{95\%}$ = [0.06,0.37], *P* = 0.007) lead to increased cognitive functioning. In contrast, we found no evidence of a causal effect of average physical activity on cognitive functioning, and no evidence of a reverse causal effect (cognitive functioning on any physical activity measures). These findings provide new evidence supporting a beneficial role of moderate and vigorous physical activity (MVPA) on cognitive functioning.

Multiple cross-sectional and longitudinal studies have shown that physical activity and cognitive functioning are strongly intertwined and decline through the course of life[1–5]. However, the evidence of causality of this relationship remains unclear. Previous results have shown that physical activity can improve cognitive functioning[6–12], but recent studies have also suggested that well-functioning cognitive skills can influence engagement in physical activity[1,13–20].

Several mechanisms could explain how physical activity, especially at moderate intensities, enhances general cognitive functioning[12,21–27]. For example, physical activity can increase brain plasticity, angiogenesis, synaptogenesis, and neurogenesis primarily through the upregulation of growth factors (e.g., brain-derived neurotrophic factor; BDNF)[23,24,26]. In addition, the repetitive activation of higher-order brain functions (e.g., planning, inhibition, and reasoning) required to engage in physical activity may contribute to the improvement of these functions[27,28]. In turn, other mechanisms could explain how cognitive functioning may affect physical activity. For example, cognitive functioning may be required to counteract the automatic attraction to effort minimization

[1]Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland. [2]Laboratory for the Study of Emotion Elicitation and Expression (E3Lab), Department of Psychology, University of Geneva, Geneva, Switzerland. [3]University for Primary Care and Public Health, University of Lausanne, Lausanne, Switzerland. [4]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [5]Department of Psychiatry, Massachusetts General Hospital, Massachusetts, Boston, MA, USA. [6]Department of Epidemiology and Biostatistics, University of Arizona, Tucson, AZ, USA. [7]Human and Evolutionary Biology Section, Department of Biological Sciences, University of Southern California, Los Angeles, CA, USA. [8]Department of Anthropology, University of Southern California, Los Angeles, CA, USA. [9]Department of Psychology, University of Arizona, Tucson, AZ, USA. [10]Department of Psychiatry, University of Arizona, Tucson, AZ, USA. [11]Evelyn F. McKnight Brain Institute, University of Arizona, Tucson, AZ, USA. [12]Arizona Alzheimer's Consortium, Phoenix, AZ, USA. [13]Population Health Laboratory, Department of Community Health, University of Fribourg, Fribourg, Switzerland. [14]School of Rehabilitation Sciences, Faculty of Health Sciences, University of Ottawa, Ottawa, ON, Canada. [15]Bruyère Research Institute, Ottawa, ON, Canada. [16]These authors contributed equally: Boris Cheval and Liza Darrous. [17]These authors jointly supervised this work: Zoltán Kutalik and Matthieu P. Boisgontier. ✉email: boris.cheval@unige.ch; liza.darrous@unil.ch; zoltan.kutalik@unil.ch; matthieu.boisgontier@uOttawa.ca

and thereby influence a person's ability to engage in physically active behavior[20,29–31]. Of note, these mechanisms are not mutually exclusive and could therefore lead to bidirectionally reinforcing relationships (i.e., positive feedback loop) between physical activity and cognitive functioning[32]. Thus, there is a mechanistic explanation theoretically supporting the associations between moderate physical activity and cognitive function.

Although these studies point to a potential mutually beneficial interplay between physical activity and cognitive functioning across the lifespan, these findings mainly stem from observational designs and analytical methods that cannot fully rule out the influence of social, behavioral, and genetic confounders[32]. While randomized controlled trials minimizing these potential confounds have been conducted[33], they were typically based on small sample sizes ($n < 100$) that can bias the estimations[33]. Critically, these trials only investigated the effect of physical activity on cognitive functioning, not the opposite. Accordingly, current evidence on the causal association between physical activity and cognitive functioning and on whether this association is one or two-way could be considered weak. Because Mendelian Randomization (MR) is less vulnerable to confounding or reverse causation than conventional approaches in observational studies[34,35], this method is particularly appropriate to address this knowledge gap.

MR is an epidemiological method in which the randomized inheritance of genetic variation is considered as a natural experiment to estimate the potential causal effect of a modifiable risk factor (exposure) on health-related outcomes in an observational design[34,35]. MR draws on the assumption that genetic variants associated with the exposure, because they are randomly allocated at conception, are less associated with other risk factors that may be confounders of the association between the exposure and the outcome, and are immune to reverse causality since diseases or health-related outcomes have no reverse effect on genetic variants. Accordingly, if an exposure (e.g., physical activity) causally affects an outcome (e.g., cognitive function), the genetic variants that influence this exposure is expected to affect the outcome to a proportional degree if no separate pathway exists by which these genetic variants can affect the outcome[32]. In other words, genetic variants associated with an exposure of interest can serve as instruments (or proxies) for estimating the causal association with an outcome (see Fig. 1 for the conceptual illustration of the MR method).

We used a newly-developed MR method showing improved power to simultaneously estimate the bidirectional causal effects between physical activity and cognitive functioning[36]. In a two-sample MR design, genetic instruments can be obtained from summary statistics of nonoverlapping large-scale genome-wide association studies (GWAS). That is, the genetic instruments for the exposure and the genetic instruments for the outcome can be obtained from separate studies[37]. This is an outstanding advantage for estimating the causal relationships between two traits (e.g., cognitive functioning and physical activity) because a trait does not necessarily need to be assessed in both samples[37]. Here, the causal estimates were modeled based on recently available summary statistics from large-scale GWAS of accelerometer-measured physical activity[38], and general cognitive functioning[39,40].

The current study focused on general cognitive functioning estimated from a battery of neuropsychological tests (e.g., N-Back working memory task, Stroop Test, Wechsler Adult Intelligence Scale)[41,42]. Although the influence of physical activity on different types of cognitive functions may differ, cognitive tests measuring these different functions yield highly correlated results in a given individual, making the assessment of general cognitive functioning highly relevant[40].

Since it has been suggested that the intensity of physical activity can be an important consideration, with moderate intensity having greater beneficial effects than vigorous intensity[43–47], we assessed whether the causal effect estimates on cognitive functioning were dependent on physical activity intensity (i.e., moderate vs. vigorous vs. average). However, if a stronger effect on cognitive function could be expected for moderate physical activity, recent studies showed that high-intensity exercise can also impact the above-mentioned mechanisms such as increased BDNF[48–50]. Here, consistent with existing literature using UK Biobank data[38,51], the fraction

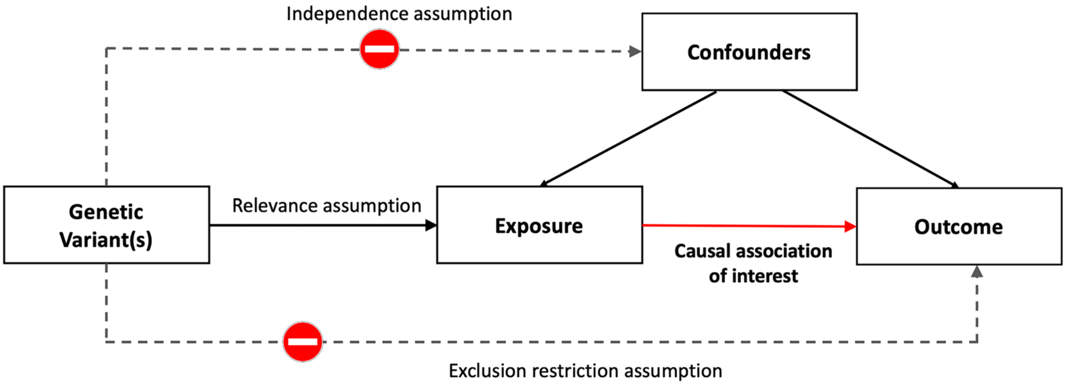


Figure 1. Conceptual illustration of the Mendelian Randomization (MR) method. The causal association of interest is between the exposure (e.g., physical activity) and the outcome (e.g., cognitive function). Relevance assumption states that the genetic instruments are strongly associated with the exposure but are not associated with the confounders. The exclusion restriction assumption states that the genetic instruments are only indirectly associated with the outcome via the exposure. Thus, the solid paths are expected to exist, while the dashed paths are expected to be nonsignificant according to the core MR assumptions.

of accelerations > 100 milli-gravities (mg) and < 425 mg was used to estimate moderate physical activity, and the fraction of accelerations ≥ 425 mg was used to estimate vigorous physical activity. Of note, as existing literature suggests reciprocal associations between physical activity and cognitive function, we applied bidirectional MR to examine the causal link from physical activity to cognitive function and from cognitive function to physical activity.

Methods

Data sources and instruments.

This study used de-identified GWAS summary statistics from original studies that were approved by relevant ethics committees. The current study was approved by the Ethics Committee of Geneva Canton, Switzerland (CCER-2019–00,065). The available summary-level data were based on 257,841 samples for general cognitive functioning and 91,084 samples for accelerometer-based physical activity. Participants' age ranged from 40 to 69 years in the UK Biobank and from 8 to 96 years in the COGENT consortium.

Physical activity.

Accelerometer-measured physical activity was assessed based on summary statistics from a recent GWAS[38], analyzing accelerometer-based physical activity data from the UK Biobank. In the UK Biobank, about 100,000 participants wore a wrist-worn triaxial accelerometer (Axivity AX3) that was set up to record data for seven days. Individuals with less than 3 days (72 h) of data or not having data in each 1-h period of the 24-h cycle or for whom the accelerometer could not be calibrated were excluded. Data for non-wear segments, defined as consecutive stationary episodes ≥ 60 min where all three axes had a standard deviation < 13 mg, were imputed. The details of data collection and processing can be found elsewhere[52]. We examined three measures derived from the three to seven days of accelerometer wear: the average acceleration in mg that includes acceleration > 0 mg, the fraction of accelerations > 100 mg and < 425 mg to estimate moderate physical activity, and the fraction of accelerations ≥ 425 mg to estimate vigorous physical activity[38]. As previous reported[51], 425 mg cut-off was chosen because it corresponds to vigorous intensity (6 METS). The GWAS for average physical activity (n_{max} = 91,084) identified 2 independent genome-wide significant SNPs (P < 5e − 09), with an SNP-based heritability of ~ 14%.

As for the other two physical activity measures, the fractions of accelerations corresponding to moderate and vigorous physical activity were obtained by running new GWAS on the decomposed acceleration data from UK Biobank using the BGENIE software[53]. The phenotype for moderate physical activity was limited to acceleration magnitudes ranging from 100 to < 425 mg, whereas vigorous physical activity was limited to acceleration magnitudes ranging from 425 to 2000 mg. These acceleration fractions were adjusted for age, sex, and the first 40 principal components (PC), and the analyzed individuals were restricted to unrelated white-British. The two datasets of average physical activity summary statistics, alongside the moderate and vigorous physical activity summary statistics, were used in Latent Heritable Confounder Mendelian Randomization (LHC-MR) to investigate the possible bidirectional effect that exists between these physical activity traits and cognitive functioning.

General cognitive functioning.

General cognitive functioning was assessed based on summary statistics from a recent GWAS combining cognitive and genetic data from the UK Biobank and the COGENT consortium (N = 257,841)[39]. The phenotypes of these cohorts are well-suited to meta-analysis because their pairwise genetic correlation has been shown to be high[40]. In the UK Biobank (n_{max} = 222,543) participants were asked to complete 13 multiple-choice questions that assessed verbal and numerical reasoning. For verbal reasoning, a typical question was "bud is to flower what child is to …?", and possible answers presented to the participants are "Grow", "Develop", "Improve", "Adult", or "Old". For numerical reasoning, a typical question was "150…137…125…114…104… what comes next?" with possible answers being "96", "95", "94", "93", or "92"[39]. The verbal and numerical reasoning score was based on the number of questions answered correctly within a two-minute time limit. Each respondent took the test up to four times. This test was designed as a measure of fluid intelligence. The phenotype consists of the mean of the standardized score across the measurement occasions for a given participant. In the COGENT consortium (n_{max} = 35,298), general cognitive function is statistically derived from a principal components analysis of individual scores on a neuropsychological test battery, such as the Verbal or spatial N-Back working memory task, Stroop Test, the Trail Making Test, or the Wechsler Adult Intelligence Scale[41]. Details on the test battery are available in the supplementary material of Davies et al.[42]. Of note, Davies et al.[42] demonstrated that two general cognitive function components extracted from different sets of cognitive tests on the same participants exhibit a high correlation, addressing the fact that different cohorts relied on different cognitive tests. Thus, the phenotype estimates overall cognitive functioning and is relatively invariant to the battery used and specific cognitive abilities assessed[54,55]. These COGENT data used to assess general cognitive functioning were also used in another GWAS study[40]. The GWAS identified 226 independent genome-wide significant SNPs, with a SNP-based heritability of ~ 20%.

Statistical analysis.

MR is a statistical approach for causal inference that can overcome the weaknesses of traditional observational studies[34,35]. MR-based effect estimates rely on three main assumptions[56], stating that genetic instruments (i) are strongly associated with the exposure (relevance assumption), (ii) are independent of confounding factors of the exposure-outcome relationship (independence assumption), and (iii) are not associated to the outcome conditional on the exposure and potential confounders (exclusion restriction assumption). Well-powered GWAS offer multiple genetic instruments that are strongly associated with exposures of interest (cognitive functioning or physical activity in our case), which validates the relevance assumption. Each of these genetic variants (instruments) provides a causal effect estimate of the exposure on the outcome, which can be in turn combined through meta-analysis using inverse-variance weighting (IVW) to obtain an overall estimate.

The second and third assumptions are less easily validated and can be violated in the case of a heritable confounder affecting the exposure-outcome relationship and biasing the causal estimate. Such confounders can give rise to instruments with proportional effects on the exposure and outcome, hence violating the Instrument Strength Independent of Direct Effect (InSIDE) assumption requiring the independence of the exposure and direct outcome effects. There have been several extensions to the common IVW method of MR analysis, including MR-Egger, which allows for directional pleiotropy of the instruments and attempts to correct the causal regression estimate. Other extensions, such as the median and mode-based estimators, assume that at least half of or the most "frequent" genetic instruments are valid/non-pleiotropic. However, despite these extensions and relaxed assumptions, all these classical MR methods are notably underpowered and still suffer from two major limitations. First, they only use a subset of markers as instruments (genome-wide significant markers), which often dilutes the true relationship between traits. Second, they ignore the presence of a potential latent heritable confounder of the exposure-outcome relationship (e.g., body mass index, educational attainment, level of physical activity at work, or material deprivation).

LHC-MR also uses GWAS summary statistics [36], but importantly, this new method appropriately uses genome-wide genetic markers to estimate bidirectional causal effects, direct heritability, and confounder effects while accounting for sample overlap. LHC-MR can be viewed as an extension of the linkage disequilibrium score regression (LDSC) [57], designed to estimate trait heritability, in that it models all genetic marker effects as random, but additionally estimates bidirectional causal effect, as well as other parameters. LHC-MR extends the standard two-sample MR by modeling a latent (unmeasured) heritable confounder that has an effect on the exposure and outcome traits. This allows LHC-MR to differentiate SNPs based on their co-association to a pair of traits and distinguish heritable confounding that leads to genetic correlation from actual causation. Thus, the unbiased bidirectional causal effect between these two traits are estimated simultaneously along with the confounder effect on each trait (Fig. 2a, b). The LHC-MR framework, with its multiple pathways through which SNPs can have an effect on the traits, as well as its allowance for null effects, make LHC-MR more precise at estimating causal effects compared to standard MR methods (i.e., MR egger, weighted median, inverse variance weighted, simple mode, and weighted mode).

The likelihood function for LHC-MR, which is derived from the mixture of different pathways through which the genome-wide SNPs can have an effect (acting on either the exposure, the outcome, the confounder, or the combinations of these three), is then optimized given random starting values for the parameters it can estimate. The optimization of the likelihood function then yields the maximum likelihood estimate (MLE) value for a set of estimated parameters, including the bidirectional causal effect between the exposure and the outcome as well as the strength of the confounder effect on each of those two traits. The standard errors of each of the parameters estimated using LHC-MR were obtained by implementing a block jackknife procedure where the SNP effects are split into blocks, and the MLE is computed again in a leave-one-block-out fashion. The variance of the estimates can then be computed from the results of the various MLE optimizations. Furthermore, the causal estimates obtained from LHC-MR are on the scale of 1 standard deviation (SD) outcome difference upon a 1 SD exposure change due to the use of standardized summary statistics for the two traits.

A sensitivity analysis in which the model was further adjusted for baseline self-reported level of physical activity at work, walking or standing at work, and the Townsend Deprivation Index was conducted.

Ethical approval. This study was approved by the Ethics Committee of Geneva Canton, Switzerland (CCER-2019–00,065).

Results

Three measures derived from accelerometer wear were used as a proxy for physical activity: average, moderate, and vigorous physical activity. These three measures were used in LHC-MR to investigate the possible bidirectional causal effects between them and cognitive functioning. The model tested was adjusted for age,

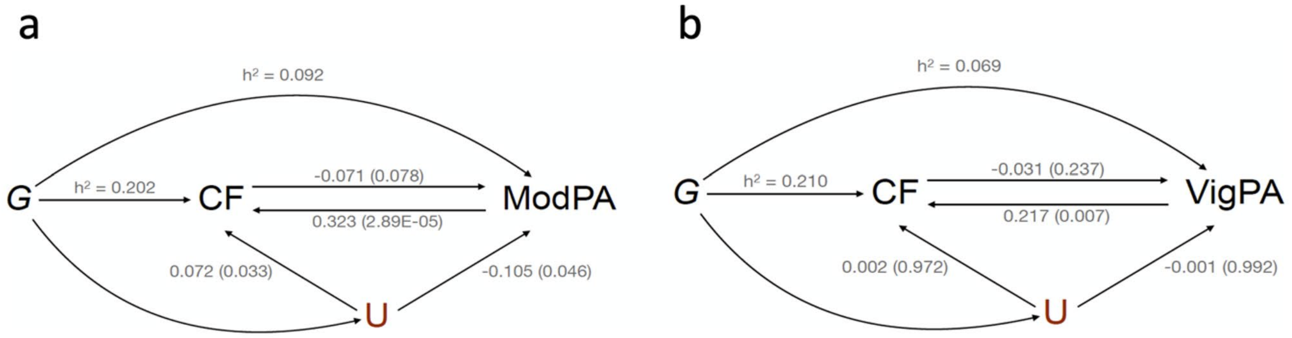


Figure 2. Visual representation of the model in Latent Heritable Confounder Mendelian Randomization (LHC-MR). G = Genetic instruments; CF = general cognitive functioning; (**a**) For moderate physical activity (ModPA); (**b**) For vigorous physical activity (VigPA); U = Latent heritable confounder; h^2 = direct heritability. Each figure includes the bidirectional causal effects between the two traits as well as the confounder effects on each of them. Coefficients are beta values. *P*-values are indicated in brackets. The models were adjusted for age, sex, genotyping chip, first ten genomic principal components, center, and season (month) of wearing an accelerometer.

