Causal Effect Estimation in Mendelian Randomisation Studies - Evaluating a Novel Bayesian Approach To Genetic Pleiotropy Versus Established Weighted Median Methodology

B233241

September 2024 - June 2025

Contents

Acknowledgements

I would like to acknowledge

Contributions

Mine others

Statement of Originality

I confirm that all work is my own except where indicated, that all sources are clearly referenced….

Word Count

Word count: 3166

1 Introduction and Background

1.1 Introduction to Mendelian Randomisation (MR)

Epidemiology is the study of determinants and distribution of disease across populations; a common epidemiological study aim is therefore to seek evidence as to whether a given exposure (e.g. cigarette smoking) may cause a given outcome (e.g. lung cancer)[1]. Logistics limit experimental interventions across large groups, so insights into associations between exposures and outcomes are gleaned from observational data of people in the population of interest. Comparing health outcomes between individuals with different levels of a particular exposure may highlight potential links, e.g. higher cancer incidence in those who smoke more is consistent with a causal role for cigarettes in carcinogenesis[1].

However, correlation does not prove causation. A key epidemiological challenge is accounting for so-called "confounding" factors; these are other variables, associated with both the exposure and the outcome of interest, which represent an alternative causal explanation for any exposure-outcome links observed[2]. If smokers also drink more alcohol than non-smokers, then an observed link between smoking and increased cancer risk could plausibly be caused by increased alcohol exposure, either partially or entirely. Another potential issue with observational data is "reverse causation", where the presumed outcome is in fact a cause of the exposure; this might be the case if a cancer diagnosis drove individuals to drink and smoke more, and data were collected without respect to exposure timings.

Mendelian randomisation (MR) is a methodology intended to support causal inference from observational data. It applies the principles of instrumental variable (IV) analysis to genetic data, performing a type of natural experiment often likened to a randomised-controlled trial (RCT)[3].

In a properly conducted RCT, causality can be inferred due to a randomisation process being used as an "instrument" to allocate different levels of exposures to different experimental groups. If groups are randomly allocated, any confounding variables which might otherwise influence exposure-outcome relationships should be evenly distributed between groups, whether these confounders are known or not. As such, there should be no systematic differences between individuals from different groups in the exposure of interest - that is, there should be no bias[4]. Statistical methods can quantify the probability that any observed outcome differences could have occurred by chance, and thereafter any outcome differences can be interpreted as caused by exposure differences. As allocation and receipt of exposures is known to precede outcome measurements, reverse causality is impossible.

In MR, naturally occurring genetic variants - "genetic instruments" – are chosen based on their known association to an exposure of interest. Provided that assumptions of IV analysis are met, random assignment of genetic variants from parents to offspring during meiosis creates randomisation analagous to that performed for an RCT – both measured and unmeasured confounders should be distributed evenly between the groups created, allowing valid causal inference after other sources of bias and random variation are accounted for[5].

1.2 Causal Effect Estimation in MR

At its simplest, the relationship between two continuous variables - an exposure X and outcome Y - can be represented as a linear model:

$$Y = \alpha + \beta X + \epsilon \tag{1}$$

where α represents all non-X determinants of Y, β is the causal effect of X on Y and ϵ is an error term. The β term is a numerical measure of strength of causal exposure-outcome association, where:

- $\beta = 0$ implies no causal link between exposure and outcome
- $\beta > 0$ implies X causes Y
- $\beta < 0$ implies X prevents Y

To estimate a causal effect using a genetic variant in an IV analysis, three key assumptions must be met[6]:

1. Relevance – the genetic variant must be associated with the exposure of interest
2. Independence – the genetic variant is independent of confounders of the relationship between exposure and outcome
3. Exclusion restriction – the genetic variant must not be associated with the outcome except via the exposure

These assumptions are represented graphically in Figure 1.

