Ten Simple Rules for conducting a Mendelian randomization study

Sarah A Gagliano Taliun[1,2,*], David M Evans[3,4,5]

[1]Faculté de Médecine, Université de Montréal, Québec, Canada.
[2]Montréal Heart Institute, Montréal, Québec, Canada.
[3]Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland Australia
[4]MRC Integrative Epidemiology Unit, University of Bristol, Oakfield House, Bristol, UK.
[5]University of Queensland Diamantina Institute, Translational Research Institute, University of Queensland, Brisbane, Queensland Australia.

* sarah.gagliano@icm-mhi.org

Introduction

Mendelian randomization (MR) is an epidemiological technique for estimating causal relationships using observational data, which has become very popular in recent years following publication of a seminal article by George Davey Smith and Shah Ebrahim in 2003 [1]. MR is a specific form of "instrumental variables" (IV) analysis (the latter being first invented by Phillip and Sewall Wright in the 1920s [2]) that uses genetic variants to proxy a modifiable variable (which we term the "exposure" variable here) in order to estimate the causal relationship between the exposure and an outcome of interest. To understand how this causal inference technique works, it is useful to think of MR as similar to a "natural" randomized controlled trial [3] where individuals are randomly assigned to groups based on the alleles that they inherit from their parents (Fig 1). MR takes advantage of Mendel's Laws of Segregation and Independent Assortment, which state that offspring inherit alleles randomly from their parents, and randomly with respect to other genes in the genome (with certain exceptions [1]). Therefore, genetic variants that are related to an exposure of interest can be used to proxy the part of the exposure variable that is independent of possible confounding influences from the environment and other traits. Providing several assumptions are satisfied (see below), and the principle of gene-environment equivalence (i.e. perturbing the exposure genetically has the same effect as perturbing the exposure by other means), statistical association between the genetic variant and the outcome is indicative of a causal relationship between the exposure and the outcome, and can be used to estimate the magnitude of the causal relationship using IV methods. Although originally developed as a way to estimate causal relationships between modifiable environmental exposures and medically relevant outcomes, in recent years, MR has been utilized in many other situations including studies of molecular biomarkers, in pharmacogenetics, in the social sciences, and in other discplines that use observational frameworks [4,5].

Given the growing number of MR studies in the literature and the increasing amount of publicly available GWAS datasets and variant-trait association summary statistics, which make such

studies feasible, we describe Ten Simple Rules for conducting a MR study. Our aim is not to provide a comprehensive and detailed overview of MR (which can be found elsewhere [4,5,6]), but rather to present a starting place for researchers to prepare to conduct, and to begin to critically evaluate existing MR studies.

Rule 1: Have a clear research question.
Specify your relationship of interest; that is to say, does having trait A (exposure) or having a particular level/dose of trait A cause trait B (outcome)? "Trait" should be interpreted broadly and could refer to, for example, a disease, an environmental exposure, a molecular biomarker and/or a quantitative trait. Often the exposure is modifiable (alcohol consumption or vitamin levels are two examples), so that there is a potential opportunity to intervene on the variable if the MR analyses provide evidence supporting a causal relationship between the exposure and the outcome. Nevertheless, investigating factors which are not easily modifiable (such as adult height or birth weight) in an MR framework can also be informative from a mechanistic perspective. The interrogation of causality by MR does not necessarily involve a single exposure and outcome pair. You could, for instance, conduct a MR-phenome-wide association study [7] looking at potential causal relationships between a single exposure and multiple outcomes, or alternatively between multiple exposures and a single outcome. Regardless of the choice of exposure(s) and outcome(s), your underlying research question must be clear.

