Phenotypic causal inference using genome-wide association study data: Mendelian randomization and beyond

Venexia M. Walker, Jie Zheng, Tom R. Gaunt, and George Davey Smith

MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK

Correspondence: venexia.walker@bristol.ac.uk

ABSTRACT (150 WORDS)

Summary statistics for genome-wide association studies (GWAS) are increasingly available for downstream analyses. Meanwhile, the popularity of casual inference methods has grown as we look to gather robust evidence for novel medical and public health interventions. This has led to the development of methods that use GWAS summary statistics for causal inference. Here, we will describe these methods in order of their escalating complexity from genetic association to extensions of Mendelian randomization that consider thousands of phenotypes simultaneously. We will also cover the assumptions and limitations of these approaches, before considering the challenges faced by researchers performing causal inference using GWAS data. GWAS summary statistics are an important data source for causal inference research that offers a counterpoint to non-genetic methods when triangulating evidence. Continued effort to address the challenges presented to those using GWAS data for causal inference will allow the full impact of these approaches to be realised.

KEYWORDS (6 WORDS)

GWAS; genetic variant; polymorphism; cause; effect; inference

INTRODUCTION

Causality is the relationship between a cause and an effect, where the cause must be at least partially responsible for the effect. (1) Evidence of causality is highly sought after in the medical literature because it increases our confidence that an intervention on a cause (e.g., risk factor) will result in a difference in the effect (e.g., disease). Consequently, causal inference methods that attempt to understand the relationships between causes and effects have become increasingly popular. (2,3)

Genetic variation refers to the natural differences in germline genomes between individuals of the same population. (4) Genome-wide association studies (GWAS) can be used to explore genetic variation as they take a hypothesis free approach to identify genetic variants associated with a given phenotype. This is typically done at the single nucleotide polymorphism (SNP) level by testing the association of millions of SNPs with the phenotype using an appropriate statistical model (for example, a linear mixed model for a continuous phenotype). (5) GWAS estimate the association of each SNP with the phenotype, not the causal effect of each SNP on the phenotype. However, there are now multiple causal inference methods that can use GWAS summary statistics as an input. There are also several instances where causal claims may be made, subject to appropriate assumptions. It should be noted that SNPs are not the only form of genetic variation between individuals. Individuals may also have 'structural variation', defined as a change of more than 50 base pairs, typically caused by insertions, deletions, duplications, inversions, or translocations. (6) Structural variation can be difficult to study using summary data as specialised algorithms are needed to detect this type of variation and they typically require access to individual level data, plus there is a lack of suitable reference data. (6,7) Therefore, while it may be possible to incorporate structural variation into an analysis, it has become commonplace to restrict to biallelic SNPs (i.e., SNPs with just two observed alleles). As such, our focus will be on the use of GWAS summary statistics concerning biallelic SNPs for causal inference.

The aim of this review is to describe methods that use GWAS data for phenotypic causal inference. With this in mind, we will describe several methods in order of their escalating causal complexity (**Figure 1**). That is, we will start by discussing genetic associations for individual SNPs or in combination to form a polygenic risk score. Next, we will consider methods that are used for understanding genetic architecture – namely, fine-mapping and colocalization. These may be considered as a progression from identifying genetic variants associated with a phenotype to attempting to identify causal genetic variants for a phenotype. Following this, we will explore how GWAS can be used to understand disease aetiology. Specifically, we will describe linkage disequilibrium (LD) score regression, Mendelian randomization (MR) and latent causal variable (LCV) models; which are downstream methods that use full GWAS summary statistics or individual level data to make causal inferences regarding two phenotypes of interest. The final section will describe methods that extend upon MR as they require more than two GWAS summary datasets and enable us to study multiple phenotypes together. These methods are MR-PheWAS, two-step MR and multivariable MR. We conclude this piece with a note on individual-level data and discussion of the challenges of using GWAS for causal inference. Key definitions used throughout this review are provided in **Table 1**.

Figure 1: Summary of the methods covered in the present review

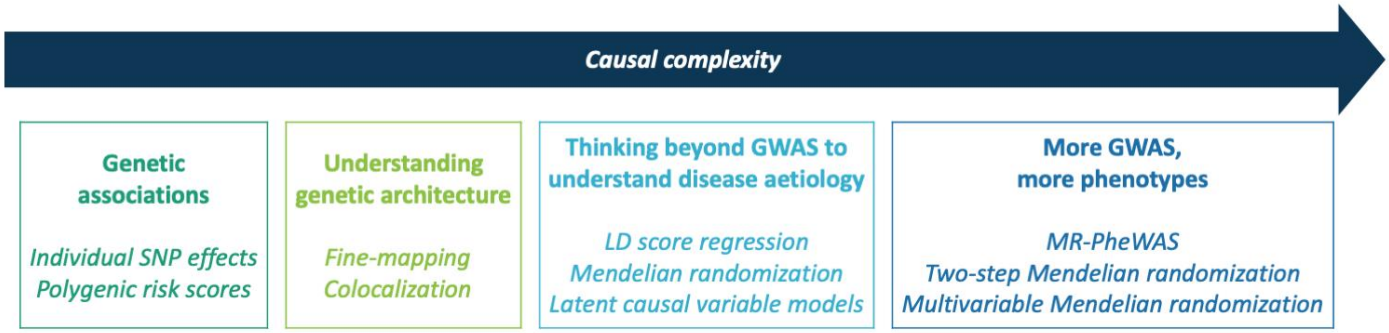


Table 1: Key definitions

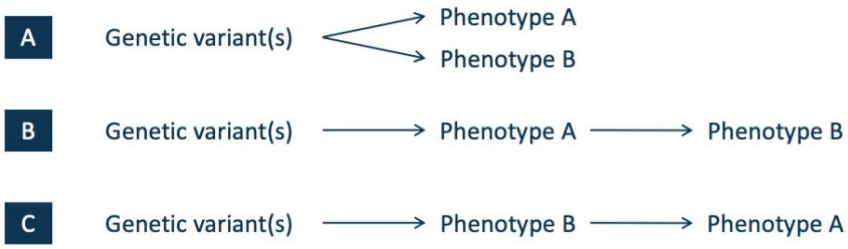
Term	Definition
Linkage disequilibrium (LD)	The non-random association of alleles between SNPs due to genomic proximity.
Polygenic risk score	The combination of genetic associations for multiple SNPs into a single metric.
Genetic architecture	The features of genetic variation that result in heritable variation in phenotypes.
Promotors and enhancers	Sequences of DNA in non-coding regions of the genome that relate to transcription of a particular gene.
Transcription factor	Proteins needed for transcription of a particular gene.
Regulatory region	A region in the genome where DNA sequences, such as promotors and enhancers, can be found.
Imputation	A process to predict unobserved genotypes in genetic data using observed genotypes.
Haplotype	The alleles inherited from a single parent.
Heritable variation	The variation between individuals caused by genetics.
Population stratification	Systematic differences in genetic variation that arise in subgroups of individuals within a sample and may be an alternative cause of non-random association of alleles.
Cryptic relatedness	The unknown inclusion of closely related individuals in genetic data.
Genetic correlation	The proportion of variance shared between two traits due to common genetic causes.
Polygenicity	The situation where the phenotype is the result of variation in multiple genes.
Instrumental variable analysis	A causal inference method that uses the relationship of the exposure and the outcome with a third variable, known as the instrument, to address confounding of the exposure-outcome association.
Reverse causation	The observation of an effect of Y on X when the direction of effect is thought to be X on Y.
Horizontal pleiotropy	The situation where a genetic variant affects the outcome through a pathway other than the exposure of interest in addition to affecting the exposure.
Vertical pleiotropy	The situation where the effect of the genetic variant through the exposure affects other phenotypes that in turn have an effect on the outcome.
Mediator	A variable that occurs on the causal path between an exposure and an outcome.
Non-collapsibility	Unwanted variation that occurs for some measures of association (e.g., odds ratios) due to covariate adjustment choices.