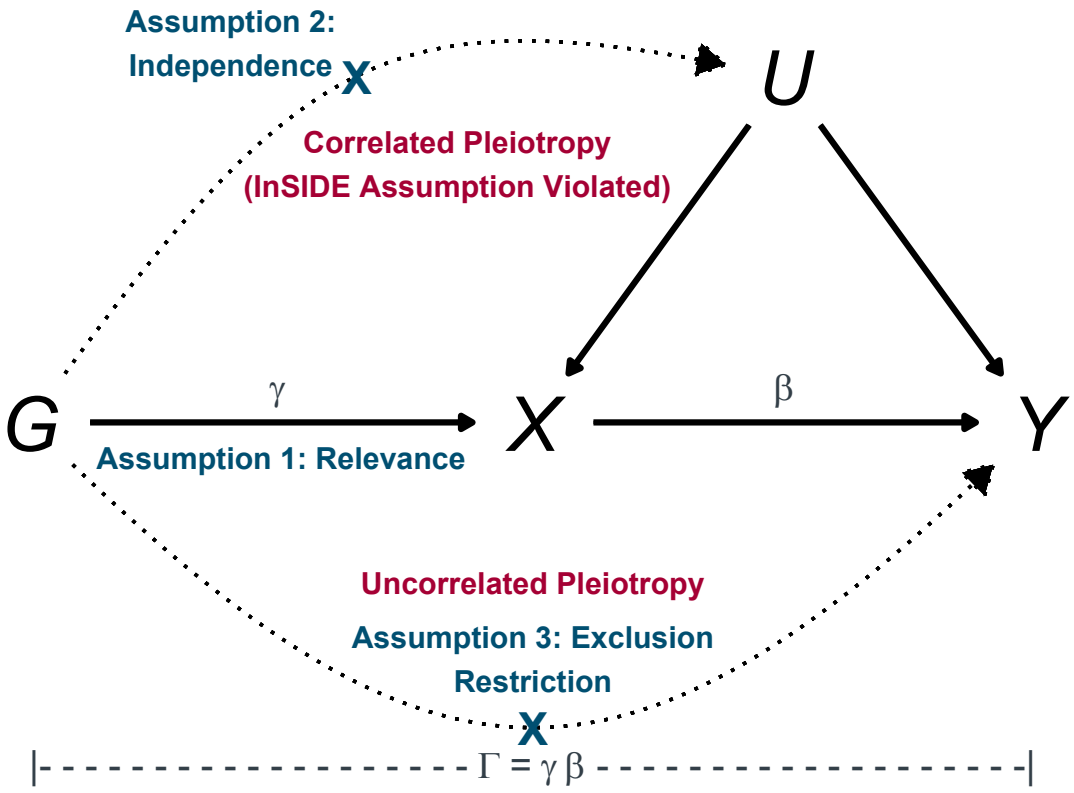

Figure 1: Causal diagram illustrating the relationships between genetic instrument G, exposure X, outcome Y and confounders of the exposure-outcome relationship U in Mendelian randomisation studies. Blue text & crosses represent key assumptions to ensure valid inference of causal effect of X on Y using G as an instrumental variable. Red text represents violations of these assumptions that may lead to invalid inference through opening of alternate causal pathways. Greek characters represent the key parameters/association coefficients to be estimated. Adapted from Burgess et al 2016[7]

Typically, MR studies estimate causal effect using a set of several genetic instruments; the causal effect estimate derived from the jth instrument is denoted $\hat{\beta}_j$. Each estimate $\hat{\beta}_j$ acknowledges there will be specific effects on the observed values of exposure and outcome given the presence of that specific genetic variable G_j under study, i.e. $\hat{\beta}_j$ is based on the observed exposure $X|G_j$ and outcome $Y|G_j$. These observed values of exposure and outcome can be described by their own linear models:

$$X|G_j = \gamma_0 + \gamma_j G_j + \epsilon_{X_j} \tag{2}$$

$$Y|G_j = \Gamma_0 + \Gamma_j G_j + \epsilon_{Y_j} \tag{3}$$

where, for exposure and outcome respectively:

- γ_0 and Γ_0 reflect base values without influence of the genetic variant
- γ_j and Γ_j are coefficients of association with the genetic variant, representing the extent to which an effect allele of G_j will perturb the value of X or Y versus the non-effect allele
- ϵ_{X_j} and ϵ_{Y_j} are error terms, containing contributions from confounders of the exposure-outcome relationship (U in the causal diagram), and all genetic variants except G_j.

