Heritable confounding in Mendelian randomization studies: structure, consequences and relevance for gene-environment equivalence

Eleanor Sanderson[1,2], Dan Rosoff[1,3,4], Tom Palmer[1,2], Kate Tilling[1,2], George Davey Smith[1,2]* and Gibran Hemani[1,2]*

1. MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK.
2. Population Health Sciences, University of Bristol, Bristol, UK.
3. Section on Clinical Genomics and Experimental Therapeutics, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda, MD, USA
4. Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK

Corresponding author: Eleanor.sanderson@bristol.ac.uk
*These authors contributed equally to the project and are joint last author

Abstract

Mendelian randomization (MR) uses genetic variants to estimate the causal effect of an exposure on an outcome in the presence of unmeasured confounding. A key assumption of MR is that the genetic variants used influence the outcome only through the exposure. Violation of this assumption undermines the gene-environment equivalence principle, which posits that modifying the exposure via genetic variation is equivalent to modifying it through environmental factors. With increasing sample sizes in genome-wide association studies genetic instruments with smaller effect sizes are being identified as associated with a trait. Through simulation studies, we demonstrate that such variants may have greater liability to act through confounders of the exposure and outcome in a MR study, biasing effect estimates. This bias acts in the same direction as the confounded associations observed in linear regression, but often with greater magnitude and acts in the same direction across all of the most commonly used MR estimation methods, potentially leading to misleading confidence in the results. We further show that the magnitude of bias escalates as the proportion of genetic instruments associated with confounders increases. Importantly, when potential heritable confounders the genetic variants act through are known and can be instrumented, unbiased causal estimates can be obtained through pre-estimation filtering or by employing multivariable MR and adjusting for the confounder. We illustrate our approach through an application to estimate the effect of C Reactive protein on type 2 diabetes using a hypothesis free approach to identify and remove the effect of potential heritable confounders.

Epidemiological studies often aim to estimate causal effects from observational data. A common method is multivariable regression, which provides an unbiased estimate if all confounders are known and measured without error, an assumption that is rarely plausible.[1] Mendelian Randomisation (MR) offers an alternative approach to causal effect estimation by using genetic variants as instruments to infer the causal effect of an exposure on an outcome, reducing bias from unmeasured confounding and measurement error.[2-4] MR is typically implemented within an instrumental variable (IV) framework, using genetic variants associated with the exposure trait as IV's.[4]

IV estimation relies on three core assumptions (illustrated in **Figure 1A**). Taken together these assumptions imply gene-environment equivalence, i.e. if they all hold then changing the exposure via genetic variation will lead to the same change in the outcome as changing the exposure via environmental change would. In MR, single nucleotide polymorphisms (SNPs) identified as associated with the exposure in genome-wide association studies (GWAS) are usually used as instruments. Typically, SNPs are selected if their association p-value with the exposure is $<5\times10^{-8}$, a Bonferroni-corrected threshold for genome-wide significance.

The polygenic nature of complex traits, for which there is substantial empirical support [5], posits that for most traits the majority of heritable variation is driven by a very large number of genetic factors, each with very small effect sizes. To discover such variants GWAS have been growing in sample size rapidly in recent years, especially through the emergence of large scale biobanks. For example, a recent GWAS of height meta-analysed data from 5.4 million individuals.[6] Correspondingly, increasing numbers of variants are being identified as associated with the traits studied in these GWAS.

The structure of "heritable confounding"

For MR analyses involving exposures that are themselves complex traits, genetic variants associated with the exposure are mediated through several phenotypic layers.[7] Germline genetic variaton influences the traits that are generally instrumented in MR through a wide variety of phenotypic factors: gene expression levels, mRNA splicing, DNA methylation, protein abundance, post-translational protein modifications, enzyme activity, etc.[8] All of these can be tissue specific, with both their expression and their downstream effects depending on corporal location. Such mediation clearly opens the door to horizontal pleiotropy, which biases MR estimates by violating assumption IV3 (see **Figure 1A**).

There are a number of different forms that horizontal pleiotropy can take that will lead to bias in MR analysis.[4] Note that whether or not horizontal pleiotropy distorts a MR analysis will depend on the outcome being studied. For some outcomes an upstream factor might influence the outcome – in which case the MR findings will be biased, but on some other outcome the same upstream factor might have no effect, and the MR analysis would be valid. This means that a variant is not valid or invalid for all MR analyses it is used in; validity can depend on the outcome being studied.

Correlated pleiotropy is used to describe pleiotropy that is correlated with the magnitude of association between the genetic variants and the exposure. A heritable confounder can generate correlated pleiotropy, as illustrated in **Figure 1B.** Heritable confounding and misspecification of the primary phenotype (i.e. the forms of pleiotropy illustrated by **Figure 1B(ii)** and **1B(iii)**) are not always consistently defined and so the distinction between them is not always clear. We refer to heritable confounding as a situation where some of the genetic instruments used are associated with a confounder of the exposure and outcome and misspecification of the primary phenotype as the situation where all of the genetic instruments act on the exposure through another trait that also has an effect on the outcome. A key issue is whether the heritable confounder is measured and known, or is an anonymous unidentified factor. In the former case an informed researcher would not take the genetic variant to be an valid instrument for the exposure. In the latter case they may do so and this has previously been categorised as misspecification of the primary phenotype[9]. However many MR studies currently do not explore the potential for heritable confounding in their analysis.

Reverse causation is a particular category of heritable confounding, in which the confounder is either the outcome itself (e.g. myocardial infarction) or the disease process leading up to the outcome (e.g. atherosclerosis). The distinction here is important, because often the actual outcome cannot generate reverse causation. Consider the situation when the outcome is death from myocardial infarction: death can clearly not influence any phenotype in the living, so reverse causation from the outcome is not possible, but heritable confounding from the disease process is. This paper considers the consequences of heritable confounding for MR studies, which relates to the correlated pleiotropy that it leads to. For example, note that the presence of correlated pleiotropy means that the InSIDE assumption of the MR Egger approach will be violated.[10] More fundamentally, heritable confounding may indicate the degree to which the key MR assumption of gene-environment equivalence fails to hold.[4, 11, 12]

In this paper we explore heritable confounding in MR studies. First, we compare the bias in MR studies to the bias from residual confounding in linear regression, showing that bias in MR estimation from a single heritable confounder can exceed the bias from the same confounder in multivariable regression. We then explore the question of whether increasing GWAS sample sizes could increase the rate of correlated pleiotropy through a simulation of a network of traits explained by set of genetic variants. We go on to evaluate how this increase in correlated pleiotropy impacts standard MR estimators in a series of simulations. In settings where potential confounders are hypothesised and measurable, we propose two approaches to mitigate bias: (1) multivariable MR (MVMR), including the confounder as an additional exposure, and (2) Steiger filtering, removing SNPs that explain more variance in the confounder than the exposure.[13] To illustrate the potential use of these approaches we apply both methods, along with existing approaches, to estimate the effect of C-reactive protein (CRP) on type 2 diabetes, using phenome-wide association studies (PheWAS) to identify potential heritable confounders.

Alternative MR methods have been developed to help identify and/or adjust for correlated pleiotropy under specific assumptions about its structure.[14-24] These approaches are often computationally intensive, using Bayesian estimation approaches applied to a large number of genetic variants. They also impose an assumption of a structure of the relationship between the

pleiotropic variants and other variants, often assuming that the plurality of the instruments are valid. The approaches we propose here are already widely used in other settings and are simple to apply. We therefore propose that they should be systematically used within MR studies to identify how sensitive the results obtained are to the most likely confounders.

