Bias from heritable confounding in Mendelian randomization studies

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) is an approach to causal inference that utilises genetic variants to obtain estimates of the causal effect of an exposure on an outcome in the presence of unobserved confounding. MR relies on a set of assumptions to obtain unbiased effect estimates, one of these assumptions is that there is no pathway from the genetic variants to the outcome that does not act through the exposure. Increasing genome-wide association study (GWAS) sample sizes for the exposure enables discovery of instrumental variables with smaller effect sizes. We illustrate through simulations how smaller effect sizes could arise from genetic variants that act through traits that have greater liability to confound an exposure-outcome relationship. When such genetic variants are selected as instruments this can bias the MR effect estimate obtained from that instrument in the same direction as the confounded association from a linear regression, but with potentially larger magnitude in the MR estimates. This bias acts in the same way across all of the most commonly used MR estimation methods leading to false confidence in the results obtained by using these methods. Through simulation we illustrate how the total bias of the MR estimates increases across a range of standard MR estimation methods as the proportion of the genetic instruments that are associated with the confounder increases. However, if such heritable confounders are known and can be instrumented, the confounder free effect estimate can be obtained using MR by applying a pre-estimation filtering, removing instruments that explain more variation in that confounder than the exposure, or by estimating effects through multivariable MR. We illustrate our approach through two applications. Firstly we estimate of the causal effect of age at menarche on type 2 diabetes, hypothesising that the MR effect estimate may be biased by confounding due to the inclusion of genetic variants associated with early life adiposity as instruments. Secondly, estimate the effect of C Reactive protein on type 2 diabetes using a hypothesis free approach to identify and remove the effect of potential confounders.

Introduction

Epidemiological studies often aim to estimate a causal effect from observational data. A common method is to estimate this using multivariable regression, which will provide an unbiased estimate of the causal effect provided all confounders are known and measured without error. This assumption is rarely plausible.[1] Mendelian Randomisation (MR) is an alternative approach to causal effect estimation which utilises genetic variants to infer causality of the exposure on the outcome that is not biased by that confounding.[2, 3] MR is most often implemented by using genetic variants associated with the exposure within an instrumental variable (IV) estimation framework.[4]

IV estimation relies on three key assumptions to be able to test for a causal effect of the exposure on the outcome, illustrated in **Figure 1A.** In MR genetic variants associated with the exposure of interest are used as IVs. These are usually single nucleotide polymorphisms (SNPs) identified as associated with the exposure trait at a predetermined level in a genome-wide association study (GWAS).[5] The level chosen is typically set to select all independent variants where the p-value for the estimated association between the variant and the exposure is smaller than $5x10^{-8}$. This level is chosen as it is the Bonferroni corrected 5% significance level, given the effective number of independent tests being conducted in a GWAS.

The polygenic nature of complex traits, for which there is substantial empirical support[6], posits that 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. The most recent GWAS of body mass index (BMI) included approximately 700,000 individuals and identified 941 independent SNPs associated with BMI.[7] The most recent GWAS of educational attainment included approximately 3 million individuals and identified nearly 4000 independent SNPs.[8] This has led to increasing numbers of SNPs being available as instruments for MR estimation.

For MR analyses involving exposures that are themselves complex traits, it is well understood that genetic variants associated with the exposure are mediated through other phenotypic layers.[9] Such mediation opens the possibility of horizontal pleiotropy if those mediating traits also have effects on the outcome through pathways independent of the exposure. If some, or all, of the variants associated with the exposure exhibit horizontal pleiotropy the MR effect estimates can be biased through violation of assumption IV3. There are a number of different forms that horizontal pleiotropy can take and so bias MR.[4] Correlated pleiotropy is used to describe pleiotropy that causes bias in MR which is correlated with the magnitude of association between the SNPs and the exposure. Correlated pleiotropy can occur through heritable confounding, misspecification of the primary phenotype[10] or reverse causation, illustrated in **Figure 1B.** The distinction between heritable confounding and misspecification of the primary phenotype (i.e. the forms of pleiotropy illustrated by **Figure 1B(ii)** and **1B(iii)**) is not always clear, especially as one may lead to the other. If the genetic variant for a heritable confounder is taken to be an instrument for the exposure that the heritable confounder influences then this misspecification of the primary phenotype will clearly lead to erroneous inference, with the extreme case being when the upstream phenotype is the outcome under investigation, as in **Figure 1B(iv).** As each of these mechanisms lead to the same bias in the resulting MR estimates, misspecification of the primary phenotype could be considered to be an

particular form of heritable confounding. Note however that for there to be bias in the MR estimation the path from the confounder to the outcome is required. If there is no effect of the confounder on the outcome that does not act through the exposure then MR estimates will not be biased. Throughout the rest of this paper we consider correlated pleiotropy driven by genetic variants associated with a confounder of the exposure and outcome being included within the set of instruments used in an MR analysis and refer to this as heritable confounding.

There remains a question of whether increasing GWAS sample sizes will lead to increased horizontal pleiotropy through heritable confounding, i.e. are smaller genetic effects for the exposure more likely to also have an independent path to the outcome for any given exposure-outcome relationship? 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 they will lead to pleiotropic bias in the SNPs associated with that trait. One implication of this is that the genetic variants with the largest effect sizes in the GWAS may be the least likely to be pleiotropic instruments.[11]

There are a number of MR methods available that are robust to violations of assumption IV3, each of which assume particular forms for the structure of the pleiotropy[4]. The most commonly used methods are MR Egger[12], weighted Median[13] and weighted Mode[14]. Weighted median and weighted mode approaches will be robust to a small proportion of the SNPs included in the estimation being associated with a confounder of the exposure and outcome. IVW and MR Egger will be biased if any of the SNPs are associated with the confounder, although the level of the bias will depend on the proportion of the SNPs used as instruments that are associated with the confounder, and the presence/absence of another confounder balancing the bias in the opposite direction. This result has been illustrated elsewhere.[15] MR Corge has been proposed under the assumption that the strongest instruments are the least pleiotropic and estimates the MR effects for sets of SNPs based on their association with the exposure.[11]

Alternative methods of MR estimation are available that adjust for correlated pleiotropy under particular assumptions of the structure the pleiotropy takes.[15-24] 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.