It can be shown that a simple causal effect estimate for the exposure on the outcome can be obtained from a single genetic instrument by the Wald method, dividing the coefficient of gene-outcome association by the coefficient of gene-exposure association, i.e.:

$$\hat{\beta}_j = \frac{\hat{\Gamma}_j}{\hat{\gamma}_j} \tag{4}$$

Each instrument may be valid or invalid, depending on it meeting the above assumptions. The overall causal effect estimate $\hat{\beta}$ from any given MR method will typically seek to pool effect estimates from several instruments so as to minimise effects of any invalid instruments included, e.g. by removing/down-weighting contributions of genetic instruments which violate one or more assumptions. This is equivalent to plotting all estimated coefficients of gene-outcome association ($\bar{\Gamma}$) versus all estimated coefficients of gene-exposure association ($\bar{\gamma}$) for the set of instruments, then using the gradient of a regression line through the points as the causal effect estimate $\hat{\beta}$; picking an MR methodology is analogous to choosing the method to draw the line of best fit (Figure 2). For binary outcomes, the causal effect estimate can be converted to an odds ratio (OR) through exponentiation, i.e.:

$$OR = e^{\hat{\beta}} \tag{5}$$

1.3 Violations to Assumptions

In practice, only the relevance assumption can be directly tested and proven. Typically, genetic variants for MR studies are selected as instruments based on Genome Wide Association Studies (GWAS), which quantify associations between genetic Single Nucleotide Polymorphisms (SNPs) and various phenotypes. Association between genetic variants and a phenotypes representing exposures of interest can be partly assured by selection using an appropriate genome-wide significance level (e.g. $p < 10^{-8}$). Statistical testing can also quantify the gene-exposure relationship; commonly used measures include the r^2 statistic, representing the proportion of variance in the exposure explained by the genotype, and the related F-statistic, which additionally accounts for the the sample size under investigation[9]. An F-statistic of ≥ 10 is generally considered to represent a strong enough gene-exposure association to consider a genetic instrument for use[2].

The assumptions of independence and exclusion restriction depend on all possible confounders of the exposure-outcome association, both measured and unmeasured; as such, these can never be proven absolutely. Various methods have been proposed to quantify and account for violations of these two additional assumptions, including the weighted median estimator, described below[8].

The main methods to avoid violations of the independence assumption relate to appropriate selection of populations studied to avoid confounding due to ancestry or population stratification. For example, in two-sample MR studies, where gene-exposure and gene-outcome coefficients are estimated from two separate GWAS studies, it is recommended to select GWAS studies performed in similar population groups (e.g. both in Western Europeans). This practice helps avoid spurious exposure-outcome associations being generated

