eGastroenterology

Attention to the misuse of Mendelian randomisation in medical research

Lanlan Chen ,[1] Adrien Guillot ,[1] Carolin Victoria Schneider [2]

To cite: Chen L, Guillot A, Schneider CV. Attention to the misuse of Mendelian randomisation in medical research. *eGastroenterology* 2025;**3**:e100187. doi:10.1136/egastro-2025-100187

► Prepublication history for this paper is available online. To view these files, please visit the journal online (https://doi.org/10.1136/egastro-2025-100187).

Received 14 January 2025
Accepted 10 February 2025

Linked

- ► https://doi.org/10.1136/egastro-2023-100042
- ► https://doi.org/10.1136/egastro-2023-100058
- ► https://doi.org/10.1136/egastro-2023-100029
- ► https://doi.org/10.1136/egastro-2024-100114
- ► https://doi.org/10.1136/egastro-2024-100126
- ► https://doi.org/10.1136/egastro-2023-100034

[1]Department of Hepatology & Gastroenterology, Charite - Universitatsmedizin Berlin, Berlin, Germany
[2]Department of Medicine III, University Hospital RWTH Aachen, Aachen, Germany

Correspondence to
Dr Lanlan Chen; lanlan.chen@charite.de and Professor Carolin Victoria Schneider; cschneider@ukaachen.de

Mendelian randomisation (MR) is a widely used method that employs genetic variants as instrumental variables (IVs, also referred to as genetic instruments below and they represent eligible genetic variants in MR) from genome-wide association studies (GWAS) to explore the causative relationship between putative risk factors (ie, exposures) and outcomes.[1] According to a Web of Science literature search conducted at the end of 2024, the number of related publications has grown exponentially, reaching 3545 in 2023 and 6607 in 2024 (figure 1). Almost simultaneously, we observed a peak of citations from 2019 to 2021, which reflected the popularity and acknowledgement of papers published in these 3 years. However, a decline in citations has been witnessed from 2022 and here are two possible explanations for it: (1) papers published in 2022 or later only have two or fewer years to be cited, and it can lead to fewer citations absolutely; (2) some papers published from 2022 are of low quality and may even be redundant analyses, thus, they are less likely to be cited. Maybe we can validate which one is correct after 3 years. Meanwhile, this surge in publications has raised concerns about the potential misuse of the method, often without sufficient critical evaluation, leading to questions about its credibility.[2] While the increasing use of MR has contributed to answering critical questions in medical research, inappropriate application of the method—for instance, through simplified two-sample MR analyses—is favoured by an apparently easy use of MR in widely available and rapidly expanding large datasets, which risk inundating the scientific community with low-quality studies.[3] Additionally, according to consulted editors, journals and expert reviewers are overwhelmed by numerous MR-related submissions on a weekly basis, an overflow that may potentially stain the method's reputation and call for standardisation. In this editorial, we aim to highlight both advantages and detrimental consequences of MR misuse and discuss key considerations for researchers, data scientists, readers and reviewers.

Initially, most MR studies were conducted using individual-level genetic/phenotypic data from well-characterised cohorts where the detailed information on genotypes, levels of exposure and outcomes for each participant is known, guaranteeing greater control over data quality and assumptions. For instance, MR has been employed to investigate the supposed protective effects of serum high-density lipoprotein cholesterol (HDL-C) on coronary artery disease (CAD). A landmark MR study by Voight *et al* challenged the long-standing belief in the protective role of HDL-C against CAD, demonstrating that genetically elevated HDL-C does not reduce the risk of CAD.[4] This finding had great implications, suggesting that the development of HDL-C-raising drugs may not translate into cardiovascular benefits. Such insights helped redirect resources in drug development towards more promising targets, such as low-density lipoprotein cholesterol (LDL-C) and proprotein convertase subtilisin/kexin type 9 (PCSK9) inhibitors, which have shown consistent efficacy in reducing cardiovascular risk. However, most of the MR publications are based on summary-level data where only the associations of genetic instruments with exposures and outcomes are known, which may lead to insufficient considerations in analysis.

