

REVIEW

Open Access



Addressing the credibility crisis in Mendelian randomization

Stephen Burgess^{1,2*}, Benjamin Woolf^{1,3,4}, Amy M. Mason^{2,5}, Mika Ala-Korpela^{6,7} and Dipender Gill^{8,9}

Abstract

Background Genome-wide association studies have enabled Mendelian randomization analyses to be performed at an industrial scale. Two-sample summary data Mendelian randomization analyses can be performed using publicly available data by anyone who has access to the internet. While this has led to many insightful papers, it has also fuelled an explosion of poor-quality Mendelian randomization publications, which threatens to undermine the credibility of the whole approach.

Findings We detail five pitfalls in conducting a reliable Mendelian randomization investigation: (1) inappropriate research question, (2) inappropriate choice of variants as instruments, (3) insufficient interrogation of findings, (4) inappropriate interpretation of findings, and (5) lack of engagement with previous work. We have provided a brief checklist of key points to consider when performing a Mendelian randomization investigation; this does not replace previous guidance, but highlights critical analysis choices. Journal editors should be able to identify many low-quality submissions and reject papers without requiring peer review. Peer reviewers should focus initially on key indicators of validity; if a paper does not satisfy these, then the paper may be meaningless even if it is technically flawless.

Conclusions Performing an informative Mendelian randomization investigation requires critical thought and collaboration between different specialties and fields of research.

Keywords Causal inference, Genetic epidemiology, Instrumental variables, Evidence synthesis, Risk of bias, Bias evaluation

Background

Mendelian randomization is an epidemiological technique that exploits the properties of genetic variants to address causal questions about the potential effect of an exposure on an outcome [1, 2]. Mendel's laws of heritability mean that, conditional on parental genotype, genetic variants should only be associated with traits that they influence [3, 4]. Given a well-mixed population, the same property should hold at the population level [5]. Empirical investigations have shown that genetic associations with unrelated traits estimated in population-based cohorts are no stronger than would be expected due to chance alone [6, 7]. This suggests a generic strategy for testing the causal effect of any exposure on any outcome by the following steps:

*Correspondence:

Stephen Burgess
sb452@medschl.cam.ac.uk

¹ Medical Research Council Biostatistics Unit, University of Cambridge, Cambridge, UK

² Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

³ School of Psychological Science, University of Bristol, Bristol, UK

⁴ MRC Integrative Epidemiology Unit at the, University of Bristol, Bristol, UK

⁵ Victor Phillip Dahdaleh Heart and Lung Research Institute, University of Cambridge, Cambridge, UK

⁶ Systems Epidemiology, Faculty of Medicine, Research Unit of Population Health, University of Oulu and Biocenter Oulu, Oulu, Finland

⁷ NMR Metabolomics Laboratory, School of Pharmacy, Faculty of Health Sciences, University of Eastern Finland, Kuopio, Finland

⁸ Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK

⁹ Sequoia Genetics, London, UK



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

1. Find genetic variants that are predictors of the exposure
2. Test whether these genetic variants associate with the outcome

The simplicity and universality of the approach is appealing [8]. Analogously to a randomized trial, inferences are made not by application of clever statistical methodology, but by exploiting random variation [9, 10]—although in the case of Mendelian randomization, this is naturally-occurring randomization rather than random allocation by a trialist [11]. However, such a simple recipe cannot provide reliable causal inferences without thoughtful application.

The Mendelian randomization approach relies on the gene–environment equivalence principle [12]. This states that selected genetic variants influence an environmental (that is non-genetic) exposure equivalently to a proposed intervention that changes the population distribution of the exposure. In practice, there are often differences between the effect of a genetic variant and a proposed intervention in terms of mechanism, magnitude, timing, and duration that imply downstream consequences are not exactly equivalent [13, 14]. The principle can be restated to require that genetic associations are informative about the presence, direction, and (to a more limited extent) the size of the effect on the outcome resulting from an intervention in the exposure.

The availability of data from genome-wide association studies (GWAS) has enabled Mendelian randomization analyses to be performed at an industrial scale [15, 16]. In particular, it has enabled two-sample summary data Mendelian randomization investigations [17]: “two-sample” indicates that genetic associations with the putative exposure and outcome come from different datasets; “summary data” indicates that analyses are performed using genetic association estimates—beta-coefficients and standard errors representing associations of the respective variants with the exposure and outcome—rather than individual-level data [18, 19]. Such association data have been released for many large consortia and biobanks [20, 21]. Anyone with access to the internet can download genetic associations with risk factors and disease outcomes and use these to implement Mendelian randomization methods [22]. Indeed, such applications of Mendelian randomization have great advantages: they are able to use large datasets published by GWAS consortia, and analyses can be made fully transparent and replicable.

However, particularly in the age of artificial intelligence, such analyses are arguably too accessible. Web-based tools have been created that simplify the task of the analyst to simply choosing the exposure and

outcome—the automated analysis is performed at the touch of a button [23]. Mendelian randomization has become an easy target for researchers who are incentivized to publish as often as they can, as well as to predatory journals which are willing to publish such articles. While the two-sample summary data approach has led to many insightful papers, it has also fuelled an explosion of poor-quality Mendelian randomization publications, which threatens to overwhelm the capacity of qualified reviewers and undermine the credibility of the whole approach.