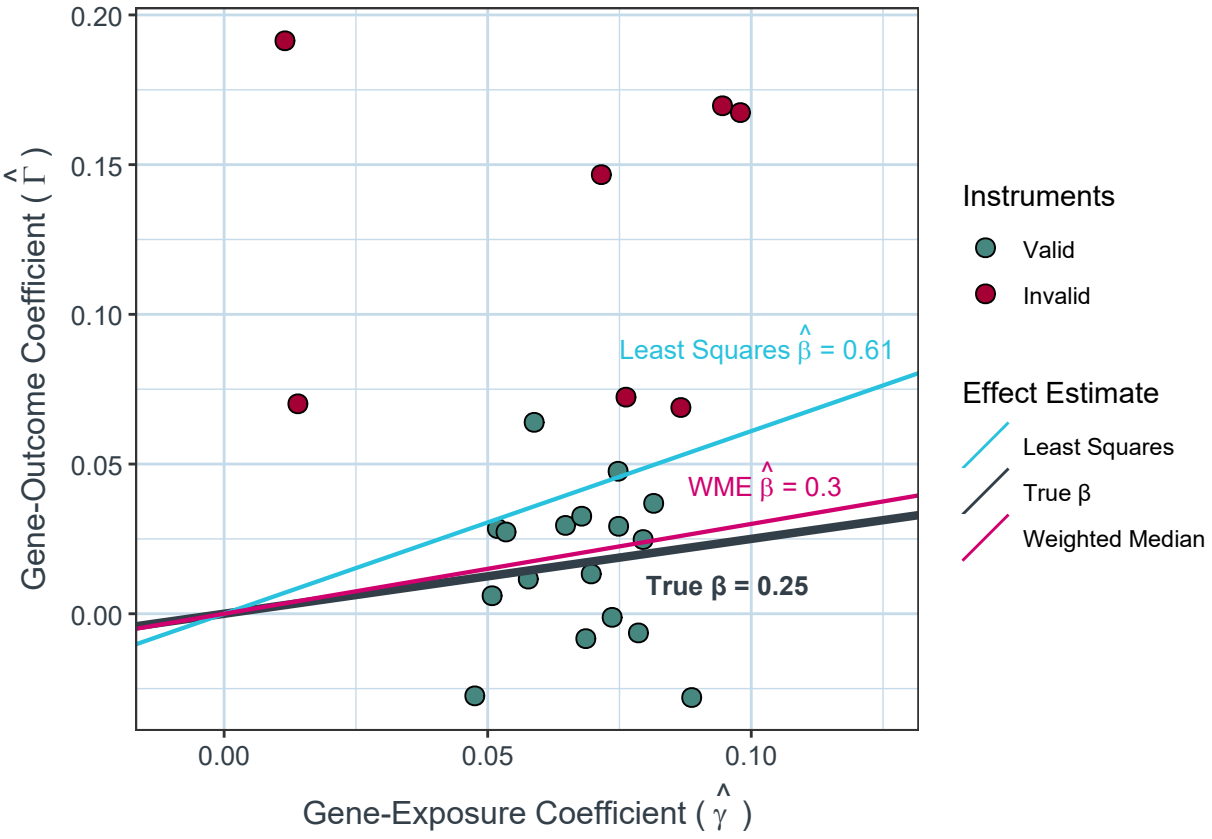

Figure 2: Simulated MR Study on 10,000 individuals using 25 genetic instruments, of which 30% are invalid (red points) and introduce directional pleiotropic effects. The true value of the exposure-outcome causal effect is 0.25 (grey line, causal effect represented by gradient). Regression using an unajusted least-squares linear model (light blue line) results in a biased estimate in the positive direction due to the influence of the invalid instruments. Using the Weighted Median Estimator method (pink line) attenuates the effects of the invalid instruments, resulting in an estimate closer to the true value. Adapted from Bowden et al 2016[8]

by confounding due to underlying differences in allele frequency, baseline disease risks etc between different populations[9].

Exclusion restriction is a particularly universal issue in MR, due to so-called (horizontal) genetic pleiotropy, where a single genetic variant may have multiple "pleiotropic" effects – i.e. it may influence several traits simultaneously. Such pleiotropic effects may be unknown and open unmeasured causal pathways between a genetic instrument and the outcome, thus potentially biasing MR estimates of the association between exposure and outcome. As pleiotropy influences outcome separate to the path involving the exposure of interest, the term "direct effects" is also used[10]. Where pleiotropic effects are in both positive and negative directions with a mean of zero - "balanced pleiotropy" - then they only add noise to causal effect estimation[11]. By contrast, "directional pleiotropy", where the mean of pleiotropic effects is non-zero, may introduce bias[8].

If such an additional causal pathway acts between gene G and outcome Y via a confounding factor U, then the magnitude of direct/overall effects of G on Y will correlate with the effects of G on X (i.e. $\Gamma \propto \gamma$), and "correlated pleiotropy" is present. If an additional causal pathway acts directly between gene G and outcome Y independent of both exposure X and confounders U, this results in "uncorrelated pleiotropy" (Figure 1). Both correlated and uncorrelated pleiotropy can introduce bias which distorts the estimate of the true causal effect. In general, correlated pleiotropy is more challenging to account for; several MR methods explicitly require an additional assumption of Instrument Strength Independent of Direct Effects (the InSIDE assumption), i.e no correlated pleiotropy to be present[12].

1.4 Weighted Median Estimator (WME)

A common approach to produce exposure-outcome causal effect estimates robust to violations of the exclusion restriction assumption is the Weighted Median Estimator (WME) method, proposed by Bowden et al[8].