MR can provide evidence of causation using observational data, which is a great advantage compared with other epidemiological approaches.[5] Each MR analysis must satisfy three core IV assumptions: (1) relevance—genetic instruments must be closely associated with the exposure; (2) independence—no shared confounders exist between genetic instruments and the outcome; and (3) exclusion restriction—genetic instruments influence the outcome solely through the exposure (no horizontal pleiotropy).[1] Additional assumptions, such as linearity between

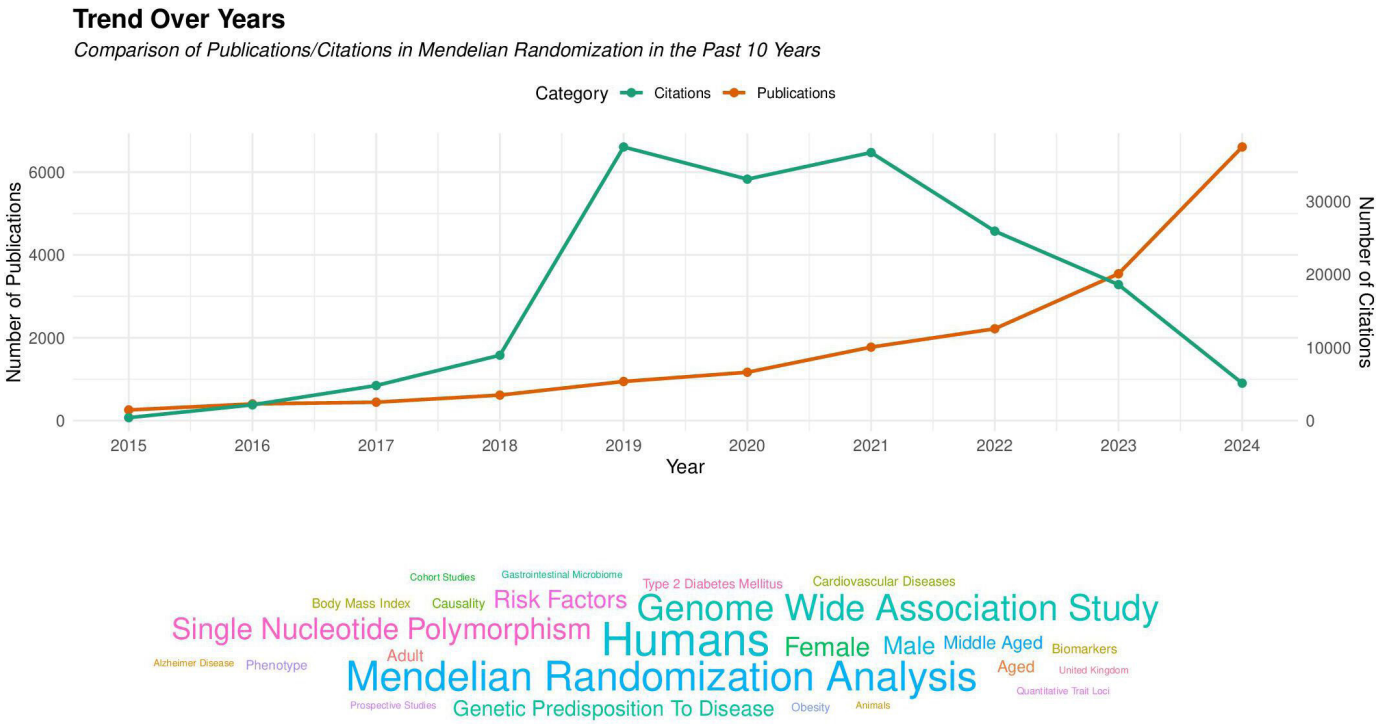

Figure 1 The comparison of publications/citations in Mendelian randomisation from 2015 to 2024. On the top, it reflects the trend of total publications and citations each year. The keywords in this field are displayed at the bottom.

exposure and outcome and the absence of interaction with mediators, are also important as most MR statistical models assume linearity.[6] Causal inference is valid only when all these assumptions are met. However, assumptions (2) and (3) are often untestable, which can lead to invalid conclusions when misapplied.[7] Therein, horizontal pleiotropy, which means the genetic instruments can directly affect the outcome via mechanisms unrelated to exposure, is an inherent weakness of this method but is not carefully evaluated in most recent publications.