Bias in IVW MR estimates compared to linear regression

In this section we explore the bias of MR compared to the equivalent linear regression in the presence of a single unmeasured heritable confounder. Although a simplification this highlights the effect the single confounder will have on each estimation. We consider a simple model of a relationship between an exposure (X) and an outcome (Y) which is confounded by a single unobserved heritable confounder (U). The exposure is associated with a single instrument (Z) that also has an effect on the outcome via the unobserved confounder. We assume that all relationships are linear and there are no interactions. We focus on the setting with a single heritable confounder for simplicity and clarity the results given here therefore only compare the results for bias from a single heritable confounder and not the overall bias that would be expected. The model we consider can be written as;

$$Y = \beta_{0y} + \beta_{xy}X + \beta_{uy}U + v_y$$
$$X = \beta_{0x} + \pi_{zx}Z + \beta_{ux}U + v_x$$
$$U = \beta_{0u} + \pi_{zu}Z + v_u$$

Where v_y, v_x and v_u are all uncorrelated normally distributed random error terms. This model is illustrated in **Figure 2A**.

We show in **Supplementary material section 1.1** that the bias in the MR estimate from a single heritable confounder will be less than the bias of the linear regression estimator from the same confounder when;

$$\frac{\beta_{uy}\pi_{zu}}{\pi_{zx} + \beta_{ux}\pi_{zu}} < \beta_{uy}\delta_{ux} \#(1)$$

Where δ_{ux} is the association between U and X. The implication of equation (1) is that the bias depends on the ratio of the pleiotropic effect of the instrument on the outcome ($\beta_{uy}\pi_{zu}$) to the effect of the instrument on the exposure ($\pi_{zx} + \beta_{ux}\pi_{zu}$). The larger the pleiotropic effect relative to the effect of the instrument on the exposure the larger the bias of the MR estimator will be.

To illustrate the relationship between the bias of linear regression and IV estimation across different levels of pleiotropy we simulated the model given in **Figure 2A** varying the effect of the instrument on the confounder but holding all other parameters constant. The mean bias of each estimator across 1000 repetitions for varying levels of confounding is given in **Figure 2B**. These results illustrate how the bias of the IV estimator increases as the pleiotropic effect via the confounder increases and can be substantially more than the bias in the purely observational linear regression. The exact position and slope of these lines will depend on the other parameters in the model.

Figures 2C and **2D** show the same bias plots as **Figure 2B** with an additional non-heritable confounder added to the model. The direction of the confounding from this additional confounder is varied from positive (**Figure 2C**) to negative (**Figure 2D**). These plots highlight how the overall bias of each estimator will depend on other factors as well as the level of heritable confounding and the bias of the estimate from MR may not always act in the same direction as the bias of the estimate from linear regression.

Network simulations

There remains a question of whether increasing GWAS sample sizes will lead to increased horizontal pleiotropy through heritable confounding. One hypothesised mechanism by which this could arise is if genetic variants that have effects on the exposure that are mediated through other traits tend to be those that are more phenotypically connected in the overall genotype-phenotype map. Such variants would be likely to have smaller effects on a trait and so only become discoverable as GWAS sample sizes increase. If any of these upstream mediating traits have another path to the outcome that does not act through the exposure then the result will be heritable confounding in MR estimation. An implication of this is that the genetic variants with the largest effect sizes in the GWAS may be the least likely to be those which lead to heritable confounding and correlated pleiotropy through an upstream phenotype.[25] This is illustrated in **Figure 3a**, where the variants with the shortest path to the exposure do not lead to correlated pleiotropy but the variant with the longest path to the exposure does.

To explore whether this is likely to be the case we performed simulations in which we consider that all hypothesised exposure-outcome relationships exist in a broader genotype-phenotype network. We use the randomDAG function in the R/dagitty package to simulate a DAG with an arbitrary number of nodes, and a fixed probability that any set of nodes is connected. Nodes with no parents in the simulated DAG are identified as genetic variants, and 500 trait pairs are selected at random, in each pair assigning one to be the hypothesised exposure and one to be the hypothesised outcome. We then identify the path length for every genetic variant for the exposure, and also determine whether they act via a confounder, or otherwise (i.e. directly or through other traits that mediate the association between the genetic variant and the exposure and have no other paths to the outcome). The effect size of the genetic variant on the exposure is β^L where L is the path length from genotype to exposure and, for simplicity, every edge in the DAG has the same effect $\beta = 0.2$. These simulations were conducted for a set of parameters of graph size N=75, 100, 125, 150 and graph density based on probability of an edge occurring ranging from p = 0.01, 0.02, …, 0.1. Each parameter combination was repeated 10 times.

The results from this simulation are given in **Figure 3.** These results illustrate that in this model instruments that act via confounders tend to have smaller effects due to more traits lying on the path from the genetic variant to the exposure. For traits where our simulation is a realistic model of the underlying biology, a key implication of this result is that as sample sizes grow for GWAS the rate of discovery of genetic instruments that are liable to act via confounders will grow. However, it is possible that this will not hold for all traits, for some traits it may be that the strongest effects are via

confounders (e.g. because those confounders have very strong effects on the exposure) and the genetic variants with smaller effects will be the unconfounded ones. It is therefore important to consider if this model is likely to hold for each exposure trait considered.

Simulations evaluating standard MR methods under heritable confounding

We simulated a model where the exposure and outcome are confounded by a third, potentially unobserved, heritable trait. In this model, the exposure has no causal effect on the outcome. The exposure is influenced by two upstream traits: one is a mediator, which simply transmits genetic effects to the exposure, while the other is a heritable confounder affecting both exposure and outcome. Note that if a different outcome were being studied it is possible that the first upstream trait mentioned above is the one that generates bias due to horizontal pleiotropy, and the second upstream trait that does not do so.

Due to the confounding, linear regression of the exposure on the outcome would give a positive estimate of the association between the exposure and outcome due to confounding. The mediator and confounder are both influenced by a separate set of genetic variants. We assume no non-heritable confounding and that all instrumental variable (IV) assumptions otherwise hold. The model corresponds to **Figure 1B(ii)** and is illustrated in **Figure 4**.

We simulated two independent samples of 100,000 individuals each. In each sample, 400 SNPs were generated from a binomial distribution with the proportion of SNPs associated with the mediator and confounder varying across the simulations—ranging from a 50% associated with each to 12.5% associated with the mediator and 87.5% associated with the confounder.

SNP effects on the mediator or confounder were drawn from a normal distribution (mean = 0, SD = 0.04), orientated so all SNPs had a positive association with the relevant trait. The confounder and mediator were generated as the sum of their respective SNP effects plus a normally distributed error term. The exposure was generated as a linear combination of the mediator, confounder, and an error term. As shown earlier, SNPs associated with a mediator may be more strongly associated with the exposure than SNPs associated with a confounder. Therefore, we downweighted the confounder – exposure association by 25% relative to the mediator – exposure association to mimic this more distant association. Results with this down weighting varied to 0% and 50% are given in the **Supplementary Figure 1** and **Supplementary Figure 2** respectively. Finally, the outcome was generated as the effect of the confounder and a random error term. The expressions for the data generation process are given in **Supplementary material section 1.2**.

GWAS summary statistics were generated by regressing each SNP on the exposure and confounder in Sample 1, and on the outcome in Sample 2. SNPs (from either set) were selected to be used as instruments in our analysis based on their association with the exposure in Sample 1. Increasing the sample size of a GWAS will decrease the standard error in the SNP-trait association proportionally across all SNPs leading to more SNPs being selected based on a fixed p-value threshold cut-off. Therefore, to model varying the sample size in the GWAS while keeping all other aspects of the data constant we varied the p-value threshold used to select SNPs for use as instruments in the MR

across our simulations with larger p-values being equivalent to a larger sample size. The same sample was used for SNP selection and exposure effect estimation, reflecting standard MR practice.

Using the summary statistics generated we applied four common univariable MR methods; Inverse Variance Weighted (IVW)[26], MR Egger[10], Weighted Median[27], and Weighted Mode[28]. We also applied Multivariable MR (MVMR)[29, 30], including the confounder as an additional exposure and SNPs associated with the confounder as additional instruments. In addition, we repeated all univariable MR methods with Steiger filtering[13] applied to exclude SNPs that explained more variation in the confounder than in the exposure.