In WME analysis, several genetic instruments are used to estimate the exposure-outcome causal effect $\hat{\beta}$. Each instrument is known to be associated with the exposure of interest, but an unknown proportion of these instruments may be invalid due to pleiotropic genetic effects. Any instrument linked to an outcome via multiple pleiotropic causal pathways will exhibit a less consistent gene-outcome association than a relationship mediated by a single pathway; this results in larger variance in causal estimates derived from invalid/pleiotropic genetic instruments versus estimates from valid instruments.

WME therefore assigns a weight to each genetic instrument's estimate of the causal effect according to the inverse of the variance of the estimate; these weighted effect estimates are used to construct a cumulative distribution function for probability of true causal effect size across the range of estimated values. The 50th percentile of this distribution can then be taken as a "weighted median estimate" of the true causal effect, theoretically producing consistent causal estimates even if up to 50% of the included information comes from invalid instruments[8]. An example of WME attenuating the effects of invalid instruments is shown in Figure 2.

1.5 Issues With WME

WME calculation methods are available via several prolific MR tools: the R packages "MendelianRandomization"[13] and "TwoSampleMR", and the MR-Base web platform[14]. However, these implement the original authors' suggested process of generating 95% confidence intervals for WME, which deviates from accepted resampling methodology:

> "We found the bootstrap confidence interval…too conservative. However, the bootstrap standard error… gave more reasonable coverage using a normal approximation (estimate ±1.96 x standard error) to form a 95% confidence interval"[8]

This modification, explicitly aiming to boost estimate precision artificially, would be expected to lead to a high Type 1 error rate, which has been a growing concern in the field of late[15].

1.6 MR-Hevo

MR-Hevo[16] is an R package which uses more typical Bayesian methodology to estimate MR causal effects and corresponding 95% confidence intervals. It uses the probabilistic programming language, Stan, to directly sample the posterior probability distribution of pleiotropic effects on the outcome, rather than assuming that this distribution can be simulated as Gaussian, as current WME implementations do. MR-Hevo also handles multiple instruments per genetic locus via scalar construction, and specifies a prior probability distribution which reflects prior knowledge that most individual genetic instruments will have only small effects on complex traits[17,18], further aiding biologically plausible inference regarding distribution of pleiotropic effects.

The main aim of this study will be to demonstrate if the WME approach gives over-confident causal estimates in the presence of pleiotropy, and whether this issue is more correctly handled by the MR-Hevo Bayesian approach.

2 References

1. Coggon D, Rose G, Barker D. Chapter 1. What is epidemiology? | The BMJ. In: The BMJ | The BMJ: Leading general medical journal Research Education Comment [Internet]. 2003 [cited 2025 Apr 29]. Available from: https://www.bmj.com/about-bmj/resources-readers/publications/epidemiology-uninitiated/1-what-epidemiology

2. Martens EP, Pestman WR, Boer A de, Belitser SV, Klungel OH. Instrumental Variables: Application and Limitations. Epidemiology [Internet]. 2006 May [cited 2025 Apr 29];17(3):260. Available from: https://journals.lww.com/epidem/fulltext/2006/05000/instrumental_variables__application_and.10.aspx

3. Hernán MA, Robins JM. Instruments for Causal Inference: An Epidemiologist's Dream? Epidemiology [Internet]. 2006 Jul [cited 2025 Apr 29];17(4):360. Available from: https://journals.lww.com/epidem/fulltext/2006/07000/instruments_for_causal_inference___an.4.aspx#JCL1-2

4. Stel VS, Dekker FW, Zoccali C, Jager KJ. Instrumental variable analysis. Nephrology Dialysis Transplantation [Internet]. 2013 Jul [cited 2025 Apr 29];28(7):1694–9. Available from: https://doi.org/10.1093/ndt/gfs310

5. Davies NM, Holmes MV, Smith GD. Reading Mendelian randomisation studies: A guide, glossary, and checklist for clinicians. BMJ [Internet]. 2018 Jul [cited 2025 Jan 7];362:k601. Available from: https://www.bmj.com/content/362/bmj.k601