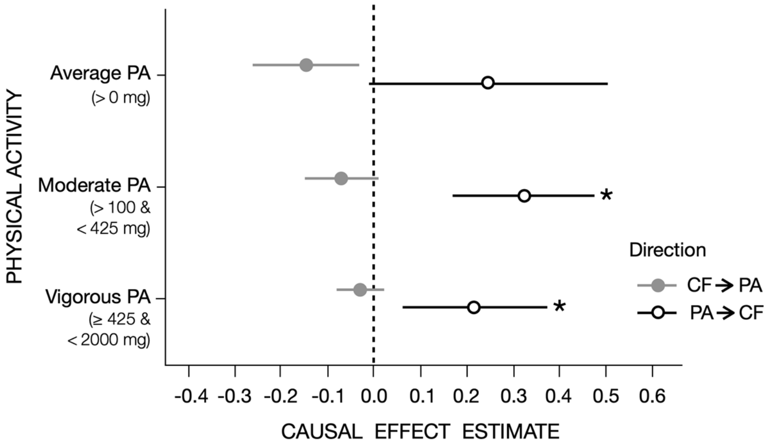


Figure 3. LHC-MR plots for the association between accelerometer-based physical activity and general cognitive functioning, *Notes*, This modified dot-and-whisker plot reports the causal estimate between general cognitive functioning (CF) as exposure and varying physical activity (PA)-related traits as outcomes. The forward (CF → PA) and reverse (PA → CF) causal estimates are shown in two different colors as dots (grey and white) with 95% CI whiskers (grey and black). Average PA = average of overall accelerations > 0 mg. Moderate PA = fraction of acceleration corresponding to moderate physical activity (> 100 mg and < 425 mg). Vigorous PA = fraction of acceleration corresponding to vigorous physical activity (≥ 425 mg and < 2000 mg). The models were adjusted for age, sex, genotyping chip, first ten genomic principal components (PC), center, and season (month) of wearing accelerometer. * = significant effect after Boneferroni correction (i.e., *P*-value < .008).

	Cognitive functioning		Physical activity		Cognitive functioning → Physical activity	Physical activity → Cognitive functioning
Parameter	**Heritability**	**t**	**Heritability**	**t**		
Average accelerometer-measured physical activity (fraction of acceleration > 0 mg)						
Estimate	0.207	0.033	0.123	− 0.011	− 0.145	0.245
P value	2.67E − 115	0.612	4.41E − 28	0.816	0.013	0.065
Moderate accelerometer-measured physical activity (fraction of acceleration > 100 mg and < 425 mg)						
Estimate	0.202	0.072	0.092	− 0.105	− 0.071	0.323
P value	1.13E − 165	0.032	5.98E − 29	0.046	0.078	2.89e − 05
Vigorous accelerometer-measured physical activity (fraction of acceleration ≥ 425 mg and < 2000 mg)						
Estimate	0.210	0.002	0.069	− 0.001	− 0.031	0.212
P value	6.75E − 157	0.972	3.95E − 25	0.992	0.237	0.007

Table 1. Latent Heritable Confounder Mendelian Randomization (LHC-MR) results for the association between accelerometer-measured physical activity and general cognitive functioning. Notes. Parameters estimates and their *P*-values were obtained from the LHC-MR optimized model with the maximum likelihood. Bidirectional associations from cognitive functioning to physical activity and from physical activity to cognitive functioning are reported. t = effect of the confounder. Bonferroni corrected α = 0.008.

sex, genotyping chip, first ten genomic principal components (PC), center, and season (month) of wearing accelerometer. The Bonferroni correction was used to control for familywise error rates, yielding an α = 0.05 / (2 directions × 3 tests) = 0.008.

Average physical activity and general cognitive functioning. LHC-MR applied to summary statistics belonging to model 1 showed no evidence for a potential causal effect of average physical activity on cognitive functioning (b = 0.245, $CI_{95\%}$ = [− 0.01,0.50], *P* = 0.065) (Table 1, Fig. 3) and no evidence for the reverse causal effect (b = − 0.145, $CI_{95\%.}$ = [− 0.26,− 0.03], *P* = 0.013 [α = 0.008]). Similarly, standard MR methods such as IVW, MR Egger, weighted median, simple mode, and weighted mode yielded non-significant causal estimates in either direction (Table 2), using 129 genome-wide significant single nucleotide polymorphisms (SNPs) as instruments for cognitive functioning and 6 SNPs for average acceleration.

Exposure	Outcome	MR method	Valid SNPs	Causal estimate	SE	*P* value
Average accelerometer-based physical activity (fraction of acceleration > 0 mg)						
Cognitive functioning	Physical activity	MR Egger	129	0.015	0.185	0.935
		Weighted median	129	−0.027	0.036	0.440
		Inverse variance weighted	129	−0.011	0.032	0.723
		Simple mode	129	−0.102	0.116	0.376
		Weighted mode	129	−0.084	0.111	0.452
Physical activity	Cognitive functioning	MR Egger	4	−2.833	1.148	0.069
		Weighted median	4	0.017	0.062	0.782
		Inverse variance weighted	4	−0.088	0.127	0.488
		Simple mode	4	0.020	0.076	0.801
		Weighted mode	4	0.023	0.074	0.770
Moderate accelerometer-based physical activity (fraction of acceleration > 100 mg and < 425 mg)						
Cognitive functioning	Physical activity	MR Egger	129	−0.054	0.181	0.766
		Weighted median	129	−0.032	0.037	0.389
		Inverse variance weighted	129	−0.012	0.032	0.710
		Simple mode	129	−0.059	0.106	0.575
		Weighted mode	129	−0.031	0.091	0.729
Physical activity	Cognitive functioning	MR Egger	106	0.325	0.319	0.310
		Weighted median	106	−0.001	0.021	0.981
		Inverse variance weighted	106	0.023	0.022	0.309
		Simple mode	106	−0.017	0.057	0.767
		Weighted mode	106	−0.010	0.050	0.837
Vigorous accelerometer-based physical activity (fraction of acceleration ≥ 425 mg)						
Cognitive FUNCTIONING	Physical activity	MR Egger	129	0.009	0.149	0.952
		Weighted median	129	0.018	0.036	0.623
		Inverse variance weighted	129	0.002	0.026	0.939
		Simple mode	129	0.021	0.097	0.829
		Weighted mode	129	0.021	0.088	0.812
Physical activity	Cognitive functioning	MR Egger	88	0.151	0.335	0.653
		Weighted median	88	−0.035	0.022	0.108
		Inverse variance weighted	88	−0.016	0.020	0.432
		Simple mode	88	−0.065	0.060	0.286
		Weighted mode	88	−0.059	0.052	0.257

Table 2. Standard Mendelian Randomization (MR) results for the association between accelerometer-based physical activity and general cognitive functioning. Causal estimates from 5 standard Mendelian Randomization (MR) methods on alternating exposure and outcome traits. For both moderate and vigorous physical activity as exposure, the cutoff was decreased to 6.33e−5 because of the low number of genome wide significant single nucleotide polymorphisms (SNPs) to use as instruments. Corrected $\alpha = 0.008$.

Moderate physical activity and general cognitive functioning. LHC-MR applied to the fraction of accelerations corresponding to moderate physical activity showed a potential positive causal effect of moderate physical activity on greater cognitive functioning ($b = 0.32$, $\text{CI}_{95\%} = [0.17,0.47]$, $P = 2.89\text{e}{-05}$) (Table 1, Fig. 3). We found no evidence for the reverse causal effect ($b = -0.071$, $\text{CI}_{95\%.} = [-0.15, 0.01]$, $P = 0.078$ [$\alpha = 0.008$]). As was found with average physical activity, there was no evidence for the presence of a heritable confounder. Standard MR methods yielded non-significant causal estimates in both directions (Table 2).

Vigorous physical activity and general cognitive functioning. LHC-MR applied to the fraction of accelerations corresponding to vigorous physical activity on cognitive functioning showed a potential positive causal effect of vigorous physical activity on greater cognitive functioning ($b = 0.22$, $\text{CI}_{95\%} = [0.06,0.37]$, $P = 0.007$) (Table 1, Fig. 3). We found no evidence for the reverse causal effect ($b = -0.031$, $\text{CI}_{95\%.} = [-0.08, 0.02]$, $P = 0.237$ [$\alpha = 0.008$]). As was found with average and moderate physical activity, there was no evidence for the presence of a heritable confounder. Of note, the coefficient of this causal effect was qualitatively weaker than of the causal effect of moderate physical activity on cognitive functioning ($b = 0.22$ vs. $b = 0.32$). Standard MR methods yielded non-significant causal estimates in both directions (Table 2).

Sensitivity analyses. We tested another model where an extra adjustment had been done for the baseline self-reported level of physical activity at work, walking or standing at work, and the Townsend Deprivation Index. LHC-MR applied to summary statistics emerging from this second model showed consistent results with that of the first model (b = 0.22, $CI_{95\%} = [-0.05,0.50]$, $P = 0.111$ and b = − 0.090, $CI_{95\%} = [-0.23,0.05]$, $P = 0.200$, respectively). Both models showed no evidence for the presence of a heritable confounder. Due to the similarity in results between these models, we did not conduct this second model on moderate and vigoruous physical activity.

Discussion

Main findings. This study used a genetically informed method that provides evidence of putative causal relations to investigate the bidirectional associations between accelerometer-based physical activity and general cognitive functioning. Drawing on large-scale GWAS, we found evidence for potential causal effects, suggesting that higher levels of moderate and vigorous physical activity lead to increased cognitive functioning. In the opposite direction, we did not observe evidence of a causal effect of cognitive functioning on physical activity. Hence, our study suggests a favorable effect of moderate and vigorous physical activity on cognitive functioning, but does not provide evidence that increased cognitive functioning promotes engagement in more physical activity.

Comparison with previous studies

Previous reviews and meta-analyses of observational studies showed a beneficial effect of physical activity on cognitive functioning [6,9,10,27]. However, the evidence arising from intervention studies was inconclusive [11,12,14–16,58]. It has been argued that these inconsistencies may primarily be attributed to the design-specific tools used to assess physical activity [14]. Specifically, many observational studies rely on self-reported measures of physical activity, whereas intervention studies often rely on accelerometer-measured physical activity, or have people exercising under monitored conditions. In other words, evidence of a favorable effect of physical activity on cognitive functioning may have emerged in observational studies because of the self-reported nature of the measures they typically used. Yet, in our study, results are based on accelerometer-assessed physical activity, thereby partially ruling out this explanation. Therefore, our findings further support the literature that demonstrated a protective role of physical activity on cognitive functioning and extend it by doing so using an accelerometer-based measure.

Of note, results obtained from LHC-MR differed from those obtained with standard MR methods. At least three key differences in the methods can explain this divergence: i) standard MR uses only genome-wide significant markers, ii) standard MR is biased in case of sample overlap (as is the case in this study) and hence their estimate may be biased towards the observational correlation, and iii) LHC-MR explicitly models correlated pleiotropy unlike standard MR. Accordingly, our results obtained from LHC-MR are expected to be more robust than those obtained from standard MR. Since LHC-MR could not find evidence for the presence of a heritable confounder, correlated pleiotropy is less likely, or there might be multiple confounders with opposite effects canceling each other out. This finding highlights that the main reason for the difference between LHC-MR and classical MR methods is statistical power. For testing the reverse causal effect (cognition on physical activity), we had numerous instruments available, ensuring that all MR-methods are well-powered and yielding the same (null effect) conclusion. The forward effect (physical activity on cognition) relied on only a few (weak) instruments, rendering classical MR methods notably underpowered. This is the type of situation in which methods such as LHC-MR, which leverage genome-wide genetic markers, are crucial to facilitate discovery. It is important to point out that while the statistical conclusion from classical and LHC-MR methods differ, their effect estimates are not significantly different, suggesting that there is no discrepancy in the results, but that they have different precision. Finally, we acknowledge that LHC-MR assumptions may be violated and results should thus still be considered cautiously. Yet, while the assumptions of LHC-MR may not hold, the assumptions of the other five methods are known not to hold because of insufficient genome-wide significant instruments.

To the best of our knowledge, our study is the first to investigate the potential causal relationship between physical activity and cognitive functioning using a genetically informed method. We are aware of only two other, non-genetic studies that examined the potential bidirectional associations between physical activity and cognitive functioning [1,13]. In contrast to the present study, those two studies observed a positive influence of cognitive functioning on physical activity. At least two factors can explain the differences in the results observed. First, both those studies are based on longitudinal assessment of the two traits, while our approach is based on a genetically instrumented causal inference technique (LHC-MR). Second, these studies draw on self-reported physical activity rather than accelerometer-measured physical activity, which may not accurately reflect the objective level of physical activity.

Our results obtained with recently-improved genetically-informed analyses (LHC-MR) highlight the potential critical role of physical activity, specifically of moderate and vigorous intensity, on cognitive functioning. However, it should be noted that the estimated effect of moderate physical activity on cognitive functioning was about 1.5 times stronger in magnitude than the effect of vigorous physical activity. To the best of our knowledge, this study is the first to assess and compare the causal relationships of moderate and vigorous physical activity with cognitive functioning using a genetically-informed method based on large-scale datasets. Alhough additional evidence is needed, this study confirms the importance to examine the extent to which the intensity of physical activity moderates the effects observed on cognitive functioning [43].

The LHC-MR method revealed two causal relations that are consistent with each other. Importantly, these findings are consistent with theoretical and experimental work explaining the mechanisms underlying the association between the physical activity and cognitive functioning. Results obtained with both the LHC and standard MR methods showed no evidence of an effect of average physical activity on cognitive functioning. This finding can likely be explained by physical activities of low intensity (i.e., < 100 mg) that are part of the average physical activity, which further suggests that physical activity should be of moderate-to-vigorous intensity to benefit cognitive functioning.

The absence of evidence for a reverse causal effect of cognitive functioning on physical activity may be partly explained by the lower power of this analysis due to smaller sample size of the GWAS of physical activity (n = 91,084) compared to the sample size of the GWAS of cognitive functioning (n = 257,841). This absence of evidence contrasts with other studies arguing that cognitive functioning is critical for supporting engagement in physical activity [20,29,30]. This difference could be explained in at least two ways. Firstly, previous studies examining the positive effect of cognitive functions on physical activity relied on self-reported physical activity, which can bias the observed associations [1,17,20]. Secondly, our study relied on general cognitive functioning, whereas previous results highlight the specific importance of inhibition resources that may be required to counteract an automatic tendency for effort minimization [20,29–31,59,60]. Therefore, future studies should investigate the specific relationships between motor inhibition and physical activity when such data is available.

Strengths and limitations

Among the strengths of the current study are the use of large-scale datasets, the reliance on instruments derived from objective measures of physical activity, and the application of a robust genetically informed method that can estimate causal effects. However, this study has several features that limit the conclusions that can be drawn. First, the measure of cognitive functioning spans multiple performance domains, which reduced the specificity of the cognitive functioning that was assessed. This feature limits our ability to evaluate the putative causal effects between specific cognitive functioning, such as motor inhibition, and physical activity. Second, MR analysis is designed to elucidate a life-long exposure effect on a life-long outcome (except in special cases when genetic factors have time-dependent effects), thus it is not suited to explore temporal aspects of these causal relationships. Third, 2-sample MR methods require that SNP effects on the exposure are homogeneous between the two samples. Here, because our two samples differ in age, we rely on the assumption that these genetic effects do not change depending on age. This assumption often turns out to be true, although there are rare exceptions [61]. It is therefore still possible that genetic variants related to physical activity and cognitive function may differ across the life course. For example, genetic variants related to cognitive development, maintenance and decline may strongly differ. Likewise, the genetic variance predicting physical activity engagement in early-life may differ from those predicting engagement in adulthood or late life. Accordingly, as the age range between the sample is not equivalent (40 to 60 years for the UK Biobank vs. 8 to 96 years in the COGENT consortium) and, most importantly, as physical activity was only assessed in the UK biobank that provides the narrowest age range, the potential differences in the genetic variants depending on individual's age may have bias the current findings. Testing to which extent age may influence the genetic variants associated with physical activity and cognitive functioning traits is thus warranted in future studies. Fourth, LHC-MR can be limited by the low heritability of traits, potentially causing bimodal/unreliable estimates. Fifth, LHC-MR assumes a single confounder (or several ones with similar effects), but a limitation exists when multiple confounders are present with similar but opposing effect directions on the traits of interest, resulting in a higher misdetection rate. Sixth, although the coefficients estimated with LHC-MR did not statistically differ from the coefficents estimated with classical MR, it is important to acknowledge that no classical MR was unable to find a significant association between physical activity and cognitive function. Accordongly, even if we can be rather confident in the estimation provided by the newly developed methods, it seems more reasonable to consider that the current findings are provisional and need to be replicated. Finally, it is worth noting that the genetic instruments were developed on a primarily white population of European ancestry, limiting the generalizability of the results.

Conclusion and policy implications

Our findings provide preliminary support for a unidirectional relation whereby higher levels of moderate and vigorous physical activity lead to improved cognitive functioning. These results underline the essential role of moderate and vigorous physical activity in maintaining or improving general cognitive functioning. Therefore, health policies and interventions that promote moderate and vigorous physical activity are relevant to improve cognitive functioning or to delay its decline.

Data availability

The datasets used for the analysis are openly available from the Neale Lab GWAS results at http://www.nealelab.is/uk-biobank and from the Social Science Genetic Association Consortium Downloads at https://www.thessgac.org/data. Only the new GWAS dataset created for the fractions of physical activity are available with permission from the UK Biobank https://www.ukbiobank.ac.uk/. The LHC-MR code is available at https://github.com/LizaDarrous/lhcMR.

Received: 30 November 2022; Accepted: 23 March 2023
Published online: 31 March 2023

References

1. Cheval, B. *et al.* Relationship between decline in cognitive resources and physical activity. *Health Psychol.* **39**, 519–528 (2020).
2. Cheval, B. *et al.* Effect of early-and adult-life socioeconomic circumstances on physical inactivity. *Med. Sci. Sports Exerc.* **50**, 476–485 (2018).
3. DiPietro, L. Physical activity in aging: Changes in patterns and their relationship to health and function. *J. Gerontol. A Biol. Sci. Med. Sci.* **56**, 13–22 (2001).
4. Levy, R. Aging-associated cognitive decline. *Int. Psychogeriatr.* **6**, 63–68 (1994).
5. Sebastiani, P. *et al.* Patterns of multi-domain cognitive aging in participants of the long life family study. *Geroscience* **42**, 1335–1350 (2020).