Results from this simulation are given in **Figure 4** and **Supplementary Tables 1** and **2**. When the total proportion of the variation in the exposure is explained by genetic variants that act on that exposure through the confounder is moderate IVW and MR Egger are biased across all p-value selection thresholds. However both weighted median and weighted mode correctly estimate the causal effect of the exposure on the outcome as their assumption about the nature of any pleiotropy is still satified (**Figure 4, Supplementary Table 1**). When almost all of the SNPs are associated with the confounder all univariable methods of estimation are biased and give results that are consistent with each other (**Figure 4, Supplementary Table 1**). With a stricter p-value cut-off, equivalent to a smaller sample size, MR Egger is very imprecise. This is expected as a stricter p-value cut off will mean fewer SNPs are included in the estimation and MR Egger has low power when the number of SNPs included in the estimation is small. In all cases both MVMR and the univariable MR approaches with Steiger filtering between the exposure and confounder applied gave unbiased estimates of the causal effect of the exposure on the outcome. However, each of these methods requires the confounder to be known and have available GWAS summary statistics.

Table 1 shows the mean number of genetic variants in each simulation that were selected for the univariable MR estimation methods (IVW, MR Egger, weighted median and weighted mode). When the mediator is associated with as many SNPs as the confounder in the data generating process the final proportion of SNPs that are associated with the confounder in the MR estimation is small due to the effects of those SNPs on the exposure being smaller.

In these simulations there is a single confounder of the exposure and outcome of interest biasing the estimates obtained. These results therefore show how the MR effect estimates are biased by that confounder in the same direction as the confounded observational association and may be substantially more biased than the observational association. In practice it is likely that there may be many potential confounders, some of which will be heritable and associated with a subset of the genetic variants selected as instruments. The overall bias in the linear regression will depend on the direction and magnitude of the bias from all of the confounders and so is likely to be substantially more than the bias in the MR estimation from confounding. With multiple potential heritable confounders both Steiger filtering and MVMR could be applied to each of them in turn to identify any bias.

Illustrative application: Estimation of the effect of C-Reactive protein on Type 2 Diabetes

In our illustrative application we considered a potential exposure that has been widely used in MR studies; C-Reactive Protein (CRP), a marker of inflammation.[31-34] A recent GWAS identified a large number of genetic variants associated with CRP across the genome, many of which may be associated due to their influence on upstream traits which influence outcomes independent of CRP, and thus be potential heritable confounders. In this example we use MR to estimate the effect of CRP on type 2 diabetes (T2D) and systematically identify potential confounders and remove SNPs more strongly associated with those confounders than CRP.

SNPs associated with CRP levels were identified from in a recent GWAS of >500,000 individuals with European genetic ancestry.[35] This GWAS identified 728 lead SNPs ($P < 5\times10^{-8}$; LD $r^2 < 0.1$) across 266 loci. To identify traits associated with these loci, we used the FUMA platform[36] and performed a trait look-up in the GWAS Catalog (accessed 10th April 2024). This yielded 33,573 SNP–trait associations among the lead SNPs or variants in strong LD ($r^2 > 0.6$). From this, we curated a set of 5,198 traits for further analysis, prioritizing biomarkers and risk factors over disease outcomes.

We used summary data MR to estimate the effect of each trait identified on CRP using data from the MRC IEU OpenGWAS platform[37]. When only one SNP was available for a trait, we used the Wald ratio; for traits with two or more SNPs, we applied inverse-variance weighted (IVW) MR; and for traits with three or more SNPs, we also included MR Egger[10], Weighted Median[27], and Weighted Mode[28] estimates. **Table 3** summarises the proportion of the analyses conducted that indicated evidence of a causal effect using a heuristic threshold to identify associations that indicate a potential causal effect. Depending on the MR method applied, 20–30% of traits with available instruments showed nominal evidence of an effect on CRP ($p < 0.05$). All MR analyses were conducted in R using the TwoSampleMR package[37, 38] with default settings.

865 traits showing evidence of an association with CRP ($p < 0.01$ using Wald ratio or IVW) were subsequently tested for their effect on T2D, again using Wald ratio or IVW and OpenGWAS data[37, 39]. Traits which showed evidence of an effect on T2D ($p < 0.01$ using Wald ratio or IVW) were retained, resulting in 235 potential heritable confounders.

We estimated the causal effect of CRP on T2D using MR with all genome-wide significant CRP SNPs ($P < 5\times10^{-8}$; $r^2 < 0.001$). To identify bias from heritable confounding, we applied Steiger filtering, removing SNPs that explained more variance in a confounder than in CRP ($p < 0.05$) for each identified potential confounder in tern. **Table 4** gives results for the 10 confounders that had the largest effect on the MR results obtained (measured by IVW p-value). Full results are provided in **Supplementary Section 2**.

Figure 5a shows a scatterplot of the SNPs used in the CRP–T2D MR analysis, with those associated with the top 10 confounders highlighted. These confounders clustered into two major groups: adiposity-related traits and protein biomarkers**.** The consistent appearance of adiposity-related traits suggests adiposity is a key confounder in the CRP–T2D relationship. Excluding SNPs

associated with these traits led to substantial attenuation of the MR effect estimate for CRP on T2D, reducing evidence for a causal relationship.

The protein traits were all associated with the same SNP (*rs429358* located in the APOE locus), raising the possibility that these associations reflect either horizontal pleiotropy or, less likely, mediation through CRP. rs429358 has a substantial influence on CRP (highlighted in **Figure 5a**, beta: 0.265, p-value 1.0×10^{-200}) and when it was included in the analysis (i.e. when the adiposity related confounders where considered but not the protein markers) the weighted mode estimator gave nominally significant effect estimates. Re-estimating the main analysis excluding only this SNP led to a substantial attenuation of the effect estimates and no evidence of a causal effect of CRP on T2D, results given in **Supplementary Table 7**.

This pattern was further supported by MR Corge analysis[25] (**Figure 5b**). MR Corge groups SNPs by the strength of their association with CRP and estimates MR effect estimates by group both individually and by combining groups cumulatively, starting with the strongest group. The group of SNPs most strongly associated with CRP (Group 1) showed no evidence of an effect on T2D. Larger, statistically significant effects only emerge as additional, weaker instruments are added. Notably, Group 1 included *rs2211320*, a variant within the CRP gene locus.

Implications for gene-environment equivalence

The gene-environment equivalence assumption is that the hypothetical counterfactual change in a genetic variant that influences an exposure would produce the same effect on the outcome as would manipulating the exposure through an intervention.[4, 11, 12] When a variant influences the exposure through an upstream phenotype, then the hypothetical counterfactual change in the variant would produce the effects of the change in the upstream phenotype (which would include a change in the exposure of interest), rather than that of changing the exposure of interest alone. The CRP example suggests that cis CRP variants may mimic differences in circulating CRP levels produced by interventions such as antisense oligonucleotides to CRP or selective CRP apheresis.[40, 41] However SNPs which influence CRP through adiposity or a range of protein biomarkers will not produce an effect that is equivalent to the selective targeting of CRP levels. The usual caveat that genetic influences are from conception onwards, whereas interventions in general occur later in life, also applies when considering the reasonableness – or otherwise – of this assumption.[2, 4, 42]

Discussion

In this paper we have shown how correlated pleiotropy due to heritable confounding can bias MR effect estimates when the genetic variants selected as instruments are associated with a confounder of the exposure and outcome of interest. We show that the bias in MR from including instruments associated with a heritable confounder could be greater than the bias in the linear regression from that confounder. In our network simulations we show that as GWAS sample sizes increase genetic variants identified as associated with an exposure are increasingly likely to be associated with an upstream confounder of the exposure and outcome. This will increase the presence of correlated pleiotropy and heritable confounding in MR effect estimates.

If a large proportion of the genetic instruments for an exposure are associated with a confounder of the exposure and outcome, correlated pleiotropy will bias all of the standard pleiotropy robust methods of estimation. In this case each pleiotropy robust method is biased in the same way as the IVW estimator, potentially leading to false confidence in the results obtained. These results are likely to particularly apply to exposures that are downstream of a phenotype (such as BMI) which is associated with many genetic variants and has an influence on a wide range of traits, or where there are likely to be relatively few genetic variants more directly associated with the exposure of interest. The results given here highlight why it is important to consider whether the IV assumptions are likely to hold even when a range of MR estimates give consistent results for the effect of the exposure on the outcome.