Rule 2: Keep in mind the core IV assumptions.
There are three core assumptions genetic variants must satisfy in order to be considered instrumental variables for testing hypotheses about whether an exposure is causally related to an outcome of interest. The first is that the genetic variants used to proxy for the exposure are robustly associated with the exposure. The second assumption is that there is no confounding (measured or unmeasured) of the genetic variants with the outcome. The third assumption is that the variants potentially influence the outcome *only* through the exposure. Only the first of these assumptions can be proven definitively. That is to say, you can obtain statistical evidence that your genetic variants are related to the exposure and compute a measure (typically the F-statistic from a regression of the exposure on the variant is used) of the strength of this association [8]. The more variance the genetic variant explains in the exposure and the larger your sample size, the more powerful your analysis and the more accurate your estimate of the causal effect of exposure on outcome. For the remaining assumptions, sensitivity analyses should be performed, if possible, to assess whether the assumptions are likely to have been violated. For example, different genetic variants exhibiting differences in the magnitude of the estimated causal effect suggests (i) the presence of horizontal pleiotropy (where a variant affects multiple phenotypes) and (ii) violation of the third IV assumption. In the case of the second IV assumption, Mendel's Laws of Segregation and Independent Assortment strongly suggest that genetic variants should be unrelated to environmental and genetic confounding variables, respectively. Empirical tests of the relationship between genetic variants and known confounders of the exposure-outcome relationship can increase confidence in the validity of this assumption, but are not definitive in that other unmeasured confounders of the exposure-outcome relationship could still be associated with the genetic variant. Additionally, other processes can generate spurious associations between the genetic variant and the outcome including population stratification, selection bias and

dynastic effects [9]. Investigators should be particularly cognizant of the potential for population stratification to reintroduce confounding into the MR analysis and take actionable steps to control for this possibility, such as including ancestry informative principal components in the statistical model.

Whilst the above three core assumptions are sufficient for testing whether an exposure causes the outcome, in order to obtain accurate point estimates of the causal effect, further (strong) assumptions regarding the form of the relationship between the genetic variant, exposure and outcome also need to be made (e.g. linearity) [10].

Rule 3: Be attentive when selecting genetic variants to be used as instruments.
Decreasing genotyping costs, the emergence of large-scale biobanks and GWAS meta-analytic consortia, and the wide-spread availability of variant-trait association summary statistics and databases in the public domain, such as MR-Base [11], have facilitated the identification and utilization of genetic instruments for MR studies. There are no set rules for selecting the "best" set of genetic instruments for an MR study. For example, a well-powered MR analysis using a single variant with a well understood mechanism of action (and unlikely to involve horizontal pleiotropy) may be a superior strategy to performing a MR study as opposed to using as many genome-wide significant variants as possible to proxy the exposure [12]. Decisions related to genetic instrument selection are made on a case-by-case basis, but guidelines have been developed to assist in this process [13]. Important considerations include strength of the variant-exposure association (the more robust the better), variant independence (genetic variants should not be in linkage disequilibrium unless this correlation is explicitly modelled in the analysis), and the likelihood of pleiotropy (which may be a rationale for variant exclusion). Selecting genetic variants that are appropriate for your sample is also key; in particular, it is important to be aware of any variants that exert ancestry-, sex-, or age-dependent effects.

Rule 4: Consider the possibility of reverse causality.
Do the genetic variants exhibit their primary association with the exposure variable as opposed to the outcome variable? For example if variable A has a large causal effect on variable B, then genetic variants primarily associated with variable A will reach genome-wide significance in a GWAS of variable B given a large enough sample. These variants could be erroneously used as instruments for estimating the causal effect of variable B on variable A when in fact their primary association is with variable A. The use of such variants would bias the results of MR analyses of the causal effect of variable B on variable A and potentially provide spurious evidence of reverse causality due to mis-specification of the primary trait. Steiger filtering [14] can be used to identify a set of genetic variants that have their primary association with the exposure of interest. If bidirectional causal relationships are a possibility (i.e. variable A causes variable B, and variable B causes variable A) then consider using "reciprocal MR" in which exposure and outcome are instrumented and MR performed in both directions [15].

Rule 5: Understand the pros and cons of using one versus two sample MR.
Perform one or two sample MR using one of the many well-documented software packages that are freely available (e.g. the "two sample MR package" in the R statistics software), or the MR-Base web utility that uses the MRC IEU OpenGWAS data infrastructure of harmonized GWAS

