METHOD ARTICLE

UPDATE Guidelines for performing Mendelian randomization investigations: update for summer 2023 [version 3; peer review: 2 approved]

Stephen Burgess[1,2], George Davey Smith[3,4], Neil M. Davies[3,5-7], Frank Dudbridge[8], Dipender Gill[9], M. Maria Glymour[10], Fernando P. Hartwig[3,11], Zoltán Kutalik[12-14], Michael V. Holmes[15,16], Cosetta Minelli[17], Jean V. Morrison[18], Wei Pan[19], Caroline L. Relton[3,4,20], Evropi Theodoratou[21,22]

[1]MRC Biostatistics Unit, University of Cambridge, Cambridge, UK
[2]BHF Cardiovascular Epidemiology Unit, University of Cambridge, Cambridge, UK
[3]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK
[4]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
[5]Division of Psychiatry, University College London, London, UK
[6]Department of Statistical Sciences, University College London, London, WC1E 6BT, UK
[7]K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, Norway
[8]Department of Health Sciences, University of Leicester, Leicester, UK
[9]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK
[10]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, USA
[11]Postgraduate Program in Epidemiology, Federal University of Pelotas, Pelotas, Brazil
[12]Swiss Institute of Bioinformatics, Lausanne, Switzerland
[13]Department of Computational Biology, University of Lausanne, Lausanne, Switzerland
[14]University Center for Primary Care and Public Health (Unisanté), Lausanne, Switzerland
[15]MRC Population Health Research Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK
[16]Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK
[17]National Heart and Lung Institute, Imperial College London, London, UK
[18]Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA
[19]Division of Biostatistics, University of Minnesota, Minneapolis, MN, USA
[20]London School of Hygiene & Tropical Medicine, Keppel Street, London, WC1E 7HT, UK
[21]Centre for Global Health, Usher Institute, University of Edinburgh, Edinburgh, UK
[22]Edinburgh Cancer Research Centre, Institute of Genetics and Cancer, The University of Edinburgh, Edinburgh, UK

v3 **First published:** 26 Nov 2019, **4**:186
https://doi.org/10.12688/wellcomeopenres.15555.1
Second version: 28 Apr 2020, **4**:186
https://doi.org/10.12688/wellcomeopenres.15555.2
Latest published: 04 Aug 2023, **4**:186
https://doi.org/10.12688/wellcomeopenres.15555.3

Open Peer Review

Approval Status ✓ ✓

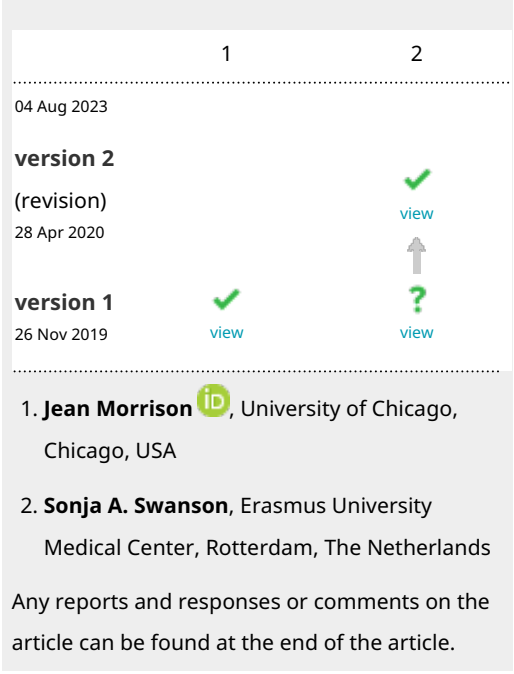
	1	2
version 3 (update)		

Abstract

This paper provides guidelines for performing Mendelian randomization investigations. It is aimed at practitioners seeking to

undertake analyses and write up their findings, and at journal editors and reviewers seeking to assess Mendelian randomization manuscripts. The guidelines are divided into ten sections: motivation and scope, data sources, choice of genetic variants, variant harmonization, primary analysis, supplementary and sensitivity analyses (one section on robust statistical methods and one on other approaches), extensions and additional analyses, data presentation, and interpretation. These guidelines will be updated based on feedback from the community and advances in the field. Updates will be made periodically as needed, and at least every 24 months.