The guidance in this article is written to help those who want to write meaningful Mendelian randomization papers, as well as to journal editors and reviewers seeking to triage and identify low-quality submissions. There is already plentiful guidance on performing and reporting Mendelian randomization investigations [24], including the Strengthening the Reporting of Observational Studies in Epidemiology using Mendelian Randomization (STROBE-MR) guidelines [25]; we would encourage journals to insist that authors complete the checklist based on these guidelines at initial submission. This is important to ensure analyses are performed accurately and to avoid errors, such as mistakes in allele harmonization [26]. However, a Mendelian randomization investigation may be perfectly written and follow these guidelines to the letter—and yet the whole study may be completely useless.

We focus here on two-sample Mendelian randomization analyses using established methods for the analysis of summarized data. Advanced methods, such as non-linear analyses [27], cross-generational analyses [28], and time-varying analyses [29], require additional assumptions and detailed considerations that could potentially lead to biased estimates if violated [30–32]. Such methods are outside the scope of this paper. However, the considerations discussed here about instrument selection, instrument validity, and interpretation are foundational, and also apply to such applications.

We consider five common pitfalls in conducting a reliable Mendelian randomization investigation: (1) inappropriate research question, (2) inappropriate choice of variants as instruments, (3) insufficient interrogation of findings, (4) inappropriate interpretation of findings, and (5) lack of engagement with previous literature. We present a short list of relevant questions relating to these points in Fig. 1 for authors to consider. While not as comprehensive as the STROBE-MR guidelines, it is more succinct and focuses on the key critical judgements that are required to assess the reliability of an investigation. It should be particularly valuable not just to authors, but also to reviewers and editors, and indeed, to eventual readers wanting to evaluate the quality of evidence provided by a Mendelian randomization publication.

1. Is the research question addressable using Mendelian randomization?
 - Is it plausible we can find genetic variants that satisfy the gene—environment equivalence assumption?
2. Is the choice of genetic variants appropriate?
 - Is the biological relevance of the variants to the exposure known?
 - If chosen based on GWAS of an external dataset, how was that GWAS conducted? How was the exposure defined and measured?
3. Have the researchers sufficiently interrogated findings?
 - Have approaches that assess the robustness of findings been performed?
 - Have results been reported clearly, accounting for multiple testing if appropriate?
4. Have findings been appropriately interpreted?
 - Are conclusions proportionate to our confidence in the validity of the instruments?
 - Have alternative causal pathways from the variants to the outcome been considered?
5. Are results adequately compared to the literature?
 - Is there sufficient discussion about coherence of findings from different approaches?
 - Is the overall conclusion presented in a triangulation framework?

Fig. 1 Key considerations when assessing the credibility of a Mendelian randomization investigation

Inappropriate research question

The instrumental variable assumptions [33] require that any genetic variant used in a Mendelian randomization investigation as an instrument must:

1. Be associated with the exposure (relevance)
2. Not be associated with the outcome via a confounding pathway (exchangeability)
3. Have no direct effect on the outcome, only potentially an indirect effect via the exposure (exclusion restriction) [34]

Only the first of these assumptions can be verified based on data. The other two assumptions cannot be formally tested and must be justified either on the basis of scientific understanding, or empirically supported based on the application of statistical methods [24].

These assumptions require the genetic variants to be specific in how they affect the exposure—there cannot be pleiotropic associations with variables on alternative causal pathways to the outcome. Associations with variables on the causal pathway from the genetic variants to the outcome via the exposure (sometimes called “vertical pleiotropy”) are allowed; associations with variables on alternative causal pathways (sometimes called “horizontal pleiotropy”) are not [35] (Fig. 2).

For some exposures, it is implausible that there are genetic variants that influence the exposure in a way that meets these requirements. A paradigmatic example of such an exposure is “use of chopsticks”—if a researcher found genetic predictors of chopstick use in a Western population, the likely explanation would be that the variants reflect demographic or socioeconomic status, rather

than a biological mechanism that affects cutlery choice [37]. Such variants would be invalid instrumental variables: first, they would be subject to population stratification, and second, even if population stratification could be addressed, they would be associated with other traits and behaviours that are more common in chopstick users. As such, a Mendelian randomization study supposedly finding evidence of an effect of chopsticks use would have to show that this effect is not attributable to the many other social and cultural factors that likely differ between the genetically defined population groups.

Another implausible exposure for use in Mendelian randomization is pollution levels [38]. Again, it is implausible that there are particular genetic variants that affect exposure to air pollution. If genetic predictors of air pollution are found, it is likely that these are markers of social status rather than representing intrinsic biological mechanisms. In some large datasets, such as UK Biobank, air pollution is not measured at an individual level, but inferred based on home address [39]. This reinforces the concern that such an analysis is actually evaluating social status, not air pollution in any specific way. Another category of implausible exposures for Mendelian randomization is gut microbiota [40]. It is implausible that there are particular genetic variants that have specific effects on individual gut microbiome species. While some genetic predictors of gut microbiota have been found, they are located in highly pleiotropic gene regions, such as the *ABO* gene region [41]. Just because a GWAS has found genetic predictors of a trait does not imply that the trait is an appropriate exposure for a Mendelian randomization investigation, nor that the genetic predictors represent valid genetic instruments.