We show that the bias from heritable confounding with a known confounder could be corrected through the application of either MVMR including the confounder, or Steiger filtering to remove any genetic variants that explain more variation in the confounder than the exposure. We therefore recommend that researchers applying MR with summary statistics generated from large GWAS studies check the association of their SNPs with the main potential confounders of the exposure and the outcome. This could be done either through a PheWAS approach as taken here or by considering a set of hypothesised potential confounders. Steiger filtering, removing SNPs that explain more variation in the confounder than the exposure, and/or MVMR also including the confounder can then be applied to the MR analysis as an additional robustness test. A substantial difference between the results obtained is an indication of bias in the original MR analysis. However, in practice it is likely that confounders will not be identifiable/measured and there will be a number of confounders acting in this way and so such an approach does not guarantee the remove of all bias from heritable confounding. There are also limitations in the application of MVMR as weak instruments could lead to a change in the effect estimates in the absence of heritable confounding and therefore it is important to assess the robustness of the MVMR estimation when applying this method.[30, 43] The application of methods such as MR Corge[25] can help to identify whether the estimates obtained are robust across SNPs with different magnitude of association with the exposure.

Other approaches that could be applied to help identify bias from heritable confounding include cis MR and methods that adjust for correlated pleiotropy. Cis MR makes use of genetic variants that are in a gene region known to affect a trait such as a protein.[44, 45] This concept can be extended to any setting where the exposure has a clear biological mechanism by which some genetic variants affect the exposure. Where possible comparing the results obtained for the MR estimation with all genetic variants to the results obtained for the genetic variants which have a known biological pathway can help to support the results or highlight the potential for bias in the MR estimate. Recently novel MR methods have been proposed that attempt to adjust for correlated pleiotropy.[14, 16-19, 46] Each of these methods depends on different assumptions about the nature of the correlated pleiotropy and applies a different approach to correct for that pleiotropy. Applying these in addition to the approaches proposed here to identify specific confounders will help to identify whether results are biased by confounding. Approaches such as MRClust that cluster the effect estimates obtained by individual SNPs can also help identify groups of SNPs that are biased by the same heritable confounder if they identify a separate cluster for those SNPs.[47]

In our illustrative example we estimate the effect of CRP on T2D, systematically identifying potential confounders through a PheWAS approaches. This application identified that the observed effect in the MR estimation was likely to be driven by heritable confounding from adiposity, an established risk factor for elevated CRP levels.[48] Notably, a single genomic region tagged by rs429358 in the *APOE* locus emerged as a particularly influential locus in this analysis and was associated with many potential confounders. The APOE locus has been associated with a range of different traits including lipid metabolism[49] and changes in BMI during adulthood[50] strongly raising the possibility that it is a heritable confounder. The prominence of this variant within the analysis underscores how increasing GWAS sample size, while increasing power to detect novel variants, also enables identification of genetic variants with more complex trait associations that are more likely to exhibit pleiotropic effects, increasing the risk of heritable confounding in MR analyses.

While our results show that the MR results can be more biased than linear regression from the exclusion of a single heritable confounder there are many other sources of bias for both types of estimation that should be considered when determining how much weight to give to the results from each type of analysis. It is likely that there will be a number of confounders of the exposure and outcome, if some of these are not associated with the SNPs used as instruments then the MR estimates may still be less biased from confounding than the observational linear regression estimates. Additionally, MR can overcome bias and effect attenuation from measurement error that occurs in linear regression. However, MR is subject to other forms of bias such as population stratification and pleiotropy of other forms which should also be considered in any analysis.

A limitation of the results here is how we have set up the structure of relationships between genetic variants and phenotypes. In our network simulations we assumed that every edge has the same strength of effect. In our simulations of the bias of the different MR estimators we have assumed the same distribution in size of the association between the relevant genetic variants and the exposure and confounder. These assumptions are relatively strong and are not likely to be realistic for every trait. For some traits the confounders may have much stronger genetic effects than the exposure, which will accentuate the biases described here. For other traits the direct effect of the genetic variants on the exposure will be much larger than for the confounders, meaning that bias of the type described here will be less prevalent. Which of these is likely to hold, along with how strongly the confounders are likely to affect the outcome, will determine how much of a problem the results described here are likely to be in any particular case. Each MR study conducted should therefore consider individually how likely there is to be bias from heritable confounding in that setting and what the main confounders causing that bias may be. Findings from MR studies should be interpreted within a triangulation of evidence framework, including falsification strategies such as the use of negative controls, which may trigger further investigation to identify the source of any observed inconsistencies.[51]

Acknowledgements

ES is funded by the Medical Research Council (UKRI077). ES, TP, GH, GDS, KT work in a unit funded by the Medical Research Council (MC_UU_00032/1, MC_UU_00032/2).

Data availability

Summary statistics used in all applied analyses are available from the MRC IEU OpenGWAS platform (doi: 10.1101/2020.08.10.244293 and https://doi.org/10.7554/eLife.34408)

Code availability

The code for the simulations and analyses is available at https://github.com/eleanorsanderson/confounding.

References

1. Davey Smith, G. and A.N. Phillips, *Correlation without a cause: an epidemiological odyssey.* International Journal of Epidemiology, 2020. **49**(1): p. 4-14.
2. Davey Smith, G. and S. Ebrahim, *'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?*.* International Journal of Epidemiology, 2003. **32**(1): p. 1-22.
3. Richmond, R. and G. Davey Smith, *Mendelian randomization: concepts and scope.* Cold Spring Harbour Perspectives in Medicine, 2021.
4. Sanderson, E., et al., *Mendelian randomization.* Nature Reviews Methods Primers, 2022. **2**(1): p. 6.
5. O'Connor, L.J., et al., *Extreme Polygenicity of Complex Traits Is Explained by Negative Selection.* The American Journal of Human Genetics, 2019. **105**(3): p. 456-476.
6. Yengo, L., et al., *A saturated map of common genetic variants associated with human height.* Nature, 2022. **610**(7933): p. 704-712.
7. Boyle, E.A., Y.I. Li, and J.K. Pritchard, *An Expanded View of Complex Traits: From Polygenic to Omnigenic.* Cell, 2017. **169**(7): p. 1177-1186.
8. Nam, Y., et al., *Harnessing artificial intelligence in multimodal omics data integration: paving the path for the next frontier in precision medicine.* Annual review of biomedical data science, 2024. **7**.
9. Davey Smith, G. and G. Hemani, *Mendelian randomization: genetic anchors for causal inference in epidemiological studies.* Hum Mol Genet, 2014. **23**(R1): p. R89-98.
10. Bowden, J., G. Davey Smith, and S. Burgess, *Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression.* International journal of epidemiology, 2015. **44**(2): p. 512-525.
11. Davey Smith, G., *Epigenesis for epidemiologists: does evo-devo have implications for population health research and practice?* International Journal of Epidemiology, 2012. **41**(1): p. 236-247.
12. Skrivankova, V.W., et al., *Strengthening the reporting of observational studies in epidemiology using mendelian randomisation (STROBE-MR): explanation and elaboration.* bmj, 2021. **375**.
13. Hemani, G., K. Tilling, and G.D. Smith, *Orienting the causal relationship between imprecisely measured traits using GWAS summary data.* PLoS genetics, 2017. **13**(11): p. e1007081.