Since 2020, this field has witnessed a surge in MR publications, with the annual number exceeding 1000 (figure 1). The accessibility of summary-level data from large-scale GWAS probably played a key role in driving this growth. These data have enabled the proliferation of simple two-sample MR designs, where genetic associations with exposures and outcomes are derived from separate datasets. While this approach has democratised MR research, allowing rapid hypothesis testing on a wide scale and providing novel insights into pathogenesis, it has also raised concerns about misuse and overall reduced study robustness. One could state that many two-sample MR studies published recently lack methodological rigour, fail to address potential biases, such as pleiotropy, and often replicate findings already accessible through online platforms. This trend has led to a flood of poorly verifiable publications, burdening peer reviewers and editors, and threatened to undermine the credibility of the MR approach as a robust tool for causal inference.

To circumvent what we see as a roadblock for future valuable MR-based studies, we outline below the critical considerations for conducting robust MR analyses, as well as recommendations for assessing MR studies for researchers, readers, reviewers and editors.

KEY REQUIREMENTS FOR CONDUCTING ROBUST MR INVESTIGATIONS

Plausibility of genetic relevance

A valid exposure must have clear genetic relevance. For example, serum HDL-C levels are influenced by 3-hydroxy-3-methylglutaryl-CoA reductase (HMGCR) variants, and genetic variants strongly affect susceptibility to metabolic dysfunction-associated steatotic liver disease (MASLD, formerly known as non-alcoholic fatty liver disease).[8 9] Conversely, some studies have used exposures with weak or non-existent genetic relevance, such as air pollution[10] or noodle consumption.[11] Since air pollution exposure is primarily environmental, it is almost impossible to choose genetic instruments directly associated with it.[12] However, air pollution can change homeostasis, and a better approach would involve examining the interaction between genetics and pollutant metabolism, as certain enzymes involved in pollutant processing are genetically influenced (table 1).[13 14]

Selection of genetic instruments

Choosing valid genetic instruments is paramount. Two strategies are commonly employed: biological mechanism-based selection and statistical significance-driven selection.[7] The latter, although prevalent, may introduce bias due to horizontal pleiotropy, as it often identifies instruments with unknown biological functions. Conceptually, pleiotropy refers to a single genetic

Table 1 Summarised key messages for conducting a qualified Mendelian randomisation study

Aspects	Key messages	Examples/notes
Plausibility of genetic relevance	▶ Exposures must have clear genetic relevance. ▶ Avoid weak genetic relevance (eg, environmental factors without genetic interaction). ▶ Focus on gene-environment interactions for environmental exposures.	▶ Valid: Serum HDL-C influenced by HMGCR variants. ▶ Weak: Environmental/behavioural exposures (eg, air pollution and noodle consumption). ▶ Suggestion: Study environment-related molecular features with genetic influence (eg, air pollutant-related metabolites).
Selection of genetic instruments	▶ Combine biological mechanism-based and statistical significance-driven strategies. ▶ Use instruments closely linked to the exposure. ▶ Optimise SNP pruning thresholds (eg, clump r^2 and window values).	▶ Mechanism-based: Specific but limited number of IVs. ▶ Statistics-driven: Risk of horizontal pleiotropy. ▶ Shorter distance between genetic variants and exposure reduces bias.
Diverse statistical models	▶ Use multiple models to ensure robustness. ▶ Choose models based on assumptions (eg, IVW for valid instruments, median-based for tolerance to violations). ▶ Apply advanced models to relax assumptions and reduce errors.	▶ Models: IVW, MR-APSS, MR-Egger, MR-CAUSE. ▶ Advanced models minimise type I errors and improve reliability.
Integration with complementary evidence	▶ Cross-validate MR findings with alternative methods. ▶ Use *in silico* data when experimental or real-world data are unavailable. ▶ Avoid overstating causality; frame results as 'genetically predicted associations.'	▶ Methods: Real-world studies, meta-analyses, colocalisation analyses, functional experiments. ▶ Example: Inverse cholesterol-cholelithiasis link validated through colocalisation.
Interpretation of results and comparison with other literature	▶ Clarify the evidence of causation derived from observational data. ▶ Compare MR results with other studies and discuss to what extent we are confident about the causal inference. ▶ Discuss potential mechanisms mediating the causal pathway.	▶ Neutrally describe results using 'genetically predicted associations', but discuss the evidence of causation. ▶ Compare the genetic effects of genetic variants with environmental factors (eg, drugs or lifestyle changes).