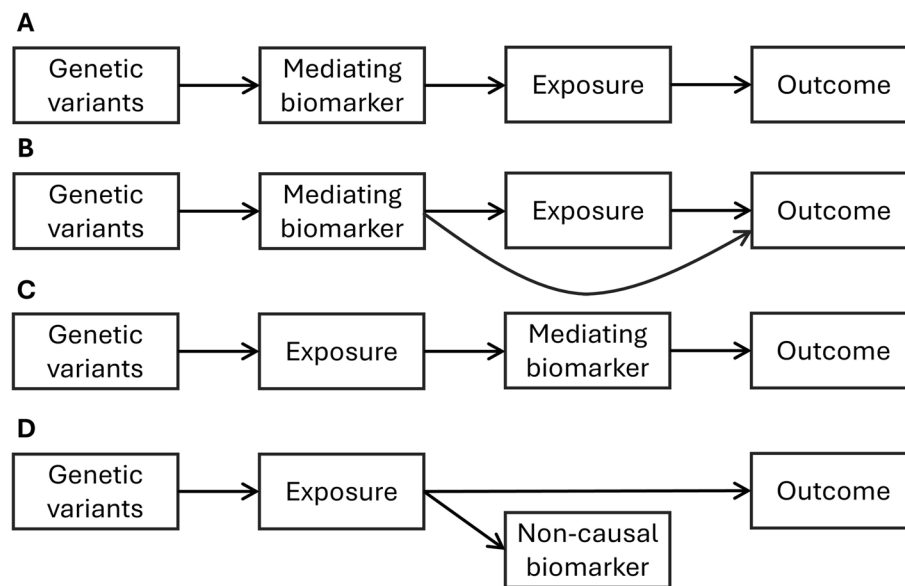


Fig. 2 Genetic associations with an exposure variable that is downstream of a mediating biomarker (diagrams A and B), or has a downstream effect on either a mediating biomarker (diagram C) or a non-causal biomarker (diagram D). In case A, the only causal pathway from the genetic variants to the outcome passes via the exposure; hence, this is an example of “vertical pleiotropy”, and the genetic variants are valid instruments. In case B, there is a causal pathway from the genetic variants to the outcome that does not pass via the exposure; hence, this is an example of “horizontal pleiotropy”, and the genetic variants are not valid instruments. In cases C and D (which also represent “vertical pleiotropy”), Mendelian randomization analyses should be conceptualized in terms of the exposure (the putative causal trait), even if measured genetic associations are expressed in terms of the biomarker. Diagram D is likely to represent the situation between genetic variants in the *IL6R* gene region, interleukin 6 signalling (exposure), and C-reactive protein (non-causal biomarker). C-reactive protein is likely to be a non-causal biomarker when considering the effect of interleukin 6 receptor inhibition on coronary heart disease [36]

If an exposure is externally or environmentally determined, or variation in the exposure is influenced purely by social and cultural factors rather than intrinsic biological mechanisms, then it is unlikely that effects of the exposure can be reliably interrogated in a Mendelian randomization design. Such traits are more likely to be subject to bias from population stratification, non-random mating patterns, and dynastic effects (that is intergenerational effects, such as when the parental genotype directly influences the offspring exposure or outcome) [42].

A counter-example to this is alcohol consumption. While alcohol consumption is partially determined by personal and environmental factors, there are biological mechanisms influencing the metabolism of alcohol that affect consumption levels, as well as exposure to alcohol in the bloodstream. Genetic variants in key regulators of these mechanisms are potential instruments for understanding the downstream effects of alcohol consumption [43]. However, care is still required to appropriately perform and interpret such analyses; we follow up this example in further sections.

Researchers should be aware that not all causal questions can appropriately be addressed in a Mendelian randomization paradigm. Journal editors and reviewers

should use their judgement to rapidly decide whether a question can plausibly be addressed by Mendelian randomization based on the abstract (or even the title) alone: is it plausible that there exist genetic variants such that the gene–environment equivalence principle holds? That is, are there likely to be genetic variants that affect the exposure in a way equivalent to the (possibly hypothetical) intervention implied by the causal question under investigation? If this is unlikely, then the investigation, even if perfectly implemented and reported, does not provide reliable evidence to address the causal question of interest.

Inappropriate choice of variants as instruments

The instrumental variable assumptions require that any causal pathway from the genetic variants to the outcome passes via the exposure under investigation. This is more plausible if the genetic variants are located in a gene region with known functional or biological relevance to the exposure [44–46]. It is less plausible for exposures that the genetic variants influence indirectly via complex causal pathways, such as educational attainment. For example, genetic variants in the *UGT1A1* gene region that encodes an enzyme regulating the synthesis of bilirubin are more plausible instruments than variants in gene

regions that are not functionally related to bilirubin, or whose function is unknown [47]. Genetic variants in the *ALDH2* and *ADH1B* gene regions are known to relate to alcohol metabolism, and hence are plausible instruments to investigate the effect of alcohol consumption [48]. If gene regions with biological relevance to the exposure are not known, Mendelian randomization can provide some evidence on the causal relevance of the exposure, but additional caution is required [49].

For a given gene region, genetic variants should be chosen based on their biological relation to the causal risk factor of interest, as far as is possible. This includes proximity to the relevant gene and functional effects on regulation of the gene or its downstream protein. For example, when investigating the effect of angiotensin converting enzyme (ACE) on risk of Alzheimer's disease, use of variants in the *ACE* gene region predicting tissue-specific gene and protein expression likely increases their plausibility as valid instruments for pharmacological perturbation of ACE at the relevant biological site [50]. In some cases, the same variant may be the lead signal for circulating protein levels, gene expression in the most relevant tissue, and levels of a downstream risk factor. In other cases, these approaches may identify different variants [51]. If these differ, careful consideration is needed to select the variant(s) that best mimic the intervention of interest.