14. Morrison, J., et al., *Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics.* Nature Genetics, 2020: p. 1-7.
15. Cheng, Q., et al., *MR-LDP: a two-sample Mendelian randomization for GWAS summary statistics accounting for linkage disequilibrium and horizontal pleiotropy.* NAR genomics and bioinformatics, 2020. **2**(2): p. lqaa028-lqaa028.
16. Xue, H., X. Shen, and W. Pan, *Constrained maximum likelihood-based Mendelian randomization robust to both correlated and uncorrelated pleiotropic effects.* The American Journal of Human Genetics, 2021. **108**(7): p. 1251-1269.
17. Xu, S., W.K. Fung, and Z. Liu, *MRCIP: a robust Mendelian randomization method accounting for correlated and idiosyncratic pleiotropy.* Briefings in Bioinformatics, 2021. **22**(5).
18. Brown, B.C. and D.A. Knowles, *Welch-weighted Egger regression reduces false positives due to correlated pleiotropy in Mendelian randomization.* The American Journal of Human Genetics, 2021. **108**(12): p. 2319-2335.
19. Cheng, Q., et al., *Mendelian randomization accounting for complex correlated horizontal pleiotropy while elucidating shared genetic etiology.* Nature Communications, 2022. **13**(1): p. 6490.
20. Cheng, Q., et al., *MR-Corr2: a two-sample Mendelian randomization method that accounts for correlated horizontal pleiotropy using correlated instrumental variants.* Bioinformatics, 2022. **38**(2): p. 303-310.
21. Lin, Z., Y. Deng, and W. Pan, *Combining the strengths of inverse-variance weighting and Egger regression in Mendelian randomization using a mixture of regressions model.* PLoS genetics, 2021. **17**(11): p. e1009922.
22. Darrous, L., N. Mounier, and Z. Kutalik, *Simultaneous estimation of bi-directional causal effects and heritable confounding from GWAS summary statistics.* Nat Commun, 2021. **12**(1): p. 7274.
23. Darrous, L., et al., *PheWAS-based clustering of Mendelian Randomisation instruments reveals distinct mechanism-specific causal effects between obesity and educational attainment.* Nat Commun, 2024. **15**(1): p. 1420.
24. Wang, J., et al., *Causal Inference for Heritable Phenotypic Risk Factors Using Heterogeneous Genetic Instruments.* PLoS genetics, 2021. **17**(6).
25. Zhang, W., et al., *MR Corge: sensitivity analysis of Mendelian randomization based on the core gene hypothesis for polygenic exposures.* Bioinformatics, 2024. **40**(11).
26. Burgess, S., A. Butterworth, and S.G. Thompson, *Mendelian Randomization Analysis With Multiple Genetic Variants Using Summarized Data.* Genetic Epidemiology, 2013. **37**(7): p. 658-665.
27. Bowden, J., et al., *Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator.* Genetic epidemiology, 2016. **40**(4): p. 304-314.
28. Hartwig, F.P., G. Davey Smith, and J. Bowden, *Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption.* International journal of epidemiology, 2017. **46**(6): p. 1985-1998.
29. Burgess, S., F. Dudbridge, and S.G. Thompson, *Re:"Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects".* American journal of epidemiology, 2015. **181**(4): p. 290-291.

30. Sanderson, E., et al., *An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings.* International journal of epidemiology, 2019. **48**(3): p. 713-727.
31. Zhu, M., et al., *C-reactive protein and cancer risk: a pan-cancer study of prospective cohort and Mendelian randomization analysis.* BMC Medicine, 2022. **20**(1): p. 301.
32. Zhou, A. and E. Hyppönen, *Vitamin D deficiency and C-reactive protein: a bidirectional Mendelian randomization study.* International Journal of Epidemiology, 2022. **52**(1): p. 260-271.
33. Han, X., et al., *Using Mendelian randomization to evaluate the causal relationship between serum C-reactive protein levels and age-related macular degeneration.* European Journal of Epidemiology, 2020. **35**(2): p. 139-146.
34. Timpson, N.J., et al., *C-reactive protein and its role in metabolic syndrome: mendelian randomisation study.* Lancet, 2005. **366**(9501): p. 1954-9.
35. Said, S., et al., *Genetic analysis of over half a million people characterises C-reactive protein loci.* Nature Communications, 2022. **13**(1): p. 2198.
36. Watanabe, K., et al., *Functional mapping and annotation of genetic associations with FUMA.* Nature Communications, 2017. **8**(1): p. 1826.
37. Elsworth, B., et al., *The MRC IEU OpenGWAS data infrastructure.* bioRxiv, 2020: p. 2020.08.10.244293.
38. Hemani, G., et al., *The MR-Base platform supports systematic causal inference across the human phenome.* eLife, 2018. **7**: p. e34408.
39. Sakaue, S., et al., *A cross-population atlas of genetic associations for 220 human phenotypes.* Nat Genet, 2021. **53**(10): p. 1415-1424.
40. Warren, M.S., et al., *Results of a proof of concept, double-blind, randomized trial of a second generation antisense oligonucleotide targeting high-sensitivity C-reactive protein (hs-CRP) in rheumatoid arthritis.* Arthritis Research & Therapy, 2015. **17**(1): p. 80.
41. Ries, W., et al., *C-Reactive Protein Apheresis as Anti-inflammatory Therapy in Acute Myocardial Infarction: Results of the CAMI-1 Study.* Front Cardiovasc Med, 2021. **8**: p. 591714.
42. Smith, G.D. and S. Ebrahim, *Mendelian randomization: prospects, potentials, and limitations.* International Journal of Epidemiology, 2004. **33**(1): p. 30-42.
43. Sanderson, E., W. Spiller, and J. Bowden, *Testing and correcting for weak and pleiotropic instruments in two-sample multivariable mendelian randomisation.* Statistics in medicine, 2021.
44. Schmidt, A.F., et al., *Genetic drug target validation using Mendelian randomisation.* Nature communications, 2020. **11**(1): p. 3255-3255.
45. Holmes, M.V., et al., *Integrating genomics with biomarkers and therapeutic targets to invigorate cardiovascular drug development.* Nature Reviews Cardiology, 2021. **18**(6): p. 435-453.
46. Qi, G. and N. Chatterjee, *Mendelian randomization analysis using mixture models for robust and efficient estimation of causal effects.* Nature Communications, 2019. **10**(1): p. 1941.
47. Foley, C.N., et al., *MR-Clust: clustering of genetic variants in Mendelian randomization with similar causal estimates.* Bioinformatics, 2020. **37**(4): p. 531-541.

48. Timpson, N.J., et al., *C-reactive protein levels and body mass index: elucidating direction of causation through reciprocal Mendelian randomization.* International Journal of Obesity, 2011. **35**(2): p. 300-308.
49. Lumsden, A.L., et al., *Apolipoprotein E (APOE) genotype-associated disease risks: a phenome-wide, registry-based, case-control study utilising the UK Biobank.* eBioMedicine, 2020. **59**: p. 102954.
50. Venkatesh, S.S., et al., *Characterising the genetic architecture of changes in adiposity during adulthood using electronic health records.* Nature Communications, 2024. **15**(1): p. 5801.
51. Gutierrez, S., M.M. Glymour, and G.D. Smith, *Evidence triangulation in health research.* Eur J Epidemiol, 2025. **40**(7): p. 743-757.
52. Mbatchou, J., et al., *Computationally efficient whole-genome regression for quantitative and binary traits.* Nat Genet, 2021. **53**(7): p. 1097-1103.
53. Yengo, L., et al., *Meta-analysis of genome-wide association studies for height and body mass index in ~ 700000 individuals of European ancestry.* Human molecular genetics, 2018. **27**(20): p. 3641-3649.