summary datasets and metadata [11,16]. One sample MR has the advantage that it is possible to confirm that genetic markers used in the analysis are independent of known confounding variables, and also permits many specialized types of MR analysis (e.g. gene by environment MR [17]; factorial MR [18]; non-linear MR [19] etc.). Potential disadvantages are that large samples may be difficult to obtain, which lowers power, and that any bias from weak instruments (that is, genetic markers that are not robustly related to the exposure in the sample under study) will be in the direction of the observational association [8]. Two sample MR (i.e. obtaining variant-exposure and variant-outcome effect sizes from two different datasets) is often advantageous in terms of statistical power, in that publicly available summary results data from large genome-wide association consortia can be used inexpensively, easily and efficiently. This approach can boost sample size and facilitate the analysis of expensive/hard to measure exposures or outcomes. However, it assumes that the different exposure and outcome datasets are ancestrally homogeneous and that the same causal process operates in both datasets. Any bias from weak instruments tends to be towards the null [4,20]. If performing two sample MR, ensure that the effect of the variant on the exposure variable and the effect of the variant on the outcome variable correspond to the same allele. Be careful if using pallindromic varants (i.e. A/T or C/G variants), so that variant-exposure and variant-outcome effect estimates correspond to the same strand. If using several independent variants, causal effect estimates can be combined by weighting them by the inverse of their variance (i.e. termed the "inverse-variance weighted (IVW) MR" method).

Rule 6: Visualize results.
Graphical visualization can be useful for checking the validity of MR assumptions. Forest plots, which display causal effect estimates across the different genetic variants, can be useful in terms of identifying outliers and potential pleiotropic variants. Funnel plots, which graph variant instrument strength (y-axis) against causal effect estimate, can be useful in identifying the presence of directional pleiotropy in the data.

Rule 7: Run sensitivity analyses to increase confidence in the validity of the results.
Whilst IVW MR is the most statistically powerful approach to combine/meta-analyze causal effect estimates [21], it assumes the complete absence of horizontal pleiotropy. Therefore, perform tests of heterogeniety to investigate whether estimates of the causal effect differ across the various genetic variants [22]. Different causal effect estimates suggest the presence of horizontal pleiotropy and can flag outlying variants for further investigation. Perform sensitivity analyses that relax the strict assumption of no horizontal pleiotropy. These different methods include, but are not limited to, random effects MR, the MR modal-based estimator [23], weighted median MR [24], MR-Egger regression [25], MR-RAPS [26], and simulation-based heterogeneity and outlier tests (e.g MR-PRESSO) [27]. Also consider assessing potential biases due to measurement error in the variant-exposure associations [28], sample overlap [29] and selection bias [30]. Consistent causal effect estimates across the different methods improves confidence in the validity of the MR results. In certain cases, it may be possible to utilize an informative gene by environment interaction to inform on the presence or absence of horizontal genetic pleiotropy. For example, in MR studies of the relationship between alcohol consumption and disease outcomes, an association between genetic variants proxying number of units of alcohol consumed per day and disease status should not be present in the subpopulation of

individuals who do not consume alcohol. Indeed, the existence of such associations may indicate the presence of horizontal pleiotropy.

Rule 8: Document code and ensure reproducibility.
Replication is essential for advancing science, and code transparency in computational research is a key step in facilitating reproducibility. There are papers to help with fostering reproducible computational research, including from the Ten Simple Rules series of *PLOS Computational Biology* [31]. With this premise in mind, code should be clear, concise, and well-documented in a manner that allows others to replicate your results. Using an online open-source code collaboration tool, such as GitHub (https://github.com), and getting your code independently tested are useful ways to share code and verify reproducibility.

Guidelines for reporting MR studies have been proposed [32]. As well as sharing code, it is helpful to document how the datasets have been constructed, such as the characteristics of the participants in the GWAS studies (especially in cases where sharing of individual level datasets is not possible), and to present detailed summary results data (effect alleles, strand, effect sizes, allele frequencies, p values etc.) for the individual genetic variants used to proxy for the exposure.

Rule 9: Carefully interpret results and acknowledge limitations.
The critical appraisal checklist available in a review paper on interpreting Mendelian randomization studies offers concrete guidance for the interpretation of results [33]. Of note, an essential point to consider when interpreting such studies is whether gene-environment equivalence is reasonable; that is, do changes caused by genotypes have the same downstream effects as if they were caused by modifiable exposures? Additionally, results from MR cannot necessarily be generalized to individuals who differ from those from which the effect sizes were derived, such as individuals from different ancestries, environments, sexes, and ages. The non-transferability of results is one reason why it is crucial to provide detailed information on the characteristics of the datasets used in the analysis, as noted in the previous rule.

Rule 10: Disseminate findings to the research community.
Now it is time to formally share the results from your efforts with colleagues and the broader research community through a scientific publication and/or a conference presentation. There is helpful advice provided in earlier articles of this Ten Simple Rules series of *PLOS Computational Biology* for disseminating research through written and oral communication [34,35].