Biological mechanisms affecting many exposures are not known. In such cases, genetic variants used as instruments may be selected solely based on their statistical association with the exposure. Such analyses are often still valuable, in that they provide a source of evidence supporting or refuting a causal effect of the exposure on the outcome. The strength of evidence provided depends on our confidence in the validity of the genetic variants as instruments. Testing genetic associations with potential confounders can provide empirical evidence supporting the validity of the variants as instruments, as can other statistical approaches, such as the application of pleiotropy-robust methods [52].

Researchers should prioritize investigating exposures using variants in gene regions that are biologically related to the exposure where possible. Journal editors and reviewers should look for a justification as to why the genetic variants in a given analysis were chosen. If this is absent, or if genetic variants are purely chosen on statistical grounds, then findings will generally be less authoritative and require a greater degree of statistical assessment.

Insufficient interrogation of findings

While the exact analysis plan will depend on the specifics of the question under investigation, availability of valid instruments, data quality, and so forth, one

recommended generic strategy for conducting Mendelian randomization analyses is as follows. First, if there are biologically informed candidate instruments, the primary analysis should be based on these variants. Second, if there are no biologically informed candidate instruments, an initial liberal analysis based on a wide range of variants is recommended. Finally, results should be interrogated further to investigate robustness to a variety of factors [24]. A null finding in a liberal analysis that includes potentially pleiotropic variants is likely to reflect a true null relationship; it is more likely that bias will lead to a false positive finding than a false negative finding [53]. However, false negative results can be just as harmful to science. Absence of evidence does not always mean evidence of absence, particularly if the analysis is underpowered, unspecific, or poorly designed.

There are many approaches for the interrogation of findings (see reference [24] for more details), including examining genetic associations with potentially pleiotropic variables [54], testing against positive and/or negative controls [55], colocalization (particularly when the finding is based on a single gene region) [56], use of pleiotropy-robust methods (particularly when the finding is based on variants from several gene regions) [52], investigation in subgroups of the population [57] (although noting such stratification can lead to collider bias [58]), investigation with a subset of variants, and multivariable Mendelian randomization [59]. No single sensitivity analysis approach is foolproof [60], and all approaches make their own assumptions. A causal effect may be present even if one or more approach does not provide supportive evidence of a causal effect (or equally, a causal effect may be absent even if one or more approach supports a causal effect). In many cases, the evidence will be equivocal; there may be evidence supporting a causal effect, but this evidence may not be fully consistent across all analyses. If there is inconsistent evidence, then it is important that results are reported clearly, without undue emphasis on significant findings. Similarly, if multiple hypotheses are tested by the investigators, this should be accounted for when interpreting findings.

As an example, the robustness of Mendelian randomization analyses with alcohol as an exposure has been tested in several ways. Analyses in East Asian populations have typically used variants in the *ALDH2* and *ADH1B* gene regions as instruments [61]. These investigations have exploited a further natural experiment by conducting analyses separately for men and women. Genetic associations with disease outcomes would not be expected in East Asian women as their alcohol consumption levels are much lower than those of men. East Asian women represents a negative control population, and null associations in women but positive associations with

men have been observed for oesophageal cancer [62] and blood pressure [63]. In European-descent populations, similar findings have been observed using a variant in the *ADH1B* gene region only and using a wider range of genetic predictors of alcohol consumption [64]. Consistent results for many outcomes have been observed across a range of robust methods, including MR-Egger, weighted median, and MR-PRESSO methods [65]. Multi-variable analyses have also been conducted adjusting for smoking behaviour, as genetic predictors of alcohol may have pleiotropic effects on smoking intensity [66].

Researchers should perform a range of approaches to investigate the robustness of findings. The reported level of confidence in conclusions should be dependent on the consistency of these results. Journal editors and reviewers should be suspicious of selective reporting of significant findings, particularly when approaches to assess the validity of findings have either not been reported, or indicate lack of support for a causal effect.

Inappropriate interpretation of findings

We have hereto assumed that the exposure measured in the Mendelian randomization analysis is the true causal agent affecting the outcome. However, this may not be the case. It is possible that a version of the gene–environment equivalence principle is true, but not for the measured exposure. It may be that the measured exposure is a biomarker that acts as a proxy measure of the true causal mechanism of action (Fig. 2).

As an example, genetic variants in the *IL6R* gene region are associated with levels of both interleukin 6 and C-reactive protein. This is plausibly an example of vertical pleiotropy, as the association with C-reactive protein is potentially a downstream consequence of the effect of interleukin 6 receptor signalling [36]. If we use genetic variants in the *IL6R* gene region in a Mendelian randomization analysis investigating the effects of interleukin 6 receptor signalling, we should come to the same conclusion whether our nominal biomarker for selecting and weighting instruments is levels of interleukin-6 receptor or levels of C-reactive protein [67]. Our estimate may be expressed in terms of change in genetically predicted interleukin-6 receptor levels or genetically predicted C-reactive protein levels, but it is the choice of variants that determines the causal question that is being addressed, not the biomarker used to select instruments for the exposure. As a further hypothetical example, suppose that we performed a Mendelian randomization analysis using genetic predictors of left leg mass. Would we be confident that any finding was truly attributable to an effect of left leg mass as opposed to adiposity or muscle mass more generally?