6. Baumgart, M. *et al.* Summary of the evidence on modifiable risk factors for cognitive decline and dementia: A population-based perspective. *Alzheimer Dement.* **11**, 718–726 (2015).
7. Blondell, S. J., Hammersley-Mather, R. & Veerman, J. L. Does physical activity prevent cognitive decline and dementia?: A systematic review and meta-analysis of longitudinal studies. *BMC Public Health* **14**, 510 (2014).
8. Hamer, M., Terrera, G. M. & Demakakos, P. Physical activity and trajectories in cognitive function: English Longitudinal Study of Ageing. *J. Epidemiol. Community Health* **72**, 477–483 (2018).
9. Morgan, G. S. *et al.* Physical activity in middle-age and dementia in later life: Findings from a prospective cohort of men in Caerphilly, South Wales and a meta-analysis. *J. Alzheimers Dis.* **31**, 569–580 (2012).
10. Sofi, F. *et al.* Physical activity and risk of cognitive decline: A meta-analysis of prospective studies. *J. Intern. Med.* **269**, 107–117 (2011).
11. Angevaren, M., Aufdemkampe, G., Verhaar, H., Aleman, A. & Vanhees, L. Physical activity and enhanced fitness to improve cognitive function in older people without known cognitive impairment. *Cochrane Database Syst. Rev.* **3**, CD005381 (2008).
12. Colcombe, S. & Kramer, A. F. Fitness effects on the cognitive function of older adults: A meta-analytic study. *Psychol. Sci.* **14**, 125–130 (2003).
13. Daly, M., McMinn, D. & Allan, J. L. A bidirectional relationship between physical activity and executive function in older adults. *Front. Hum. Neurosci.* **8**, 1044 (2015).
14. Sabia, S. *et al.* Physical activity, cognitive decline, and risk of dementia: 28 year follow-up of Whitehall II cohort study. *Brit. Med. J.* **357**, j2709 (2017).
15. Snowden, M. *et al.* Effect of exercise on cognitive performance in community-dwelling older adults: Review of intervention trials and recommendations for public health practice and research. *J. Am. Geriatr. Soc.* **59**, 704–716 (2011).
16. Young, J., Angevaren, M., Rusted, J. & Tabet, N. Aerobic exercise to improve cognitive function in older people without known cognitive impairment. *Cochrane Database Syst. Rev.* **4**, CD005381 (2015).
17. Lindwall, M. *et al.* Dynamic associations of change in physical activity and change in cognitive function: Coordinated analyses of four longitudinal studies. *J. Aging Res.* **2012**, 793598 (2012).
18. Cheval, B. *et al.* Higher inhibitory control is required to escape the innate attraction to effort minimization. *Psychol. Sport Exerc.* **51**, 101781 (2020).
19. Cheval, B. *et al.* Cognitive functions and physical activity in aging when energy is lacking. *Eur. J. Ageing* **19**, 533–544 (2022).
20. Cheval, B. *et al.* Cognitive resources moderate the adverse impact of poor neighborhood conditions on physical activity. *Prev Med* **126**, 105741 (2019).
21. Colzato, L. S., Kramer, A. F. & Bherer, L. Editorial special topic: Enhancing brain and cognition via physical exercise. *J. Cogn. Enhanc.* **2**, 135–136 (2018).
22. Roig, M., Nordbrandt, S., Geertsen, S. S. & Nielsen, J. B. The effects of cardiovascular exercise on human memory: A review with meta-analysis. *Neurosci. Biobehav. Rev.* **37**, 1645–1666 (2013).
23. Cotman, C. W. & Berchtold, N. C. Exercise: A behavioral intervention to enhance brain health and plasticity. *Trends Neurosci.* **25**, 295–301 (2002).
24. Hillman, C. H., Erickson, K. I. & Kramer, A. F. Be smart, exercise your heart: Exercise effects on brain and cognition. *Nat. Rev. Neurosci.* **9**, 58–65 (2008).
25. Lisanne, F., Hsu, C. L., Best, J. R., Barha, C. K. & Liu-Ambrose, T. Increased aerobic fitness is associated with cortical thickness in older adults with mild vascular cognitive impairment. *J. Cogn. Enhanc.* **2**, 157–169 (2018).
26. Cotman, C. W., Berchtold, N. C. & Christie, L.-A. Exercise builds brain health: Key roles of growth factor cascades and inflammation. *Trends Neurosci.* **30**, 464–472 (2007).
27. Raichlen, D. A. & Alexander, G. E. Adaptive capacity: An evolutionary neuroscience model linking exercise, cognition, and brain health. *Trends Neurosci.* **40**, 408–421 (2017).
28. Frith, E. & Loprinzi, P. Physical activity and individual cognitive function parameters: Unique exercise-induced mechanisms. *JCBPR* **7**, 92–106 (2018).
29. Cheval, B. *et al.* Behavioral and neural evidence of the rewarding value of exercise behaviors: A systematic review. *Sports Med.* **48**, 1389–1404 (2018).
30. Cheval, B. *et al.* Avoiding sedentary behaviors requires more cortical resources than avoiding physical activity: An EEG study. *Neuropsychologia* **119**, 68–80 (2018).
31. Cheval, B. *et al.* Higher inhibitory control is required to escape the innate attraction to effort minimization. *Psychol. Sport Exerc.* **51**, 101781 (2020).
32. Choi, K. W. *et al.* Assessment of bidirectional relationships between physical activity and depression among adults: A 2-sample mendelian randomization study. *JAMA Psychiatry* **76**, 399–408 (2019).
33. Northey, J. M., Cherbuin, N., Pumpa, K. L., Smee, D. J. & Rattray, B. Exercise interventions for cognitive function in adults older than 50: A systematic review with meta-analysis. *Br. J. Sports Med.* **52**, 154–160 (2018).
34. Davies, N. M., Holmes, M. V. & Smith, G. D. Reading Mendelian randomisation studies: A guide, glossary, and checklist for clinicians. *BMJ* **362**, k601 (2018).
35. Byrne, E. M., Yang, J. & Wray, N. R. Inference in psychiatry via 2-sample mendelian randomization—from association to causal pathway?. *JAMA Psychiatry* **74**, 1191–1192 (2017).
36. Darrous, L., Mounier, N. & Kutalik, Z. Simultaneous estimation of bi-directional causal effects and heritable confounding from GWAS summary statistics. *Nat. Commun.* **12**, 1–15 (2021).
37. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**, 34408 (2018).
38. Klimentidis, Y. C. *et al.* Genome-wide association study of habitual physical activity in over 377,000 UK biobank participants identifies multiple variants including CADM2 and APOE. *Int. J. Obes.* **42**, 1161–1176 (2018).
39. Lee, J. J. *et al.* Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment. *Nat. Genet.* **50**, 1112 (2018).
40. Davies, G. *et al.* Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat. Commun.* **9**, 1–16 (2018).
41. Trampush, J. W. *et al.* GWAS meta-analysis reveals novel loci and genetic correlates for general cognitive function: A report from the COGENT consortium. *Mol. Psychiatry* **22**, 336–345 (2017).
42. Davies, G. *et al.* Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N= 112 151). *Mol. Psychiatry* **21**, 758–767 (2016).
43. Szuhany, K. L., Bugatti, M. & Otto, M. W. A meta-analytic review of the effects of exercise on brain-derived neurotrophic factor. *J. Psychiatry Res.* **60**, 56–64 (2015).
44. Bosch, B. M., Bringard, A., Ferretti, G., Schwartz, S. & Iglói, K. Effect of cerebral vasomotion during physical exercise on associative memory, a near-infrared spectroscopy study. *Neurophotonics* **4**, 041404 (2017).
45. Chang, Y.-K., Labban, J. D., Gapin, J. I. & Etnier, J. L. The effects of acute exercise on cognitive performance: A meta-analysis. *Brain Res.* **1453**, 87–101 (2012).
46. Suwabe, K. *et al.* Rapid stimulation of human dentate gyrus function with acute mild exercise. *PNAS* **115**, 10487–10492 (2018).
47. Bosch, B. M. *et al.* A single session of moderate intensity exercise influences memory, endocannabinoids and brain derived neurotrophic factor levels in men. *Sci. Rep.* **11**, 14371 (2021).

48. Antunes, B. M., Rossi, F. E., Teixeira, A. M. & Lira, F. S. Short-time high-intensity exercise increases peripheral BDNF in a physical fitness-dependent way in healthy men. *Eur. J. Sport. Sci.* **20**, 43–50 (2020).
49. Piepmeier, A. T. *et al.* A preliminary investigation of acute exercise intensity on memory and BDNF isoform concentrations. *Eur. J. Sport. Sci.* **20**, 819–830 (2020).
50. Rentería, I. *et al.* Short-term high-Intensity interval training increases systemic brain-derived neurotrophic factor (BDNF) in healthy women. *Eur. J. Sport Sci.* **20**, 516–524 (2020).
51. Hildebrand, M., Van Hees, V. T., Hansen, B. H. & Ekelund, U. Age group comparability of raw accelerometer output from wrist-and hip-worn monitors. *Med. Sci. Sports Exerc.* **46**, 1816–1824 (2014).
52. Doherty, A. *et al.* Large scale population assessment of physical activity using wrist worn accelerometers: The UK biobank study. *PLoS One* **12**, e0169649 (2017).
53. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
54. Johnson, W., te Nijenhuis, J. & Bouchard, T. J. Jr. Still just 1 g: Consistent results from five test batteries. *Intelligence* **36**, 81–95 (2008).
55. Panizzon, M. S. *et al.* Genetic and environmental influences on general cognitive ability: Is ga valid latent construct? *Intelligence* **43**, 65–76 (2014).
56. Lawlor, D. A., Harbord, R. M., Sterne, J. A., Timpson, N. & Davey Smith, G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* **27**, 1133–1163 (2008).
57. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
58. Erickson, K. I. *et al.* Physical activity, cognition, and brain outcomes: A review of the 2018 physical activity guidelines. *Med. Sci. Sports Exerc.* **51**, 1242–1251 (2019).
59. Cheval, B. *et al.* Inhibitory control elicited by physical activity and inactivity stimuli: An EEG study. *Motiv. Sci.* **7**, 386–389 (2021).
60. Cheval, B. & Boisgontier, M. P. The theory of effort minimization in physical activity. *Exerc. Sport Sci. Rev.* **49**, 168–178 (2021).
61. Winkler, T. W. *et al.* The influence of age and sex on genetic associations with adult body size and shape: A large-scale genome-wide interaction study. *PLoS Genet.* **11**, e1005378 (2015).

Acknowledgements

B.C. is supported by an Ambizione grant (PZ00P1_180040) from the Swiss National Science Foundation (SNSF). M.P.B. is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC; RGPIN-2021-03153), the Canada Foundation for Innovation (CFI), and the Banting Research Foundation. Y.C.K. is supported by the National Institute of Health (R01 HL136528). The authors are thankful to Gail Davies for providing useful comments on the manuscript.

Author contributions

B.C., M.P.B. conceived and designed the study. L.D., Z.K. analyzed the data. B.C., M.P.B., L.D, Z.K. drafted the manuscript. All authors critically appraised the manuscript, worked on its content, and approved its submitted version.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to B.C., L.D., Z.K. or M.P.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Appendix C: PheWAS-based clustering of Mendelian Randomisation instruments reveals distinct mechanism-specific causal effects between obesity and educational attainment

This article is presented in chapter 3.1.

PheWAS-based clustering of Mendelian Randomisation instruments reveals distinct mechanism-specific causal effects between obesity and educational attainment

Liza Darrous[1,2,3,†], Gibran Hemani[4,5], George Davey Smith[4,5], and Zoltán Kutalik[1,2,3,†]

[1]University Center for Primary Care and Public Health, University of Lausanne, Switzerland
[2]Swiss Institute of Bioinformatics, Lausanne, Switzerland
[3]Department of Computational Biology, University of Lausanne, Lausanne, Switzerland
[4]Medical Research Council Integrative Epidemiology Unit, Population Health Sciences, University of Bristol, Bristol, United Kingdom
[5]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom
[†] *Correspondence should be addressed to darrous.liza@gmail.com or zoltan.kutalik@unil.ch*

Abstract

Mendelian Randomisation (MR) estimates causal effects between risk factors and complex outcomes using genetic instruments. Pleiotropy, heritable confounders, and heterogeneous causal effects violate MR assumptions and can lead to biases. To alleviate these, we propose an approach employing a Phenome-Wide association Clustering of the MR instruments (PWC-MR) and apply this method to revisit the surprisingly large apparent causal effect of body mass index (BMI) on educational attainment (EDU): $\widehat{\alpha}$ = -0.19 [-0.22, -0.16].

First, we clustered 324 BMI-associated genetic instruments based on their association with 407 traits in the UK Biobank, which yielded six distinct groups. The subsequent cluster-specific MR revealed heterogeneous causal effect estimates on EDU. A cluster enriched for socio-economic indicators yielded the largest BMI-on-EDU causal effect estimate ($\widehat{\alpha}$ = -0.49 [-0.56, -0.42]) whereas a cluster enriched for body-mass specific traits provided a more likely estimate ($\widehat{\alpha}$ = -0.09 [-0.13, -0.05]). Follow-up analyses confirmed these findings: within-sibling MR ($\widehat{\alpha}$ = -0.05 [-0.09, -0.01]); MR for childhood BMI on EDU ($\widehat{\alpha}$ = -0.03 [-0.06, -0.002]); step-wise multivariable MR ($\widehat{\alpha}$ = -0.05 [-0.07, -0.02]) where socio-economic indicators were jointly modelled.

In-depth examination of the BMI-EDU causal relationship demonstrated the utility of our PWC-MR approach in revealing distinct pleiotropic pathways and confounder mechanisms.

1 Introduction

Genome-wide association studies[1] (GWASs) have identified many genetic variants associated with multiple complex phenotypes, aiding us in annotating single nucleotide polymorphisms (SNPs) and their functions, as well as identifying putative causal genes. As sample sizes of GWASs increase, more SNP associations are revealed which improve various downstream analyses such as polygenic score prediction, pathway- and tissue-enrichment, and causal inference[2, 3].

Mendelian Randomisation[4, 5] (MR), an approach generally applied through the use of genetic variants/SNPs as instrumental variables (IVs) to infer the causal relationship between an exposure or a risk factor X and an outcome Y, has become increasingly used thanks to well-powered GWASs from which hundreds of genetic associations with heritable exposures can be used as IVs.

MR has three major assumptions concerning the genetic variant G used as an instrument: (1) Relevance – G is strongly associated with the exposure. (2) Exchangeability – there is no confounder of the G-outcome relationship. (3) Exclusion restriction – G affects the outcome only through the exposure. Each instrument provides a causal effect estimate, which can then be combined with others using an inverse variance-weighting[6] (IVW) method to obtain an estimate of the total causal effect of the exposure on the outcome. This estimate is more reliable than observational associations due to it being more protected against unmeasured confounding and reverse causality, provided that the core conditions are met.

Thanks to well-powered GWASs, we have also discovered that most genetic instruments are highly pleiotropic[7], i.e. associated to more than a single trait. This has also been shown in phenome-wide association studies (PheWASs), where associations between a SNP and a large number of phenotypes are tested. The situation where a genetic variant influences multiple traits, but there is a primarily associated trait which mediates all other trait associations, is referred to as vertical pleiotropy. On the other hand, genetic variants that affect some traits through pathways other than the primary trait (the exposure) – a phenomena known as horizontal pleiotropy – are in direct violation of the exclusion restriction assumption and could lead to biased causal effect estimates. However, if the InSIDE assumption [8](Instrument Strength is Independent of the Direct Effect on the outcome) holds and the direct SNP effects are on average null, then IVW will yield consistent causal effect estimates. There have been MR extensions to IVW such as MR-Egger to produce less biased causal effect estimates if the InSIDE assumption holds and direct effects are not null on average. Note that violation of the InSIDE assumption leads to correlated pleiotropy, which can severely bias causal effect estimates. Such a phenomenon may emerge as a result of a heritable confounder of the exposure-outcome relationship and has been modelled in the past[9, 10].

Well-powered GWAS may also provide confounded genetic associations through dynastic effects[3, 11], assortative mating[12, 13], and population stratification[14]. These phenomena can introduce correlation between an instrument and confounding factors, such as parental/partner traits or genetic ancestry leading to a violation of the exchangeability assumption and biased causal effect estimates. This type of confounding can be resolved when using family-based study designs[15, 16] such as sibling-pair studies. Since genetic differences between sibling pairs are due to independent and random meiotic events, these effects are unaffected by population stratification and other potential confounders influencing the phenotype. This and other emerging family-based designs have been used to obtain unbiased heritability estimates, validate GWAS results and test for unbiased causal effect estimates using MR[17, 18].

Another factor that can lead to complications in MR studies is the presence of heterogeneous causal effects emerging due to distinct biological mechanisms: various subtypes of the exposure (e.g. subcutaneous vs visceral adiposity) or different biological pathways through which the exposure impacts the outcome (e.g. interaction between the exposure and other factors). To date, horizontal pleiotropy, confounding of genetic associations, and heterogeneous causal effect have been largely treated as distinct mechanisms in MR modelling. However, what they have in common is that they can lead to variable causal effects estimated depending on the group of IVs used in the MR.

To address this, we introduce in this paper our approach of PheWAS-driven clustering of instrumental variables (PWC-MR) and test the resulting clusters for distinct pathways or mechanisms that could underlie the overall causal effect of the exposure. Throughout the paper, we demonstrate the approach through the example of estimating the causal effect of body mass index (BMI) on educational attainment (EDU). This relationship has been analysed extensively in the past and family studies have shown that an apparent strong effect of higher BMI on lower educational attainment is shrunk to near zero when using family studies[17]. One explanation is that offspring BMI is influenced by parental alleles associated with parental (rearing) behaviour, which in turn modify the environment of the offspring. Such parental traits act as a confounder of the offspring genotype-EDU relationship, hence violate the exchangeability assumption of MR. Moreover, they confound the BMI-EDU association in the tested sample, violating the exclusion-restriction assumption and inducing correlated pleiotropy (see Figure 1). Thus, it is plausible that some of the detected IV clusters arise through parental genetic confounding which may manifest statistically as horizontal pleiotropy. To test this, we ran a systematic confounder search and probed the causal effect of the exposure conditional on candidate confounder traits.