6. Lousdal ML. An introduction to instrumental variable assumptions, validation and estimation. Emerging Themes in Epidemiology [Internet]. 2018 Jan [cited 2025 Apr 29];15:1. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5776781/

7. Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants. Epidemiology (Cambridge, Mass) [Internet]. 2016 Nov [cited 2024 Oct 22];28(1):30. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC5133381/

8. Bowden J, Smith GD, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genetic Epidemiology [Internet]. 2016 Apr [cited 2024 Oct 22];40(4):304. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC4849733/

9. Richmond RC, Smith GD. Mendelian Randomization: Concepts and Scope. Cold Spring Harbor Perspectives in Medicine [Internet]. 2022 Jan [cited 2024 Oct 22];12(1):a040501. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC8725623/

10. Hemani G, Bowden J, Smith GD. Evaluating the potential role of pleiotropy in Mendelian randomization studies. Human Molecular Genetics [Internet]. 2018 May [cited 2024 Oct 23];27(R2):R195. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC6061876/

11. Morrison J, Knoblauch N, Marcus JH, Stephens M, He X. Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. Nature Genetics [Internet]. 2020 Jul [cited 2025 May 22];52(7):740–7. Available from: https://www.nature.com/articles/s41588-020-0631-4

12. Grant AJ, Burgess S. A Bayesian approach to Mendelian randomization using summary statistics in the univariable and multivariable settings with correlated pleiotropy. The American Journal of Human Genetics [Internet]. 2024 Jan [cited 2025 May 20];111(1):165–80. Available from: https://www.cell.com/ajhg/abstract/S0002-9297(23)00433-0

13. Yavorska OO, Burgess S. MendelianRandomization: An R package for performing Mendelian randomization analyses using summarized data. International Journal of Epidemiology [Internet]. 2017 Dec [cited 2024 Oct 23];46(6):1734–9. Available from: https://doi.org/10.1093/ije/dyx034

14. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. Loos R, editor. eLife [Internet]. 2018 May [cited 2025 Jan 7];7:e34408. Available from: https://doi.org/10.7554/eLife.34408

15. Stender S, Gellert-Kristensen H, Smith GD. Reclaiming mendelian randomization from the deluge of papers and misleading findings. Lipids in Health and Disease [Internet]. 2024 Sep [cited 2025 Jan 7];23(1):286. Available from: https://doi.org/10.1186/s12944-024-02284-w

16. McKeigue PM, Iakovliev A, Spiliopoulou A, Erabadda B, Colhoun HM. Inference of causal and pleiotropic effects with multiple weak genetic instruments: Application to effect of adiponectin on type 2 diabetes [Internet]. medRxiv; 2024 [cited 2024 Oct 23]. Available from: https://www.medrxiv.org/content/10.1101/2023.12.15.23300008v2

17. Park JH, Wacholder S, Gail MH, Peters U, Jacobs KB, Chanock SJ, et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. Nature genetics [Internet]. 2010 Jun [cited 2024 Oct 23];42(7):570. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC4615599/

18. Piironen J, Vehtari A. Sparsity information and regularization in the horseshoe and other shrinkage priors. Electronic Journal of Statistics [Internet]. 2017 Jan [cited 2025 Jan 7];11(2):5018–51. Available from: https://projecteuclid.org/journals/electronic-journal-of-statistics/volume-11/issue-2/Sparsity-information-and-regularization-in-the-horseshoe-and-other-shrinkage/10.1214/17-EJS1337SI.full

A Appendix A: List of Abbreviations

MR Mendelian randomisation

B Appendix B: Simulation Code

B.0.1 Generating Data and Models

The data generating model used was from Appendix 3 of Bowden et al (ref); the relevant section describing their model is reproduced below:

"...

$$U_i = \sum_{j=1}^{J} \phi_j G_{ij} + \epsilon_i^U \tag{6}$$

$$X_i = \sum_{j=1}^{J} \gamma_j G_{ij} + U_i + \epsilon_i^X \tag{7}$$

$$Y_i = \sum_{j=1}^{J} \alpha_j G_{ij} + \beta X_i + U_i + \epsilon_i^Y \tag{8}$$

for participants indexed by $i = 1, ..., N$*, and genetic instruments indexed by* $j = 1, ..., J$*.*