GENETIC ASSOCIATIONS

As noted in the introduction, the raw output from a GWAS is the genetic associations of millions of SNPs with a phenotype of interest. Genetic associations are a blunt tool that may represent several things, including different types of genuine causal effect (**Figure 2**) or an induced association due to LD or population stratification. Consequently, a genetic association is not inherently causal. However, we

anticipate that SNPs with a causal effect on a phenotype will be robustly associated with that phenotype. It is on this basis that we use genetic associations obtained from GWAS for causal inference and prioritize SNPs that are strongly associated with a phenotype in GWAS for further causal investigation. Furthermore, data collected from routine genotyping and used for GWAS reflect germline genetic variation. That is, differences in SNPs between individuals that are a result of the random allocation of SNPs to individuals from their parents at conception. This adds a temporal element to genetic associations as we know that any observed genetic variation precedes the clinical manifestation of the phenotype of interest. Again, this alone is insufficient to claim causality but is a feature of genetic associations that is exploited in some causal inference methods that will be discussed in this review.

Figure 2: Illustration of the situations where genetic associations with two phenotypes may be observed



In the situation where a GWAS provides evidence that multiple SNPs are associated with a phenotype, there is the opportunity to derive a polygenic risk score. Polygenic risk scores combine the genetic associations for multiple SNPs into a single metric. This can then be used to calculate the risk of a given phenotype in a novel group of individuals (i.e., individuals who did not contribute to the GWAS used to derive the polygenic risk score). In their simplest form, polygenic risk scores can be calculated as the weighted sum of SNPs associated with the phenotype of interest. However, several methods have been proposed to improve their prediction. For instance, LDpred is a Bayesian method that uses reference data on linkage disequilibrium and genetic architecture as a prior to increase polygenic risk score accuracy. (8) Regardless of how the polygenic risk score is obtained, like the genetic associations for individual SNPs that the score is composed from, a polygenic risk score is not inherently causal. This is because the polygenic risk score will also reflect the underlying causes of a phenotype – for instance, a polygenic risk score derived for C-reactive protein is likely to reflect factors such as body mass index and smoking. Despite this, the use of SNPs determined at conception to construct a polygenic risk score does infer some inherent causality. Polygenic risk scores have therefore been used as an alternative input for some causal inference methods, such as MR. This has been demonstrated many times in the literature for several different applications, including the identification of mediators and potential novel targets for cardiometabolic disease. (9) Note that careful consideration should be given to the p-value threshold for inclusion of SNPs into a polygenic risk score, especially if it is to be used as an input for other methods. For example, a polygenic risk score composed of genome-wide significant SNPs should produce similar estimates to an inverse variance weighted MR using the same SNPs. However, MR conducted using a polygenic risk score derived from a more lenient p-value threshold may produce a different estimate and warrants further sensitivity analyses. (10)

UNDERSTANDING GENETIC ARCHITECTURE

Genetic architecture is a term used to describe the features of genetic variation that result in heritable variation in phenotypes. (11) Characteristics of interest include how many SNPs affect a phenotype, how common those SNPs are at the population level, the effect size of those SNPs, and SNP interactions. (11) Below we describe two methods that attempt to understand aspects of genetic architecture: fine-mapping and colocalization. Both methods use multiple SNPs within a specific region of the genome to identify causal variants. In the case of fine-mapping, the causal variant is for a single phenotype, whereas for colocalization, the method is trying to determine whether the causal variant is shared between

phenotypes or specific to a single phenotype. These methods therefore advance upon genetic associations from GWAS and polygenic risk scores as they are explicitly attempting to map causal genetic variants.

Fine-mapping

Fine-mapping considers the relationship of multiple SNPs in a region with a phenotype to identify causal variants for that phenotype. Consequently, this approach assumes that at least one causal variant for the phenotype exists in the region under study. (12) Fine-mapping may be performed in a number of ways: overlapping SNPs identified in a GWAS with functional elements, such as promotors and enhancers; determining allele-specific SNP effects that indicate function; identifying SNPs that relate to transcription factors; and calculating SNP effects on regulatory region activity. (13) Each of these fine-mapping approaches requires a strong understanding of the biology related to the region you wish to study. For this reason, fine-mapping is often performed as a follow-up analysis once regions of interest have been identified, usually from a GWAS. For instance, fine-mapping using high-density imputation and islet-specific epigenome maps was implemented to further interrogate loci identified in the 2018 DIAGRAM type 2 diabetes GWAS of almost 900,000 European individuals. (14)

A key limitation of fine-mapping is that it is reliant on a high density of SNPs to be either directly genotyped, or imputed well, within the region of interest. Further direct genotyping of participants has time and cost implications for a study and is simply not feasible in many situations, such as studies using secondary data, while imputation is subject to its own limitations. (12) In addition, relevant knowledge of transcription factors or promotors and enhancers is essential for some fine-mapping approaches. However, these genetic features can be cell-type and/or disease specific. (15) Consequently, fine-mapping may not be informative in settings where cell-type and/or disease specific data is unavailable. (13)

Colocalization

Colocalization is a Bayesian method to determine whether two phenotypes share a causal variant within a specified region of the genome. (16–18) The method quantifies the likelihood of a shared causal variant by calculating the posterior support for five hypotheses:

- H_0 : No association with phenotype A or phenotype B
- H_1 : Association with phenotype A but not phenotype B
- H_2 : Association with phenotype B but not phenotype A
- H_3 : Association with phenotypes A and B, different causal variants
- H_4 : Association with phenotypes A and B, same causal variant

The greater the posterior probability for hypothesis H_4, the greater the evidence to support colocalization of the two phenotypes.