Figures

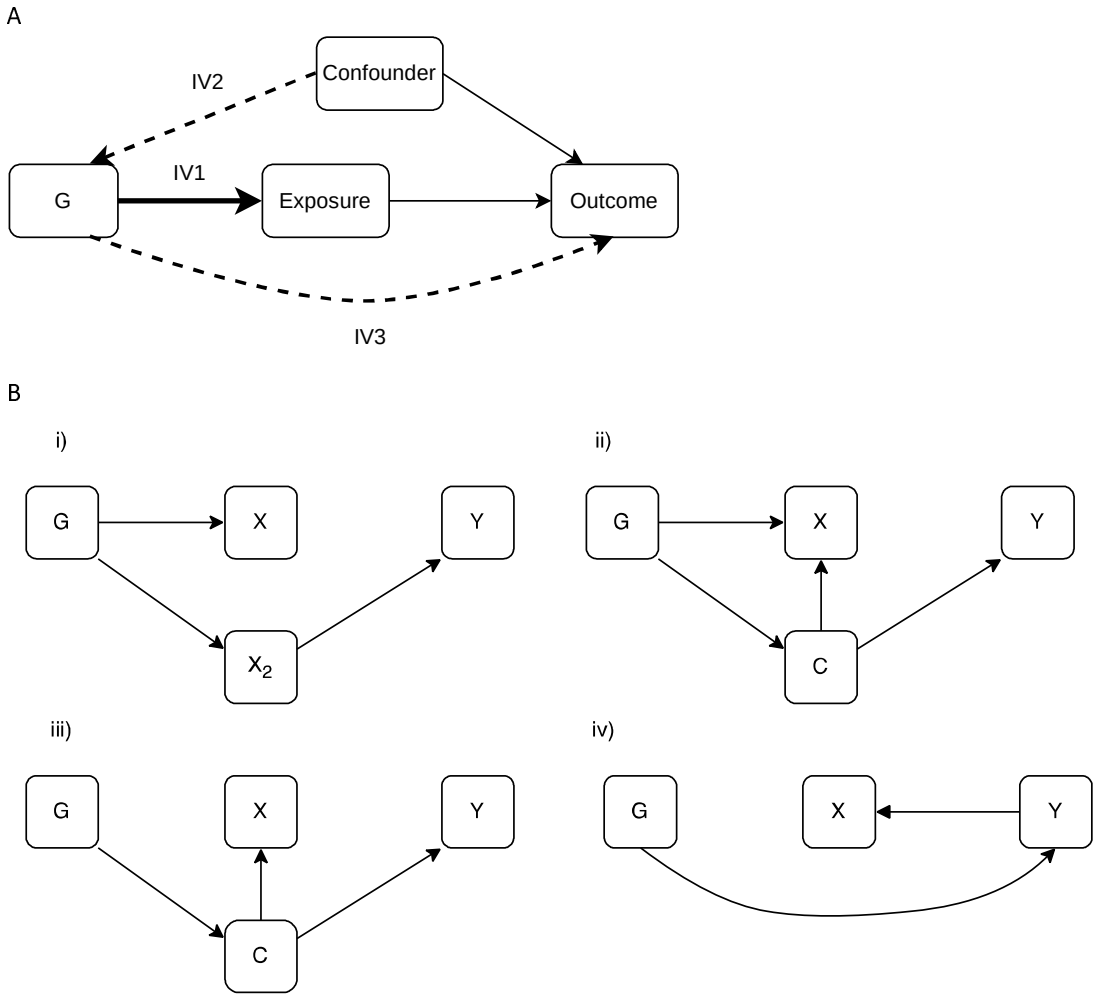


Figure 1: Instrumental Variable assumptions and potential mechanisms that can generate correlated pleiotropy and bias MR.

Figure illustrating the IV assumptions and the different mechanisms that can cause correlated pleiotropy in MR studies in the absence of a causal effect of the exposure on the outcome. X is the exposure of interest, Y is the outcome, C is a confounder of the exposure and outcome. G is a set of SNPs associated with the exposure, some of which may be associated through another phenotype.

A. Illustration of the main IV assumptions; (1) the IV should be associated with the exposure, (2) there should be no confounding of the IV and the outcome and (3) there should be no pathway from the IV to the outcome that doesn't act through the exposure. The dashed lines represent associations that should not exist for the IV assumptions to be satisfied.

B. i) Uncorrelated horizontal pleiotropy, some of the SNPs in G are associated with Y through X_2 ii) and iii) show forms of heritable confounding and mis-specification of the primary phenotype. In ii) some of the SNPs in G are associated the exposure through the confounder, in iii) ALL of the SNPs in G are associated with the exposure through the confounder. iv) reverse causation.

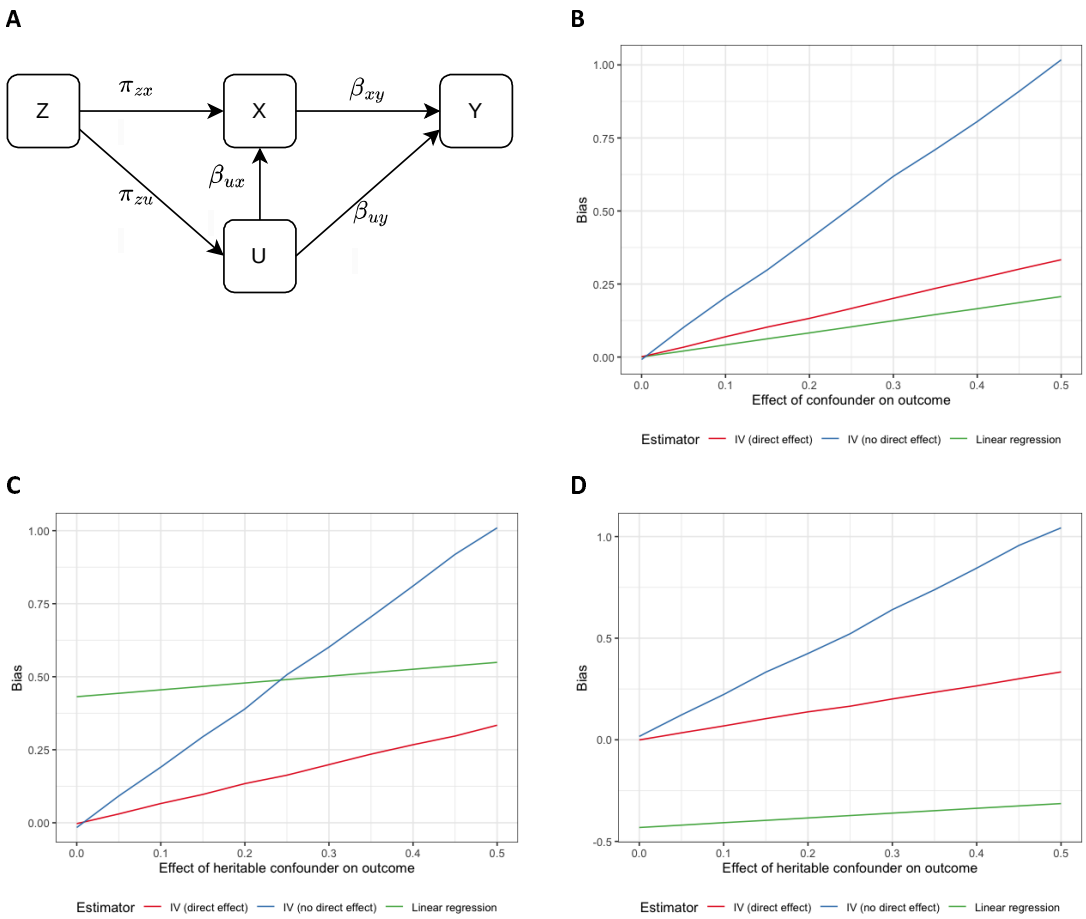


Figure 2 – The Bias of linear regression and IV estimators when instruments have a pleiotropic effect through a single heritable confounder.