2 Results

2.1 Overview of the method

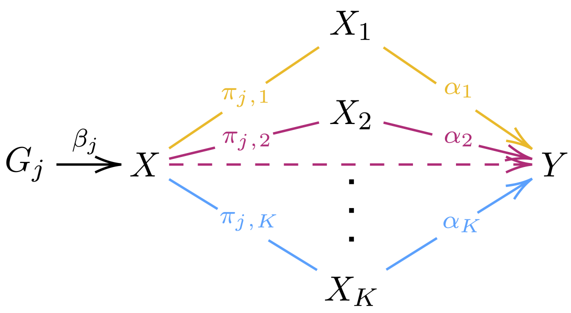
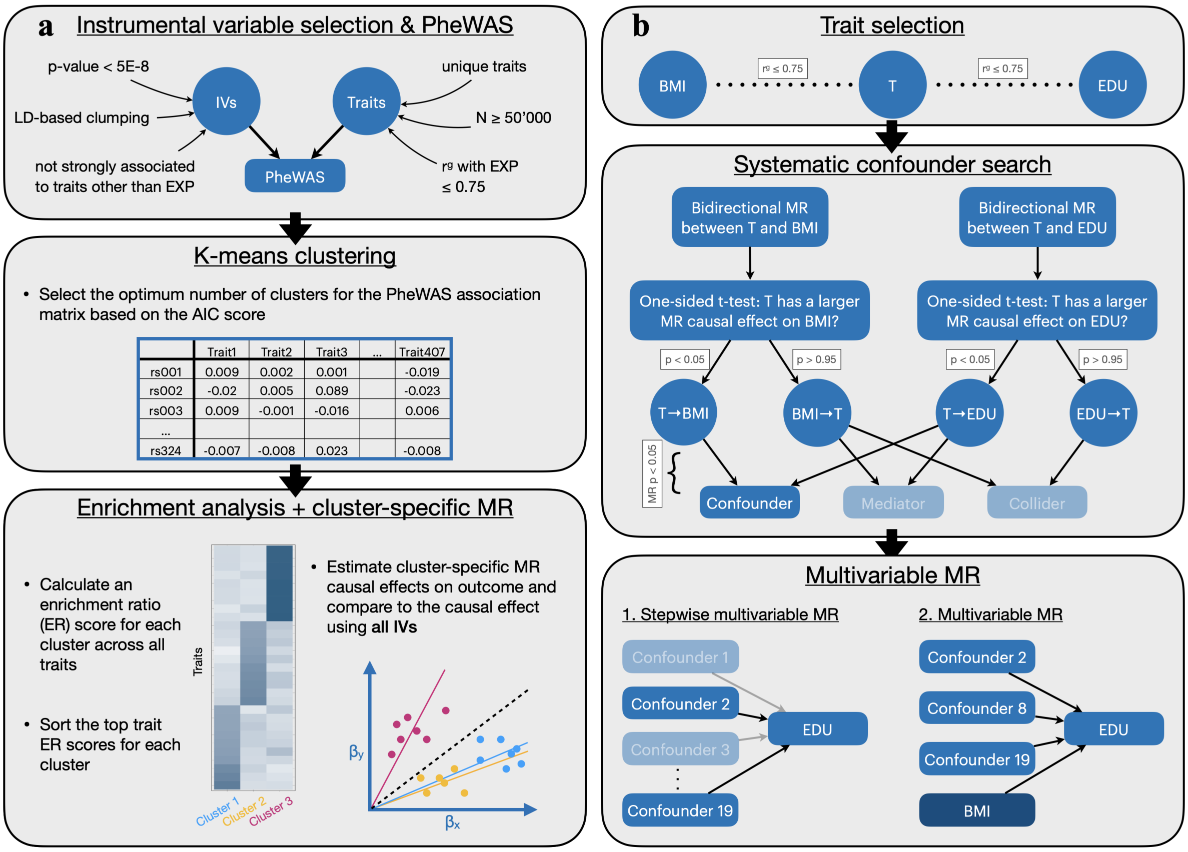


Figure 1: Directed Acyclic Graph (DAG) illustrating the complex relationship between exposure and outcome. G_j represents genetic instrument j with an effect β_j on exposure X. Exposure X is associated with outcome Y through K possible pathways of mediation or confounding denoted through the various $X_1 ... X_K$ elements. The associations between the main exposure and the various elements denoted by the π arrows purposely do not show directionality to allow for both mediators and confounders. The causal effects on outcome Y are denoted by $\alpha_1, \alpha_2, ..., \alpha_K$.

Horizontal/correlated pleiotropy, confounded genetic associations, and mechanism-specific causal effects all lead to heterogeneous MR causal effect estimates. In PWC-MR, we attempt to investigate all these possible biases simultaneously by informatively clustering the various IVs and testing the resulting groups for distinct pathways or mechanisms underlying the overall causal effect as illustrated in Figure 1.

We applied the PWC-MR approach to investigate potential horizontal pleiotropic effects (emerging due to heritable confounders, dynastic effects, genetic subtypes of obesity and other pleiotropic mechanisms, see Figure 1) of BMI on educational attainment. The analysis focused on grouping the IVs of the exposure by running a PheWAS-based clustering to reveal distinct mechanisms or pathways underlying their overall effect on the outcome (Figure 2**a**). This was done by obtaining the standardised PheWAS association of the BMI IVs across a filtered set of 408 traits, and running a k-means clustering on the resulting matrix. This resulted in six clusters of IVs for BMI, which were then annotated by traits based on the association of the clustered SNPs with each trait. Specifically, for each cluster-trait pair we computed the average explained variance of the trait by the SNPs of the given cluster. This yielded an enrichment ratio (ratio of the average explained variances) for each cluster-trait pair, and we chose the top ten traits with the highest enrichment ratio for each cluster as representatives. Furthermore, the causal effect of each cluster's IVs on education was calculated and compared against each other and that of the causal effect obtained using all BMI IVs.



- Select the optimum number of clusters for the PheWAS association matrix based on the AIC score

	Trait1	Trait2	Trait3	...	Trait407
rs001	0.009	0.002	0.001		-0.019
rs002	-0.02	0.005	0.089		-0.023
rs003	0.009	-0.001	-0.016		0.006
...					
rs324	-0.007	-0.008	0.023		-0.008

- Calculate an enrichment ratio (ER) score for each cluster across all traits
- Sort the top trait ER scores for each cluster
- Estimate cluster-specific MR causal effects on outcome and compare to the causal effect using **all IVs**

Figure 2: Flow diagram representing how the PWC-MR approach aims to disentangle causal effect between trait pairs from confounding or pleiotropy, as well as systematically search for confounders of the trait pair.
Panel **a** represents the main steps of the PWC-MR method: (i) Instrument selection and PheWAS; (ii) Informative IV clustering using K-means; and (iii) Enrichment analysis and cluster specific MR. Panel **b** represents a complimentary approach to PWC-MR where a systematic candidate confounder trait search is performed. These candidate confounder traits are defined as having an effect on both the exposure and the outcome. A stepwise multivariable MR (MVMR) of the candidate confounder traits is performed to select those with a strong effect on the outcome. These are then added with the primary exposure (BMI) to a standard MVMR and the multivariable causal effect on the outcome (EDU) is estimated. Acronyms: T - trait, p: t-test p-value; MR p: MR p-value.

To complement our findings from the clustering-based analysis, we explored (i) the BMI-EDU causal relationship using sib-regression SNP effect sizes[18], (ii) the childhood BMI-EDU causal relationship, (iii) replacing the outcome trait with systolic blood pressure (SBP), and finally (iv) the potential role of each of the filtered set of traits as a confounder of the BMI-EDU relationship.
We implemented the latter one by systematically running bidirectional MR between each of the traits and either BMI or EDU as outcome, then classifying the traits depending on their bidirectional associations with both BMI and EDU. The resulting set of candidate confounder traits was further analysed for its potential to bias the causal effect of BMI on EDU. To assess this, we ran stepwise MVMR and finally calculated the causal effect of BMI on EDU conditional on the surviving set of candidate confounder traits of the BMI-EDU relationship (illustrated in Figure 2**b**).

To further understand the emerging clusters, we sought to uncover tissue-specific mechanisms. To do this, we performed a colocalisation analysis of the BMI and gene expression association signals at each locus around (±400kb) the 324 BMI IVs. For the gene expression association we used eQTL data from both adipose and brain tissue. This yielded a proportion of brain-vs-adipose colocalised IVs for each cluster.

2.2 PheWAS-based K-means clustering and trait identification

After identifying 324 genome-wide significant SNPs as IVs for BMI, and selecting 407 filtered traits to run PheWAS on, we obtained a standardised effect matrix of the 324 IVs on the 407 traits. Normalising the matrix by IVs and running K-means clustering on it revealed that six clusters yielded the lowest AIC score (Supplementary Figure S1) when compared to varying the number of clusters from two to 50. The number of SNPs in each of the six clusters were: $32, 98, 35, 41, 69, 49$ respectively (Supplementary Table 2).

Next, we computed an enrichment ratio (ER) to identify with which traits the SNPs in each cluster were strongly associated. The overall ER value between clusters was roughly centred around 1, however clusters #2, #3, #4, and #6 had some large ER values (see Supplementary Figure S2). Visualising the top 10 enriched traits in each cluster and their ER values in Figure 3 and Supplementary Table 3, we see that cluster #2 is strongly enriched for lean mass traits such as 'Trunk fat-free mass' and 'Whole body fat-free mass'.
Similarly, cluster #3 seemed to mostly be enriched for blood- and body stature-related traits such as 'Platelet count' and 'Standing height', while cluster #4 was enriched for traits related to socio-economic position (SEP) such as 'Job involves heavy manual or physical work', 'Time spent outdoors in summer', and 'Fluid intelligence score'. Lastly cluster #6 was enriched for food supplements/nutrients such as 'Folate' and 'Potassium'.

2.2.1 Causal effect estimate per cluster

To test whether the clusters had different causal effects on a selected outcome than the overall causal effect (using all IVs), we computed the IVW causal effect estimate of each cluster on education using cluster-specific IVs. As seen in Figure 4**a** and Supplementary Table 4, the causal effect estimates between the different clusters are significantly heterogeneous (Q-test value = 130.61, p-value $< 10^{-300}$). Clusters #2 and #5 had the smallest causal effect estimates of -0.09 (p-value $= 1.23 \times 10^{-}5$) and -0.12 (p-value $= 5.22 \times 10^{-}6$) respectively, where cluster #2 was enriched for lean-mass traits. These estimates are consistent with those obtained from within-family studies, which are relatively immune to confounding (see section 2.3.1). By contrast, clusters #1 and #4 had the largest negative causal effect estimates of -0.44 (p-value $= 7.78 \times 10^{-}20$) and -0.49 (p-value $= 1.63 \times 10^{-}44$) respectively, where cluster #4 was strongly enriched for SEP-related traits.
All the clusters were less heterogeneous than the group of all the IVs combined (see 'Avg_het' in Supplementary Table 4).

2.3 Post hoc analyses

To test the robustness of the PWC-MR results, we performed four additional analyses. First, we used the same exposure and outcome, but the MR analysis was based on sib-regression-based SNP effect sizes instead of SNP effects from GWAS of unrelated samples. Second, we replaced the exposure with childhood BMI and estimated its causal effect on EDU. Third, we replaced the outcome, EDU, with SBP. Finally, we executed a systematic search for confounders to include in a multivariable MR analysis.

2.3.1 Sib-regression MR

In Howe et al.(2022)[18], within-sibship (within-family) meta-analysed GWAS estimates were generated from 178,086 siblings across 19 cohorts. Using these effect estimates, MR was performed with BMI as exposure on multiple traits, including educational attainment. They used 418 independent and genome-wide significant genetic variants for BMI, and estimated its effect

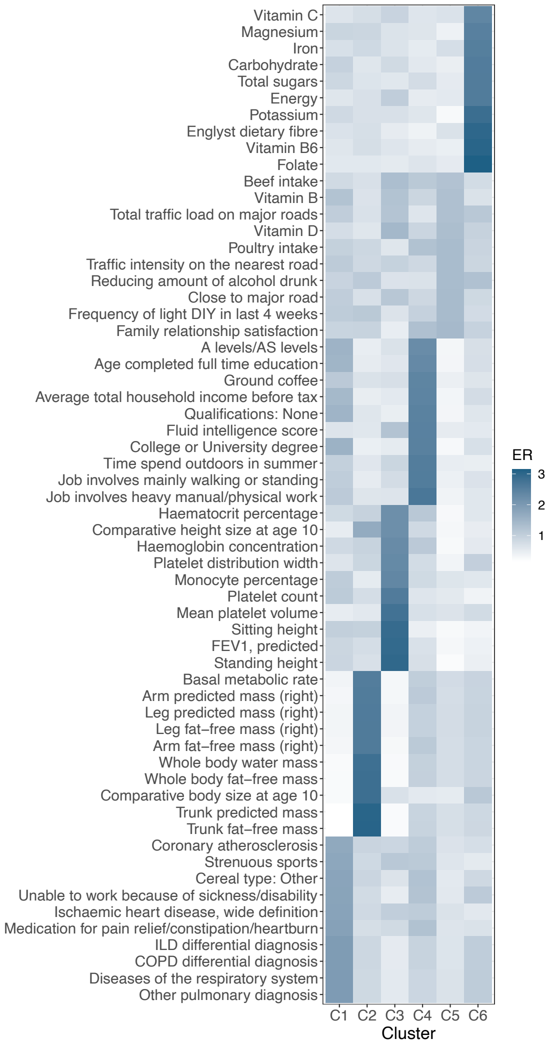


Figure 3: **Heatmap of the enrichment ratio of the top 10 traits in each cluster.** K-means clustering of BMI revealed six clusters with the following trait enrichment ratios.

on EDU using IVW to be -0.05 (95% CI: $-0.09, -0.01$).

They also used jackknife to estimate the standard error of the difference between the sib-regression MR estimate and that of the GWAS of unrelated samples MR estimate: -0.19 (95% CI: $-0.22, -0.16$). Using the difference Z-score to generate a p-value for heterogeneity between the two estimates revealed a significant difference with a p-value < 0.001.

2.3.2 Causal effect of childhood BMI on Educational attainment

We used the UK Biobank trait 'Comparative body size at age 10' as a proxy for childhood BMI – a measure that has been validated against measured BMI in childhood[19, 20] – for the exposure trait. Childhood BMI is presumed to be less influenced by SEP compared to adult BMI and hence we expect the causal effect estimate on EDU to have less confounding bias. For this trait, we had 171 genome-wide significant SNPs that we used as IVs for the analysis. Of these, 16 SNPs were more strongly associated to traits other than childhood BMI and were thus excluded from further analysis. The standardised effect matrix of the remaining 155 SNPs across 461 traits was normalised with respect to the SNPs, and then clustered into four clusters (yielding optimal AIC), each containing $37, 42, 32, 44$ IVs, respectively (Supplementary Figure

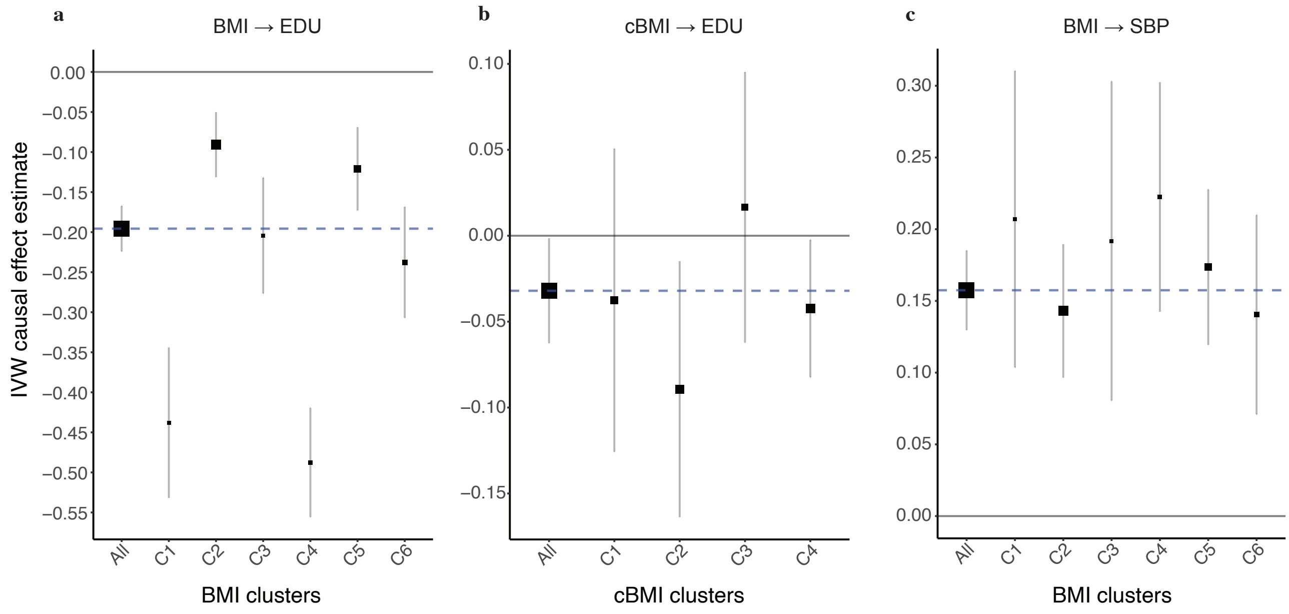


Figure 4: **Forest plot of IVW causal effect estimate on outcome using either all exposure IVs (All) or cluster-specific IVs (C1..C4/C6).** Panel **a** shows causal effect estimates of adult BMI on EDU, panel **b** proxy of childhood BMI (cBMI) on EDU, and panel **c** adult BMI on SBP. Horizontal error bars represent the 95% confidence interval. The blue vertical line represents the causal effect estimated using all BMI/cBMI IVs. Box sizes of clusters represent the proportion of the number of IVs in each cluster to the total.

S3, Supplementary Table 5).

Analysing the trait enrichment for each cluster revealed only two clusters with high ER values: clusters #2 and #4 (Supplementary Figure S4, Supplementary Table 6). Cluster #2 had only two traits with ERs greater than 2, which were 'Number of fluid intelligence questions attempted within time limit' and 'Fluid intelligence score', whereas cluster #4 was highly enriched for body-measurement/fat-mass traits such as 'Waist circumference' and 'Whole body fat mass' (see Supplementary Figure S5). However, calculating the IVW causal effect estimate for each cluster and comparing it to the estimate calculated using all IVs revealed homogeneous causal effect estimates with a Q-statistic of 3.84 (p-value of 0.43) as seen in Figure 4**b** and Supplementary Table 7. Cluster #2 had a causal effect estimate of -0.09 (95% CI:-0.1638, -0.0148), and cluster #4 had a causal effect estimate of -0.04 (95% CI:-0.0823, -0.0024). Noteworthy is the finding that the IVs of cluster #2 were more heterogeneous than all the IVs combined. Thus, we obtained a massively attenuated causal effect of BMI on EDU, when childhood BMI is used as an exposure. Reassuringly, no strongly SEP-enriched cluster emerged and the cluster specific causal effects were homogeneous.

2.3.3 Causal effect of BMI on SBP

To find further evidence that our approach does not always reveal distinct causal effects when the causal effect is non-null, we replaced EDU with SBP as outcome. Namely, we tested a well-established non-null causal relationship that is hypothesised to not be biased by pleiotropy or confounding: BMI's effect on SBP. Using the same six clusters previously obtained for BMI, we calculated the estimated causal effect of each of the clusters compared to using all the IVs combined on SBP. This revealed a homogeneous set of causal effect estimates (Q-test value of 4.49, p-value = 0.61), with the IVW estimate using all IVs being 0.15 (p-value = $1.09 \times 10^{-}28$) as seen in Figure 4**c** and Supplementary Table 8.