HDL-C, high-density lipoprotein cholesterol; HMGCR, 3-hydroxy-3-methylglutaryl-CoA reductase; IVs, instrumental variables; IVW, inverse variance weighted; MR, Mendelian randomisation; MR-APSS, Mendelian randomisation-accounting for pleiotropy and sample structure; MR-CAUSE, Mendelian randomisation-causal analysis using summary effect estimates; MR-Egger, Mendelian randomisation-Egger regression; SNP, single nucleotide polymorphism.

variant influencing multiple traits or phenotypes, and it can be categorised into vertical pleiotropy and horizontal pleiotropy.[1] In MR, vertical pleiotropy (not problematic) occurs when a genetic variant influences the outcome indirectly through its effect on the exposure (causal pathway).[1] Horizontal pleiotropy (problematic) occurs when the genetic variant directly affects the outcome through different mechanisms.[1] The biological mechanism-based selection is a suitable way to reduce bias from horizontal pleiotropy; however, it may not yield sufficient IVs to perform various sensitivity analyses. Thus, there is a trade-off between the two IV selection approaches. Generally, combining both strategies while considering different thresholds for pruning correlated single nucleotide polymorphisms (SNPs) (eg, varying clump r^2 from 0.01 to 0.001 and window values from 1000 kilobases to 10 000 kilobases) can enhance validity.[15] The term 'pruning' refers to the process of selecting independent genetic variants to serve as instruments by removing those that are in strong linkage disequilibrium (LD).[15] If two variants are in high LD (defined by the r^2 in a given window size in the genome), one is removed (usually the one with weaker association with the exposure). Additionally, instruments closely linked to the exposure—with shorter distances from genetic variants to the exposure—are preferable to minimise bias.[7]

Diverse statistical models

Different statistical models accommodate various assumptions and hypotheses in MR. For example, the inverse variance-weighted (IVW) model assumes all instruments meet IV assumptions, while the median-based model tolerates violations in up to 50% of instruments.[16] Recent advancements, such as MR-APSS, MR-Egger, MR-CAUSE and mode-based models, relax certain assumptions and reduce type I errors.[17] Researchers are encouraged to use multiple models to ensure robust results.

Integration with complementary evidence

Evidence from MR should be corroborated using alternative methods with different underlying assumptions, such as real-world studies, meta-analyses, colocalisation analyses or functional experiments. For example, the inverse association between serum cholesterol and cholelithiasis discovered via MR was strengthened by colocalisation and cohort validation.[18] Although it is difficult to acquire large-scale individual-level data and performing function experiments can take significant time and

effort, researchers can integrate various *in silico* data and methods to corroborate MR results.

Interpretation of results and comparison with other literature

The interpretation of results involves evaluating causation evidence from observational data and comparing MR findings with other studies to assess confidence in causal inference.[3] Importantly, MR is a method of causal inference and can provide evidence of causation from observational data, measuring the associations of SNPs with exposure and outcome separately. Thus, we better use 'genetically predicted association' when presenting results to clarify the causal relationship between exposure and outcome, which is the advantage of MR. Potential mechanisms mediating causal pathways should be discussed, highlighting intermediary factors.[19] Additionally, the study should better compare genetic effects with environmental factors, such as drugs or lifestyle changes, to provide context of gene-environment equivalence and insights into potential interventions or public health implications.[5]

RECOMMENDATIONS FOR READERS, REVIEWERS AND EDITORS

For readers

1. Ensure MR studies adhere to established guidelines, such as STROBE-MR.[19]
2. Verify that genetic instruments are biologically plausible and satisfy MR assumptions.
3. Be cautious of overinterpretation and avoid studies that make strong causal claims without robust sensitivity analyses.
4. Compare MR findings with evidence from other methodologies, such as randomised controlled trials or observational studies.

For editors and reviewers

1. Reject low-quality two-sample MR studies lacking additional evidence or context. Here are the tips to identify the 'low-quality' ones: (1) the levels of exposure are mainly determined by environmental factors and are unlikely to be affected by genetic variants, such as air pollution and noodle consumption; (2) the potential confounders are not clearly defined and the horizontal pleiotropy is not carefully assessed by different methods; (3) only using low-quality GWAS summary statistics (eg, small sample size) from public databases without additional analyses and data. However, we still appreciate MR analysis derived from a reputable GWAS consortium, which can provide high-level evidence of causation.
2. Demand exact justifications for instrument selection and comprehensive sensitivity analyses.
3. Scrutinise methodologies, particularly for novel MR approaches, ensure robustness through control tests.
4. Discourage redundant MR studies that add little value.
5. Encourage interdisciplinary submissions integrating genetics, biology and clinical perspectives.