Here we propose two alternative approaches to estimation in the presence of correlated pleiotropy when the confounder or confounders are hypothesised and measured; including the confounder in a multivariable MR (MVMR) and removing the SNPs associated with the confounder through confounder-based Steiger filtering.[25] Each approach relies on knowledge of the likely confounders, however when this is unknown the other traits associated with the genetic instruments, and so most likely to be confounders, can be identified through PheWAS analysis.[26] MVMR and outcome-based Steiger filtering are simple to implement and already widely used in MR analyses for general

pleiotropy. The approaches we propose here are an adaptation of the standard application of those approaches to specifically correct for correlated pleiotropy providing a simple approach to explore and mitigate the potential for bias from heritable confounding in MR studies.

The remainder of this paper is structured as follows; we first compare the bias from heritable confounding in IV estimation to the bias from residual confounding using linear regression. We 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 then present a simulation study to illustrate the bias induced by heritable confounding and how MVMR and confounder-based Steiger filtering can correct the bias. We apply these methods to two settings. Firstly; we consider estimation of the causal effect of age at menarche on risk of type 2 diabetes hypothesising that the IVW estimates are biased by correlated pleiotropy from childhood adiposity. Secondly; we consider estimation of the effect of C-reactive protein (CRP) on type 2 diabetes using phenome-wide approaches to systematically identify potential heritable confounders.

Results

Bias in IVW MR estimates compared to linear regression

We consider a simple model of a relationship between an exposure (X) and an outcome (Y) which is confounded by a single unobserved confounder (U). The exposure is associated with a single instrument (Z) that also has an effect on the outcome via the unobserved confounder. This single instrument could be considered to be a genetic risk score constructed from all SNPs identified as strongly associated with the exposure in a GWAS. 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 however in any application there would also be a number of non-heritable confounders as well as potentially more than one heritable confounder. 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 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;

$$\frac{\beta_{uy}\pi_{zu}}{\pi_{zx} + \beta_{ux}\pi_{zu}} < \beta_{uy}\delta_{ux}$$

As the bias in the 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 ($\pi_{zx} + \beta_{ux}\pi_{zu}$). This shows that the larger

the pleiotropic effect as a proportion of the effect of the instrument on the exposure the larger the bias of the IV estimator will be.

Under a simpler model of heritable confounding, where U is a confounder of X and Y and has its own independent genetic profile in which Z influences only one path, i.e. $\pi_{zx} = 0$ the bias term for the IV estimator reduces to $\frac{\beta_{uy}}{\beta_{ux}}$. The implication of this is that if the genetic variants influence the exposure X only through the confounder U (i.e. they have no direct effect on X) the bias in the MR estimate from that confounder will always be greater than the bias in the observational association from the same confounder.

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 from 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.

Increasing GWAS sample size may increase correlated pleiotropy

A genetic variant inducing correlated pleiotropy must influence a trait that is a parent to both exposure and outcome. We hypothesised that such a genetic variant would generally be associated with more traits upstream of the exposure than valid instruments, and therefore would have relatively smaller effects on the exposure than valid instruments. The consequence of such a relationship would be that increasing GWAS sample sizes to identify more instruments of smaller effect sizes could lead to increasing bias in MR estimates.

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. The network represents a directed acyclic graph where all nodes are traits except those that have no parents, which are genetic variants. The simulation proceeds by randomly choosing two traits, one as exposure and one as outcome, and then evaluating a) which variants are instruments for the exposure, and whether or not the instruments act through confounders in the graph, and b) the path length from instrument to exposure. The length of the edge between any pair of nodes has been fixed to be the same length for all edges.

Figure 3 illustrates 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 an 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.

Simulating the impact of genetic confounding on MR estimates

We simulated a model with an exposure and outcome which are confounded by a third (potentially unobserved) heritable trait. Here the exposure has no causal effect on the outcome however, 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 exposure is influenced by two traits, one simply mediates genetic effects onto the trait (i.e. influences on X), whereas the other is a heritable confounder of the exposure and the outcome. Both mediator and confounder are each associated with a set of genetic variants. This model is illustrated in **Figure 4**, which corresponds to the scenario illustrated in **Figure 1B(ii)**. We have included only a single heritable confounder in these models, no non-genetically influenced confounders and have assumed that all other IV assumptions hold.

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 (/suspected) and have GWAS summary statistics available.

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.

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.

Example: Estimation of the effect of age at menarche on risk of type 2 diabetes

We illustrate the potential for genetic variants associated with a confounder to bias MR estimation through an example estimating the causal effect of age at menarche on risk of type 2 diabetes (T2D). Adiposity is an established risk factor for age at menarche and SNPs associated with age at menarche have been implicated as acting through body mass index (BMI).[27] Here, we explore whether the MR effect estimate of age at menarche on risk of T2D is confounded by childhood adiposity. The hypothesised DAG is given in **Figure 5** and results are given in **Table 2**. Univariable MR analysis shows an estimated causal effect of age at menarche on type 2 diabetes using IVW [odds ratio for a category increase in age at menarche on the risk of type 2 diabetes: 0.76, 95% confidence interval: 0.67 to 0.87]. However, this result is not consistent across the robust MR methods.

We applied Steiger filtering in which 13 of 183 instruments for age at menarche appeared to primarily influence childhood adiposity at age 10 were removed, leading to an attenuated effect estimate [OR: 0.90, 95% CI 0.81 to 1.00]. The SNPs removed by filtering are highlighted in **Figure 6(a)**. Including childhood adiposity in a MVMR more substantially attenuates the effect estimate obtained [OR: 0.98, 95% CI: 0.87 to 1.11]. This is supported by the MR Corge plot (**Figure 6(b)**) which shows the estimated effect for different groups of SNPs, divided up by how strongly they are associated with the exposure, and shows no effect for the SNPs most strongly associated with age at menarche but a protective effect appearing as more peripheral SNPs are included in the estimation. As childhood adiposity is an established risk factor for both age at menarche and – through adulthood BMI - type 2 diabetes, these results are likely to reflect the model given in **Figure 2b** and suggest that the result seen in the naïve IVW estimation is likely to be biased by confounding from SNPs associated with childhood adiposity. As has been shown elsewhere, there is an overlap in the SNPs associated with childhood and adulthood adiposity.[28]

Example: Heritable confounding in genetic variants associated with C-Reactive Protein

In our second application we considered another potential exposure that has been widely used as an exposure in MR studies; C-Reactive Protein (CRP), a marker of inflammation.[29-32] A recent GWAS has identified a large number of genetic variants associated with CRP across the genome, many of which may be associated with CRP due to their influence on traits which are upstream of CRP. In this example we use MR to estimate the effect of CRP on T2D and then systematically identify potential confounders and remove SNPs more strongly associated with those confounders than CRP.