Conclusion
MR uses genetic variant-trait associations to estimate the causal effect of an exposure variable on an outcome. Originally developed to estimate causal relationships between modifiable environmental exposures and medically relevant outcomes, the scope of the MR paradigm has widened to include applications in fields as diverse as molecular biology, pharmacology and the social sciences. When conducted appropriately and its results triangulated with substantive knowledge and results using other research methodologies, MR can be a powerful tool for informing causality.

Acknowledgements

Figure legends

Fig 1. Similarities between the Mendelian randomization (MR) study design and a randomized controlled trial.

References

1. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003;32(1):1-22. doi: https://doi.org/10.1093/ije/dyg070

2. Stock JH, Trebbi F. Retrospectives: Who Invented Instrumental Variable Regression? J Econ Perspect. 2003;17(3):177-194. doi: https://doi.org/10.1257/089533003769204416

3. Davey Smith G, Holmes MV, Davies NM, Ebrahim S. Mendel's laws, Mendelian randomization and causal inference in observational data: substantive and nomenclatural issues. Eur J Epidemiol. 2020;35:99-111. doi: https://doi.org/10.1007/s10654-020-00622-7

4. Zheng J, Baird D, Borges M-C, Bowden J, Hemani G, Haycock P, et al. Recent developments in Mendelian randomization studies. Curr Epidemiol Rep. 2017;4(4):330-345. doi: https://doi.org/10.1007/s40471-017-0128-6

5. Evans DM, Davey Smith G. Mendelian randomization: new applications in the coming age of hypothesis-free causality. Annu Rev Genomics Hum Genet. 2015;16:327-350. doi: https://doi.org/10.1146/annurev-genom-090314-050016

6. Burgess S, Davey Smith G, Davies NM, Dudbridge F, Gill D, Glymour MM, et al. Guidelines for performing Mendelian randomization investigations [version 2; peer review: 2 approved]. Wellcome Open Res. 2020;4:186. doi: https://doi.org/10.12688/wellcomeopenres.15555.2

7. Millard LAC, Davies NM, Timpson NJ, Tilling K, Flack PA, Davey Smith G. MR-PheWAS: hypothesis prioritization among potential causal effects of body mass index on many outcomes, using Mendelian randomization. Sci Rep. 2015;5:16645. doi: https://doi.org/10.1038/srep16645

8. Burgess S, Thompson SG, CRP CHD Genetics Collaboration. Avoiding bias from weak instruments in Mendelian randomization studies. Int J Epidemiol. 2011;40(3):755-764. doi: https://doi.org/10.1093/ije/dyr036

9. Davies NM, Howe LJ, Brumpton B, Havdahl A, Evans DM, Davey Smith G. Within family Mendelian randomization studies. Hum Mol Genet. 2019;28(R2):R170-R179. doi: https://doi.org/10.1093/hmg/ddz204

10. Hartwig FP, Bowden J, Wang L, Davey Smith G, Davies NM. Average causal effect estimation via instrumental variables: the no simultaneous heterogeneity assumption. arXiv:2010. 10017v1. [Preprint]. 2020 [cited 2021 June 5]. Available from: https://arxiv.org/abs/2010.10017

11. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018;30;7:e34408. doi: https://doi.org/10.7554/eLife.34408

12. Ligthart S, Vaez A, Võsa U, Stathopoulou MG, de Vries PS, Prins BP, et al. Genome analyses of >200,000 individuals identify 58 loci for chronic inflammation and highlight pathways that link inflammation and complex disorders. AJHG. 2018;103(5):691-706. doi: https://doi.org/10.1016/j.ajhg.2018.09.009

13. Swerdlow DI, Kuchenbaecker KB, Shah S, Sofat R, Holmes MV, White J, et al. Selecting instruments for Mendelian randomization in the wake of genome-wide association studies. Int J Epidemiol. 2016;45(5):1600-1616. doi: https://doi.org/10.1093/ije/dyw088

14. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits Using GWAS summary data. PLOS Genet. 2017;13:e1007081. doi: https://doi.org/10.1371/journal.pgen.1007081

15. Timpson NJ, Nordestgaard BG, Harbord RM, Zacho J, Frayling TM, Tybjærg-Hansen A, et al. (2011) C-reactive protein levels and body mass index: elucidating direction of causation through reciprocal Mendelian randomization. Int J Obes (Lond). 2011;35(2):300-308. doi: https://doi.org/1038/ijo.2010.137