The error terms $\epsilon_i^U, \epsilon_i^X$ *and* ϵ_i^Y *were each drawn independently from standard normal distributions. The genetic effects on the xposure* j *are drawn from a uniform distribution between 0.03 and 0.1. Pleiotropic effects* α_j *and* ϕ_j *were set to zero if the genetic instrument was a alid instrumental variable. Otherwise (with probability 0.1, 0.2, or 0.3):*

1. In Scenario 1 (balanced pleiotropy, InSIDE satisfied), the α_j *parameter was drawn from a uniform distribution between* -0.2 *and 0.2.*

2. In Scenario 2 (directional pleiotropy, InSIDE satisfied), the α_j *parameter was drawn from a uniform distribution between 0 and 0.2.*

3. In Scenario 3 (directional pleiotropy, InSIDE not satisfied), the ϕ_j *parameter was drawn from a uniform distribution between* -0.2 *and 0.2.*

The causal effect of the exposure on the outcome was either $\beta X = 0$ *(null causal effect) or* $\beta X = 0.1$ *(positive causal effect). A total of 10 000 simulated datasets were generated for sample sizes of N = 10 000 and 20 [sic] participants. Only the summary data, that is genetic associations with the exposure and with the outcome and their standard errors as estimated by univariate regression on the genetic instruments in turn, were used by the analysis methods. In the two-sample setting, data were generated on 2N participants, and genetic associations with the exposure were estimated in the first N participants, and genetic associations with the outcome in the second N participants."*[8]

To reproduce this model, code was written in R to generate the relevant participant level data. First, a function (`simulate_MR_data`) was written which included arameters specified by Bowden et al, and also to allow testing of data simulation:

This initial simulation function generated data in the following format:

A function (`extract_models`) was then written to create linear models from each dataset generated as per Bowden et al:

These models generated estimates of the coefficient of gene:exposure association (`coeff_G_X`), coefficient of gene:outcome association (`coeff_G_Y`), and the elevant standard errors of these estimates. The values of parameters inputted were also returned to aid in further testing of data/model generation, i.e. actual ene:exposure associations (`gamma`), pleiotropic effects of invalid instruments (`alpha`), additional pleiotropic effects when InSIDE assumption not satified `phi`), causal effect of exposure on outcome (`beta`) and the proportion of invalid genetic instruments with pleiotropic effects on the outcome (`prop_invalid`).

B.0.2 Testing Generation of Data and Models

A series of test plots were used to verify that data were simulated as intended under the various conditions specified by input parameters. Test plots were not reated for the parameters `n_participants`, `n_instruments` or `n_datasets`, as the functioning of these parameters could be readily inferred from the structure of he datasets outputted, as above.

The `prop_invalid` parameter specifies the proportion of invalid genetic instruments simulated, i.e. the proportion of genetic instruments affecting the outcome via irect/pleiotropic effects, and thus not solely via the exposure of interest. If simulated correctly, increasing the value of `prop_invalid` should increase the umber of instruments with pleiotropic effects, i.e. instruments with `alpha` $=/= 0$. With random error terms set to 0 and no causal effect present (i.e. `rand_error = ALSE` and `causal_effect = FALSE`), the estimated gene:outcome coefficient estimated using any given instrument will equal the pleiotropic effects of that instrument i.e. `coeff_G_Y = alpha`), and therefore will only be non-zero for invalid instruments with non-zero pleiotropic effects on the outcome . Plotting `coeff_G_Y` gainst `alpha` for simulated data with no causal effect or random error should therefore yield a graph where

- For valid instruments: gene:outcome coefficient = alpha = 0
- For invalid instruments: gene:outcome coefficient = alpha $=/= 0$, with values spread uniformly between `alpha_min` and `alpha_max`

Similarly, with random error terms set to 0 and no causal effect present, gene:exposure coefficients estimated for each instrument should exactly match the actual alues simulated, i.e. `coeff_G_X = gamma` for all instruments:

For the next phase of testing, a function (`plot_GY_GX`) was written to plot the coefficients for gene:exposure versus gene:outcome as estimated using the reviously created linear models:

C Appendix C: Citation Search Strategy