To identify potential confounders we first made use of the FUMA platform[33] to identify other traits that have been associated with the genomic loci in a recent large GWAS of CRP.[34] For each of these traits we conducted an MR estimation of the effect of that trait on CRP and for those traits that showed potential evidence of an effect on CRP we also estimated their effect on T2D. Any traits that also showed evidence of an effect on T2D were taken forward and included as a potential confounder. For each potential confounder we removed SNPs that explained more variation in the confounder than CRP and re-estimated our MR results.

The genetic signature of CRP was comprised of 728 lead SNPS (SNPs associated with CRP levels at $p< 5x10^{-8}$ and LD $r^2 < 0.1$) in 266 genomic loci and we identified 33,573 SNP -> trait associations among these lead SNPs or related SNPs in strong LD (LD $r^2 > 0.6$) in the GWAS Catalog. We selected 5198 traits as potential confounders and estimated their effect on CRP with MR. These traits can be broadly categorised into: anthropometry measures related to height and weight, bone mineral density, glycaemic measures, proteins, metabolites and smoking behaviour traits. Between 20 and 30% of all traits for which instruments were available showed evidence of an effect on CRP (at $p< 0.05$), depending on the MR estimation method applied. **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.

We identified 870 traits to take forward (IVW or Wald $p<0.01$) and from these 325 potential confounders ($p<0.01$ for an effect of the trait on T2D using IVW or a wald ratio). **Table 4** gives results for the 10 confounders that had the largest effect on the MR results obtained (measured by p-value). Full results for each step are given in **Supplementary material section 2**. **Figure 7(a)** shows a scatter plot of the SNPS used in the MR estimation of CRP on T2D with the SNPs identified by any of the top 10 confounders highlighted.

The potential confounders can be divided into two groups, a set of proteins and different measures of adiposity. Each of the proteins was associated with the same influential SNP (rs429358) and so further investigation would be warranted to determine if these are confounders of CRP and T2D or if they are traits that have an effect on CRP but only influence T2D through CRP, which would make this SNP a valid instrument for CRP. However, the presence of 4 adiposity related traits in the top list strongly suggests that a number of the SNPs associated with CRP are confounded by adiposity. Once these are removed the MR effect estimates attenuate substantially and there is limited evidence of a causal effect of CRP on T2D. Our MR Corge plot (**Figure 7(b))** supports this, the group of SNPs most strongly associated with CRP show no evidence of an effect of CRP on T2D with the effect increasing as the number of SNPs included in the analysis increases. The CRP SNP group indicated by MR Corge as being core variants for CRP levels (i.e. group 1) included a variant within the *CRP* locus (rs2211320).

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, under a plausible model of biological effects, how genetic variants associated with confounders of an exposure and outcome may be associated with more traits upstream of the exposure than genetic variants which are associated with the exposure but not confounders of the exposure and outcome. The implication of this is that, as GWAS sample sizes increase these confounding genetic variants are increasingly likely to be detected as strongly associated with the exposure. This will increase the presence of correlated pleiotropy in MR effect estimates which use genetic variants strongly associated with the exposure in a GWAS as instruments for that exposure. Through simulations where there is a mixture of valid instruments or those acting via confounders, we show that bias caused by a specific confounder in MR estimation could be larger than the bias in the observation association caused by the same confounder and that this problem is likely to increase as sample sizes increase. Although it is likely that there will also be non-heritable confounding biasing the results from linear regression these results highlight how MR results may be increasingly biased as sample sizes for GWAS grow.

If a large proportion of the genetic instruments 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 relatively few genetic variants directly associated with the exposure of interest. In these cases a high proportion of the genetic variants used may be associated with one or more heritable confounders, increasing the overall bias in the MR estimates. 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.

In our first illustrative application we estimated the effect of age at menarche on risk of Type 2 diabetes. The standard IVW effect estimate showed a negative causal effect with later age at menarche estimated to be protective for risk of type 2 diabetes. This effect however attenuated in all of the robust methods as well as when Steiger filtering or MVMR was applied to adjust for childhood adiposity. Childhood adiposity is an established risk factor for both age at menarche[26, 35, 36] and type 2 diabetes, through adulthood adiposity, [28, 37-40] and these analyses support the effect observed in the IVW estimation being driven by a set of SNPs that are associated with the confounder childhood adiposity being included as instruments for age at menarche. In our second example we estimate the effect of CRP on type 2 diabetes, systematically identifying potential confounders through PheWAS approaches. This application similarly 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.[41] Notably, a single genomic region tagged by rs429358 in the *APOE* locus emerged as a particularly influential locus in this analysis. Given the well-documented pleiotropic effects of rs429358, including its association with APOE and lipid metabolism,[42] it is likely that its apparent effect on CRP is mediated by lipid-related processes, suggesting a strong potential for confounding through metabolic pathway. 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.

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 through a PheWAS analysis. Any genetic variant-trait associations found should then be checked for whether the trait involved is likely to be a confounder.[43] 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. The application of methods such as MR Corge[11] can help to identify whether the estimates obtained are robust across SNPs with different strength 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.[15, 17, 19, 20, 46, 47] 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.

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.

Methods

Network simulations

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 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.

Simulations evaluating MR methods

We generated the data in this model for two separate samples. In each sample 400 SNPs were generated using a binomial distribution. The proportion of these SNPs associated with the mediator and the confounder varied across the simulations, varying between 50% associated with each trait to 12.5% associated with the mediator and 87.5% associated with the confounder. The SNP-trait (i.e. mediator/confounder) association was drawn from a normal distribution with mean 0 and standard deviation of 0.04 for each SNP, orientated so all SNPs had a positive association with the relevant trait.