2.3.4 Systematic confounder search and MVMR analysis

Given our suspicion that the large BMI-EDU causal effect is driven by heritable confounders, we performed a systematic search to reveal traits that may be potential confounders. As described in the Methods section, the strength of the bidirectional effect of the traits on either the exposure or the outcome determined their categorisation. This led to the identification of 19 traits that were found to be candidate confounder traits (Supplementary Table 9). Matching the 19 confounder traits from this analysis to their respective ERs across the six clusters from the previous analysis revealed higher ERs in cluster #1 and cluster #4 (associated with SEP-related traits), both of which also had the largest negative causal effects on EDU (Supplementary Figure S6). It is worth noting that the traits labelled as candidate confounders were predominantly environmental exposures, such as 'Exposure to tobacco smoke outside home' and 'Transport type for commuting to job workplace: Cycle'.
Furthermore, these candidate confounder traits are attributed as *candidate* or *potential* confounders since they are most likely only genetic correlates of the true confounding traits of the BMI-EDU relationship and do not act as true confounders themselves.

To investigate the possible biasing effect that potential confounder traits can have on the causal relationship of BMI on EDU, we ran a stepwise MVMR on these 19 candidate confounder traits (Supplementary Table 9). During the creation of the Z-score matrix of SNPs and traits, only twelve traits had at least three genome-wide significant and independent SNPs whose effects could be used in the analysis, leaving us with a total of 683 SNPs across these twelve traits and BMI. The twelve traits were: 'Time spent watching television (TV)', 'Usual walking pace', 'Past tobacco smoking', 'Cereal type: Muesli', 'Frequency of tiredness / lethargy in last 2 weeks', 'Frequency of depressed mood in last 2 weeks', 'Public transport', 'Walking for pleasure', 'Weekly usage of mobile phone in last 3 months', 'Eating eggs, dairy, wheat, sugar', 'Symptoms, signs and abnormal clinical and laboratory findings', and 'Average weekly beer plus cider intake'. Of these, only the first four had a significant causal effect estimate on EDU ($p\text{-value} < 0.05/12$) based on stepwise MVMR, and were subsequently used as exposures alongside BMI in a standard MVMR analysis.

To ensure the strength of the IVs used in the MVMR analysis, we calculated the conditional F-statistic and the MVMR causal effect estimate of BMI given various combinations of the four remaining candidate confounder traits. We saw the expected trend of a decreasing conditional F-statistic with the addition of traits and their IVs to the analysis (see Supplementary Figure S7). We note that the causal effect estimate of BMI on EDU decreases when any combination of the candidate confounder traits is used with BMI as exposure in comparison to the univariable MR causal effect estimate of BMI on EDU (-0.19, p-value $= 7.11 \times 10^{-}41$). We settled on the combination of candidate confounder traits yielding a conditional F-statistic for BMI of 10.19, for which the corresponding causal effect estimates are reported in Table 1 below. This choice was a compromise between two sources of biases: weak instrument bias *vs* upward bias due to omitting relevant confounders.

Trait	Description	α estimate	SE	P-value
1070	Time spent watching television (TV)	-0.2771	0.0256	4.63E-25
1249	Past tobacco smoking	0.1592	0.0218	7.85E-13
1468_4	Cereal type: Muesli	0.2930	0.0383	7.96E-14
21001	Body mass index (BMI)	-0.0455	0.0106	2.07E-05

Table 1: MVMR analysis results of BMI and three candidate confounder traits on education. α: causal effect estimate.

2.4 Relationship with other approaches

2.4.1 Comparison against MR-Clust

Other known IV clustering methods include MR-Clust[21], which attempts to cluster variants with similar causal effect estimates together following the hypothesis that exposures can affect an outcome by distinct causal mechanisms to varying extents. MR-Clust also accounts for the possibility of spurious clusters by assigning IVs with uncertain causal effect estimates to 'null' or 'junk' clusters.

We compared the PWC-MR clustering of BMI IVs against that of MR-Clust with EDU as the outcome. The MR-Clust results revealed two main clusters as well as a 'null' cluster. Cluster #1 had 35 SNPs, 13 of which had an inclusion probability greater than 80%. Cluster #2 had 171 SNPs, 36 of which had an inclusion probability greater than 80%, and the remaining 142 SNPs were categorised into the 'null' cluster as seen in Supplementary Figure S8. The mean causal effect estimate of SNPs in cluster #1 was -0.496, whereas it was -0.246 for cluster #2. Searching for trait associations for the SNPs in each of the clusters revealed that body measurement traits like 'Arm fat mass' or 'Body fat percentage' are associated to SNPs in both clusters, while SEP-related traits such as 'Fluid intelligence score' or 'Time spent watching television' were associated to more SNPs in cluster #1 than in cluster #2.

Comparing the SNP clustering between the PWC-MR method against that of MR-Clust in Table 2 below, we see that cluster #1 in MR-Clust, which seems to be more strongly enriched for SEP traits than cluster #2, has SNPs that were similarly clustered in clusters #1 and #4 using PWC-MR, matching their large negative causal effect of BMI on EDU. However, the same distinct comparison cannot be made for SNPs in cluster #2 of MR-Clust.

Of the 12 Fisher's exact tests performed to examine the contingency of SNPs in the two separate sets of clusters, only four tests revealed a significant association: SNPs in cluster #1 of MR-Clust were significantly associated with SNPs in clusters #1, #2 (lean-mass traits), #4 (SEP-related traits) and #5 of the PWC-MR clustering.

Given the differences between the two methods (where PWC-MR performs informative clustering of IVs based on external data, and then measures the MR causal effect estimates per cluster compared to MR-Clust that clusters IVs based on the magnitude of their MR causal effects) we see a more biologically meaningful separation of SNPs using PWC-MR shedding light on the various mechanisms through which BMI can act on EDU.

PWC-MR MR-Clust	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6
Cluster1	13	1	0	13	0	5
Cluster2	15	38	21	26	32	29
Null	4	59	14	2	37	15

Table 2: **Cross table of BMI IVs clustered using PWC-MR and MR-Clust.**

2.4.2 Colocalisation analysis

With the aim of finding supporting evidence for the k-means clustering and enrichment analysis, we ran a genetic colocalisation analysis on BMI IVs and two types of tissue: subcutaneous adipose and brain, the results of which can be found in Supplementary Tables 12 and 13 respectively.

	Cluster1	**Cluster2**	**Cluster3**	**Cluster4**	**Cluster5**	**Cluster6**
Adipose	9	9	14	3	6	5
Brain	3	3	4	1	2	4
Both	1	2	1	1	4	4
Neither	29	77	36	23	53	47

Table 3: Cross table indicating the number of genes whose expression colocalises in adipose/brain tissue with BMI. The colocalisation exercise was performed at loci-defined BMI IVs falling into particular clusters. Colocalisation was defined as the posterior probability of both GWAS and eQTL being associated is ≥ 0.8 in either brain or adipose tissue or both.

Running a set of Fisher's tests to compute the overlap between the membership of the SNPs in the six clusters and their tissue of colocalization did not reveal any association between clusters and tissues.

3 Discussion

We have developed a method that performs informative clustering of IVs by utilising their association with a large number of traits. Our use of PheWAS data to guide the clustering of IVs has revealed distinct mechanisms by which exposure effects could act on outcomes. For our exposure, BMI, six distinct clusters were obtained through optimal K-means clustering. These clusters had well-defined trait enrichments, with clusters matching SEP-related, substrate, and body measurement traits. Estimating individual causal effects of each cluster on EDU as an outcome revealed heterogeneous causal effect estimates which allowed us to further strengthen our suspicion that the MR estimate for the causal effect of BMI on EDU is upward biased when using population-based SNP effect size estimates due to confounding.
We note from MR analysis run using within-sibling GWAS data[18] that the causal effect estimate between BMI and EDU is -0.05 (95% CI: $-0.09, -0.01$), which is smaller than the causal effect estimate seen using population based GWAS data (-0.19, 95% CI: $-0.22, -0.16$). Investigating the various mechanisms or pathways through which BMI could have a causal effect estimate on EDU through trait-enrichment analysis has revealed notable causal effect estimates from two clusters: one with a strongly negative MR estimate whose trait enrichment reflects shared mechanisms with socio-economic factors, and another cluster with close to zero causal effect estimate enriched for lean-mass traits. MR has typically presented bias due to heterogeneous causal effects emerging via distinct pathways, and bias due to confounding of the instrument-outcome association as being separate mechanisms. Here, we have illustrated that a pheWAS-based clustering approach can classify instruments into clusters, some of which correspond to different pathways, while others include IVs that are primarily confounder-associated. Our results have two major implications: 1) The lean-mass-related IV cluster indicated a more plausible, close to zero causal effect of BMI on EDU. 2) We revealed that the SEP-related IVs leading to an apparent, sizeable negative effect of BMI on EDU, possibly overestimating the true underlying causal effect.

In order to substantiate our findings, we performed several follow-up analyses. First, sib-regression based MR of BMI on EDU recapitulated the close-to-zero causal effect obtained for the body-mass specific cluster of IVs. This indicates that many IVs for adult BMI (from population-based GWAS) represent indirect (parental/dynastic) effects, which are associated rather with a rearing-related parental trait and not primarily with offspring BMI. Second, replacing adult BMI with childhood BMI (much less associated with SEP) as exposure in the PWC-MR analysis confirmed a negligible causal effect estimate (-0.03, p-value = 0.04), and the four emerging clusters showed homogeneous causal effect estimates indicating the lack of confounding or biasing effects. This comparison was supported by the growing evidence showing that genetic variants have varying effects on BMI or body size at different stages of life[22, 23], and that the UK Biobank proxy trait 'Comparative body size at age 10' captures childhood BMI well[19]. Noteworthy is the fact that the childhood BMI proxy we use is a coarsened trait in comparison to true childhood BMI, and thus the true MR causal effect estimate is likely to be overestimated. We have explored this further with our own simulation in Supplementary Methods 1.2. One of the four clusters was strongly enriched for body-measurement/fat-mass traits whereas the second most strongly enriched cluster had only two mildly enriched SEP-related traits. This finding means that as opposed to adult BMI, childhood BMI genetics are unrelated to childhood (i.e. parental) SEP. Furthermore, out of the 41 adult BMI IVs that make up cluster #4 (SEP-related traits), only three were found to be in LD with childhood BMI IVs.

In Howe et al. (2022), assortative mating, dynastic effects and population stratification were all considered candidate mechanisms for biased population-based GWAS effect estimates. Given our observations, a possible explanation is a dynastic effect of parental SEP traits acting as a confounder on the offspring's BMI and EDU in adulthood (as seen in Supplementary Figure S9). This effect is direct on the offspring's adulthood EDU but could affect the offspring's adult BMI indirectly through either of two ways: (i) Parental SEP has a direct effect on the offspring's SEP as an adult, which in turn has an effect on offspring adult/late BMI[24], or (ii) parental SEP – as a determinant of childhood social circumstances – may have an effect through this on the offspring's adult BMI.

To explore the relevance of the obtained six clusters of IVs, we replaced EDU with SBP as the outcome of interest since within-sibling GWAS MR results showed no difference when compared to population GWAS MR results, indicating that there seems to be no bias in the causal effect estimate due to pleiotropy or confounding. Our analysis revealed that for the six clusters attributed to BMI, their causal effect estimate on SBP was homogeneous with the estimate using all SNPs (0.16, p-value = $1.09 \times 10^{-}28$). As there is no significant heterogeneous effects and all the cluster causal effects agree, we can conclude that there is no other confounding effects biasing the causal effect estimate. It is reassuring to note that our PWC-MR approach does not always seek to identify distinct causal effects, confirming that confounding mechanisms are specific to certain exposure-outcome pairs.

Finally, our systematic confounder search coupled with stepwise MVMR has pinpointed TV watching, muesli eating, and past tobacco smoking as three candidate confounder traits that could bias standard MR analysis of the BMI-EDU relationship: upon accounting for these three traits, BMI exhibits a strongly attenuated causal effect on EDU, comparable to that of cluster #2 and the sib-regression MR estimate. We acknowledge the fact that past tobacco smoking is unlikely to have an effect on EDU retroactively, similar to TV watching and other later-in-life traits, which we all consider to be acting as confounder-proxies or correlates of parental SEP. We have explored this further in Supplementary Methods 1.3 by introducing 'Smoking Initiation' into the candidate confounder traits.

Comparing our method to other IV clustering methods such as MR-Clust does not reveal strong concordance in the findings. MR-Clust takes as input the association effects of the exposure and outcome as well as their association standard errors and attempts to cluster the exposure IVs based on the possible similarity between each IV's causal effect on the outcome. When using BMI and EDU as exposure and outcome respectively, MR-Clust revealed two main clusters alongside a null cluster. Both of the clusters were enriched for a variety of traits including body-measurement traits, both lean- and fat-mass, as well as SEP-related traits. The causal effect estimates of both clusters were strongly negative, similar to using all IVs in an MR analysis for this trait pair.
The most apparent difference between the clustering of our method and that of MR-Clust is our use of external information (PheWAS data of the exposure IVs and various other traits) to reveal possible pathways and mechanisms through which the exposure manifests, independently of any outcome. While MR-Clust clusters the individual MR causal effects of IVs on a specific outcome based on their magnitude.

Another comparable clustering method by Grant et al.[25] uses genetic variant associations with a set of traits to identify groups of IVs with similar biological mechanisms. Their method, NAvMix, uses a directional clustering algorithm and includes a noise-cluster to increase robustness to outliers. NAvMIX is demonstrated on BMI IVs and their associations to nine lifestyle or cardio-metabolic traits that have been previously shown to be related to BMI. Their results revealed 5 distinct clusters where they were able to identify a metabolically healthy obesity cluster that also had a small MR causal effect on coronary heart disease (CHD). However, we were unable to run their method using our data due to convergence issues arising when the number of traits used for PheWAS association increases. This comparison also highlights that the traits we include in the pheWAS analysis (and the subsequent clustering) have an important role in which biological mechanisms we can detect. For example, our analysis did not pick up the metabolically healthy obesity cluster, potentially because waist-to-hip ratio and other subcutaneous-vs-visceral fat proxy-traits were not included among the 407 selected phenotypes due to our filtering on genetic correlation with BMI ($r^g < 0.75$). Without such filtering, PWC-MR reveals 5 clusters with significantly heterogeneous causal effects on EDU. These five clusters are very similar to the original six, with the original cluster #1 getting diffused into the other clusters. Reassuringly, the cluster that is strongly enriched for SEP-related traits has a large negative causal effect estimate of -0.53 (95% CI: $-0.59, -0.48$), whereas the cluster that is most enriched for body-measurement/fat-mass traits still had an attenuated causal effect of -0.10 (95% CI: $-0.14, -0.06$).

Furthermore, we attempted to consolidate our findings of the k-means clustering and enrichment analysis by running a genetic colocalisation analysis on the 324 clustered BMI IVs and both subcutaneous adipose and brain tissue. Unfortunately, we do not find an association between the cluster memberships of the IVs and their signal colocalization in brain or adipose tissue, possibly due to high false negative rates of colocalization combined with low eQTL sample sizes.

Our method has its own set of limitations: first, we are limited by the availability of traits with PheWAS data to support our informative clustering of IVs. This may lead to a failure in identifying key pathways and thus missing clusters representing important subgroups (mediator/sub-phenotype/confounder). Second, although it is not the most ideal handling of data, our binary traits are treated as continuous ones in our analysis. In large samples, linear and logistic regression effect estimates correlate very strongly and hence, it is likely that this choice did not impact the clustering[26]. Third, although we have attempted to minimise the arbitrary choice of parameters in our analysis, the genetic correlation threshold that determines which traits are too similar to the exposure and outcome trait is arbitrarily set at 0.75 for BMI and EDU and could be modified, but the emerging clusters may change as a consequence. Similarly, some p-value thresholds and type I error rate control was set at 5%, which may be viewed as arbitrary. Fourth, the identified potential confounder traits used in the MVMR analysis act as simple proxies for true confounders. For example, exposure to current tobacco smoking or TV watching can be highly (genetically) correlated to the same or a similar exposure during early life (or even proxy a parental trait), hence it is rather the earlier version of the exposure which is likely to be the true confounder. Our proxy confounders were simply nuisance variables, their only role was to see the remaining causal effect of BMI on EDU upon conditioning on them. Fifth, while for the BMI-EDU relationship we had several lines of evidence pinpointing cluster #2 as the one yielding the most likely correct causal effect estimate, in general, we might not be able to decide which cluster(s) provide biologically meaningful causal effect estimate(s) and which ones may be linked to confounders. Lastly, we acknowledge that there are several other tests[27] that could be used in place of a t-test when excluding SNPs more strongly associated to other traits than our exposure or different MR methods used in our systematic confounder search, however both of these were simple exclusion or pre-selection steps that have very little impact on the outcome of the results.

To conclude, we found that the classical MR estimate based on population GWAS leads to an overestimation of the BMI-EDU causal effect and identified an lean-mass-specific subgroup of IVs (cluster #2) that, we believe, yields a much more reliable causal effect estimate. Still, we are uncertain whether this effect is exactly zero, or is just strongly attenuated. Our analysis also revealed that the unrealistically large standard MR estimate was driven by IVs that likely violate the pleiotropy assumption via being also linked to SEP. The attenuated estimate provided by our

PWC-MR approach (cluster #2) is compatible with both the estimate based on sib-regression summary statistics (P-values difference = 0.161) and the MVMR estimate (p-diff = 0.476), all of which are based on adulthood phenotypes. However, the estimate obtained for childhood BMI is slightly more attenuated than that of the PWC-MR method (p-diff 0.024).

Equipping the MR toolkit with a range of different analytical strategies is critical for improving insights into epidemiological questions, and PWC-MR offers a number of features that compliment other approaches: (i) it does not require summary statistics from within-family GWAS, which are typically scarce and available in much smaller samples and for a limited set of phenotypes (ii) it does not rely on association data from an early exposure, which face similar limitations as within-family GWAS (iii) in contrast to MVMR, which estimates a single causal effect, PWC-MR provides multiple causal effect estimates, some of which may reflect confounder effects, and others heterogeneous mechanisms of action, overall revealing biological insight that can be used in follow-up research.

4 Methods

4.1 Instrumental variable selection and PheWAS

As our primary analysis, we aimed to investigate the potential pleiotropy-patterns emerging from the grouping of IVs that are strongly associated with an exposure of interest, as outlined in Figure 2**a**. With BMI selected as the exposure trait, we obtained IVs from the Neale group's UK Biobank GWAS analysis[28] (data sources can be found in Supplementary Table 1) by filtering for genome-wide significant SNPs (i.e. association p-value less than 5×10^{-8}) followed by linkage disequilibrium (LD)-based clumping using the TwoSampleMR R package[29] with the following parameters: $clump\_kb = 10,000$, $clump\_r2 = 0.001$, $pop =$ "EUR" to obtain independent IVs.

This left us with 348 BMI-associated IVs, for which we ran PheWASs with $1,480$ traits from the Neale group UK Biobank GWAS analysis[28]. We extracted for each trait and for each SNP the association effect and the corresponding standard error, creating a data matrix of 348 SNPs by $1,480$ traits. For the $1,480$ traits, we also extracted details such as variable type, origin and complete sample size, among others.