Colocalization may be thought of as a locus specific consideration of SNPs as the approach is applied to a specific genomic region. For this reason, colocalization has proven valuable as a follow-up analysis for other analyses, such as MR. For example, it has been used to increase confidence in putatively causal effects between proteins and disease-related phenotypes, with a view to identifying potential therapeutic targets. (19) Several extensions to colocalization have also been developed. For instance, while the method originally required summary statistics derived from distinct samples, an extension allows common controls to be used to explore the relationship between case-control phenotypes. (20) This has been successfully demonstrated in a study of four autoimmune diseases (type 1 diabetes, rheumatoid arthritis, celiac disease, and multiple sclerosis), which found 33 genomic regions associated with two or more of the disorders. (20) Furthermore, the extensions ‘multiple trait colocalization analysis’ (MOLOC) and ‘hypothesis prioritisation for multi-trait colocalization’ (HyPrColoc) have also been developed, which allow consideration of more than two phenotypes in the same model. (21,22) MOLOC has previously been

applied to consider expression and methylation quantitative traits loci simultaneously with a disease phenotype of interest but can be too computationally intensive when more than four traits are considered at the same time. (21) HyPrColoc is a more computationally efficient method in these settings as it considers a reduced set of putative causal configurations to approximate the posterior probability for hypothesis H_4. It has previously been applied to study the shared genetic associations of coronary heart disease with 14 relevant risk factor, comorbidity and social traits. (22)

Colocalization requires certain assumptions to be met. Firstly, like many methods that use multiple GWAS, the GWAS summary statistics must be sampled from the same population. This ensures that patterns in LD are the same in each sample. Secondly, the method requires that the causal variant has either been directly genotyped or well imputed in all datasets used for the analysis so that it may be identified. Finally, it is assumed that there is only a single causal variant in the region of the genome under study. (22) However, methodological advances mean that in the presence of multiple conditionally distinct association signals, this assumption can be relaxed. For instance, an approach named 'pairwise conditional colocalization' (PWCoCo) tests all conditionally independent SNPs for one phenotype against all conditionally independent association signals for the other phenotypes and so overcomes the limitations of using marginal associations in conventional colocalization analyses. (19) An alternative approach named 'Sum of Single Effects' (SuSiE) also relaxes this assumption – this time through conditioning on the causal signal being considered while evaluating multiple causal variants simultaneously.

There are several limitations of colocalization analysis. As a Bayesian method, colocalization requires priors to be specified to calculate the posterior support for each of the five hypotheses described above. Different priors may lead to different posteriors and potentially impact the inferences made from this approach. Furthermore, colocalization is only as powerful as the summary statistics provided to it. This means the method may not perform well if one or both phenotypes under consideration are derived from small samples. Similarly, if the assumption that the causal variant has either been directly genotyped or well imputed has been violated than inferences may be misleading. Finally, in situations of perfect LD, you can only determine whether the same haplotype, not variant, is causal using colocalization approaches. As haplotypes do not necessarily translate across populations that have different LD structures, this may limit the generalizability of findings.

THINKING BEYOND GWAS TO UNDERSTAND DISEASE AETIOLOGY

This section covers three methods that use genome-wide genetic information to make causal inferences, typically for two phenotypes of interest: LD score regression, MR, and LCV models. These methods can be implemented using either full GWAS summary statistics or individual level data and attempt to characterise the relationship between the phenotypes they consider (**Figure 2**). This section is followed by a section that covers extensions to these methods, all of which make use of more than two GWAS.

LD score regression

LD score regression may be considered as a genome-wide consideration of SNPs as opposed to colocalization that, as described above, is a locus specific consideration of SNPs. For this approach, genome-wide summary statistics are regressed on LD score, which indicates how likely a SNP is to tag other SNPs in close proximity. This allows separation of genome-wide inflation of test statistics due to polygenicity from bias due to population stratification and cryptic relatedness. (23,24) Extensions to LD score regression have increased the applications of this method to include partitioning heritability by functional annotation and estimating the genetic correlation between different complex phenotypes. (25,26) However, LD score regression is not directed with respect to the phenotypes it considers and so cannot be used for causal inference between them. Instead, genetic correlation obtained from LD score regression can suggest shared genetic aetiology between the phenotypes, which may underpin causality.

LD score regression can be performed easily using LD Hub, a free tool that provides access to a centralized database of GWAS summary statistics and automates the regression component of the analysis. (23) Consequently, LD score regression is often included as a sensitivity analysis alongside other causal inference methods that use GWAS data. For instance, the genetic correlation of cleft lip/palate and educational attainment was presented alongside MR estimates to better understand whether there was a genetic predisposition to low educational attainment among individuals with cleft lip/palate. (27)

The main limitation of LD score regression is the potential for bias induced when there is correlation between the variance explained by a SNP and the LD score for that SNP. The extent of this correlation differs between phenotypes. However, when correlation is present, the intercept and slope obtained from the LD score regression may be altered and the ability to discriminate genome-wide inflation of test statistics due to polygenicity from bias due to population stratification and cryptic relatedness impacted. LD score regression is also suspectable to attenuation bias when LD score is poorly measured. This bias can be accounted for using a dis-attenuation factor if the variance of the measurement error is known. (24)

MR

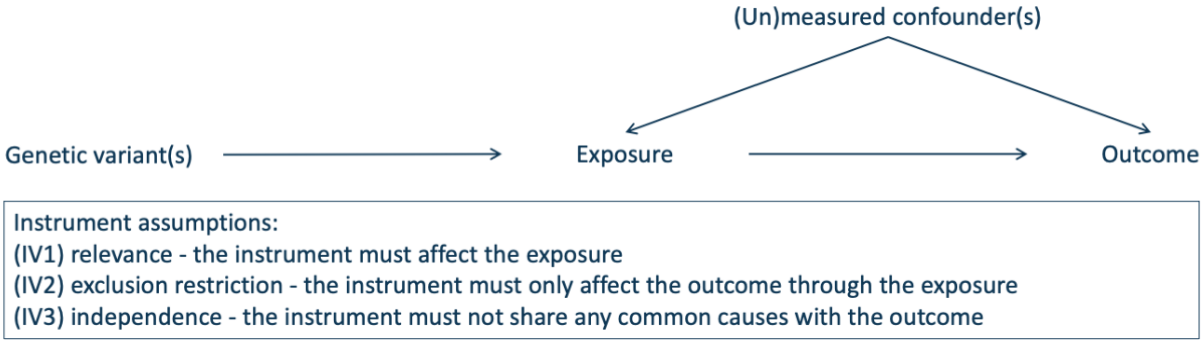
Earlier we established that differences in SNPs between individuals result from the random allocation of SNPs to individuals from their parents at conception. Individuals with the same parents are therefore equally likely to receive a given SNP, meaning the allocation of SNPs is a natural form of randomization. It was on this basis that the concept of MR was initially introduced as a within-sibship approach. (28,29) However, the scarcity of data to implement MR in this way meant a more pragmatic approach using population data had to be used and it was demonstrated that this reasonably matched expectations regarding genetic variant-phenotype associations. (30,31) Recently large-scale family-based data resources have become available, allowing analyses to be carried out within the originally proposed framework. (32,33)