The confounder was generated as the sum of all the confounder - SNP effects plus a random error term. The mediator was generated as the sum of all mediator – SNP effects plus a random error term. The exposure was generated as the sum of the mediator, the confounder and a random 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 down weighted the confounder – exposure association by 25% relative to the mediator – exposure association (by multiplying them by

0.75) to mimic this more distant association. Results with this down weighting varied (to 50% and 0%) are given in the supplementary material. 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**.

In one of the samples generated we estimated the association between all of the SNPs and each of the exposure and confounder to generate GWAS summary statistics. This was repeated in the second sample for the SNPs and the outcome. We then select all SNPs (from either set) that were significantly associated with the exposure in the first sample at a defined p-value threshold to estimate the effect of the exposure on the outcome using summary-data MR methods. Following what is currently standard practice in summary-data MR the same sample was used for the SNP-exposure association in estimation as had been used to identify the SNPs associated with the exposure. We applied four common univariable MR methods of estimation IVW[48], MR Egger[12], weighted median[13] and weighted mode[14]. We additionally estimated the model using multivariable MR (MVMR)[49, 50] including the confounder as an additional exposure, including all SNPs significantly associated with the confounder at the same p-value threshold. All univariable methods of estimation were repeated with Steiger filtering[25] applied between the exposure and confounder to remove SNPs that explain more variation in the confounder than the exposure.

Increasing the sample size of a GWAS will decrease the standard error in the SNP-trait association proportionally across all of the SNPs. This will lead 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.

All models were simulated with a sample size of 100,000 for each of the exposure and outcome samples and 1000 repetitions.

Estimation of the effect of age at menarche on type 2 Diabetes

SNP – exposure associations were extracted for age at menarche for data on recalled age at which periods started for women in UK Biobank (n = 243,944) obtained through the IEU GWAS pipeline[51, 52]. SNP – exposure associations were obtained for childhood adiposity for a previously conducted GWAS using data on self-reported childhood adiposity for men and women in UKBiobank.[28] SNP – outcome associations for Type 2 diabetes were taken from a large recent GWAS in a European population.[53] All data used were downloaded from MRC IEU OpenGWAS[54].

We applied the univariable summary data estimation approaches; IVW, MR Egger, weighted median and weighted mode and additionally MVMR adjusting for BMI. For each approach we selected all SNPs associated with the relevant exposure at genome-wide significance (p-value for the SNP-exposure association $< 5x10^{-8}$). For the MVMR we selected all SNPs associated with either age at menarche or childhood adiposity at the same threshold. Each univariable method was applied twice, once including all SNPs and once with Steiger filtering applied to remove SNPs that explained more variation in childhood adiposity than age at menarche. Steiger filtering was not applied to the MVMR

estimation due to the inclusion of childhood adiposity as an exposure. Selected SNPs were then clumped to retain only independent SNPs ($r^2 < 0.001$ based on a reference LD panel). All analyses were done in R using the 'TwoSampleMR' package with default parameters for SNP selection and clumping applied.[55]

We additionally applied MR Corge[11] to the estimation of age at menarche on T2D (with no filtering applied). MR Corge groups the SNPs selected as instruments into groups based on how 'core' they are likely to be to the trait of interest and applies MR methods to each group individually and then cumulatively. We grouped the SNPs into 10 groups based on the size of the association between the SNP and CRP (i.e. the GWAS beta).

Heritable confounding in genetic variants associated with C-Reactive Protein

SNPs associated with CRP levels were identified from in a recent GWAS of >500,000 individuals with European genetic ancestry.[34] This GWAS identified 266 loci as associated with CRP, 211 of which were novel. We made use of the FUMA platform[33] to identify other traits which have also been associated these loci. We first identified the lead SNPs the CRP loci (independent SNPs LD $r^2 < 0.1$, and GWAS P-values $< 5\times10^{-8}$) and performed a look up in the GWAS Catalog[56] (look up was performed on April 10th, 2024) on 728 lead SNPs and 3,077 independent SNPs in strong LD with the lead SNPs (GWAS P-values $< 5\times10^{-8}$ and LD $r^2 > 0.1$ with the respective lead SNP).

The resulting list of associated traits was curated to prioritise biomarkers and risk factors over diseases. From this curated list we conducted a summary-data MR analysis of each trait using summary data from the MRC IEU OpenGWAS platform.[54] on to CRP levels. Where only one SNP was available as associated with the exposure trait of interest the Wald ratio was calculated. For traits with 2 or more SNPs the IVW MR effect estimate was obtained and for traits with 3 or more SNPs the MR Egger[12], Weighted Mode[14] and Weighted Median[13] estimates were also obtained.

We selected traits with p-value < 0.01 for their MR estimated effect on CRP using IVW (>1 SNPs available as instruments for the trait) or Wald ratio (1 SNP). For each of these traits we estimated the effect of the trait on Type 2 Diabetes using IVW (>1SNP) or Wald ratio (1 SNP) and summary data from the MRC IEU OpenGWAS platform.[53, 54] All traits that had a p-value < 0.01 for their estimated effect on T2D were retained as potential confounders.

We estimated the effect of CRP on T2D using MR firstly with all SNPs identified as associated with CRP ($p<5x10^{-8}$ and $R^2<0.001$). We then re-estimated the MR effect estimates with the CRP SNPs used as instruments Steiger filtered to remove those that explained more variation in the potential confounder than CRP for each potential confounder individually ($p<0.05$ for the SNP explaining more variation in the confounder than CRP). All MR analyses were done in R using the 'TwoSampleMR' package with default parameters for SNP selection and clumping applied.[55]

We additionally applied MR Corge[11] to the estimation of CRP on T2D (with no filtering applied). MR Corge groups the SNPs selected as instruments into groups based on how 'core' they are likely to be to the trait of interest and applies MR methods to each group individually and then cumulatively.

We grouped the SNPs into 10 groups based on the size of the association between the SNP and CRP (i.e. the GWAS beta).