4.1.1 Quality control

We removed traits from the PheWAS data matrix that had missing association effects as well as duplicates (keeping the most recent version). Furthermore, we filtered out traits for which the effective sample size was less than $50,000$ due to their low power of association, leaving us with 424 traits.
Using genetic correlation data from the Neale group[28], we further removed traits that had a high genetic correlation with BMI, i.e. the exposure, ($r^g > 0.75$), to avoid obvious repetitions of traits closely related to it. The resulting association effect data matrix of 348 SNPs and 407 traits was then standardised (SNP effects are on a SD/SD scale) and used for further analysis. Note that for simplicity, effect sizes for binary traits were treated as those of continuous traits.

In order to test for invalid IVs, we performed a trait-wide variant of Steiger-filtering[30]. Specifically, for each SNP, we tested if any of the traits had a significantly stronger (in terms of explained variance) association compared to that of the exposure. The significance threshold for this one-sided t-test was corrected for using the total number of traits remaining (p-value $< 0.05/407$). This revealed 24 SNPs more strongly associated to traits other than BMI (such as 'Whole body water mass', 'Basal metabolic rate' and 'Sitting height') that were then removed from further analysis.

4.2 K-means clustering and trait identification

With the aim of discovering distinct meaningful groups of SNPs among the 324 IVs, we proceeded with the clustering of the SNPxTrait association effect matrix using the K-means algorithm[31]. Taking the absolute standardised effects matrix, we normalised the data frame with respect to the SNPs such that the variance of the SNP effects across all the traits equalled 1. We used the absolute effects to cluster, in order to ensure that negatively correlated traits were considered similar by the Euclidean distance based similarity measure of the k-means clustering. We then compared the performance of the clustering with different number of clusters ranging from two to 50, by measuring the Akaike Information Criterion (AIC) score (for further model selection criteria, see Supplementary Methods 1.1). After finding the number of clusters with the lowest AIC score (six clusters), we proceeded with the assignment of each SNP to one of the six clusters.

In order to identify traits that were particularly associated to SNPs in each of the six clusters, we computed an enrichment ratio (ER) in the following way:
For each trait t, we calculated the per-SNP average squared effect in a given cluster j, denoted as $\sigma^2_{j,t}$. Given that SNP i belongs to cluster j, $\sigma^2_{j,t}$ was calculated as follows:

$$\sigma^2_{j,t} = \frac{1}{|c_j|} \sum_{i \in c_j} \beta^2_{i,t} \tag{1}$$

where $\beta^2_{i,t}$ represents the squared standardised effect of SNP i on trait t (not normalised across traits), c_j represents the set of SNPs in cluster j and $|c_j|$ its cardinality. We then normalised these per-SNP average squared effects for each cluster relative to the total effect across all clusters (K) to obtain the enrichment ratio (ER), $R_{j,t}$:

$$R_{j,t} = \frac{\sigma^2_{j,t}}{\frac{1}{K} \sum_{k=1}^{K} \sigma^2_{k,t}} \tag{2}$$

where K is the total number of clusters. For each cluster (j), traits were then prioritised by the (highest) value of ER ($R_{j,t}$).

4.2.1 Causal effect estimate per cluster

We measured the cluster-specific IVW causal effect estimate on the outcome (EDU) using the standardised SNP effects in each cluster, and then compared these estimates to the causal effect estimate using all SNPs. We used the TwoSampleMR R package[29] for this analysis, and although we use two-sample MR techniques despite having a close to complete sample overlap, this does not lead to substantial biases[32]. Measures of IV heterogeneity were calculated using the Cochran's Q-statistic[33] for the IVW method for each cluster. Furthermore, average cluster-heterogeneity (per-IV variance) was also calculated for each cluster from the above-mentioned parameter.

As sensitivity analyses, PWC-MR was repeated twice, once with a different exposure trait (replacing BMI with childhood BMI), and another with a different outcome trait (replacing EDU with systolic blood pressure).

4.3 Systematic confounder search

In order to decide which of the emerging clusters represent genetic confounding or true biological heterogeneity, we systematically searched for BMI-EDU confounders. To do this, we investigated the bi-directional causal effects that each trait had on both the exposure and the chosen outcome.
Firstly, an extra filtering step was done where traits that were highly genetically correlated with the outcome ($r^g > 0.75$) were removed from the total 407 traits of the previous analysis.
Then, we ran a bidirectional MR for the remaining traits using the TwoSampleMR R package[29], and obtained four sets of causal effect measurements per trait (bidirectional, two different outcome traits - BMI and EDU). To select bidirectional causal effect estimates from those calculated by the different methods in the TwoSampleMR package[29] (Weighted median, Inverse variance weighted, Simple mode, and Weighted mode), we ordered the p-values of the causal effect estimates for the four different methods and selected the estimate of the second most significant method to ensure that at least one other method supports the causal claim.

The next step was to identify the direction of causality. To do so, we performed a one-sided t-test to compare the strengths of the estimated causal effects between the trait and the exposure, BMI. More precisely,

$$t_{A,B} := \frac{|\widehat{\alpha}_{A\to B}| - |\widehat{\alpha}_{B\to A}|}{\sqrt{SE^2_{A\to B} + SE^2_{B\to A}}} \tag{3}$$

where A and B denote the examined traits, $\widehat{\alpha}_{A\to B}$ the causal effect estimate from A on B and $SE_{A\to B}$ the corresponding standard error. The one-sided P-value is then calculated as $P = \Phi(t_{A,B})$: if $P < 0.05$ the $B \to A$ causal effect is nominally significantly larger, while if $P > 0.95$, the $A \to B$ direction is dominant. For all the p-values in between, it was challenging to assign a direction in which the causal effect was stronger, and thus these traits were not further categorised. The p-value thresholds we apply are not intended to suggest that there is a transition point at which the meaning of associations change, rather we use these as a heuristic that is required to control type I error rate at an arbitrary (5%) threshold. We further tested varying one-sided p-value thresholds of more stringent ($P < 0.01$, $P > 0.99$) and more lenient nature ($P < 0.1$, $P > 0.9$), the results of which are found in Supplementary Tables 10 and 11.

The same procedure was repeated to explore the relationship between the traits and the outcome trait (EDU). This allowed us to classify traits into candidate confounders, mediators, colliders and other categories (as seen in the middle panel of Figure 2**b**). For example, a confounder was defined as a trait with a significantly larger effect on both exposure and outcome than the reverse. We then focused only on the confounders which can distort MR estimates and filtered them further to make sure that they have at least a nominally significant MR estimate (p-value < 0.05) on both BMI and EDU. We were lenient in our categorisation of candidate confounder traits as adding potentially irrelevant traits would not bias the multivariable causal effect of BMI in the next step. As our aim was not to reduce the total causal effect to the unmediated part (possible by including mediators in an MVMR) but to correctly estimate it, mediators were not considered further. Similarly, the inclusion of colliders into an MVMR does not alter the exposure's causal effect as previously seen[34], thus they too were not considered further. The same holds for traits with a direct effect on either the exposure or the outcome only.
Furthermore, to test how compatible the two lines of analysis were, we examined the cluster-specific enrichment ratio values for the set of candidate confounder traits we obtained.

4.3.1 Multivariable MR

Focusing on the candidate confounder traits resulting from the systematic search that could bias the causal effect estimate between the exposure-outcome pair, we first ran a stepwise multivariable MR (MVMR) (adapted from the bGWAS R package[35]) with them as exposures to test their effect on our chosen outcome, EDU.
To do this, we created a Z-score matrix combining all genome-wide significant SNPs (p-value less than 5×10^-8) and their Z-scores for each of the 19 candidate confounder traits and BMI, such that each SNP had an effect that is genome-wide significant for at least one of the candidate traits.
To obtain independent SNPs, we performed rank-based clumping. For this, we first ranked the absolute Z-scores across all SNPs for each trait (in descending order), and then for each SNP we obtained the highest (best) rank across traits, which was used as an importance score during the clumping process (LD-clumped $clump\_kb = 5,000$, $clump\_r2 = 0.01$). We then further filtered out traits that had less than three instruments remaining. Note that any SNPs that fall in the HLA region (6p21.3) were removed for being strongly associated with multiple immune-related traits.
Using this Z-score matrix without our primary exposure (BMI) as input for step-wise MVMR, we obtained a final list of candidate confounder traits with significant multivariable causal effects (p-value $< 0.05/12$) on our chosen outcome (EDU).

Then, to minimise weak instrument bias when running MVMR, we calculated the conditional F-statistic for our primary exposure (BMI) given each of the surviving traits and their different combinations. Finally we ran standard MVMR using the combination of traits that produced a conditional F-statistic[36] ≥ 10 (for BMI), and examined the multivariable causal effect of BMI on EDU.

4.4 Relation to other approaches

4.4.1 Comparison against MR-Clust

We compared the k-means clustering of BMI IVs against another IV clustering method called MR-Clust[21], which requires as input the unstandardised SNP effects on both the exposure and the outcome, as well as the standard error of the SNP on each. To do so, we performed a Fisher's exact test to examine the frequency distribution of SNPs in each of the k-means clusters against the MR-Clust clusters.

4.4.2 Colocalisation analysis

To further interpret the findings of the IV clustering, we sought to test if specific patterns of colocalisation in different tissue types appear for the different IV clusters.

To do this, we reran the steps detailed in Leyden et al.[37] for the 324 BMI IVs used in this work. For each IV, we tested for genetic colocalisation between the BMI GWAS data and the gene expression (eQTL) data of both subcutaneous adipose and brain tissue (data sources can be found in Supplementary Table 1). For each SNP tested, we took a margin of 200kb up- and downstream, and used the coloc R package[38] to test the SNP's colocalisation with each gene found in that region, once using brain eQTL data, and another colocalisation using adipose eQTL data. We declared colocalisation if the posterior probability of the model sharing a single causal variant was larger than 80%. For each of the aforementioned clusters, we investigated if the IVs were more strongly enriched for or depleted in one tissue or the other using Fisher's exact test.

Acknowledgements

This research has been conducted using the UK Biobank Resource under Application Number 16389. Z.K. was funded by the Swiss National Science Foundation (310030_189147 and 32003B_173092). GH is funded by the Wellcome Trust and Royal Society [208806/Z/17/Z]. GH and GDS work within the MRC Integrative Epidemiology Unit at the University of Bristol (MC_UU_00011/1, MC_UU_00032/01). Payments were made to the institution. For computations, we used the CHUV HPC cluster. We are grateful for the useful discussions on MVMR with Eleanor Sanderson.

Author contributions

L.D. and Z.K. conceived and designed the project. Z.K. supervised all statistical analyses. L.D. implemented the research and performed the analyses. L.D. and Z.K. prepared the first draft of the manuscript. L.D., Z.K., G.H. and G.D.S. contributed to the review and editing of the manuscript.

Competing interests

The authors declare no competing interests.

Data availability

The origin and unique identifier of each of the summary statistics data used is referenced in Supplementary Table 1. The UK Biobank summary statistics data used in this study can be downloaded from `http://www.nealelab.is/uk-biobank`. BMI meta-analyzed GWAS and adipose meta-analyzed cis-eQTL can be obtained from Leyden et. al 2022[37], and the brain cis-eQTL data can be downloaded from `https://yanglab.westlake.edu.cn/software/smr/#DataResource`.

Code availability

The source code for this work can be found on `https://github.com/LizaDarrous/PheWAS-cluster`.

References

[1] Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., and Posthuma, D. (2021). Genome-wide association studies. Nature Reviews Methods Primers *1*, 59.

[2] Yengo, L., Vedantam, S., Marouli, E., Sidorenko, J., Bartell, E., Sakaue, S., Graff, M., Eliasen, A. U., Jiang, Y., Raghavan, S., et al. (2022). A saturated map of common genetic variants associated with human height. Nature *610*, 704–712.

[3] Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T. A., Bowers, P., Sidorenko, J., Karlsson Linnér, R., et al. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet *50*, 1112–1121.

[4] Sanderson, E., Glymour, M. M., Holmes, M. V., Kang, H., Morrison, J., Munafò, M. R., Palmer, T., Schooling, C. M., Wallace, C., Zhao, Q., et al. (2022). Mendelian randomization. Nature Reviews Methods Primers *2*, 6.

[5] Davey Smith, G. and Hemani, G. (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Human Molecular Genetics *23*, R89–R98.

[6] Burgess, S., Butterworth, A., and Thompson, S. G. (2013). Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol *37*, 658–665.

[7] Watanabe, K., Stringer, S., Frei, O., UmićevićMirkov, M., de Leeuw, C., Polderman, T. J. C., van der Sluis, S., Andreassen, O. A., Neale, B. M., and Posthuma, D. (2019). A global overview of pleiotropy and genetic architecture in complex traits. Nature Genetics *51*, 1339–1348.

[8] Bowden, J., Davey Smith, G., and Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through egger regression. Int J Epidemiol *44*, 512–525.

[9] Morrison, J., Knoblauch, N., Marcus, J. H., Stephens, M., and He, X. (2020). Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. Nat Genet *52*, 740–747.

[10] Darrous, L., Mounier, N., and Kutalik, Z. (2021). Simultaneous estimation of bi-directional causal effects and heritable confounding from gwas summary statistics. Nature Communications *12*, 7274.

[11] Young, A. I., Benonisdottir, S., Przeworski, M., and Kong, A. (2019). Deconstructing the sources of genotype-phenotype associations in humans. Science *365*, 1396–1400.

[12] Robinson, M. R., Kleinman, A., Graff, M., Vinkhuyzen, A. A. E., Couper, D., Miller, M. B., Peyrot, W. J., Abdellaoui, A., Zietsch, B. P., Nolte, I. M., et al. (2017). Genetic evidence of assortative mating in humans. Nature Human Behaviour *1*, 0016.

[13] Howe, L. J., Lawson, D. J., Davies, N. M., St. Pourcain, B., Lewis, S. J., Davey Smith, G., and Hemani, G. (2019). Genetic evidence for assortative mating on alcohol consumption in the uk biobank. Nature Communications *10*, 5039.

[14] Haworth, S., Mitchell, R., Corbin, L., Wade, K. H., Dudding, T., Budu-Aggrey, A., Carslake, D., Hemani, G., Paternoster, L., Davey Smith, G., et al. (2019). Apparent latent

structure within the uk biobank sample has implications for epidemiological analysis. Nat Commun *10*, 333.

[15] Davies, N. M., Howe, L. J., Brumpton, B., Havdahl, A., Evans, D. M., and Davey Smith, G. (2019). Within family mendelian randomization studies. Hum Mol Genet *28*, R170–R179.

[16] Benyamin, B., Visscher, P. M., and McRae, A. F. (2009). Family-based genome-wide association studies. Pharmacogenomics *10*, 181–190.

[17] Brumpton, B., Sanderson, E., Heilbron, K., Hartwig, F. P., Harrison, S., Vie, G. Å., Cho, Y., Howe, L. D., Hughes, A., Boomsma, D. I., et al. (2020). Avoiding dynastic, assortative mating, and population stratification biases in mendelian randomization through within-family analyses. Nature Communications *11*, 3519.

[18] Howe, L. J., Nivard, M. G., Morris, T. T., Hansen, A. F., Rasheed, H., Cho, Y., Chittoor, G., Ahlskog, R., Lind, P. A., Palviainen, T., et al. (2022). Within-sibship genome-wide association analyses decrease bias in estimates of direct genetic effects. Nature Genetics *54*, 581–592.

[19] Richardson, T. G., Sanderson, E., Elsworth, B., Tilling, K., and Davey Smith, G. (2020). Use of genetic variation to separate the effects of early and later life adiposity on disease risk: mendelian randomisation study. BMJ *369*.

[20] Brandkvist, M., Bjørngaard, J. H., Ødegård, R. A., Åsvold, B. O., Davey Smith, G., Brumpton, B., Hveem, K., Richardson, T. G., and Vie, G. Å. (2020). Separating the genetics of childhood and adult obesity: a validation study of genetic scores for body mass index in adolescence and adulthood in the HUNT Study. Human Molecular Genetics *29*, 3966–3973.

[21] Foley, C. N., Mason, A. M., Kirk, P. D. W., and Burgess, S. (2020). MR-Clust: clustering of genetic variants in Mendelian randomization with similar causal estimates. Bioinformatics *37*, 531–541.

[22] Alves, A. C., Silva, N. M. G. D., Karhunen, V., Sovio, U., Das, S., Taal, H. R., Warrington, N. M., Lewin, A. M., Kaakinen, M., Cousminer, D. L., et al. (2019). Gwas on longitudinal growth traits reveals different genetic factors influencing infant, child, and adult bmi. Science Advances *5*, eaaw3095.

[23] Richardson, T. G., Power, G. M., and Davey Smith, G. (2022). Adiposity may confound the association between vitamin d and disease risk - a lifecourse mendelian randomization study. Elife *11*.

[24] Blane, D., Hart, C. L., Davey Smith, G., Gillis, C. R., Hole, D. J., and Hawthorne, V. M. (1996). Association of cardiovascular disease risk factors with socioeconomic position during childhood and during adulthood. BMJ *313*, 1434–1438.

[25] Grant, A. J., Gill, D., Kirk, P. D. W., and Burgess, S. (2022). Noise-augmented directional clustering of genetic association data identifies distinct mechanisms underlying obesity. PLOS Genetics *18*, 1–24.

[26] Pedersen, E. M., Agerbo, E., Plana-Ripoll, O., Steinbach, J., Krebs, M. D., Hougaard, D. M., Werge, T., Nordentoft, M., Børglum, A. D., Musliner, K. L., et al. (2022). Adult: An efficient and robust time-to-event gwas. medRxiv.

[27] Brown, B. C. and Knowles, D. A. (2021). Welch-weighted egger regression reduces false positives due to correlated pleiotropy in mendelian randomization. Am J Hum Genet *108*, 2319–2335.

[28] Neale Lab (2018). UK BioBank - round 2. `http://www.nealelab.is/uk-biobank/`.

[29] Hemani, G., Zheng, J., Elsworth, B., Wade, K., Baird, D., Haberland, V., Laurin, C., Burgess, S., Bowden, J., Langdon, R., et al. (2018). The mr-base platform supports systematic causal inference across the human phenome. eLife *7*, e34408.

[30] Hemani, G., Tilling, K., and Davey Smith, G. (2017). Orienting the causal relationship between imprecisely measured traits using gwas summary data. PLOS Genetics *13*, 1–22.

[31] Hartigan, J. A. and Wong, M. A. (1979). Algorithm as 136: A k-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics) *28*, 100–108.

[32] Mounier, N. and Kutalik, Z. (2022). Bias correction for inverse variance weighting mendelian randomization. bioRxiv.

[33] Bowden, J., Del Greco M, F., Minelli, C., Davey Smith, G., Sheehan, N., and Thompson, J. (2017). A framework for the investigation of pleiotropy in two-sample summary data mendelian randomization. Stat Med *36*, 1783–1802.