MR can be implemented as an instrumental variable (IV) analysis that uses one or more germline SNPs as an instrument for an exposure to estimate the causal effect of that exposure on an outcome. (28,34) Assuming the genetic variant(s) selected as the instrument satisfy three assumptions, the resulting MR estimate should not be subject to non-genetic confounding or reverse causation. The three instrument assumptions (**Figure 3**) are: (IV1) relevance - the instrument must affect the exposure (note: this instrument does not need to have a causal effect on the exposure to be valid), (IV2) independence - the instrument must not share any common causes with the outcome, and (IV3) exclusion restriction - the instrument must only affect the outcome through the exposure. Under these three instrument assumptions, it is possible to obtain bounds for a causal effect. A fourth assumption, typically monotonicity (i.e., the exposure is an always increasing or always decreasing function of the instrument) or homogeneity (i.e., the effect of the exposure on the outcome is the same for all individuals), is required to identify a point estimate within these bounds. (2,35) Furthermore, if we are to use the estimates obtained from MR in practice, we must expect the gene-environment equivalence to hold. That is, we must expect that the change in a phenotype that would result from changing someone's genotype is comparable to the change in a phenotype that would result from changing someone's environmental exposure. (36) The assumptions described above, and the plausibility of gene-environment equivalence are critical for valid causal inference and therefore are a major component of the STROBE-MR guidelines for reporting MR studies (https://www.strobe-mr.org/). (37,38)

The only instrument assumption that can be empirically verified is relevance (IV1). One of the easiest ways to demonstrate this assumption has been met is to select SNPs that are known to associate with an exposure based on strong, preferably replicated, signals identified in GWAS. However, it should be noted that SNPs shown to have a robust association with the exposure of interest that have not been identified from GWAS are equally suitable for this purpose. Furthermore, we cannot know from the GWAS data alone whether the SNP(s) selected to instrument the exposure have a causal effect on the exposure, though this

is not a specific requirement for the SNP(s) to be valid instrument(s). Instrument SNPs are extracted from an independent sample that contains individual level information on both the exposure and outcome for one-sample MR to be applied. Alternatively, the instrument SNPs can be extracted from a second GWAS that relates to the outcome and two-sample MR applied. (39) Two-sample MR therefore removes the need for individual level data and the identification of a data source that contains both the exposure and the outcome. However, the data used to select the instruments should not contain the same participants as the data used to define the outcome (i.e., there should be no sample overlap) as this can result in winner's curse bias. (39) The potential impact of this bias on MR estimates can be quantified using a free, online calculator. (40) The use of GWAS summary data for MR has led to the development of resources such as the IEU OpenGWAS database, which provides free access to complete GWAS summary statistics for thousands of phenotypes, allowing easy implementation of two-sample MR approaches. (41,42)

Figure 3: Instrument assumptions required for MR



MR has multiple applications from the investigation of risk factors for disease to the prediction of adverse drug effects. (28,43) One notable example of this method is its application to study low-density lipoprotein (LDL) cholesterol and coronary heart disease. SNPs have been selected to instrument LDL cholesterol in different ways for different studies and MR consistently reports a protective effect of this lipid fraction on coronary heart disease. For instance, Ference et al selected nine SNPs from six genes identified in a 2010 GWAS of blood lipids. (44,45) Meanwhile, White et al used 185 SNPs identified by Willer et al in the 2013 Global Lipids Genetics Consortium GWAS. (46,47) More recently, SNPs have been limited to those occurring within specific genes as a proxy for LDL cholesterol reduction via specific drug targets. For example, SNPs within the 3-Hydroxy-3-Methylglutaryl-CoA Reductase (HMGCR) gene have previously been used as a proxy for statins, which target this protein. The MR studies using a restricted version of the more general LDL cholesterol instrument have confirmed the protective effect of statins on coronary heart disease, as seen in randomized controlled trials. (48–50) An important consideration when comparing MR estimates with other sources of evidence, such as randomized controlled trials, is that estimates generally reflect the long term differences in exposure across life that the SNPs used as instruments relate to. (51,52) This differs from other estimates – for example, if you calculate the effect of a drug, exposure will typically be much shorter or may occur at a specific time point in the life course.

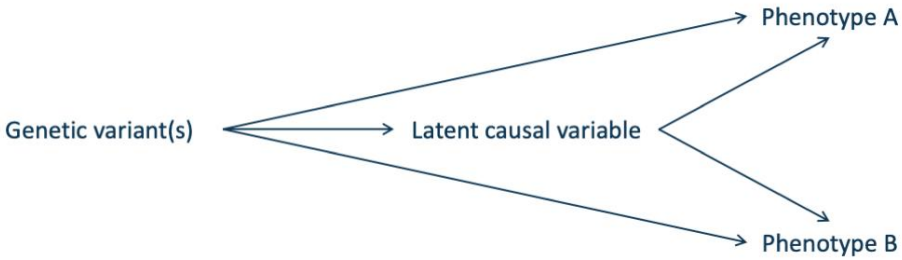
MR is subject to several limitations. Among the most concerning is horizontal pleiotropy, which violates the instrument assumptions and can bias MR estimates. Horizontal pleiotropy occurs when SNPs selected to instrument an exposure relate to multiple traits on different pathways that all affect the outcome, distorting the association of the exposure with the outcome. (34) At present, the best approach to address horizontal pleiotropy is to perform sensitivity analyses using MR methods that are potentially more robust to certain types of pleiotropy, such as the weighted median and MR-Egger approaches. (53,54) In addition to this, MR estimates may be biased by genomic confounding as the method is only robust to non-genetic confounding of the exposure-outcome association. Finally, MR – particularly when conducted in the single sample setting – can be subject to weak instrument bias, which can violate the first instrument assumption of relevance. These limitations, plus many more, are discussed in detail elsewhere. (51,52,55)

Concerns regarding the limitations of MR, particularly the potential violation of the instrument assumptions, has led to the development of multiple sensitivity analyses. (E. Sanderson, M. M. Glymour, M. V. Holmes, K. Hyunseung, J. Morrison et al., *manuscript in review*) Most of these sensitivity analyses use weaker assumptions in place of the core IV assumptions and require multiple instrument SNPs to have been used for the main analysis. (36) Examples of sensitivity analyses include MR using a Robust Adjusted Profile Score (MR RAPS), which is designed to address violation of the relevance (IV1) assumption by allowing weak instruments to be included in the analysis. (56) There are also several 'pleiotropy robust methods' that address violation of the exclusion restriction (IV3) assumption including the weighted median (53); the weighted mode (57); Steiger filtering (58); MR TRYX (59); MR PRESSO (60); MR-Egger (54); and MR GxE (61,62) approaches. Given the wealth of MR approaches available and the diversity in their strengths and limitations, inclusion of multiple MR approaches is considered the best approach to demonstrating robust results when reporting an MR study. (63)

LCV models

LCV models interrogate the relationship between two phenotypes by quantifying the genetic causality proportion (GCP) of phenotype A with a latent variable under the model assumed in **Figure 4**. (64) Phenotype A is considered 'fully genetically causal' for phenotype B if it is in perfect genetic correlation with the latent variable – this corresponds to the maximum GCP of one. The GCP has a minimum value of zero, which indicates 'limited partial genetic causality'. Therefore, GCP values between zero and one can be interpreted as indicating the extent of partial genetic causality between the two phenotypes. The key assumption of LCV models is that the joint effect size distribution for the two phenotypes is the sum of two independent distributions. The first of these distributions represents the shared genetic component of the model (i.e., the effect through the latent variable) and the second distribution represents all other effects. This assumption has been shown to be strictly weaker than the exclusion restriction (IV3) assumption required for MR. LCV models have previously been applied in the literature to investigate the relationship between brain structure and problematic alcohol use on that basis that variability in brain structure phenotypes may predispose individuals to problematic alcohol use. (65)

Figure 4: An illustration of an LCV model



There are several limitations of LCV models, some of which are shared with the other causal inference methods discussed in this section, such as the risk of confounding due to population stratification. These limitations are well covered elsewhere. (64) Other important limitations that warrant mentioning here are that the method is constrained to the model presented in **Figure 4**. As a result, the model assumes by design that there is a single latent variable that acts between the two phenotypes and that the effects between the variables act in the directions shown in this figure. Consequently, they have limited utility when bidirectional causal effects exist. For example, LCV models would be inappropriate to study the relationship between BMI and smoking, which has previously been shown to be bidirectional using MR. (66) A further related limitation of LCV models is that they can only consider two phenotypes at a time. This means the following section focuses exclusively on extensions of MR, which can consider more than two phenotypes at a time.