16. Elsworth BL, Lyon M, Alexander T, Liu Y, Matthews P, Hallett J, et al. The MRC IEU OpenGWAS data infrastructure. BioRxiv 244293 [Preprint]. 2020 [cited 2021 June 5]. Available from: https://www.biorxiv.org/content/10.1101/2020.08.10.244293v1 doi: https://doi.org/10.1101/2020.08.10.244293

17. Chen L, Smith GD, Harbord RM, Lewis SJ. Gene by environment MR Alcohol intake and blood pressure: a systematic review implementing a Mendelian randomization approach. PLoS Med. 2008;5(3):e52. doi: https://doi.org/10.1371/journal.pmed.0050052

18. Rees JMB, Foley CN, Burgess S. Factorial MR Factorial Mendelian randomization: using genetic variants to assess interactions. Int J Epidemiol. 2020;49(4):1147-1158. doi: https://doi.org/10.1093/ije/dyz161

19. Silverwood RJ, Holmes MV, Dale CE, Lawlor DA, Whittaker JC, Davey Smith G, et al. (2014) Non-linear MR (Testing for non-linear causal effects using a binary genotype in a Mendelian randomization study: application to alcohol and cardiovascular traits. Int J Epidemiol. 43(6):1781-90. doi: https://doi.org/10.1093/ije/dyu187

20. Inoue A, Solon G. Two-sample instrumental variables estimators. Rev Econ Stat. 2010;92:557-561. doi: https://doi.org/10:1162/REST_a_00011

21. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG, EPIC- InterAct Consortium. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. Eur J Epidemiol. 2015;30:543-552. doi: https://doi.org/10.1007/s10654-015-0011-z

22. Bowden J, Hemani G, Davey Smith G. Invited commentary: detecting individual and global horizontal pleiotropy in Mendelian randomization-a job for the humble heterogeneity statistic? Am J Epidemiol. 2018;187(12):268-2685. doi: https://doi.org/10.1093/aje/kwy185

23. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. Int J Epidemiol. 2017;46(6):1985-1996. doi: https://doi.org/10.1093/ije/dyx102

24. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genet Epidemiol. 2015;40(4):304-314. doi: https://doi.org/10.10002gepi.21965

25. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44:512-525. doi: https://doi.org/10.1093/ije/dyv080

26. Zhao Q, Wang J, Hemani G, Bowden J, Small DS. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. Ann Statist. 2020;48(3):1742-1769. doi: https://doi.org/10.1214/19-AOS1866

27. Verbanck M, Chen C-Y, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet. 2018;50:693-698. doi: https://doi.org/10.1038/s41588-018-0099-7

28. Bowden J, Del Greco MF, Minelli C, Zhao Q, Lawlor DA, Sheehan NA, et al. Improving the accuracy of two-sample summary-data Mendelian randomization: moving beyond the NOME assumption. Int J Epidemiol. 2019;48(3):728-742. doi: https://doi.org/10.1093/ije/dyy258

29. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. Genet Epidemiol. 2016;40(7):597-608. doi: https://doi.org/10.1002/gepi.21998

30. Gkatzionis A, Burgess S. Contextualizing selection bias in Mendelian randomization: how bad is it likely to be? Int J Epidemiol. 2019;48(3):691-701. doi: https://doi.org/10.1093/ije/dyy202

31. Sandve GK, Nekrutenko A, Taylor J, Hovig E. Ten Simple Rules for Reproducible Computational Research. PLoS Comput Biol. 2013;9(10):e1003285. doi: https://doi.org/10.1371/journal.pcbi.1003285

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

7:e27857v1. [Preprint]. 2019 [cited 2021 June 5]. Available from:
https://peerj.com/preprints/27857/ doi: https://doi.org/10.7287/peerj.preprints.27857v1

33. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ. 2018;362:k601. doi: https://doi.org/10.1136/bmj.k601

34. Zhang W. Ten Simple Rules for Writing Research Papers. PLoS Comput Biol. 2014;10(1):e1003453. doi: https://doi.org/10.1371/journal.pcbi.1003453

35. Bourne PE. Ten simple rules for making good oral presentations. PLOS Comput Biol. 2007;3:e77. doi: https://doi.org/10.1371/journal.pcbi.0030077