[34] Sanderson, E., Davey Smith, G., Windmeijer, F., and Bowden, J. (2019). An examination of multivariable mendelian randomization in the single-sample and two-sample summary data settings. Int J Epidemiol *48*, 713–727.

[35] Mounier, N. and Kutalik, Z. (2020). bGWAS: an R package to perform Bayesian genome wide association studies. Bioinformatics *36*, 4374–4376.

[36] Sanderson, E., Spiller, W., and Bowden, J. (2021). Testing and correcting for weak and pleiotropic instruments in two-sample multivariable mendelian randomization. Statistics in Medicine *40*, 5434–5452.

[37] Leyden, G. M., Shapland, C. Y., Davey Smith, G., Sanderson, E., Greenwood, M. P., Murphy, D., and Richardson, T. G. (2022). Harnessing tissue-specific genetic variation to dissect putative causal pathways between body mass index and cardiometabolic phenotypes. The American Journal of Human Genetics *109*, 240–252.

[38] Giambartolomei, C., Vukcevic, D., Schadt, E. E., Franke, L., Hingorani, A. D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet *10*, e1004383.

Supplementary Information

Supplementary Figures

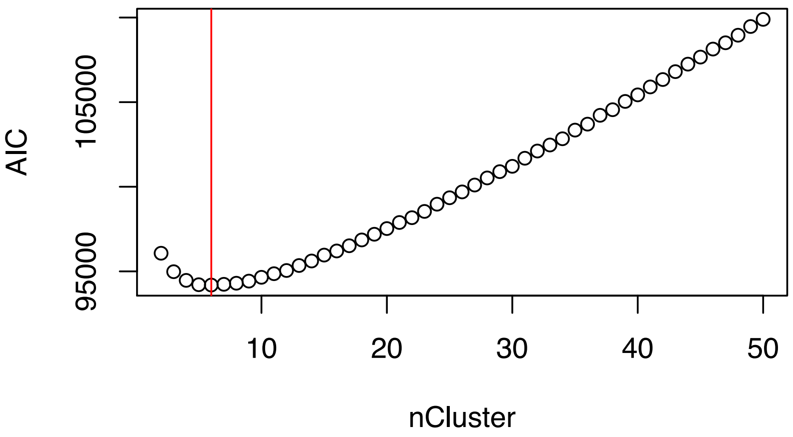


Figure S1: Dot plot representing the corresponding Akaike Information Criterion scores across varying K-means centres for BMI. K-means centres vary from 2 to 50 clusters. The red vertical line represents the number of centres/cluster with the lowest score.

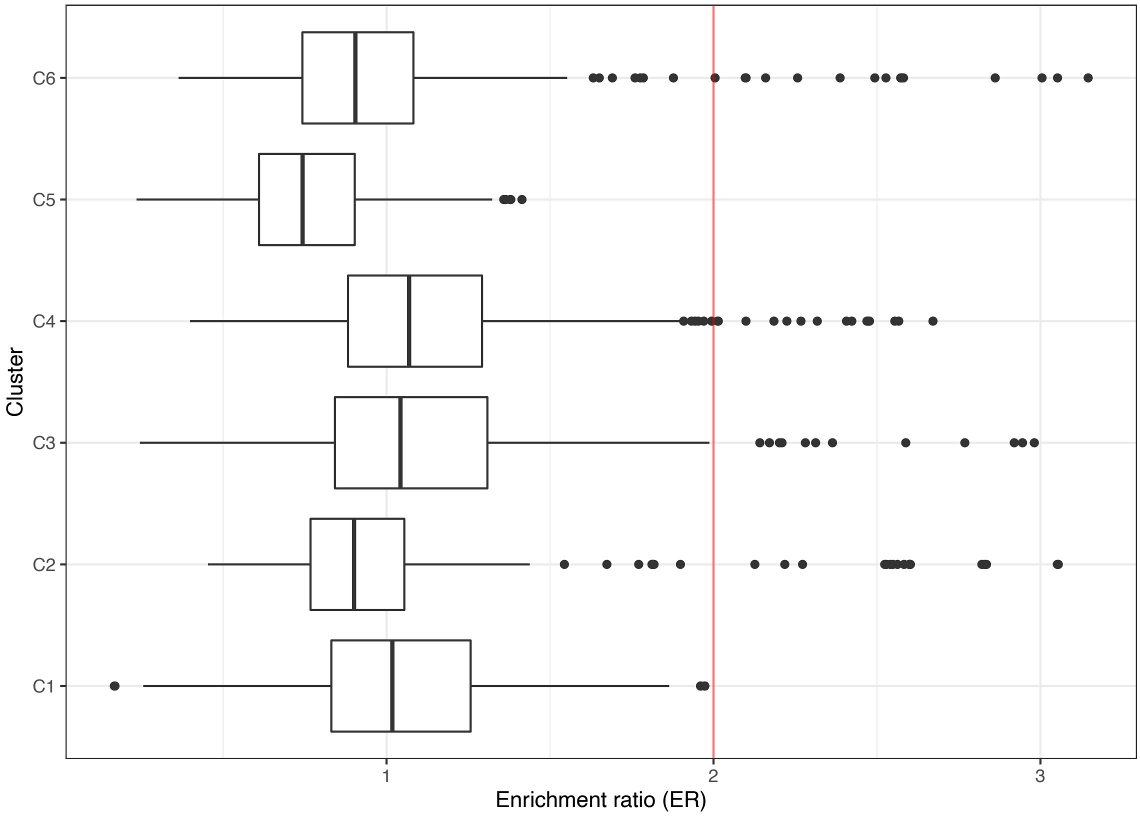


Figure S2: Boxplot showing the enrichment ratio of all traits in each cluster. BMI IVs have been clustered into 6 clusters using K-means. The enrichment ratio of each trait calculated using the cluster-specific IVs is shown in the barplot. In the boxplots, the lower and upper hinges correspond to the first and third quartiles, the middle bar corresponds to the median, whereas the upper whisker is the largest data point smaller than 1.5∗ inter-quartile range above the third quartile. The lower whisker is defined analogously.

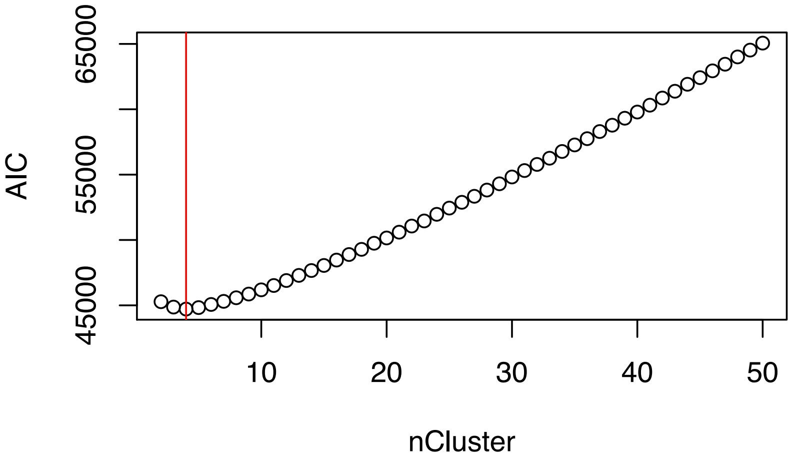


Figure S3: Dot plot representing the corresponding Akaike Information Criterion scores across varying K-means centres for child BMI. K-means centres vary from 2 to 50 clusters. The red vertical line represents the number of centres/cluster with the lowest score.

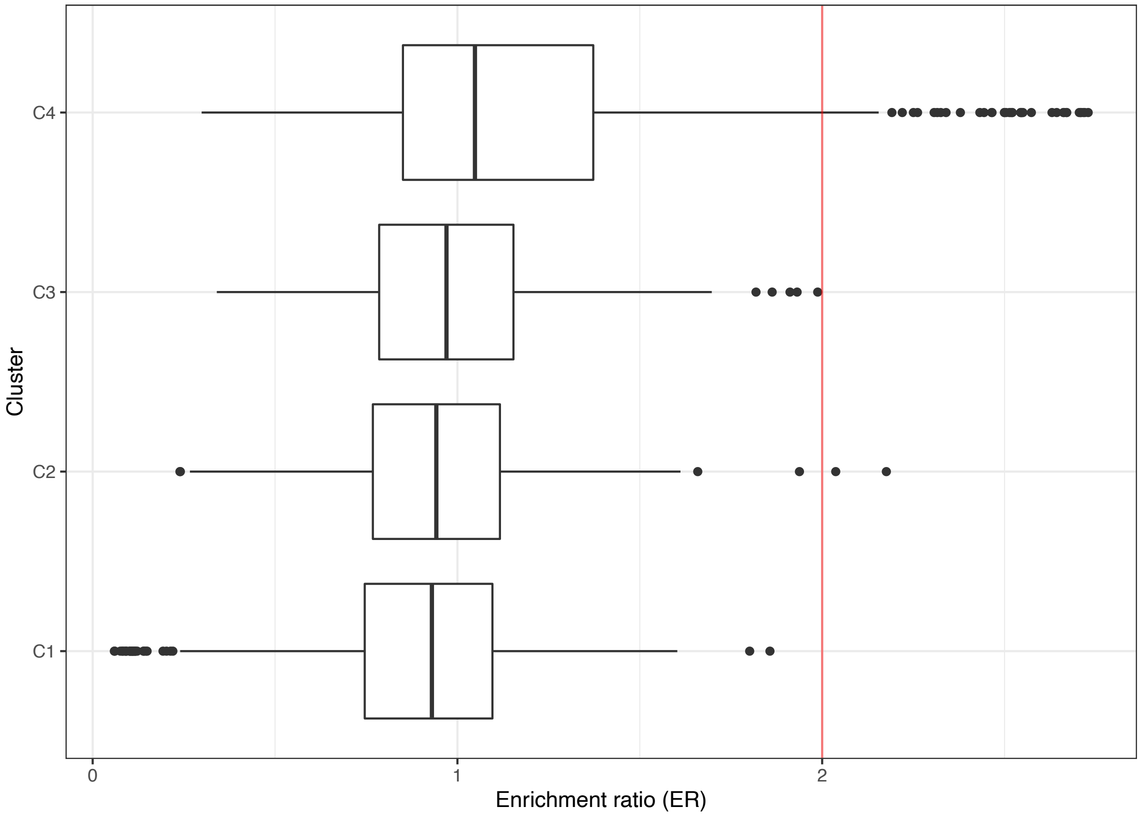


Figure S4: Boxplot showing the enrichment ratio of all traits in each cluster. Child BMI IVs have been clustered into 4 clusters using K-means. The enrichment ratio of each trait calculated using the cluster-specific IVs is shown in the barplot. In the boxplots, the lower and upper hinges correspond to the first and third quartiles, the middle bar corresponds to the median, whereas the upper whisker is the largest data point smaller than 1.5∗ inter-quartile range above the third quartile. The lower whisker is defined analogously.

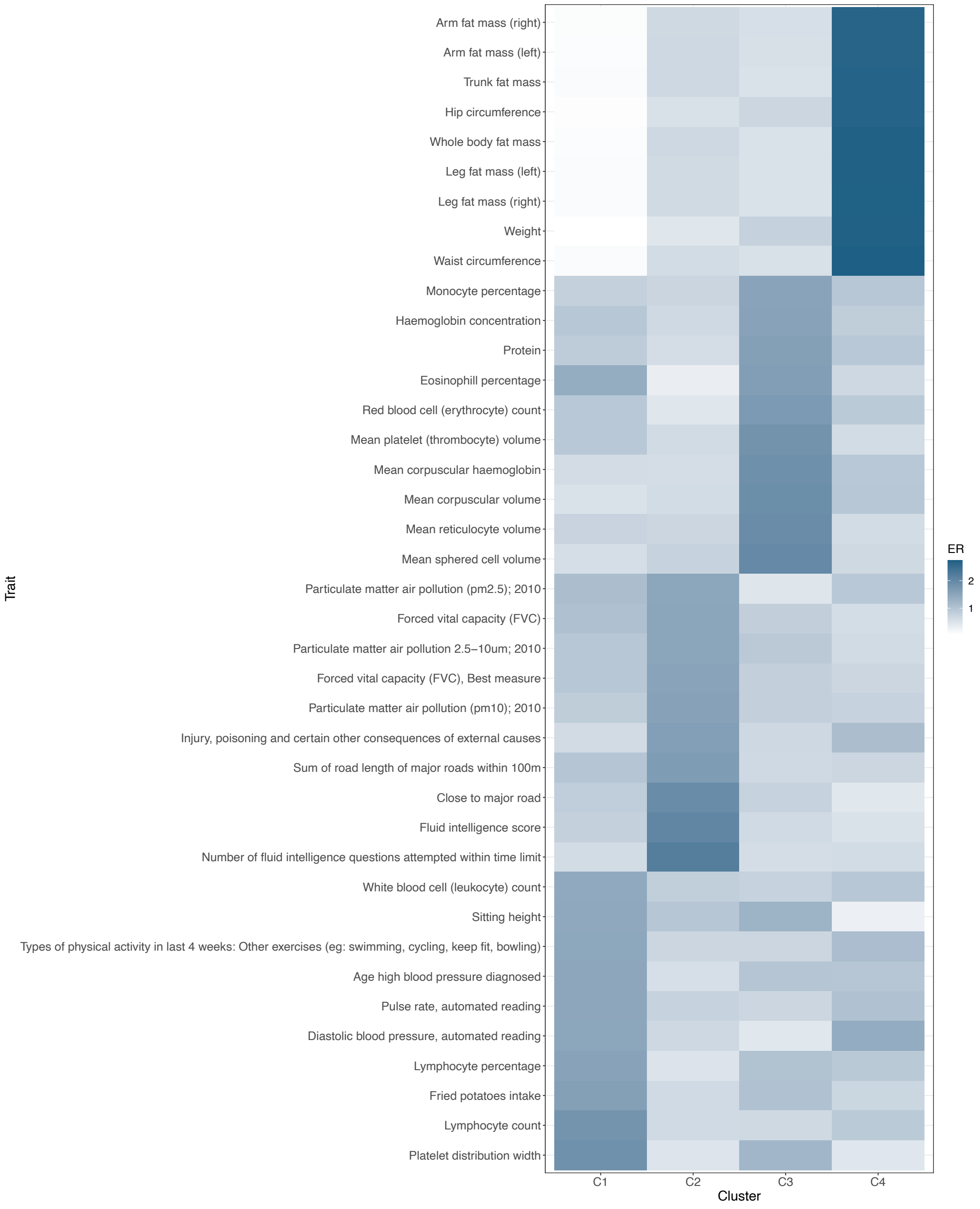


Figure S5: **Heatmap of the enrichment ratio of the top 10 traits in each cluster.** Body size at age 10 is used as a proxy exposure trait for child BMI. K-means clustering revealed 4 clusters with the following trait enrichment ratios.

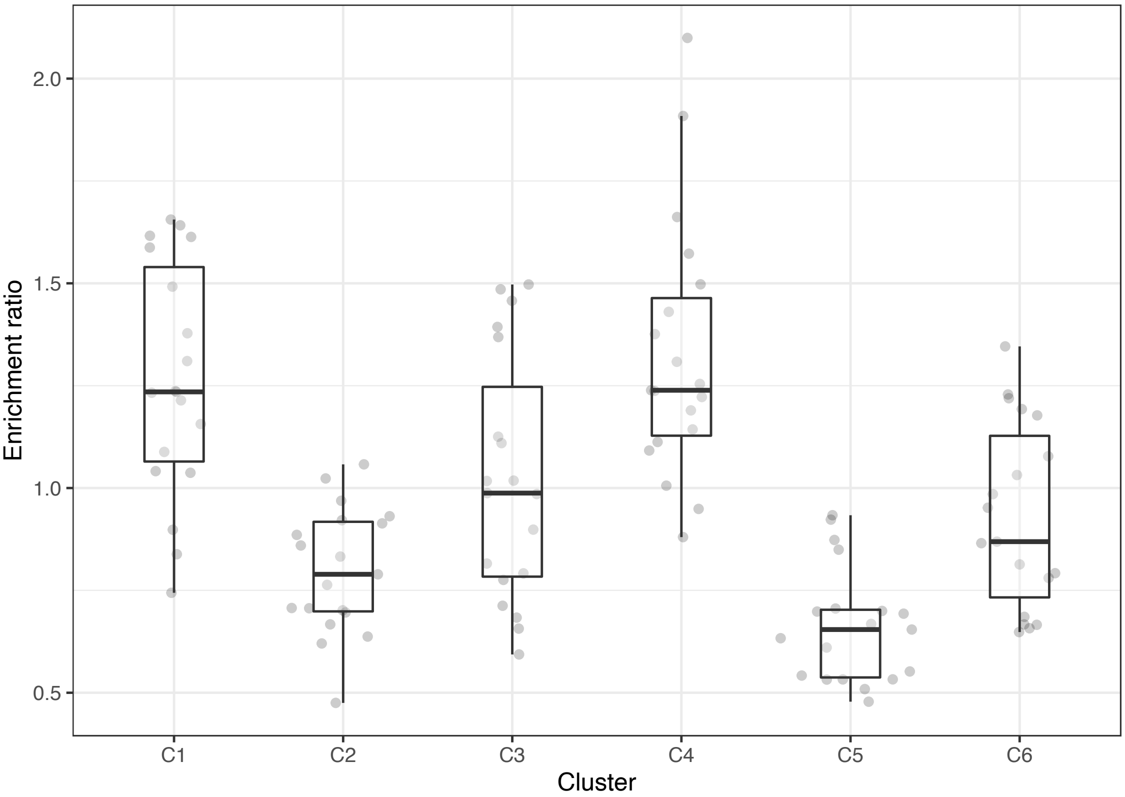


Figure S6: **Boxplot showing the ER for confounder traits across the clusters.** Confounder traits were categorised in a systematic search. In the boxplots, the lower and upper hinges correspond to the first and third quartiles, the middle bar corresponds to the median, whereas the upper whisker is the largest data point smaller than 1.5∗ interquartile range above the third quartile. The lower whisker is defined analogously.