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. Uffelmann, E., et al., *Genome-wide association studies.* Nature Reviews Methods Primers, 2021. **1**(1): p. 59.
6. 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.
7. 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.
8. Okbay, A., et al., *Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals.* Nature genetics, 2022. **54**(4): p. 437-449.
9. 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.
10. 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.
11. Zhang, W., et al., *MR Corge: sensitivity analysis of Mendelian randomization based on the core gene hypothesis for polygenic exposures.* Bioinformatics, 2024. **40**(11).
12. 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.

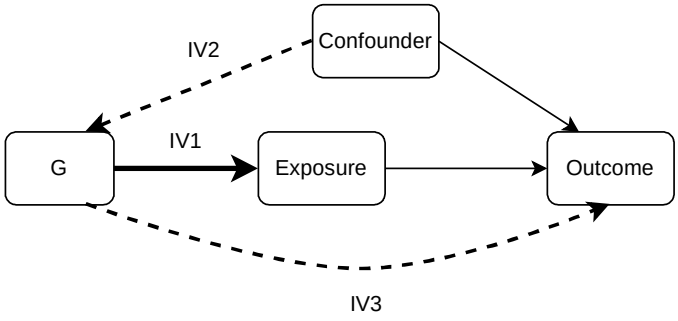
13. Bowden, J., et al., *Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator.* Genetic epidemiology, 2016. **40**(4): p. 304-314.
14. Hartwig, F.P., G. Davey Smith, and J. Bowden, *Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption.* International journal of epidemiology, 2017. **46**(6): p. 1985-1998.
15. Morrison, J., et al., *Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics.* Nature Genetics, 2020: p. 1-7.
16. 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.
17. 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.
18. Xu, S., W.K. Fung, and Z. Liu, *MRCIP: a robust Mendelian randomization method accounting for correlated and idiosyncratic pleiotropy.* Briefings in Bioinformatics, 2021.
19. 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.
20. Cheng, Q., et al., *Mendelian randomization accounting for complex correlated horizontal pleiotropy while elucidating shared genetic etiology.* Nature Communications, 2022. **13**(1): p. 6490.
21. 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.
22. 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.
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. 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.
25. 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.
26. Millard, L.A.C., et al., *MR-PheWAS: hypothesis prioritization among potential causal effects of body mass index on many outcomes, using Mendelian randomization.* Scientific Reports, 2015. **5**(1): p. 16645.
27. Kentistou, K.A., et al., *Understanding the genetic complexity of puberty timing across the allele frequency spectrum.* Nat Genet, 2024. **56**(7): p. 1397-1411.
28. Richardson, T.G., et al., *Use of genetic variation to separate the effects of early and later life adiposity on disease risk: mendelian randomisation study.* bmj, 2020. **369**.

29. 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.
30. 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.
31. 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.
32. Timpson, N.J., et al., *C-reactive protein and its role in metabolic syndrome: mendelian randomisation study.* Lancet, 2005. **366**(9501): p. 1954-9.
33. Watanabe, K., et al., *Functional mapping and annotation of genetic associations with FUMA.* Nature Communications, 2017. **8**(1): p. 1826.
34. Said, S., et al., *Genetic analysis of over half a million people characterises C-reactive protein loci.* Nature Communications, 2022. **13**(1): p. 2198.
35. Vabistsevits, M., et al., *Deciphering how early life adiposity influences breast cancer risk using Mendelian randomization.* Communications Biology, 2022. **5**(1): p. 337.
36. Mumby, H.S., et al., *Mendelian randomisation study of childhood BMI and early menarche.* Journal of obesity, 2011. **2011**.
37. Wainberg, M., et al., *Homogeneity in the association of body mass index with type 2 diabetes across the UK Biobank: A Mendelian randomization study.* PLOS Medicine, 2019. **16**(12): p. e1002982.
38. Lyall, D.M., et al., *Association of Body Mass Index With Cardiometabolic Disease in the UK Biobank: A Mendelian Randomization Study.* JAMA Cardiology, 2017. **2**(8): p. 882-889.
39. Holmes, Michael V., et al., *Causal Effects of Body Mass Index on Cardiometabolic Traits and Events: A Mendelian Randomization Analysis.* The American Journal of Human Genetics, 2014. **94**(2): p. 198-208.
40. Dale, C.E., et al., *Causal associations of adiposity and body fat distribution with coronary heart disease, stroke subtypes, and type 2 diabetes mellitus: a Mendelian randomization analysis.* Circulation, 2017. **135**(24): p. 2373-2388.
41. 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.
42. 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.
43. Yang, Q., et al., *Exploring and mitigating potential bias when genetic instrumental variables are associated with multiple non-exposure traits in Mendelian randomization.* European Journal of Epidemiology, 2022.
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. 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).

47. 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.
48. 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.
49. 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.
50. 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.
51. Elsworth, B., et al., *MRC IEU UK Biobank GWAS pipeline version 2.* University of Bristol, 2018. **10**.
52. Sudlow, C., et al., *UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age.* PLoS Med, 2015. **12**(3): p. e1001779.
53. Sakaue, S., et al., *A cross-population atlas of genetic associations for 220 human phenotypes.* Nat Genet, 2021. **53**(10): p. 1415-1424.
54. Elsworth, B., et al., *The MRC IEU OpenGWAS data infrastructure.* bioRxiv, 2020: p. 2020.08.10.244293.
55. Hemani, G., et al., *The MR-Base platform supports systematic causal inference across the human phenome.* eLife, 2018. **7**: p. e34408.
56. Sollis, E., et al., *The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource.* Nucleic Acids Res, 2023. **51**(D1): p. D977-d985.
57. Mbatchou, J., et al., *Computationally efficient whole-genome regression for quantitative and binary traits.* Nat Genet, 2021. **53**(7): p. 1097-1103.