MORE GWAS, MORE PHENOTYPES

Until this point, we have focussed on using up to two GWAS to make causal inferences regarding either a single phenotype or the effect of one phenotype on another. This section builds upon this by summarizing methods that use more than two GWAS. Specifically, we will cover three extensions of MR: MR-PheWAS (one exposure phenotype, multiple outcome phenotypes), two-step MR (one exposure phenotype, one intermediate phenotype, and one outcome phenotype), and multivariable MR (multiple exposure phenotypes, one outcome phenotype).

MR-PheWAS

MR-PheWAS, which stands for MR phenome-wide association study, is an extension of MR that involves performing a hypothesis-free scan of potentially causal effects of a single exposure on multiple outcomes. This method can either be applied by defining all outcomes in a single sample (for example, using PHESANT in UK Biobank (67)) or by identifying a different GWAS for each outcome (for example, MR-EVE (68)). In the case of MR-EVE, instruments were obtained for over 2400 traits and MR systematically performed for each trait against all other traits. The final results from MR-EVE are available from EpiGraphDB (69) and can be used for four different applications: (i) to identify confounders of an exposure-outcome association, (ii) to identify intermediates of an exposure-outcome association, (iii) to identify reverse intermediates (i.e. intermediates that act outcome > intermediate > exposure, rather than exposure > intermediate > outcome) of an exposure-outcome association, and (iv) to identify colliders of an exposure-outcome association. (69) Effects identified in MR-PheWAS approaches share the same limitations as those identified from conventional MR. In addition, studying multiple outcomes in a MR-PheWAS comes with a multiple testing burden, though the use of the same instrument to study the effect of an exposure on multiple (particularly related) outcomes can aid interpretation. For example, the MR-PheWAS approach has been used to identify causal effects of body mass index in UK Biobank. After accounting for multiple testing, this analysis identified over 580 associations that included both previously identified effects, such as higher body mass index increasing the risks of diabetes, and novel effects, such as higher body mass index reducing the risk of four nervousness/anxiety phenotypes. (70)

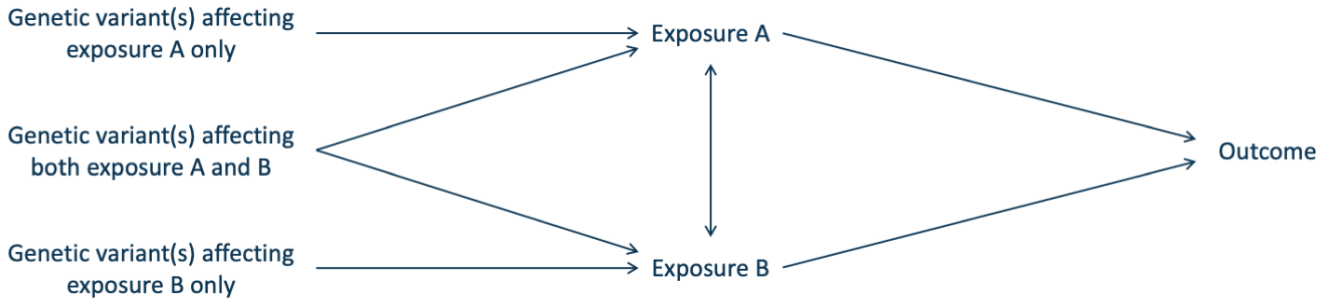
Two-step MR

Two-step MR refers to analyses that involve performing two MR analyses in sequence with the second MR analysis using the outcome from the first MR analysis as its exposure. This approach can be extended further to explore networks of relationships between phenotypes and, in that case, may then be referred to as ‘network MR’. Two-step MR was initially introduced in the epigenetic literature as ‘two-step epigenetic MR’ and contributes to our understanding of the mediating role of DNA methylation. (71) However, the approach is generalizable and can easily be applied to understand the mediating role of other (preferably continuous) phenotypes. For instance, two-step MR has previously been applied to study the mediating role of liability to type 2 diabetes when studying the causal effect of common risk factors for cardiovascular disease. (72,73) Two-step MR is appealing for mediation analysis as it allows estimation of the indirect effect of the exposure (i.e., the effect of the exposure via the mediator) on the outcome using the product of coefficients method. The product of coefficients method involves multiplying the estimate from the first step MR for the effect of the exposure on the mediator by the estimate from the second step MR for the effect of the mediator on the outcome and updating the confidence intervals accordingly. Recently, a hybrid approach that uses multivariable MR (discussed below) for the second step MR to estimate the direct effect of the mediator (i.e., the effect of the mediator independent of the exposure) on the outcome has been proposed. The advantage of this hybrid approach is that it ensures the coefficient for the effect of the mediator on the outcome included in the product is independent of the exposure. (74)

Multivariable MR

Multivariable MR is an MR approach used to calculate the effects of multiple exposures on a single outcome simultaneously (**Figure 5**). By including multiple exposures in the same model, this approach can calculate the direct effect of each exposure (i.e., the effect of each exposure independent of all the other exposures in the model) on the outcome. Multivariable MR therefore requires distinct SNPs for each exposure to be identified. Like many of the MR methods discussed in this review, multivariable MR can be applied in the individual level data setting using a single sample of data (with instruments selected from independent samples) or in the summary level data setting using full summary statistics from multiple GWAS. The consideration of multiple exposures at the same time means that multivariable MR is a particularly useful approach when you want to study multiple related phenotypes and separate their effects on an outcome from each other. For instance, this approach has previously been used to calculate the independent causal effects of education and intelligence on liability to Alzheimer's disease. (75) It is also useful in situations where the independence and exclusion restriction assumptions required for univariate MR are violated. (76)

Figure 5: Illustration of multivariable MR design



Multivariable MR can also be used to study mediation by including the proposed exposure and mediator in the model and applying the difference of coefficient method to calculate the indirect effect of the exposure (i.e., the effect of the exposure via the mediator) on the outcome. (74) To apply the difference of coefficient method, you must perform both a univariate MR analysis of the exposure on the outcome to calculate the total effect (i.e., the effect of the exposure via all pathways to the outcome) and a multivariable MR to calculate the direct effect of the exposure on the outcome. The direct effect can then be subtracted from the total effect to obtain an estimate of the indirect effect. Again, the confidence intervals for the indirect effect estimate must be updated accordingly. This approach to calculating mediating effects has been demonstrated multiple times in the literature, including to study the extent to which adult body size mediates the influence of early life adiposity on breast cancer, cardiovascular disease and adult systemic metabolism. (77–79)

