3. Methods

B233241

2025-01-07

Contents

Word count: 1507

Simulation Study

To evaluate the performance of MR-Hevo causal estimation relative to WME, the precision and consistency of both methods were quantified using simulated datasets with known parameter values.

Data Simulation

To aid comparability with existing methods and literature, the simulation methodology of the original WME exposition was reproduced based on published models and parameters in Appendix 3 of its supplementary materials[1]. Full details of simulation reproduction, including code and validation of outputs, is presented in **??**.

In brief, simulations were created based on three different scenarios, each representing a common set of assumptions about underlying data used for MR, and each increasingly challenging to the performance of any given MR causal estimation methodology:

1. Balanced pleiotropy, InSIDE assumption satisfied - A proportion of invalid genetic instruments are present and introduce pleiotropic effects uncorrelated with the instrument strength; these pleiotropic effects are equally likely to be positive as negative with a mean value = 0, thus introducing noise into the estimation of causal effect.

2. Directional pleiotropy, InSIDE assumption satisfied - A proportion of invalid genetic instruments are present and introduce pleiotropic effects uncorrelated with the instrument strength; these pleiotropic effects are positive only, with a mean value > 0, thus biasing the causal effect estimate in a positive direction.

3. Directional pleiotropy, InSIDE assumption not satisfied - A proportion of invalid genetic instruments are present and introduce pleiotropic effects correlated with the instrument strength through action via a confounder; these pleiotropic effects are positive only, with a mean value > 0, thus potentially biasing the causal effect estimate in a positive direction to an even greater extent than Scenario 2.

1,000 simulated datasets of participant-level data were generated for every combination of each scenario and each the following simulation parameters:

- Proportion of invalid instruments: 0%, 10%, 20% or 30%
- Number of participants: $n = 10,000$ or $n = 20,000$
- Causal effect: null ($\beta = 0$) or positive ($\beta = 0.1$)

The same set of 25 simulated genetic instruments were used across all datasets, with the status of each as valid/invalid determined by random draw per instrument at the start of each simulation run of 1,000 datasets.

Genotypes were simulated as for a two-sample setting: where number of particpants was $n = 10,000$, 20,000 genotypes were simulated - 10,000 for the cohort used to estimate gene-exposure association ($\hat{\gamma}$), and a separate cohort of 10,000 used to estimate gene-outcome association ($\hat{\Gamma}$). Parameter values for effect allele frequency were not specified by Bowden et al, though initial testing showed values around 0.5 produced WME causal effect estimates closest to published values when other parameters were matched[1]. As such, effect allele frequencies were assigned per instrument from a uniform distribution between 0.4 to 0.6. Each effect allele freqency thus generated per instrument was then used as a probability to assign each simulated participant effect alleles for each instrument via two draws from a binomial distribution.

Analysis of Simulated Data

Each dataset generated was analysed using both WME and MR-Hevo methods, via functions from the `TwoSampleMR` and `mrhevo` packages, respectively[2,3]. Results were aggregated per group of 1,000 simulated datasets corresponding to a particular combination of scenario and parameter values. This resulted in one meta-analysis reported per combination of scenario/parameter values, each including 1,000 simulated MR studies using the same 25 genetic instruments in the same population. Aggregated measures for both WME and MR-Hevo per meta-analysis were mean causal effect estimate; mean standard error of the causal effect estimate; and causality report rate, i.e. percentage of simulated studies reported as showing a non-null causal effect, either by p-value <0.05 (WME), or by a 95% credible interval for causal effect estimate not including 0 (MR-Hevo).

Results of the above aggregations were tabulated as per Tables 2 and 3 of Bowden et al[1] to allow direct comparisons of both methods versus each other and versus the published characteristics of existing MR causal estimation methods.

Re-Analysis of Published Data

To investigate the potential implications of any differences in performance between WME and MR-Hevo methods, a selection of published studies resporting causal effect estimates using the WME method was re-analysed. A sample size of 10 published studies was decided as a pragmatic compromise between the scope of this study and the need to check consistency of any observed differences. In the original Bowden et al simulation studies, the WME causal estimation method was shown to generate a false-positive report rate of $\geq 30\%$ with relatively minor violations of relevant assumptions[1]; therefore, even this relatively small sample of 10 studies might be expected to demonstrate differences between methods if the MR-Hevo approach is as appropriately conservative as its creators propose.