Figures

A



B

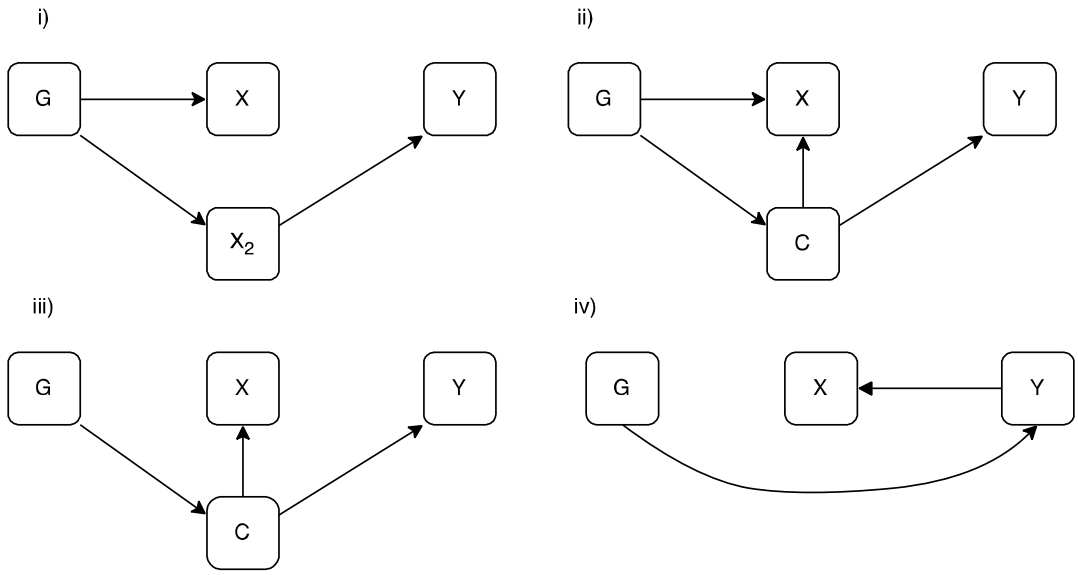
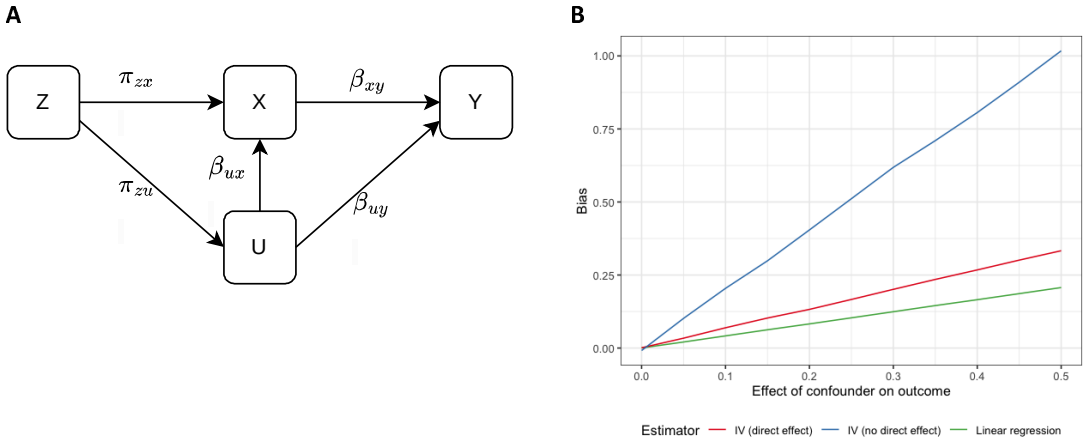


Figure 1: Instrumental Variable assumptions and potential mechanisms that can generate correlated pleiotropy and bias MR.

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.



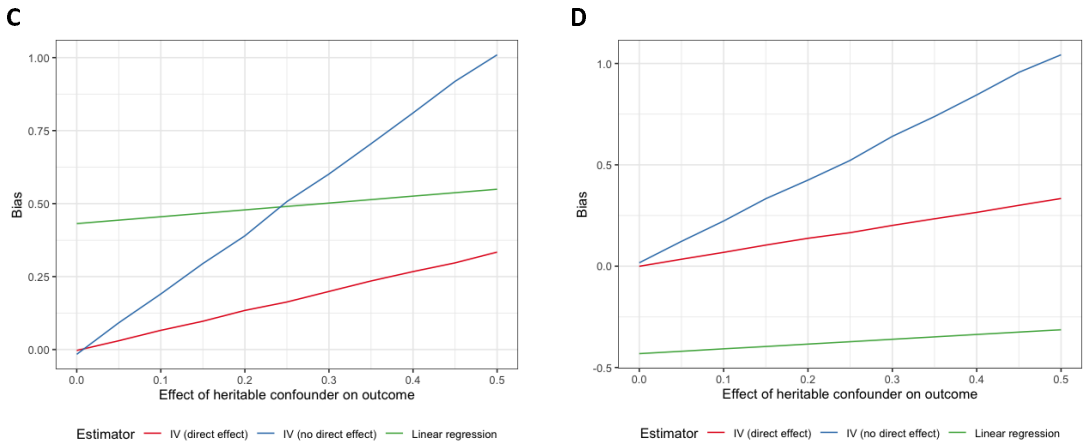


Figure 2 – The Bias of linear regression and IV estimators when instruments have a pleiotropic effect through a single heritable confounder.