By prioritising quality, transparency and contextual relevance, readers, editors and reviewers can uphold the credibility of MR research.

IN A NUTSHELL: RECLAIMING THE POTENTIAL OF MR

MR remains a powerful tool for causal inference in epidemiology. However, its misuse, particularly in simplified two-sample analyses and unvalidated methodologies, threatens its credibility. By adhering to standards and emphasising thoughtful application, the scientific community can consolidate MR's reputation as a strong approach for studying causal inference.

Contributors LC drafted the manuscript. AG and CVS revised the manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Commissioned; internally peer reviewed.

Data availability statement No data are available.

ORCID iDs
Lanlan Chen http://orcid.org/0000-0002-1075-5592
Adrien Guillot http://orcid.org/0000-0002-6002-9986
Carolin Victoria Schneider http://orcid.org/0000-0002-6728-9246

REFERENCES

1 Sanderson E, Glymour MM, Holmes MV, *et al*. Mendelian randomization. *Nat Rev Methods Primers* 2022;2:6.
2 Stender S, Gellert-Kristensen H, Smith GD. Reclaiming mendelian randomization from the deluge of papers and misleading findings. *Lipids Health Dis* 2024;23:286.
3 Burgess S, Woolf B, Mason AM, *et al*. Addressing the credibility crisis in Mendelian randomization. *BMC Med* 2024;22:374.
4 Voight BF, Peloso GM, Orho-Melander M, *et al*. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet* 2012;380:572–80.
5 Burgess S, Davey Smith G, Davies NM, *et al*. Guidelines for performing Mendelian randomization investigations: update for summer 2023. *Wellcome Open Res* 2019;4:186.
6 Lawlor DA, Harbord RM, Sterne JAC, *et al*. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med* 2008;27:1133–63.
7 Burgess S, Cronjé HT. Incorporating biological and clinical insights into variant choice for Mendelian randomisation: examples and principles. *egastro* 2024;2:e100042.
8 Tacke F, Horn P, Wai-Sun Wong V, *et al*. EASL–EASD–EASO Clinical Practice Guidelines on the management of metabolic dysfunction-associated steatotic liver disease (MASLD). *J Hepatol* 2024;81:492–542.
9 Rendel MD, Vitali C, Creasy KT, *et al*. The common p.Ile291Val variant of ERLIN1 enhances TM6SF2 function and is associated with protection against MASLD. *Med* 2024;5:963–80.
10 Song Q, Pan J, Pan M, *et al*. Exploring the relationship between air pollution, non-alcoholic fatty liver disease, and liver function indicators: a two-sample Mendelian randomization analysis study. *Front Endocrinol (Lausanne)* 2024;15:1396032.

11 Park S, Liu M. A Positive Causal Relationship between Noodle Intake and Metabolic Syndrome: A Two-Sample Mendelian Randomization Study. *Nutrients* 2023;15:2091.
12 Chen J, Wang H, Fu S. Rethinking methodology and data integrity in Mendelian randomization: Insights from air pollution and MASLD research. *J Hepatol* 2024.
13 Ran S, Zhang J, Tian F, *et al*. Association of metabolic signatures of air pollution with MASLD: Observational and Mendelian randomization study. *J Hepatol* 2024.
14 Ran S, Zhang J, Lin H. Reply to: 'Rethinking methodology and data integrity in Mendelian randomization: Insights from air pollution and MASLD research'. *J Hepatol* 2024.
15 Luo S, Zheng M-H, Wong VW-S, *et al*. Drug-target Mendelian randomisation applied to metabolic dysfunction-associated steatotic liver disease: opportunities and challenges. *Egastro* 2024;2:e100114.
16 Bowden J, Davey Smith G, Haycock PC, *et al*. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol* 2016;40:304–14.
17 Hu X, Cai M, Xiao J, *et al*. Benchmarking Mendelian randomization methods for causal inference using genome-wide association study summary statistics. *Am J Hum Genet* 2024;111:1717–35.
18 Chen L, Qiu W, Sun X, *et al*. Novel insights into causal effects of serum lipids and lipid-modifying targets on cholelithiasis. *Gut* 2024;73:521–32.
19 Skrivankova VW, Richmond RC, Woolf BAR, *et al*. Strengthening the Reporting of Observational Studies in Epidemiology Using Mendelian Randomization: The STROBE-MR Statement. *JAMA* 2021;326:1614–21.