A related issue, particularly for binary exposures, is that genetic variants increase liability to the exposure, but do not necessarily increase the exposure [68]. For example, most individuals having genetic variants associated with increased schizophrenia risk do not themselves have clinically diagnosed schizophrenia [69]. Genetic variants that predispose individuals to increased alcohol consumption do not increase exposure to alcohol in populations of non-alcohol drinkers. Genetic variants shown to predispose individuals to greater COVID-19 risk did not increase exposure to COVID-19 in pre-pandemic datasets.

Mendelian randomization is serendipitous in nature; we exploit what is available. We cannot control which genetic variants are available for our analysis, or what these genetic variants do. The gene–environment equivalence principle requires to first understand how the genetic variants operate, and express our causal question in terms of the tools that are available. This implies that a simple conclusion statement such as “the exposure has a causal effect on the outcome” may not be appropriate.

Returning to the example of alcohol, genetic variants that increase alcohol consumption may have effects relating to social aspects of alcohol consumption as well as biological aspects. Those who drink more alcohol in Western societies are likely to spend more time in licenced establishments, and potentially have greater exposure to environmental tobacco smoke. Another complication is distinguishing between alcohol consumption and exposure to high alcohol levels. For caffeine, genetic associations with coffee consumption and circulating plasma caffeine levels are not all concordant. This can be explained as some genetic variants that increase caffeine metabolism lead to lower circulating caffeine levels, but greater coffee consumption, as fast caffeine metabolizers tend to consume more coffee to get the same physiological response [70]. Variants in the *ALDH2* and *ADH1B* gene regions affect alcohol consumption via different biological pathways. While the rs671 variant in *ALDH2* decreases alcohol consumption, it impairs the metabolism of alcohol, meaning that carriers who drink alcohol have greater exposure to acetaldehyde, a known carcinogen [71]. Hence, the associations of the rs671 variant may be misinterpreted if investigators focus on associations with alcohol consumption level. Correct interpretation of Mendelian randomization analyses requires appreciation of the broad social context of alcohol consumption and understanding of the biological effects of the variants.

Researchers should think carefully about the identity of the underlying causal risk factor or mechanism evaluated in their analysis; this may differ from the measured variable used as the exposure. Journal editors and reviewers should be sceptical about strong causal claims. Jumps of

logic from factual statements such as “genetic predictors of the exposure were associated with the outcome” (or equivalently “genetically predicted exposure levels were associated with the outcome”) to subjective inferences such as “therefore, we believe the exposure is a cause of the outcome” should only be made when they can be justified [72]. If it is implausible that an exposure could be altered in a specific way by a genetically regulated mechanism, then it may be that the nominal exposure is a biomarker for a wider mechanism, not the literal causal risk factor.

Lack of engagement with previous literature

Mendelian randomization cannot by itself demonstrate or prove the existence of a causal effect. Indeed, the aim of a Mendelian randomization investigation is often to provide supportive or suggestive evidence to encourage further research, including the establishment of a randomized trial. As such, it is important to weigh evidence from Mendelian randomization against that from other approaches, including epidemiological data, trial findings, and basic science experiments. Triangulation is a framework for evidence synthesis that considers evidence from various sources that make different assumptions, and hence the validity of these assumptions will be orthogonal [60, 73]. Evidence from different approaches making different assumptions can provide a more compelling case for a causal effect, or can help enhance the specificity of evidence. By showing that evidence for a causal effect is stronger or weaker in certain circumstances (such as different populations, different times, or different subgroups), we can improve our understanding of the causal mechanism.

In the case of alcohol, while there are no large-scale long-term randomized trials investigating the impact of alcohol consumption on disease outcomes [74], there are randomized trials exploring the effects of drinking alcohol in the short term, and many mechanistic studies into the effects of alcohol. While several observational epidemiological studies have shown lower risk of cardiovascular disease amongst light drinkers compared to non-drinkers [75], Mendelian randomization analyses have not supported evidence of a protective effect of increased alcohol at any level of alcohol consumption [61, 76]. A Mendelian randomization investigation into the effect of alcohol consumption should explore reasons for discrepancies from results from conventional observational analyses. A potential explanation is that the non-drinker category contains both never-drinkers and former-drinkers, and the observational elevated risk in non-drinkers is due to former-drinkers.

Researchers should compare findings from their Mendelian randomization investigation to those from

lab-based experiments, functional genomic studies, observational epidemiological associations, and clinical trials. Results should be appraised in a triangulation framework indicating the extent to which they strengthen or weaken the evidence for a causal effect of the exposure. Journal editors and reviewers should hold authors to high standards and ensure that findings are adequately compared to those from previous Mendelian randomization investigations and other approaches.

Conclusions

Mendelian randomization can be applied in an uncritical, algorithmic way to obtain findings and generate publications [77]. Policing such outputs is an impossible task requiring far more resources than it takes to create the publications, and the onus should be on authors to perform thoughtful and well-justified analyses. Journal editors at reputable journals should be able to spot low-effort submissions without wasting precious peer review resources. Reviewers should focus not only on whether technical aspects of a submission are present, but also on key indicators that require critical judgement: whether the causal question can plausibly be addressed by Mendelian randomization, whether the choice of variants is justified, whether there has been sufficient interrogation of findings (assessment of internal validity), whether any inferred causal interpretation is appropriate (assessment of external validity), and how this finding supports or refutes aspects of the wider literature. Performing such an investigation requires close collaboration between those with biological, clinical, sociological, genetic, and statistical expertise to understand the plausibility of the assumptions and to perform and interpret analyses appropriately.