A NOTE ON INDIVIUDUAL-LEVEL DATA

This review has primarily focused on causal inference using summary-level GWAS data. However, in the current era, it is possible to access individual-level on a massive scale from resources such as the UK Biobank (80). These resources provide access to both individual-level genome-wide genetic associations and detailed clinical phenotyping from questionnaires and/or electronic health records. Many of the methods discussed in this review, such as MR and its extensions, can readily be applied in the individual-level data setting. Furthermore, the availability of individual-level data provides new opportunities. For instance, a novel gene-by-environment (GxE) MR-PheWAS approach has been proposed that combines MR-PheWAS with the use of GxE interactions and has successfully been applied to interrogate the causal effect of smoking heaviness on facial aging. (81)

CHALLENGES

Using GWAS data for causal inference presents several challenges. Firstly, most GWAS to date have been conducted in individuals of European ancestry. This means most causal inference using GWAS data to date has also been conducted in individuals of European ancestry. Availability of GWAS in multiple ethnicities can allow the identification of ancestry-specific and cross-ancestry signals that are informative for disease. (82) Resources, such as the Million Veteran Program, are enabling access to large groups of individuals from non-European ancestries, including African Americans and Hispanic Americans. (83) However, until GWAS in non-European ancestries become widespread, causal inference using this type of data will be limited.

Many of the methods discussed in this review require access to full GWAS summary statistics. However, it is not uncommon for summary statistics to be limited to the genome-wide significant results reported in a GWAS publication. This issue has previously been highlighted in a blog post by the NHGRI-EBI GWAS Catalog, who provide a database of GWAS summary statistics, and report that only 7% of GWAS studies for cancer traits provide full summary statistics. (84,85) This seriously constrains causal inference work that can be carried out using GWAS data in the cancer field.

A further challenge for causal inference using GWAS data is the types of traits that have been studied using GWAS approaches. This limits the causal inference questions that can be answered using this type of data. For instance, it has been recognised that most GWAS to date relate to disease incidence traits, rather than disease progression traits. (86) This means that we are often unable to make causal inferences regarding risk and interventions once an individual has a disease.

GWAS of binary phenotypes can also be problematic for causal inference. Firstly, they often represent a dichotomization of a latent continuous trait. This can make the effect of binary traits difficult to interpret clinically as they represent 'liability to' the binary phenotype of interest and can violate the assumptions of some methods – for example, the exclusion restriction (IV3) assumption of MR. (87) Secondly, the mathematics underlying some causal inference methods relies on traits being continuous. When this is not the case, non-collapsibility can become an issue. For instance, the use of a binary mediator in two-step MR for mediation is limited by the non-collapsibility of odds ratios. (73,74)

For GWAS to have utility in downstream causal inference analyses, they must be well documented and easily accessible. Consequently, there have been recommendations made regarding common storage formats (for example, the variant call format – an open file format that incorporates metadata at the start of the file) and standardized, informative headings for data (for example, 'effect allele' instead of 'reference allele'). (88,89) Databases that bring together multiple GWAS in a single place, such as the previously mentioned IEU OpenGWAS and NHGRI-EBI GWAS Catalog databases, have led to improvements in both documentation and access. (42,84) However, continued focus on these issues going forward will allow us to fully realise the potential for causal inference using GWAS data.

GWAS are able to identify 'small but common' effects of individual SNPs, while 'large but rare effects' are likely to be of greater clinical impact for patients. Despite this, small but common effects remain important for our understanding of disease aetiology. (5) In light of this, we must recognise what can be realistically achieved when using GWAS data and draw any causal inferences accordingly.

CONCLUSIONS

Many opportunities exist to use GWAS data for causal inference. As demonstrated here, these methods may be considered as escalating in their causal complexity from the humble genetic association to extensions of MR, such as MR-PheWAS, which allow consideration of thousands of phenotypes simultaneously. The main challenges facing those attempting to perform causal inference using GWAS data

relate to representation - in terms of ancestry and disease state (progression versus incidence) – and the availability of full summary statistics to implement these methods. As the community addresses these challenges, causal inference using GWAS data can increase its impact and continue to contribute to the triangulation of evidence for causal questions in the medical and public health literature. (90,91)

REFERENCES

1. Rothman KJ, Greenland S. Causation and Causal Inference in Epidemiology. Am J Public Health. 2005 Jul 1;95(S1):S144–50.

2. Hernán M, Robins J. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2020.

3. Davey Smith G, Richmond R, Pingault J-B. Combining Human Genetics and Causal Inference to Understand Human Disease and Development. Cold Spring Harbor Perspectives in Medicine; 2021.

4. Karczewski KJ, Martin AR. Analytic and Translational Genetics. Annual Review of Biomedical Data Science. 2020;3(1):217–41.

5. Frayling T. Genome-wide association studies: the good, the bad and the ugly. Clin Med (Lond). 2014 Aug;14(4):428–31.

6. Ho SS, Urban AE, Mills RE. Structural variation in the sequencing era. Nat Rev Genet. 2020 Mar;21(3):171–89.

7. Klarin D, Tsao PS, Damrauer SM. Genetic Determinants of Peripheral Artery Disease. Circulation Research. 2021 Jun 11;128(12):1805–17.

8. Vilhjálmsson BJ, Yang J, Finucane HK, Gusev A, Lindström S, Ripke S, et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. The American Journal of Human Genetics. 2015 Oct 1;97(4):576–92.

9. Ritchie SC, Lambert SA, Arnold M, Teo SM, Lim S, Scepanovic P, et al. Integrative analysis of the plasma proteome and polygenic risk of cardiometabolic diseases. bioRxiv. 2021 Feb 19;2019.12.14.876474.

10. Richardson TG, Harrison S, Hemani G, Davey Smith G. An atlas of polygenic risk score associations to highlight putative causal relationships across the human phenome. Loos R, McCarthy MI, editors. eLife. 2019 Mar 5;8:e43657.

11. Timpson NJ, Greenwood CMT, Soranzo N, Lawson DJ, Richards JB. Genetic architecture: the shape of the genetic contribution to human traits and disease. Nat Rev Genet. 2018 Feb;19(2):110–24.

12. Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. Nat Rev Genet. 2018 Aug;19(8):491–504.

13. Broekema RV, Bakker OB, Jonkers IH. A practical view of fine-mapping and gene prioritization in the post-genome-wide association era. Open Biology. 10(1):190221.

14. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nature Genetics. 2018 Nov;50(11):1505–13.

15. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. Nature. 2014 Mar;507(7493):455–61.

16. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. PLOS Genetics. 2014 May 15;10(5):e1004383.

17. Wallace C. Eliciting priors and relaxing the single causal variant assumption in colocalisation analyses. PLOS Genetics. 2020 Apr 20;16(4):e1008720.