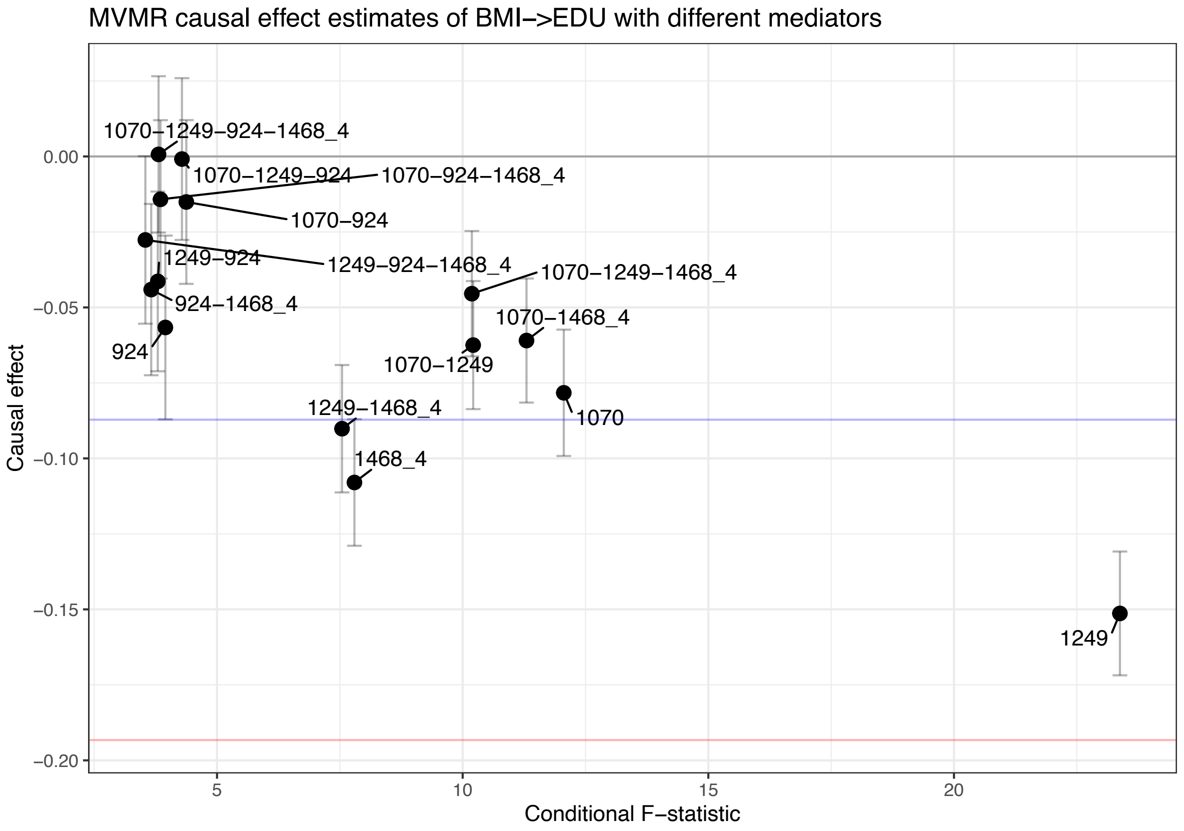


Figure S7: Dot plot showing the causal effect estimate of BMI on EDU conditional on various combinations of three candidate confounder traits. The error bars represent the 95% CI. The blue horizontal line represents the observational correlation between BMI and EDU, whereas the red horizontal line represents the univariate causal effect estimate of BMI on EDU. Trait 1070: 'Time spent watching television (TV)', trait 924: 'Usual walking pace', trait 1249: 'Past tobacco smoking', 1468_4: 'Cereal type: Muesli'.

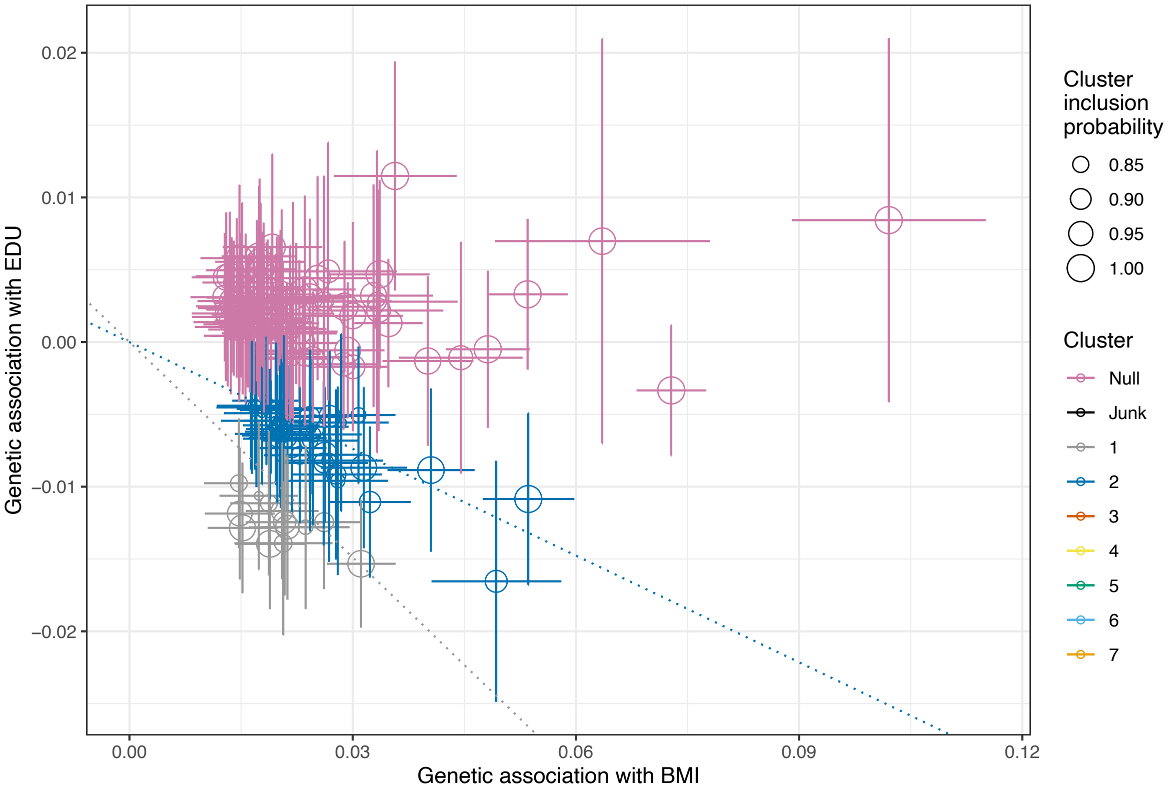


Figure S8: Dot plot showing the genetic association of IVs with the exposure: BMI, and the outcome: EDU. The exposure IVs have been clustered using MR-Clust based on their similarity in causal effect estimates. MR-Clust has revealed 2 main clusters for BMI's causal effect on EDU as well as a 'null' cluster. The IVs plotted have a cluster inclusion probability greater than or equal to 80%. The slopes represent the causal effect estimate of each cluster.

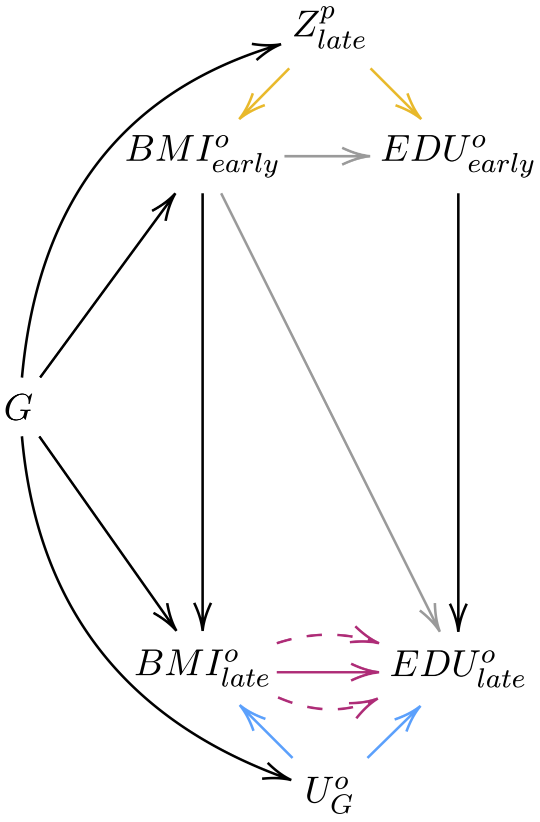


Figure S9: **Directed Acyclic Graph (DAG) illustrating the relationship between BMI and EDU.** The DAG involves early and later-in-life (late) versions of BMI as the exposure trait and EDU as the outcome trait. G represents genetic instruments, U^o_G represents a heritable confounder acting on the trait pair, whereas Z represents a parental trait involved in exerting dynastic effects. The superscripts p and o stand for parental and offspring respectively, and the dashed arrows from X to Y represent the different biological mechanisms through which a causal effect can emerge.Grey arrows represent possible causal pathways between the early traits as well as early BMI and late EDU.

Supplementary Methods

1.1 Different model selection criteria and additional number of clusters

In order to test for multiple model selection criteria, we tested for the optimal cluster number using both AIC (as shown in the manuscript) and Bayesian information criterion (BIC). Using BIC, we end up with 2 clusters being the optimal for BMI SNPs with heterogeneous causal effects on EDU (cluster 1 = -0.13 (6.61E-16), cluster 2 = -0.34 (6.15E-44)), and their enrichment reflects a clear distinction between enrichment for lean-mass and body related traits in cluster 1 and a mixed bag of trait enrichment for cluster 2 including lung/height/blood and SES-proxy traits.

This result is due to BIC introducing a stronger penalty term, $k \times \log(n)$, where k is the number of clusters and n corresponds to the number of (independent) samples used. However, in our case n represents the number of traits, which are highly correlated. Also, the more traits are used to cluster the SNPs, the more clusters we expect to obtain as they allow for a more fine-grain resolution of the underlying biological mechanisms. For these reasons, we do not believe that BIC is an appropriate measure to quantify clustering fit in this situation. Therefore, the BIC-based selection of optimal cluster number does not alter the main message/result, and only leads to coarser grain clusters.

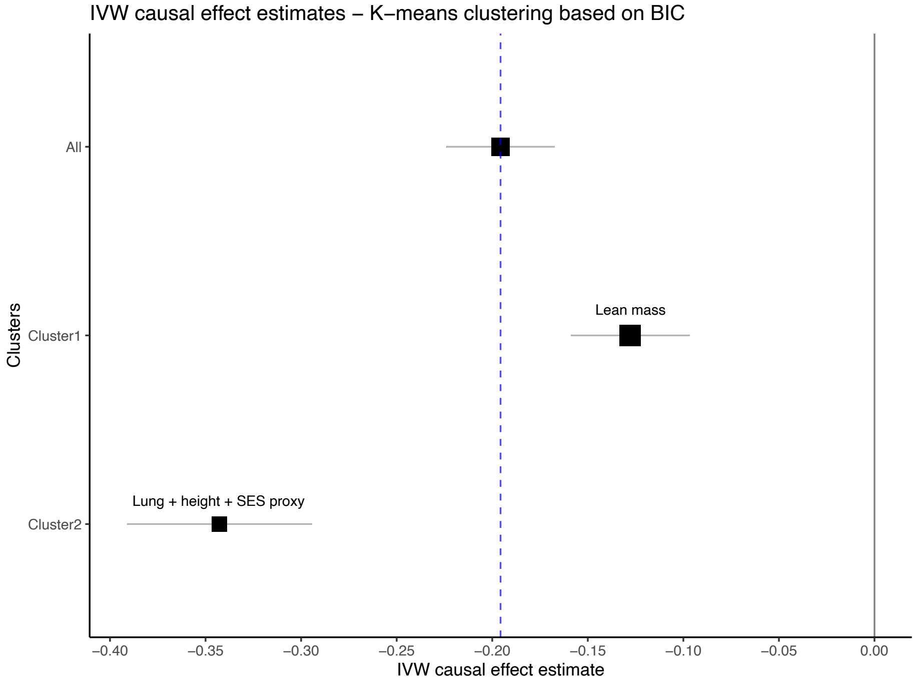


Figure S10: **Causal effect estimates of BIC-clustered BMI SNPs on educational attainment.**

On the other hand, we also tried to forcibly increase the number of clusters to 8 in the hopes of achieving more distinction in enrichment. We observed similar heterogeneous causal effects on EDU, where the smallest and largest effects were from clusters enriched for lean mass and SES-related traits respectively. As for the rest of the clusters, another 2 were strongly enriched for food supplements and a mix of height/blood/lung measurement traits, another was enriched for a mix of diseases and three other clusters had low enrichments for miscellaneous traits.

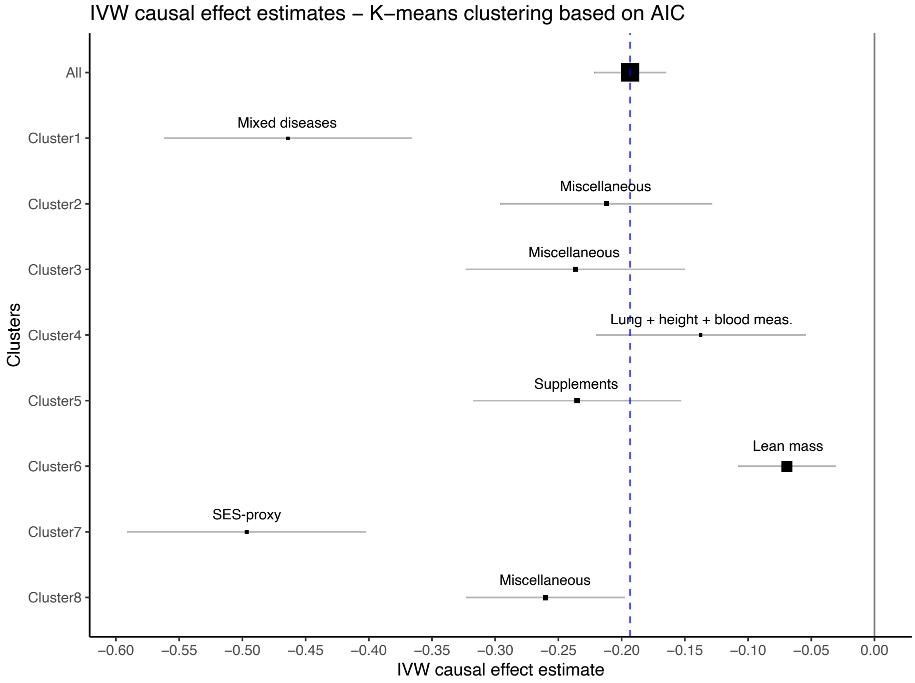


Figure S11: Causal effect estimates of clustered (forced 8 clusters) BMI SNPs on educational attainment.

1.2 Using a coarsened variable as an exposure for MR: Comparative body size at age 10

To validate our comparison between the magnitude of effect estimates for adult and childhood BMI, given that childhood BMI was proxied by coarsened variable (Comparative body size at age 10), we ran the following analysis:

We simulated polygenic risk score (PRS) to explain 10% of childhood BMI and added Gaussian noise to generate childhood BMI values for 350,000 individuals. Individuals were then split into three categories, matching the proportion of plumper and skinnier subjects in the UK Biobank data. We then normalised this coarsened/trichotomized phenotype to have a variance of 1 (mimicking our original analysis). Both the real and the coarsened childhood BMI were regressed onto the PRS. Next, we simulated a continuous EDU score with true childhood BMI having a small (-0.1) causal effect on it. Finally, we ran MR for both the coarsened and the true childhood BMI on EDU, and compared the magnitudes of the causal effects of 100 different runs (figure below).

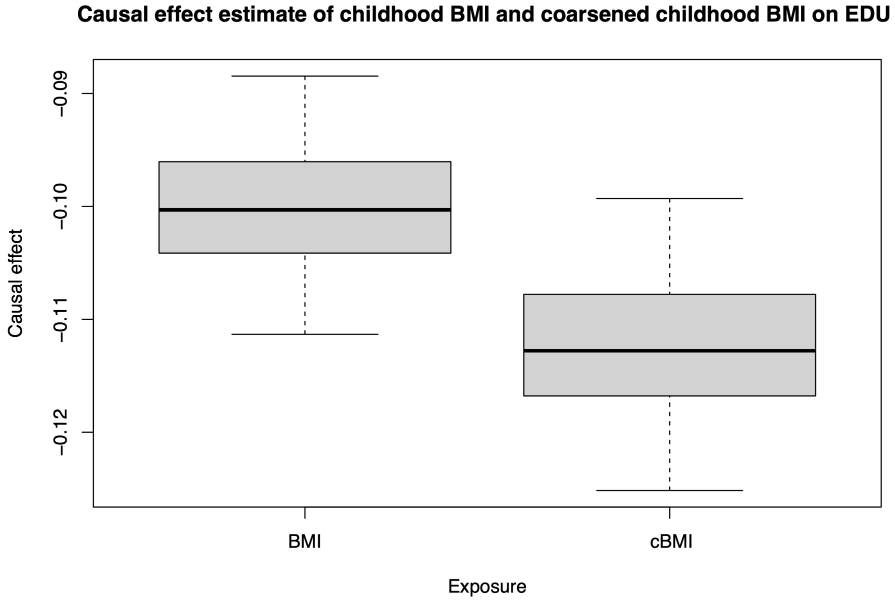


Figure S12: **Causal effect estimate of childhood BMI and coarsened childhood BMI on EDU. True causal effect of childhood BMI on EDU is -0.1.**

As seen in the results above, the causal effect estimates of BMI and coarsened BMI (cBMI) on EDU are comparable, with a slight (10%) increase of the average causal effect of cBMI in comparison to BMI's effect. This indicates that using a coarsened version of childhood BMI may have led to a slight overestimation of the causal effect, therefore the true childhood BMI on EDU effect may be even smaller than the estimated one (-0.03, p-value = 0.04). Furthermore, we see that 1 SD change in cBMI is equivalent to 0.9 SD change in BMI, assuring us of the robustness of our results and data used.

1.3 Past tobacco smoking as a candidate confounder of the BMI-EDU relationship

Despite it being a candidate confounder trait, past tobacco smoking is unlikely to have a retroactive effect on education (or an effect at all, unlike education's effect on smoking). To further investigate this, we added the trait Smoking Initiation (GWAS obtained from Saunders et al. 2022), which on average occurs around the age of 17 in the UK population, to the MVMR analysis. We repeated first the stepwise-MVMR, obtained 'Smoking initiation', 'Time spent watching television (TV)', 'Cereal type: Muesli', and 'Usual walking pace' as candidate confounder traits with significant causal effects on EDU. Note that smoking initiation replaced past tobacco smoking in this step, as it no longer had a strong causal effect on EDU. Adding BMI to this set of exposures and then calculating its conditional F-statistic with their various combination, we discover that the combination of the first three traits give a conditional F-statistic ≥ 10 (12.53) and that BMI's conditional causal effect is severely attenuated, as shown in the table below:

Smoking initiation, as seen, has a significantly negative causal effect on education, but we would like to iterate that it, as well as the other candidate confounder traits are not necessarily true confounders, but are very likely to be proxies for a confounding parental environment/trait.

Phenotype	Description	α estimate	SE	P
SmkInit	Smoking initiation	-0.1358	0.0122	7.66E-27
1070	Time spent watching television (TV)	-0.2617	0.0238	2.91E-26
1468_4	Cereal type: Muesli	0.2920	0.0341	5.39E-17
21001_irnt	Body mass index (BMI)	-0.0383	0.0103	2.01E-04

Table S1: MVMR analysis results of BMI and three candidate confounder traits on education. α: causal effect estimate.