Abbreviations

ACE	Angiotensin converting enzyme
GWAS	Genome-wide association study/studies
STROBE-MR	Strengthening the Reporting of Observational Studies in Epidemiology using Mendelian Randomization

Acknowledgements

We thank the attendees of the 6th Mendelian randomization conference at the University of Bristol for their feedback on this work.

Authors' contributions

SB wrote the initial draft of the manuscript. DG supervised the work. All authors revised the submission critically for important intellectual content.

Authors' Twitter handles

@stevesphd, @barwoolf, @amymariemason, @mak_sysepi, @dpsg108.

Funding

SB and BW are supported by the Wellcome Trust (225790/Z/22/Z) and the United Kingdom Research and Innovation Medical Research Council (MC_UU_00002/7). MAK was supported by a research grant from the Sigrid Juselius Foundation, the Finnish Foundation for Cardiovascular Research, and the Research Council of Finland (grant no. 357183). AMM is supported by core funding from the British Heart Foundation (RG/18/13/33946), Cambridge

British Heart Foundation Centre of Research Excellence (RE/18/1/34212), British Heart Foundation Chair Award (CH/12/2/29428), and by Health Data Research UK. This research was supported by the National Institute for Health Research Cambridge Biomedical Research Centre (BRC-1215-20014, NIHR203312). The views expressed are those of the authors and not necessarily those of the National Institute for Health Research or the Department of Health and Social Care.

Availability of data and materials

Not applicable; this manuscript does not contain primary data or analyses.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

Not applicable; this manuscript does not contain primary data or analyses.

Consent for publication

Not applicable; this manuscript does not contain primary data or analyses.

Competing interests

The authors declare no competing interests.

Received: 8 July 2024 Accepted: 3 September 2024

Published online: 11 September 2024

References

- 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.
- Burgess S, Thompson SG. Mendelian randomization: methods for causal inference using genetic variants. 2nd ed. New York: Chapman & Hall/CRC; 2021.
- Davey Smith G, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. *Int J Epidemiol*. 2004;33(1):30–42.
- Zheng J, Baird D, Borges MC, Bowden J, Hemani G, Haycock P, et al. Recent developments in Mendelian randomization studies. *Curr Epidemiol Rep*. 2017;4(4):330–45.
- Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey SG. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med*. 2008;27(8):1133–63.
- Davey Smith G, Lawlor DA, Harbord R, Timpson N, Day I, Ebrahim S. Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLoS Med*. 2007;4(12):e352.
- Taylor M, Tansey KE, Lawlor DA, Bowden J, Evans DM, Davey Smith G, Timpson NJ. Testing the principles of Mendelian randomization: opportunities and complications on a genomewide scale. *bioRxiv*. 2017:124362.
- Davey SG. Random allocation in observational data: how small but robust effects could facilitate hypothesis-free causal inference. *Epidemiology*. 2011;22(4):460–3.
- Hingorani A, Humphries S. Nature's randomised trials. *Lancet*. 2005;366(9501):1906–8.
- Thanassoulis G, O'Donnell CJ. Mendelian randomization: nature's randomized trial in the post-genome era. *JAMA*. 2009;301(22):2386–8.
- Swanson SA, Tiemeier H, Ikram MA, Hernán MA. Nature as a trialist?: deconstructing the analogy between Mendelian randomization and randomized trials. *Epidemiology*. 2017;28(5):653–9.
- Sanderson E, Glymour MM, Holmes MV, Kang H, Morrison J, Munafò MR, et al. Mendelian randomization. *Nat Rev Methods Primers*. 2022;2(1):6.
- Burgess S, Butterworth A, Malarstig A, Thompson SG. Use of Mendelian randomisation to assess potential benefit of clinical intervention. *BMJ*. 2012;345:e7325.
- Ference BA. How to use Mendelian randomization to anticipate the results of randomized trials. *Eur Heart J*. 2018;39(5):360–2.
- 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(7):543–52.
- Hemani G, Bowden J, Haycock P, Zheng J, Davis O, Flach P, et al. Automating Mendelian randomization through machine learning to construct a putative causal map of the human phenome. *bioRxiv*. 2017:173682.
- Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol*. 2013;37(7):658–65.
- Thompson JR, Minelli C, Fabiola Del Greco M. Mendelian randomization using public data from genetic consortia. *Int J Biostat*. 2016;12(2):20150074.
- Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan N, Thompson J. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat Med*. 2017;36(11):1783–802.
- Staley JR, Blackshaw J, Kamat MA, Ellis S, Surendran P, Sun BB, et al. PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics*. 2016;32(20):3207–9.
- Elsworth B, Lyon M, Alexander T, Liu Y, Matthews P, Hallett J, et al. The MRC IEU OpenGWAS data infrastructure. *bioRxiv*. 2020;2020.08.10.244293.
- Davey Smith G, Ebrahim S. Mendelian randomisation at 20 years: how can it avoid hubris, while achieving more? *Lancet Diabetes Endocrinol*. 2024;12(1):14–7.
- Hemani G, Zheng J, Elsworth B, Wade K, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. *eLife*. 2018;7:e34408.
- Burgess S, Davey Smith G, Davies NM, Dudbridge F, Gill D, Glymour MM, et al. Guidelines for performing Mendelian randomization investigations. *Wellcome Open Res*. 2020;4:186.
- Skrivankova VV, Richmond RC, Woolf BAR, Yarmolinsky J, Davies NM, Swanson SA, et al. Strengthening the reporting of observational studies in epidemiology using Mendelian randomization: the STROBE-MR statement. *JAMA*. 2021;326(16):1614–21.
- Hartwig FP, Davies NM, Hemani G, Davey SG. Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. *Int J Epidemiol*. 2016;45(6):1717–26.
- Tian H, Mason AM, Liu C, Burgess S. Relaxing parametric assumptions for non-linear Mendelian randomization using a doubly-ranked stratification method. *PLoS Genet*. 2023;19(6):e1010823.
- Evans DM, Moen G-H, Hwang L-D, Lawlor DA, Warrington NM. Elucidating the role of maternal environmental exposures on offspring health and disease using two-sample Mendelian randomization. *Int J Epidemiol*. 2019;48(3):861–75.
- Sanderson E, Richardson TG, Morris TT, Tilling K, Davey SG. Estimation of causal effects of a time-varying exposure at multiple time points through multivariable Mendelian randomization. *PLoS Genet*. 2022;18(7):e1010290.
- Wade KH, Hamilton FW, Carslake D, Sattar N, Davey Smith G, Timpson NJ. Challenges in undertaking nonlinear Mendelian randomization. *Obesity*. 2023;31(12):2887–90.
- Burgess S. Violation of the constant genetic effect assumption can result in biased estimates for non-linear Mendelian randomization. *Hum Hered*. 2023;88(1):79–90.
- Tian H, Burgess S. Estimation of time-varying causal effects with multivariable Mendelian randomization: some cautionary notes. *Int J Epidemiol*. 2023;52(3):846–57.
- Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol*. 2000;29(4):722–9.
- Labrecque J, Swanson SA. Understanding the assumptions underlying instrumental variable analyses: a brief review of falsification strategies and related tools. *Curr Epidemiol Rep*. 2018;5(3):214–20.
- Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet*. 2014;23(R1):R89–98.
- Holmes MV, Ala-Korpela M, Davey SG. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. *Nat Rev Cardiol*. 2017;14(10):577–90.
- Hamer D, Sirota L. Beware the chopsticks gene. *Mol Psychiatry*. 2000;5(1):11–3.