18. Wallace C. A more accurate method for colocalisation analysis allowing for multiple causal variants. bioRxiv. 2021 May 17;2021.02.23.432421.

19. Zheng J, Haberland V, Baird D, Walker V, Haycock PC, Hurle MR, et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. Nature Genetics. 2020 Sep 7;1–10.

20. Fortune MD, Guo H, Burren O, Schofield E, Walker NM, Ban M, et al. Statistical Colocalization of Genetic Risk Variants for Related Autoimmune Diseases in the Context of Common Controls. Nat Genet. 2015 Jul;47(7):839–46.

21. Giambartolomei C, Zhenli Liu J, Zhang W, Hauberg M, Shi H, Boocock J, et al. A Bayesian framework for multiple trait colocalization from summary association statistics. Bioinformatics. 2018 Aug 1;34(15):2538–45.

22. Foley CN, Staley JR, Breen PG, Sun BB, Kirk PDW, Burgess S, et al. A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. Nat Commun. 2021 Feb 3;12(1):764.

23. Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics. 2017 Jan 15;33(2):272–9.

24. Bulik-Sullivan BK, Loh P-R, Finucane H, Ripke S, Yang J, Patterson N, et al. LD Score Regression Distinguishes Confounding from Polygenicity in Genome-Wide Association Studies. Nat Genet. 2015 Mar;47(3):291–5.

25. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet. 2015 Nov;47(11):1228–35.

26. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015 Nov;47(11):1236–41.

27. Dardani C, Howe LJ, Mukhopadhyay N, Stergiakouli E, Wren Y, Humphries K, et al. Cleft lip/palate and educational attainment: cause, consequence or correlation? A Mendelian randomization study. International Journal of Epidemiology. 2020 Aug 1;49(4):1282–93.

28. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003 Jan 2;32(1):1–22.

29. Davey Smith G, Holmes MV, Davies NM, Ebrahim S. Mendel's laws, Mendelian randomization and causal inference in observational data: substantive and nomenclatural issues. Eur J Epidemiol. 2020;35(2):99–111.

30. Smith GD, Lawlor DA, Harbord R, Timpson N, Rumley A, Lowe GDO, et al. Association of C-Reactive Protein With Blood Pressure and Hypertension. Arteriosclerosis, Thrombosis, and Vascular Biology. 2005 May 1;25(5):1051–6.

31. Davey Smith G, Lawlor DA, Harbord R, Timpson N, Day I, Ebrahim S. Clustered Environments and Randomized Genes: A Fundamental Distinction between Conventional and Genetic Epidemiology. PLOS Medicine. 2007 Dec 11;4(12):e352.

32. Davies NM, Howe LJ, Brumpton B, Havdahl A, Evans DM, Davey Smith G. Within family Mendelian randomization studies. Human Molecular Genetics. 2019 Nov 21;28(R2):R170–9.

33. Brumpton B, Sanderson E, Heilbron K, Hartwig FP, Harrison S, Vie GÅ, et al. Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. Nat Commun. 2020 Jul 14;11(1):3519.

34. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet. 2014 Sep 15;23(R1):R89–98.

35. Labrecque J, Swanson SA. Understanding the Assumptions Underlying Instrumental Variable Analyses: a Brief Review of Falsification Strategies and Related Tools. Curr Epidemiol Rep. 2018 Sep 1;5(3):214–20.

36. Richmond RC, Davey Smith G. Mendelian Randomization: Concepts and Scope. Cold Spring Harb Perspect Med. 2021 Aug 23;a040501.

37. Skrivankova VW, Richmond RC, Woolf BAR, Davies NM, Swanson SA, VanderWeele TJ, et al. Strengthening the reporting of observational studies in epidemiology using mendelian randomisation (STROBE-MR): explanation and elaboration. BMJ. 2021 Oct 26;375:n2233.

38. Skrivankova VW, Richmond RC, Woolf BAR, Yarmolinsky J, Davies NM, Swanson SA, et al. Strengthening the Reporting of Observational Studies in Epidemiology Using Mendelian Randomization: The STROBE-MR Statement. JAMA. 2021 Oct 26;326(16):1614–21.

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

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

41. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. eLife. 2018 May 30;7:e34408.

42. Elsworth B, Lyon M, Alexander T, Liu Y, Matthews P, Hallett J, et al. The MRC IEU OpenGWAS data infrastructure. 2020 Aug 10 [cited 2020 Nov 5]; Available from: http://biorxiv.org/lookup/doi/10.1101/2020.08.10.244293

43. Walker VM, Davey Smith G, Davies NM, Martin RM. Mendelian randomization: a novel approach for the prediction of adverse drug events and drug repurposing opportunities. Int J Epidemiol. 2017 Dec 1;46(6):2078–89.

44. Ference BA, Yoo W, Alesh I, Mahajan N, Mirowska KK, Mewada A, et al. Effect of Long-Term Exposure to Lower Low-Density Lipoprotein Cholesterol Beginning Early in Life on the Risk of Coronary Heart DiseaseA Mendelian Randomization Analysis. J Am Coll Cardiol. 2012 Dec 25;60(25):2631–9.

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

46. White J, Swerdlow DI, Preiss D, Fairhurst-Hunter Z, Keating BJ, Asselbergs FW, et al. Association of Lipid Fractions With Risks for Coronary Artery Disease and Diabetes. JAMA Cardiol. 2016 Sep 1;1(6):692.

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

48. Swerdlow DI, Preiss D, Kuchenbaecker KB, Holmes MV, Engmann JEL, Shah T, et al. HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: evidence from genetic analysis and randomised trials. Lancet. 2015 Jan 24;385(9965):351–61.

49. Ference BA, Robinson JG, Brook RD, Catapano AL, Chapman MJ, Neff DR, et al. Variation in PCSK9 and HMGCR and Risk of Cardiovascular Disease and Diabetes. New England Journal of Medicine. 2016 Dec 1;375(22):2144–53.

50. Holmes MV, Smith GD. Revealing the effect of CETP inhibition in cardiovascular disease. Nat Rev Cardiol. 2017 Nov;14(11):635–6.

51. Holmes MV, Ala-Korpela M, Davey Smith G. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. Nature Reviews Cardiology. 2017 Oct;14(10):577–90.

52. Davey Smith G, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. International Journal of Epidemiology. 2004 Feb 1;33(1):30–42.

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

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

55. Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. Statist Med. 2008 Apr 15;27(8):1133–63.

56. Zhao Q, Wang J, Hemani G, Bowden J, Small DS. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. arXiv:180109652 [math, stat] [Internet]. 2018 Jan 29 [cited 2018 Jul 2]; Available from: http://arxiv.org/abs/1801.09652

57. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. Int J Epidemiol. 2017 Dec 1;46(6):1985–98.

58. Hemani G, Tilling K, Smith GD. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. PLOS Genetics. 2017 Nov 17;13(11):e1007081.

59. Cho Y, Haycock PC, Sanderson E, Gaunt TR, Zheng J, Morris AP, et al. Exploiting horizontal pleiotropy to search for causal pathways within a Mendelian randomization framework. Nat Commun. 2020 Feb 21;11(1):1010.