A. *A simple model with a single IV associated with an exposure and a confounder. This is used in obtaining the expressions for bias given above and for generating the simulation results in (b).*
B. *Bias from estimation by linear regression and two-stage least squares for the model in (a) estimated assuming U is unobserved and there is no other confounding. Linear Regression gives the bias of the observational association, IV (direct effect) sets and gives the bias of the IV estimator, IV (no direct effect) sets and gives the bias of the IV estimator. The relative positions of these plots depends on the other parameters in the model and the assumption of no other (non-heritable) confounding.*
C. *Bias plots given in B with the addition of a positive non-heritable confounder of the exposure and outcome.*
D. *Bias plots given in B with the addition of a negative non-heritable confounder of the exposure and outcome.*

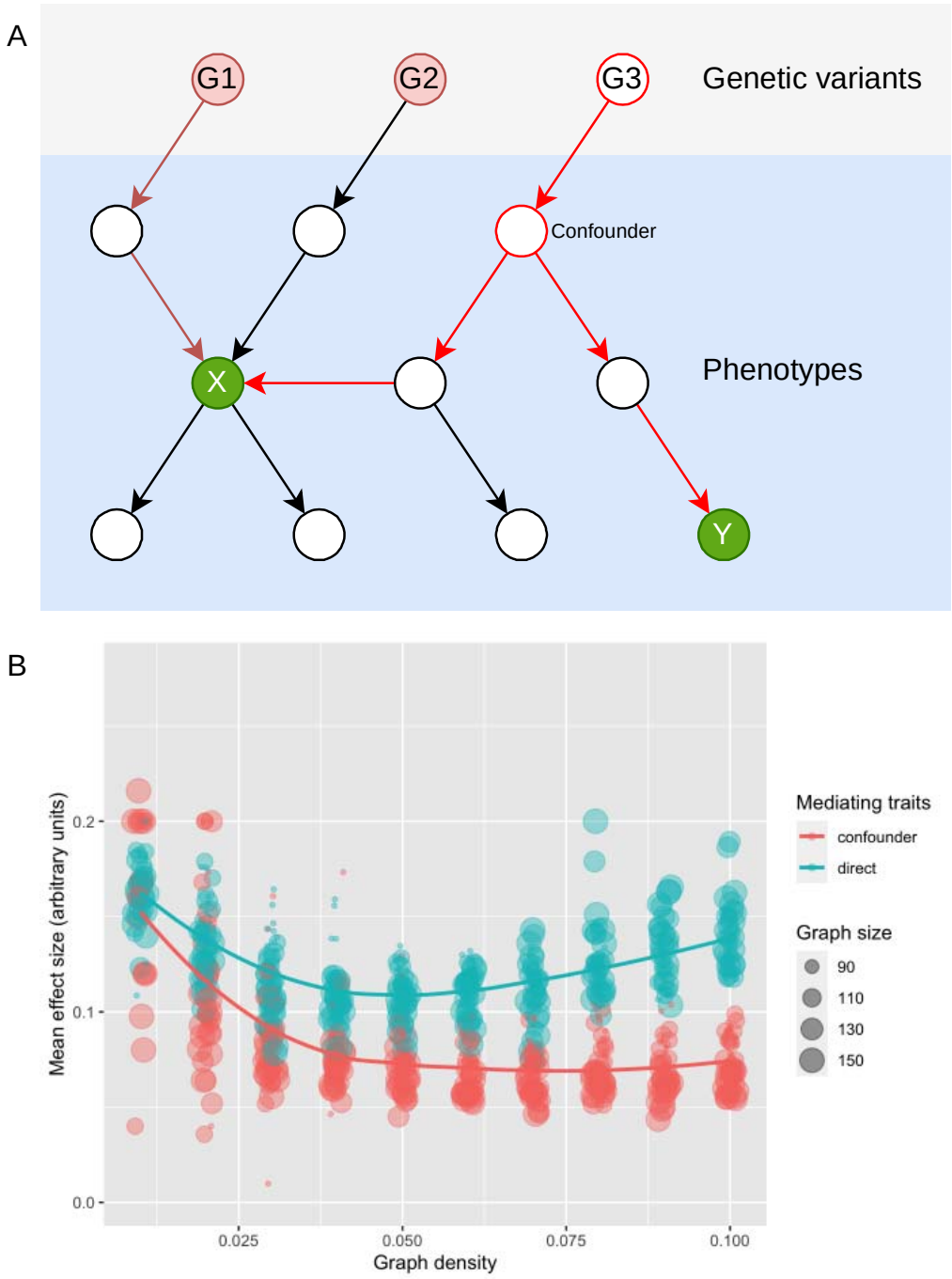


Figure 3. Network-based simulations of the origin of instruments.

A. Illustration of a random network. For each possible trait pair we assign an exposure (X) and an outcome (Y). In this simplified example instrument G1 is considered to act 'directly' because there is no confounding path between X and Y, whereas G2 is considered to act via a confounding path. The effect size of each instrument is determined based on all edges having a uniform effect of 0.1.

B. The mean effect of instrument on the exposure (y-axis) across simulations (x-axis and point size), stratified by whether they act 'directly' or via confounding paths. We observe that in general direct instruments have larger effects than those acting via confounders.

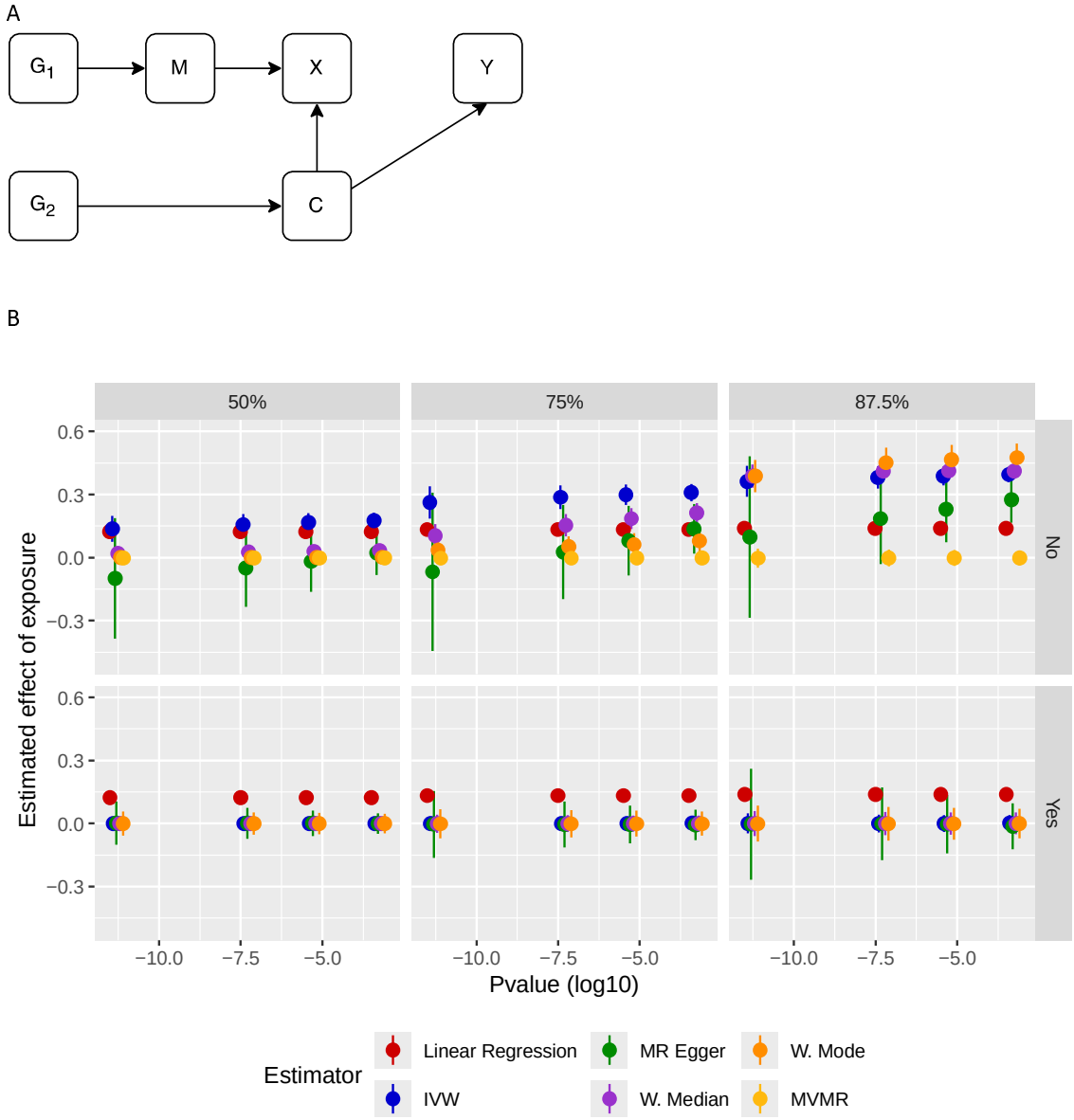
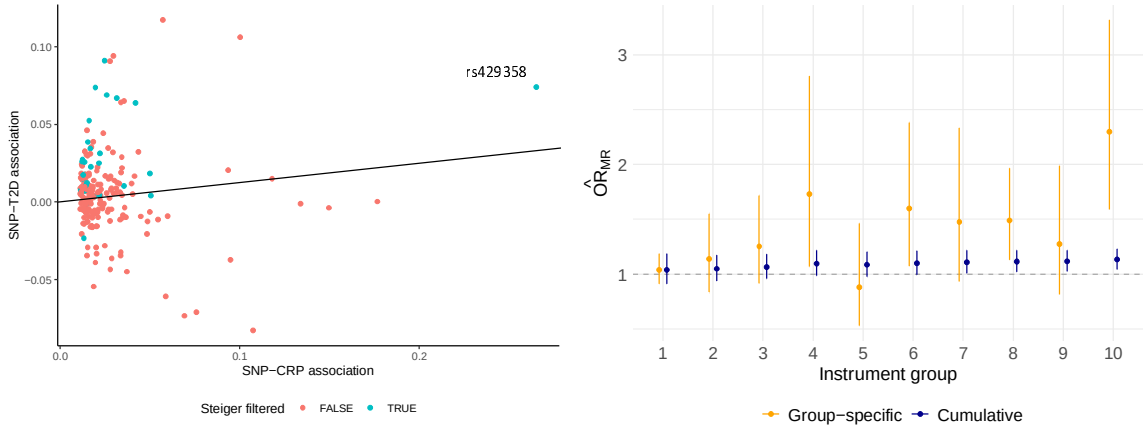