38. Au Yeung SL, Gill D. Concerns over using the Mendelian randomization design to investigate the effect of air pollution. *Sci Total Environ.* 2024;917: 170474.
39. UK Biobank Small Area Health Statistics Unit. UK Biobank – environmental exposures – metadata. 2014. Available at <https://biobank.ndph.ox.ac.uk/ukb/ukb/docs/EnviroExposEst.pdf>. (Last Accessed 2024–03–08).
40. Hatcher C, Richenberg G, Waterson S, Nguyen LH, Joshi AD, Carreras-Torres R, et al. Application of Mendelian randomization to explore the causal role of the human gut microbiome in colorectal cancer. *Sci Rep.* 2023;13(1):5968.
41. Zhernakova DV, Wang D, Liu L, Andreu-Sánchez S, Zhang Y, Ruiz-Moreno AJ, et al. Host genetic regulation of human gut microbial structural variation. *Nature.* 2024;625(7996):813–21.
42. Morris TT, Davies NM, Hemani G, Smith GD. Population phenomena inflate genetic associations of complex social traits. *Sci Adv.* 2020;6(16):eaay0328.
43. van de Luitgaarden IAT, van Oort S, Bouman EJ, Schoonmade LJ, Schrieks IC, Grobbee DE, et al. Alcohol consumption in relation to cardiovascular diseases and mortality: a systematic review of Mendelian randomization studies. *Eur J Epidemiol.* 2022;37(7):655–69.
44. Gill D, Georgakis MK, Walker VM, Schmidt AF, Gkatzionis A, Freitag DF, et al. Mendelian randomization for studying the effects of perturbing drug targets. *Wellcome Open Res.* 2021;6:16.
45. Schmidt AF, Finan C, Gordillo-Marañón M, Asselbergs FW, Freitag DF, Patel RS, et al. Genetic drug target validation using Mendelian randomisation. *Nat Commun.* 2020;11(1):3255.
46. Karjalainen MK, Karthikeyan S, Oliver-Williams C, Sliz E, Allara E, Fung WT, et al. Genome-wide characterization of circulating metabolic biomarkers. *Nature.* 2024;628(8006):130–8.
47. Seyed Khoei N, Carreras-Torres R, Murphy N, Gunter MJ, Brennan P, Smith-Byrne K, et al. Genetically raised circulating bilirubin levels and risk of ten cancers: a Mendelian randomization study. *Cells.* 2021;10(2):394.
48. Holmes MV, Dale CE, Zuccolo L, Silverwood RJ, Guo Y, Ye Z, et al. Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data. *BMJ.* 2014;349:g4164.
49. Burgess S, Cronjé HT. Incorporating biological and clinical insights into variant choice for Mendelian randomisation: examples and principles. *eGastroenterology.* 2024;2(1):e100042.
50. Ryan DK, Karhunen V, Su B, Traylor M, Richardson TG, Burgess S, et al. Genetic evidence for protective effects of angiotensin-converting enzyme against Alzheimer disease but not other neurodegenerative diseases in European populations. *Neurol Genet.* 2022;8(5):e200014.
51. Yang C, Farias FH, Ibanez L, Suhay A, Sadler B, Fernandez MV, et al. Genomic atlas of the proteome from brain, CSF and plasma prioritizes proteins implicated in neurological disorders. *Nat Neurosci.* 2021;24(9):1302–12.
52. Slob EAW, Burgess S. A comparison of robust Mendelian randomization methods using summary data. *Genet Epidemiol.* 2020;44(4):313–29.
53. VanderWeele TJ, Tchetgen Tchetgen EJ, Cornelis M, Kraft P. Methodological challenges in Mendelian randomization. *Epidemiology.* 2014;25(3):427–35.
54. 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.* 2017;28(1):30–4.
55. Sanderson E, Richardson TG, Hemani G, Davey SG. The use of negative control outcomes in Mendelian randomization to detect potential population stratification. *Int J Epidemiol.* 2021;50(4):1350–61.
56. Zuber V, Grinberg NF, Gill D, Manipur I, Slob EAW, Patel A, et al. Combining evidence from Mendelian randomization and colocalization: review and comparison of approaches. *Am J Hum Genet.* 2022;109(5):767–82.
57. van Kippersluis H, Rietveld CA. Pleiotropy-robust Mendelian randomization. *Int J Epidemiol.* 2018;47(4):1279–88.
58. Burgess S. “C-reactive protein levels and risk of dementia”: subgroup analyses in Mendelian randomization are likely to be misleading. *Alzheimers Dement.* 2022;18(12):2732–3.