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.*

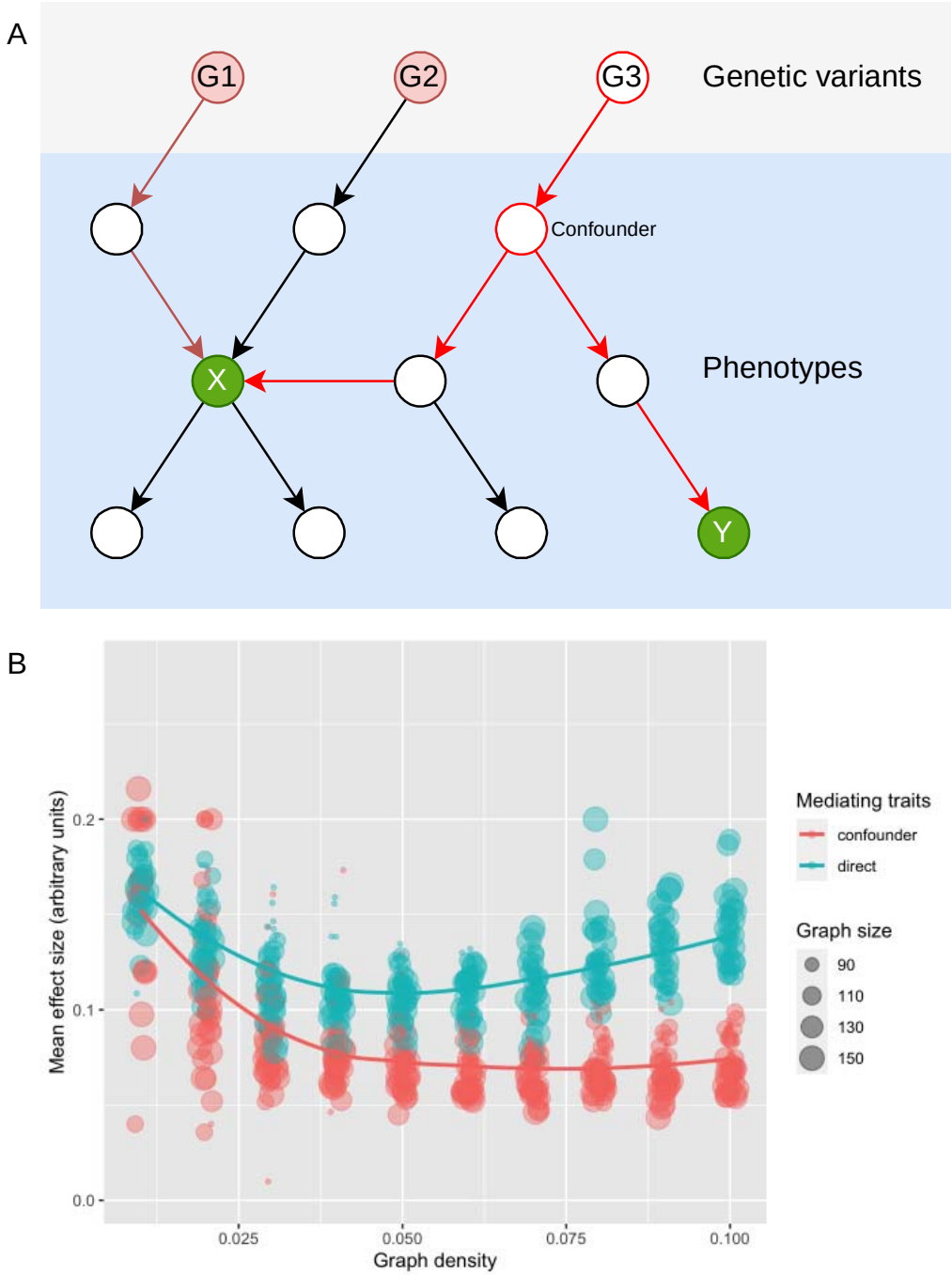


Figure 3. Network-based simulations of the origin of instruments.

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.

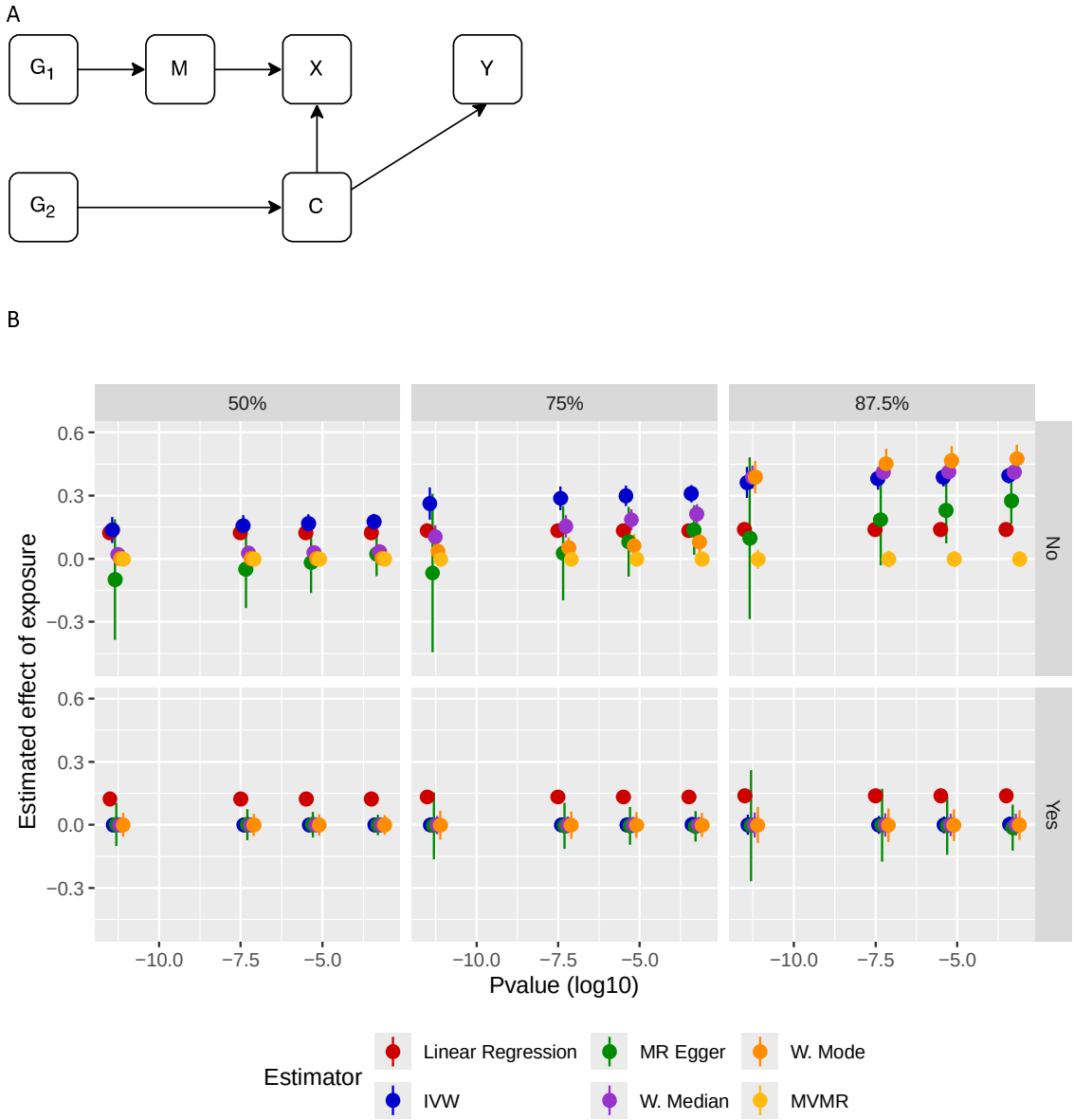


Figure 4 – Simulation of bias from a single heritable confounder in linear regression and MR.