To estimate the upper bound of the potential impact of MR-Hevo versus existing WME methodology, studies were chosen for re-analysis based on their number of citations in the wider MR literature. Compared to studies with few or no citations, highly-cited studies would be expected to have a larger impact on their respective fields if their conclusions were to change. In addition, highly-cited works will typically have been submitted to more scrutiny than less-cited works - both during peer review whilst under consideration by journals likely to produce highly-cited works, and from the wider scientific community following the

widespread dissemination evidenced by a high citation count. As such, it would be expected that highly-cited works are likely to be free of significant methodological flaws which may impede interpretation of any re-analysis.

Citation Search

The Scopus search platform [@] was used on 15/04/2025 to retrieve all articles citing the original weighted median estimator exposition paper[1]. The articles returned were sorted by the number of times each article itself had been cited, and the resulting list was saved to RIS format in blocks of ten articles for upload into the Covidence evidence synthesis platform. Abstracts were screened by a single reviewer (B233241), starting with the most cited article and proceeding in descending order of citation count, against the following inclusion and exclusion criteria:

Inclusion criteria:

- Original two-sample MR study
- Able to determine samples' ancestry sufficient to establish presence/potential degree of participant overlap between groups
- Reporting $\geq$ 20 human genetic instruments relating to exposure
- Reports details of effect/non-effect alleles
- Regression coefficients and standard errors and/or confidence intervals available for each genetic instrument used
- Uses Weighted Median Estimator

Exclusion criteria:

- Methodology paper, review article, editorial or letter
- English full-text not accessible

Where eligibility could not be determined from abstract screening alone, full texts were retrieved and screened against the same criteria. Screening of abstracts and full texts was undertaken in blocks of ten articles, until the target of ten included studies for reanalysis had been reached.

Where an article reported multiple exposure-outcome associations, data were only extracted for the association with the highest number of genetic instruments available, or else for the first reported association where several were based on the same number of instruments. Data were extracted from full texts of included studies using a standardised data collection template, which included publication details, citation count, primary study question, degree of participant overlap between groups, number/details of genetic instruments used, effect estimates/standard errors calculated, and conclusion regarding causality as determined by the weighted median estimator method.

Data Manipulation and Analysis

All simulations, data manipulations and data analyses were performed in R version 4.4.1 (2024-06-14 ucrt)[4].

For the simulation study, full details of computation are available in Appendix **??**.

For citation search data, a standardised data collection form was Microsoft Excel[5] to create .csv files for subsequent analysis in R; Excel's "Get Data" function was also used to extract tables of genetic instruments where these were presented in non-csv format (e.g. pdf).

Data cleaning for citation search data was primarily undertaken using the Tidyverse suite of R packages[6]. A full list of packages used can be found in Appendix @(ref:appendix-pkg).

Data were manually screened at summary level and relevant features were extracted. Data were checked for completeness, consistency, duplicate values and plausibility. Data were transformed to an appropriate data type, and encoding of genetic variables was standardised into a single format. Accounting for missing genetic data was not required, as studies were only included if details of genetic instruments used were reported in full. It was noted during early testing that MR-Hevo functions do not operate correctly when zero values are present in coefficients of genetic association or their standard errors; such zero values were therefore re-coded as an arbitrarily low value of 10^{-100}.

Ethical Approval

The protocol for this work has been reviewed and approved by the Usher Masters Research Ethics Group (UMREG) at the University of Edinburgh, Ethics ID: UM241126. Due to the nature of the project, using simulated and publically available data only, no significant ethical issues were foreseen, and sponsorship was deemed unnecessary by the UMREG reviewing panel.

1. 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/

2. 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

3. 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

4. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2024. Available from: https://www.R-project.org/

5. Microsoft Corporation. Microsoft Excel [Internet]. 2018. Available from: https://office.microsoft.com/excel

6. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the tidyverse. Journal of Open Source Software. 2019;4(43):1686.