Figure 4 – Simulation of bias from a single heritable confounder in linear regression and MR.

A. The model used to generate the data in the simulation. X is the exposure of interest, Y is the outcome, C is a (potentially observable) confounder of the exposure and outcome, M is a mediator upstream of X. G1 is a set of SNPs associated with the mediator, G2 is a set of SNPs associated with the confounder. SNPs in G1 may be associated with the exposure through the mediator and SNPs in G2 may be associated with the exposure through the confounder. There is no causal effect of the exposure on the outcome.

B. B Results for the simulations varying; the proportion of the SNPs in G2 and whether Steiger filtering is applied. Total number of SNPs – 400, % given is the % in G2, the rest are valid SNPs for X (i.e. G1). All SNPs that are significantly associated with X in the simulated data are included in the univariable MR estimation and all SNPs significantly associated with X or C are included in the MVMR estimation. There is no causal effect of X on Y, the effect of M on X is 1, of C on Y is 0.4 and of C on X is 0.75. n=10000, reps=1000. The effect of X on Y is estimated by; Linear regression, IVW, MR Egger, weighted median, weighted mode and MVMR including C as an additional

exposure. Steiger filtered results includes all univariable estimators with Steiger filtering applied to remove SNPs that explain more variation in C than X.



(a) (b)

Figure 5 – MR results from univariable MR estimation in estimation of the effect of CRP on Type 2 Diabetes *(a) scatter plot of SNP-exposure and SNP-outcome associations (in log odds) for the SNPs used in the MR estimation of the effect of age at menarche on type 2 diabetes (T2D). All SNPs are genome-wide significantly associated with CRP, those SNPs that showed a stronger association with at least one of the top 10 potential confounders identified are highlighted in blue. The fitted line shows the IVW effect estimate using all SNPS. (b) MR Corge per group IVW effect estimates (in odds ratios) for the effect of CRP on T2D. SNPs are grouped evenly by their association with CRP (18 or 19 SNPs per group). Group 1 has the strongest association with CRP. Yellow data points show the estimate and 95% confidence interval estimated using only the SNPs in that group, blue data points show the estimate and 95% confidence interval estimated using SNPs from that group and all groups that are more strongly associated with the exposure. Full results are given in Supplementary Table 8.*

Tables

Table 1 – The number of SNPs selected in each simulation.

P-value	*No. SNPs*	*Exposure*	*Confounder*	*% for confounder*
50% of simulated SNPs associated with confounder				
$5x10^{-12}$	57	39	18	31%
$5x10^{-8}$	97	61	36	37%
$5x10^{-6}$	131	79	52	40%
$5x10^{-4}$	182	104	78	43%
87.5% of simulated SNPs associated with confounder				
$5x10^{-12}$	42	10	32	76%
$5x10^{-8}$	79	15	64	80%
$5x10^{-6}$	112	20	92	82%
$5x10^{-4}$	164	27	137	84%

The number of SNPs selected as associated with the exposure (out of the 400 generated) based on the p-value cutoff used for selection. Exposure gives the number of those that were generated as associated with the exposure and Confounder the number of those that were generated as associated with the confounder but selected as associated with the exposure. Numbers are means across all iterations, rounded to the nearest whole number.

Table 2 – Proportion of traits that show evidence of a causal effect on C-Reactive Protein in MR

	No of tests	**% P < 0.05**	**% P < 0.01**
Wald ratio/IVW	3770	21.7%	13.9%
MR Egger	1480	27.5%	19.2%
Weighted median	1478	29.0%	19.6%
Weighted mode	1478	22.7%	13.7%

Results from MR estimation of the effect of up to 3770 traits on CRP. Traits were selected from those strongly associated with at least one lead SNP in a recent CRP GWAS. Wald ratio was used for traits with 1 SNP available as an instrument and IVW was used for traits with 2 or more SNPs available as instruments. MR Egger, Weighted median and Weighted mode where additionally applied for all traits with 3 or more SNPs available as instruments.

Table 3 – MR effect estimates for CRP on T2D, with all SNPS included and with SNPs more strongly associated with each of the top 10 confounders removed.

Potential confounder	***SNPs removed***	***IVW beta (95% CI)***	***W. Median beta (95% CI)***	***W. Mode beta (95% CI)***	***MR Egger beta (95% CI)***
None	0	1.13 (1.04 1.23)	1.09 (1.01 1.17)	1.11 (1.06 1.16)	0.96 (0.91 1.14)
Tubulin-specific chaperone A	8	1.08 (0.98 1.19)	0.99 (0.93 1.06)	1.01 (0.96 1.07)	0.88 (0.76 1.01)
Leucine-rich repeat neuronal protein 1	7	1.08 (0.98 1.19)	0.99 (0.93 1.06)	1.01 (0.96 1.07)	0.88 (0.76 1.01)
Protein S100-A13	7	1.08 (0.99 1.19)	0.99 (0.93 1.06)	1.01 (0.96 1.06)	0.88 (0.76 1.01)
PDZK1-interacting protein 1	3	1.09 (0.99 1.19)	0.99 (0.93 1.06)	1.01 (0.96 1.07)	0.88 (0.77 1.01)
CUB and zona pellucida-like domain-containing protein 1	3	1.09 (0.99 1.19)	0.99 (0.93 1.06)	1.01 (0.96 1.07)	0.88 (0.77 1.01)
Killer cell immunoglobulin-like receptor 2DL5A	2	1.09 (0.99 1.19)	0.99 (0.93 1.06)	1.01 (0.96 1.07)	0.88 (0.77 1.01)
*Body mass index**	22	1.08 (1.00 1.16)	1.03 (0.95 1.11)	1.12 (1.07 1.17)	1.03 (0.93 1.14)
Waist circumference	16	1.09 (1.01 1.17)	1.05 (0.98 1.13)	1.12 (1.07 1.17)	1.03 (0.93 1.13)
Body fat percentage	20	1.09 (1.01 1.17)	1.05 (0.98 1.13)	1.12 (1.07 1.17)	1.03 (0.93 1.13)
*Body mass index***	12	1.09 (1.01 1.17)	1.05 (0.98 1.13)	1.12 (1.07 1.17)	1.03 (0.93 1.13)

*MR estimates for the effect of CRP on T2D. The first row shows effect estimates with all SNPS included (n=223). Subsequent rows are with SNPs that explain more variation in one of the top 10 confounders than they do in CRP removed from the estimation. BMI appears twice in the list from different GWAS in OpenGWAS *BMI from Mbatchou et al. (2021)[52] **BMI from Yengo et al. (2018)[53]*