59. Sanderson E, Davey Smith G, Windmeijer F, Bowden J. An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. *Int J Epidemiol.* 2019;48(3):713–27.
60. Lawlor DA, Tilling K, Davey SG. Triangulation in aetiological epidemiology. *Int J Epidemiol.* 2016;45(6):1866–86.
61. Millwood IY, Walters RG, Mei XW, Guo Y, Yang L, Bian Z, et al. Conventional and genetic evidence on alcohol and vascular disease aetiology: a prospective study of 500 000 men and women in China. *Lancet.* 2019;393(10183):1831–42.
62. Lewis SJ, Davey SG. Alcohol, ALDH2, and esophageal cancer: a meta-analysis which illustrates the potentials and limitations of a Mendelian randomization approach. *Cancer Epidemiology and Prevention Biomarkers.* 2005;14(8):1967–71.
63. Cho Y, Shin S-Y, Won S, Relton CL, Davey Smith G, Shin M-J. Alcohol intake and cardiovascular risk factors: a Mendelian randomisation study. *Sci Rep.* 2015;5(1):18422.
64. Lankester J, Zanetti D, Ingelsson E, Assimes TL. Alcohol use and cardiometabolic risk in the UK Biobank: a Mendelian randomization study. *PLoS One.* 2021;16(8):e0255801.
65. Larsson SC, Burgess S, Mason AM, Michaëlsen K. Alcohol consumption and cardiovascular disease. *Circ Genom Precis Med.* 2020;13(3):e002814.
66. Yuan S, Chen J, Ruan X, Sun Y, Zhang K, Wang X, et al. Smoking, alcohol consumption, and 24 gastrointestinal diseases: Mendelian randomization analysis. *eLife.* 2023;12:e84051.
67. Rosa M, Chignon A, Li Z, Boulanger M-C, Arsenaault BJ, Bossé Y, et al. A Mendelian randomization study of IL6 signaling in cardiovascular diseases, immune-related disorders and longevity. *NPJ Genom Med.* 2019;4(1):23.
68. Burgess S, Labrecque JA. Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. *Eur J Epidemiol.* 2018;33:947–52.
69. Davey Smith G, Munafo M. Does schizophrenia influence cannabis use? How to report the influence of disease liability on outcomes in Mendelian randomization studies. 2019. Available at <https://targ.blogs.bristol.ac.uk/2019/01/07/does-schizophrenia-influence-cannabis-use-how-to-report-the-influence-of-disease-liability-on-outcomes-in-mendelian-randomization-studies/>. (Last Accessed 2024–03–08).
70. Woolf B, Cronjé HT, Zagkos L, Larsson SC, Gill D, Burgess S. Comparison of caffeine consumption behavior with plasma caffeine levels as exposures in drug-target Mendelian randomization and implications for interpreting effects on obesity. *Am J Epidemiol.* 2024. In press.
71. Im PK, Yang L, Kartsonaki C, Chen Y, Guo Y, Du H, et al. Alcohol metabolism genes and risks of site-specific cancers in Chinese adults: an 11-year prospective study. *Int J Cancer.* 2022;150(10):1627–39.
72. Burgess S, O'Donnell CJ, Gill D. Expressing results from a Mendelian randomization analysis: separating results from inferences. *JAMA Cardiol.* 2021;6(1):7–8.
73. Munafo MR, Davey SG. Robust research needs many lines of evidence. *Nature.* 2018;553(7689):399–401.
74. Spiegelman D, Lovato LC, Khudyakov P, Wilkens TL, Adebamowo CA, Adebamowo SN, et al. The Moderate Alcohol and Cardiovascular Health Trial (MACH15): design and methods for a randomized trial of moderate alcohol consumption and cardiometabolic risk. *Eur J Prev Cardiol.* 2020;27(18):1967–82.
75. Wood AM, Kaptoge S, Butterworth AS, Willeit P, Warnakula S, Bolton T, et al. Risk thresholds for alcohol consumption: combined analysis of individual-participant data for 599912 current drinkers in 83 prospective studies. *Lancet.* 2018;391(10129):1513–23.
76. Biddinger KJ, Emdin CA, Haas ME, Wang M, Hindy G, Ellinor PT, et al. Association of habitual alcohol intake with risk of cardiovascular disease. *JAMA Netw Open.* 2022;5(3):e223849-e.
77. Burgess S, Davey SG. How humans can contribute to Mendelian randomization analyses. *Int J Epidemiol.* 2019;48(3):661–4.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.