Keywords
Mendelian randomization, guidelines, genetic epidemiology, causal inference



Corresponding author: Stephen Burgess (sb452@medschl.cam.ac.uk)

Author roles: Burgess S: Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Davey Smith G**: Conceptualization, Writing – Review & Editing; **Davies NM**: Conceptualization, Writing – Review & Editing; **Dudbridge F**: Conceptualization, Writing – Review & Editing; **Gill D**: Conceptualization, Writing – Review & Editing; **Glymour MM**: Conceptualization, Writing – Review & Editing; **Hartwig FP**: Conceptualization, Writing – Review & Editing; **Kutalik Z**: Writing – Review & Editing; **Holmes MV** : Conceptualization, Writing – Review & Editing; **Minelli C**: Conceptualization, Writing – Review & Editing; **Morrison JV**: Writing – Review & Editing; **Pan W**: Writing – Review & Editing; **Relton CL**: Conceptualization, Writing – Review & Editing; **Theodoratou E**: Conceptualization, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: Grant information is now as follows: Stephen Burgess is supported by the Wellcome Trust (225790/Z/22/Z) and the United Kingdom Research and Innovation Medical Research Council (MC_UU_00002/7). Dipender Gill is funded by the Wellcome Trust. Evropi Theodoratou is supported by a CRUK Career Development Fellowship (C31250/A22804). George Davey Smith, Neil M Davies, Fernando Hartwig, and Caroline Relton work within a Unit funded by the Medical Research Council (MC_UU_00011/1, MC_UU_00011/5). Neil Davies is supported by a Future Research Leaders grant funded by the Economics and Social Research Council (ES/N000757/1) and a Norwegian research Council grant number 295989. Michael Holmes works in a unit that receives funding from the UK Medical Research Council and is supported by a British Heart Foundation Intermediate Clinical Research Fellowship (FS/18/23/33512) and the National Institute for Health Research Oxford Biomedical Research Centre. Wei Pan is supported by NIH grants (R01 AG065636, RF1 AG067924, U01 AG073079). This research was supported by the National Institute for Health Research Cambridge Biomedical Research Centre (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.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Burgess S, Davey Smith G, Davies NM *et al.* **Guidelines for performing Mendelian randomization investigations: update for summer 2023 [version 3; peer review: 2 approved]** Wellcome Open Research 2023, **4**:186 https://doi.org/10.12688/wellcomeopenres.15555.3

First published: 26 Nov 2019, **4**:186 https://doi.org/10.12688/wellcomeopenres.15555.1

REVISED **Amendments from Version 2**

This is an update to the guidelines for performing Mendelian randomization investigations to reflect updates in the literature over the past three years – both advances in technologies and datasets providing more opportunities for advanced analyses, and methodological innovations enabling broader and more reliable analyses. As stated in the original publication, we will continue to revise these guidelines periodically. Notable changes in this update are: 1) addition of three new co-authors (Zoltán Kutalik, Jean Morrison, Wei Pan) to better reflect the diversity and geographical spread of thought leaders in MR, 2) new paragraphs on within-family analyses, 3) updated advice on drug-target MR analyses, including on variant choice and use of colocalization, 4) revised discussion of methods including recently published robust methods for MR, 5) substantial revision to the section on other sensitivity analyses and a renewed emphasis on triangulation of evidence, 6) a new section "Extensions and additional analyses", and 7) a general edit of the text to improve accuracy and clarity. We will continue to revisit these guidelines as applied practice shifts and the methodological literature develops.