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

61. Spiller W, Slichter D, Bowden J, Davey Smith G. Detecting and correcting for bias in Mendelian randomization analyses using Gene-by-Environment interactions. International Journal of Epidemiology. 2019 Jun 1;48(3):702–12.

62. Smith GD. Mendelian Randomization for Strengthening Causal Inference in Observational Studies: Application to Gene × Environment Interactions. Perspect Psychol Sci. 2010 Sep 1;5(5):527–45.

63. Davey Smith G, Davies NM, Dimou N, Egger M, Gallo V, Golub R, et al. STROBE-MR: Guidelines for strengthening the reporting of Mendelian randomization studies. PeerJ Preprints. 2019 Jul 15;

64. O'Connor LJ, Price AL. Distinguishing genetic correlation from causation across 52 diseases and complex traits. Nat Genet. 2018 Dec;50(12):1728–34.

65. Hatoum AS, Johnson EC, Agrawal A, Bogdan R. Brain structure and problematic alcohol use: a test of plausible causation using latent causal variable analysis. Brain Imaging and Behavior [Internet]. 2021 Jul 21 [cited 2021 Sep 20]; Available from: https://doi.org/10.1007/s11682-021-00482-z

66. Carreras-Torres R, Johansson M, Haycock PC, Relton CL, Davey Smith G, Brennan P, et al. Role of obesity in smoking behaviour: Mendelian randomisation study in UK Biobank. BMJ. 2018 May 16;361:k1767.

67. Millard LA, Davies NM, Gaunt TR, Davey Smith G, Tilling K. Software Application Profile: PHESANT: a tool for performing automated phenome scans in UK Biobank. International Journal of Epidemiology. 2018 Feb 1;47(1):29–35.

68. Hemani G, Bowden J, Haycock P, Zheng J, Davis O, Flach P, et al. Automating Mendelian randomization through machine learning to construct a putative causal map of the human phenome. bioRxiv. 2017 Aug 23;173682.

69. Liu Y, Elsworth B, Erola P, Haberland V, Hemani G, Lyon M, et al. EpiGraphDB: a database and data mining platform for health data science. Bioinformatics. 2021 May 1;37(9):1304–11.

70. Millard LAC, Davies NM, Timpson NJ, Tilling K, Flach PA, Davey Smith G. MR-PheWAS: hypothesis prioritization among potential causal effects of body mass index on many outcomes, using Mendelian randomization. Scientific Reports. 2015 Nov 16;5:16645.

71. Relton CL, Davey Smith G. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. Int J Epidemiol. 2012 Feb;41(1):161–76.

72. Gill D, Zuber V, Dawson J, Pearson-Stuttard J, Carter AR, Sanderson E, et al. Risk factors mediating the effect of body-mass index and waist-to-hip ratio on cardiovascular outcomes: Mendelian randomization analysis. 2020 Jul 16 [cited 2020 Jul 29]; Available from: http://medrxiv.org/lookup/doi/10.1101/2020.07.15.20154096

73. Walker V, Vujkovic M, Carter AR, Davies NM, Udler M, Levin M, et al. Separating the direct effects of risk factors for atherosclerotic cardiovascular disease from those mediated by type 2 diabetes. medRxiv. 2021 Aug 7;2021.08.05.21261658.

74. Carter AR, Sanderson E, Hammerton G, Richmond RC, Davey Smith G, Heron J, et al. Mendelian randomisation for mediation analysis: current methods and challenges for implementation. Eur J Epidemiol. 2021 May 1;36(5):465–78.

75. Anderson EL, Howe LD, Wade KH, Ben-Shlomo Y, Hill WD, Deary IJ, et al. Education, intelligence and Alzheimer's disease: evidence from a multivariable two-sample Mendelian randomization study. International Journal of Epidemiology. 2020 Aug 1;49(4):1163–72.

76. Sanderson E, Davey Smith G, Windmeijer F, Bowden J. An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. International Journal of Epidemiology. 2019 Jun 1;48(3):713–27.

77. Richardson TG, Sanderson E, Elsworth B, Tilling K, Smith GD. Use of genetic variation to separate the effects of early and later life adiposity on disease risk: mendelian randomisation study. BMJ. 2020 May 6;369:m1203.

78. Richardson TG, Mykkänen J, Pahkala K, Ala-Korpela M, Bell JA, Taylor K, et al. Evaluating the direct effects of childhood adiposity on adult systemic metabolism: a multivariable Mendelian randomization analysis. International Journal of Epidemiology. 2021 Mar 30;

79. Power GM, Tyrrell J, Frayling TM, Davey Smith G, Richardson TG. Mendelian Randomization Analyses Suggest Childhood Body Size Indirectly Influences End Points From Across the Cardiovascular Disease Spectrum Through Adult Body Size. Journal of the American Heart Association. 2021 Sep 7;10(17):e021503.

80. Allen NE, Sudlow C, Peakman T, Collins R, UK Biobank. UK biobank data: come and get it. Sci Transl Med. 2014 Feb 19;6(224):224ed4.

81. Millard LAC, Munafò MR, Tilling K, Wootton RE, Smith GD. MR-pheWAS with stratification and interaction: Searching for the causal effects of smoking heaviness identified an effect on facial aging. PLOS Genetics. 2019 Oct 31;15(10):e1008353.

82. Wyss AB, Sofer T, Lee MK, Terzikhan N, Nguyen JN, Lahousse L, et al. Multiethnic meta-analysis identifies ancestry-specific and cross-ancestry loci for pulmonary function. Nat Commun. 2018 Jul 30;9(1):2976.

83. Gaziano JM, Concato J, Brophy M, Fiore L, Pyarajan S, Breeling J, et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. Journal of Clinical Epidemiology. 2016 Feb;70:214–23.

84. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. 2017 Jan 4;45(Database issue):D896–901.

85. Why we need more freely available cancer GWAS summary statistics [Internet]. Open Targets Blog. 2021 [cited 2021 Jul 30]. Available from: https://blog.opentargets.org/open-sharing-of-cancer-summary-statistics/

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

87. Tudball MJ, Bowden J, Hughes RA, Ly A, Munafò MR, Tilling K, et al. Mendelian randomisation with coarsened exposures. Genetic Epidemiology. 2021;45(3):338–50.

88. Lyon MS, Andrews SJ, Elsworth B, Gaunt TR, Hemani G, Marcora E. The variant call format provides efficient and robust storage of GWAS summary statistics. Genome Biology. 2021 Jan 13;22(1):32.

89. Wootton RE, Sallis HM. Let's call it the effect allele: a suggestion for GWAS naming conventions. International Journal of Epidemiology. 2020 Oct 1;49(5):1734–5.

90. Munafò M, Davey Smith G. Robust research needs many lines of evidence. Nature [Internet]. 2018 Jan 23 [cited 2018 Jan 23]; Available from: http://www.nature.com/articles/d41586-018-01023-3

91. Munafò MR, Higgins JPT, Smith GD. Triangulating Evidence through the Inclusion of Genetically Informed Designs. Cold Spring Harb Perspect Med. 2021 Aug 2;11(8):a040659.