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.

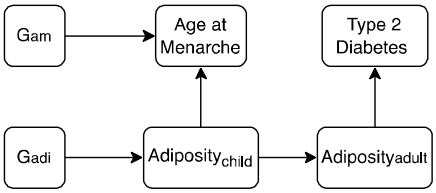
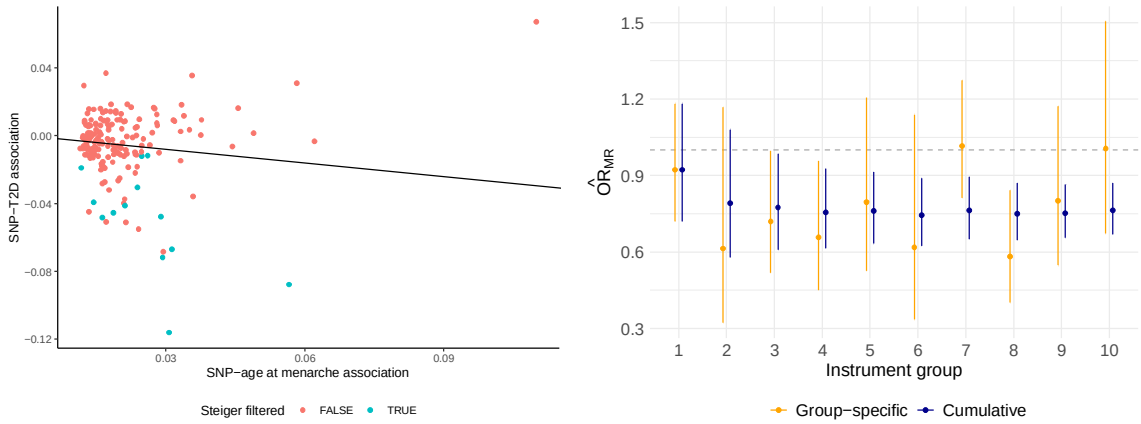
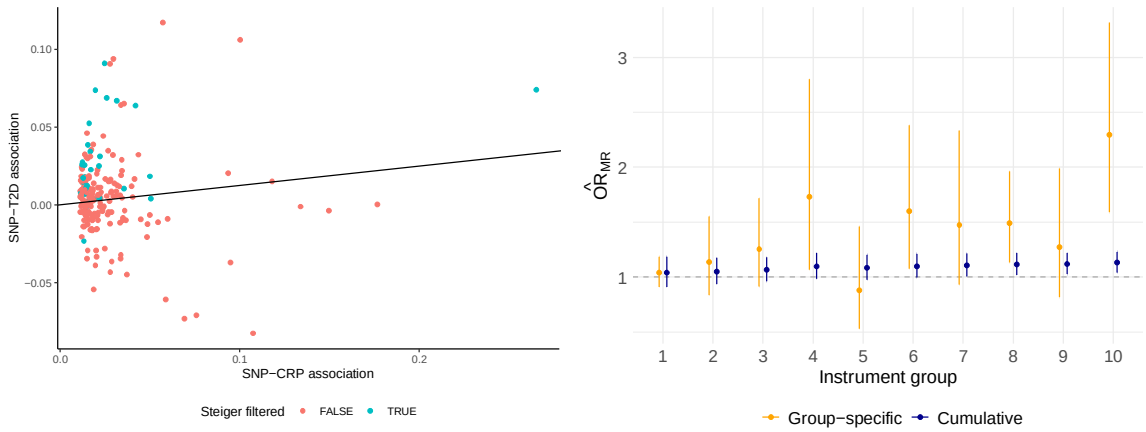


Figure 5 – Illustrative diagram for the hypothesised relationships between adiposity, age at menarche and Type 2 Diabetes. *G_{am} is a set of genetic variants associated with age at menarche, G_{adi} is a set of genetic variants associated with childhood adiposity.*



(a) *(b)*

Figure 6 – MR results from univariable MR estimation in estimation of the effect of age at menarche 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 age at menarche, those SNPs that showed a stronger association with childhood adiposity than age at menarche are highlighted in blue. The fitted line shows the IVW effect estimate using all SNPS. (b) MR Corge per group IVW effect estimates for the effect of age at menarche on T2D (in Odds ratios). SNPs are grouped evenly by their association with age at menarche (18 or 19 SNPs per group). Group 1 has the strongest association with age at menarche. 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 7.*



(a) (b)

Figure 7 – 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 – MR effect estimates for age at menarche on Type 2 diabetes.

	All SNPs included			***Steiger filtering applied***		
	OR	**95% CI**	**p-value**	**OR**	**95% CI.**	**p-value**
IVW	0.76	[0.67, 0.87]	$5.8x10^{-5}$	0.90	[0.81, 1.00]	0.04
MR Egger	0.87	[0.57, 1.15]	0.24	1.11	[0.85, 1.45]	0.45
Weighted median	0.94	[0.84, 1.06]	0.30	0.95	[0.84, 1.07]	0.41
Weighted mode	0.98	[0.84, 1.14]	0.78	0.98	[0.84, 1.05]	0.79
MVMR	0.98	[0.87, 1.11]	0.71	-	-	-

OR: Effect estimates given in odds ratios for an increase in type 2 diabetes risk for a category increase in age at menarche. Univariable MR - 183 SNPs, with Steiger filtering – 170 SNPs, MVMR 274 SNPs.

Table 3 – 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 4 – MR effect estimates for CRP on T2D, with all SNPS included and with SNPs more strongly associated with each of the top 10 confounders than CRP removed individually

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)[57] **BMI from Yengo et al. (2018)[7]*