Any further responses from the reviewers can be found at the end of the article

The aim of this paper is to provide guidelines for performing Mendelian randomization investigations. It is written both for practitioners seeking to undertake analyses and write up their findings, and for journal editors and reviewers seeking to assess Mendelian randomization manuscripts. These guidelines are deliberately written as suggestions and recommendations rather than as prescriptive rules, as we believe that there is no recipe or single "right way" to perform a Mendelian randomization investigation. Best practice will depend on the aim of the investigation and the specific exposure and outcome variables. However, we believe these guidelines will help investigators to consider the key issues in designing, undertaking and presenting Mendelian randomization analyses. These guidelines will be updated based on feedback from the community and advances in the field. Updates will be made periodically as needed, and at least every 24 months.

These guidelines are complementary to the STROBE-MR recommendations on reporting Mendelian randomization investigations[1,2]. Here, we provide advice on which analyses to perform in a Mendelian randomization investigation, whereas the STROBE-MR guidelines focus on reporting the analyses chosen by the investigators. We assume a familiarity with the basic concepts of Mendelian randomization and genetic epidemiology, such as pleiotropy and linkage disequilibrium[3–6]. We use the term "exposure" to refer to the proposed causal factor, and "outcome" to refer to the trait or disease that the exposure is hypothesized to influence.

Flowcharts highlighting some of the key analytic steps and choices for investigators are provided as Figure 1 and Figure 2, and a one-page checklist summarizing these guidelines written for reviewers of Mendelian randomization analyses is provided as Figure 3. The guidelines are divided into ten sections: motivation and scope, data sources, choice of genetic variants, variant harmonization, primary analysis, supplementary and sensitivity analyses (one section on robust statistical methods and one on other approaches), extensions and additional analyses, data presentation, and interpretation. Software to implement the statistical methods is referenced in Table 1.

1. Motivation and scope

Mendelian randomization uses genetic variants to assess causal relationships using observational data. A genetic variant can be considered as an instrumental variable for a given exposure if it satisfies the instrumental variable assumptions: 1) it is associated with the exposure, 2) it is not associated with the outcome due to confounding pathways, and 3) it does not affect the outcome except potentially via the exposure[7,8].

Before embarking on a Mendelian randomization analysis, investigators should consider the aims of their investigation and the primary hypotheses of interest. There are many potential motivations for using Mendelian randomization, and the motivation should influence decisions on how to perform the analysis, and how to arrange and present its results. The objective of a Mendelian randomization analysis is a test of a causal hypothesis, and sometimes additionally an estimate of a causal effect[9]. The straightforward statement of the causal hypothesis is that interventions on the exposure variable will affect the outcome. If the genetic associations with the exposure vary with time, then there are some nuances in terms of what causal hypotheses can be tested[10]; we discuss the impact of time-varying relationships between variables in Section 10.

If a Mendelian randomization investigation is performed primarily to assess whether an exposure has a causal effect on an outcome, then estimating the size of the causal effect of the exposure on the outcome is less important and may even be unnecessary[8,11]. Priorities in such an analysis are to find genetic variants that satisfy the instrumental variable assumptions and to test their associations with the outcome in the largest available dataset that is relevant to the causal question of interest. Investigators may be able to find mediating traits downstream of the exposure that both help understand the mechanistic pathways from the exposure to the outcome, and provide modifiable targets for intervention in order to influence the outcome.

In contrast, if investigators seek to estimate the quantitative impact on the outcome of a proposed intervention in the exposure[12], then further questions become more important, such as how well the genetic variant proxies the specific intervention, whether genetic associations with the exposure are estimated in a relevant population, and whether the relationships between variables are linear and homogeneous in the population[13]. However, as we discuss in Section 10, causal estimates from Mendelian randomization should always be interpreted with caution. Alternatively, if investigators simply want to assess whether traits share common genetic predictors (potentially implying shared aetiological mechanisms), then an analytic approach that assesses shared heritability (such as LD-score regression[14] or bivariate genome-based restricted maximum likelihood [GREML][15]) may be preferable to conducting a Mendelian randomization investigation.

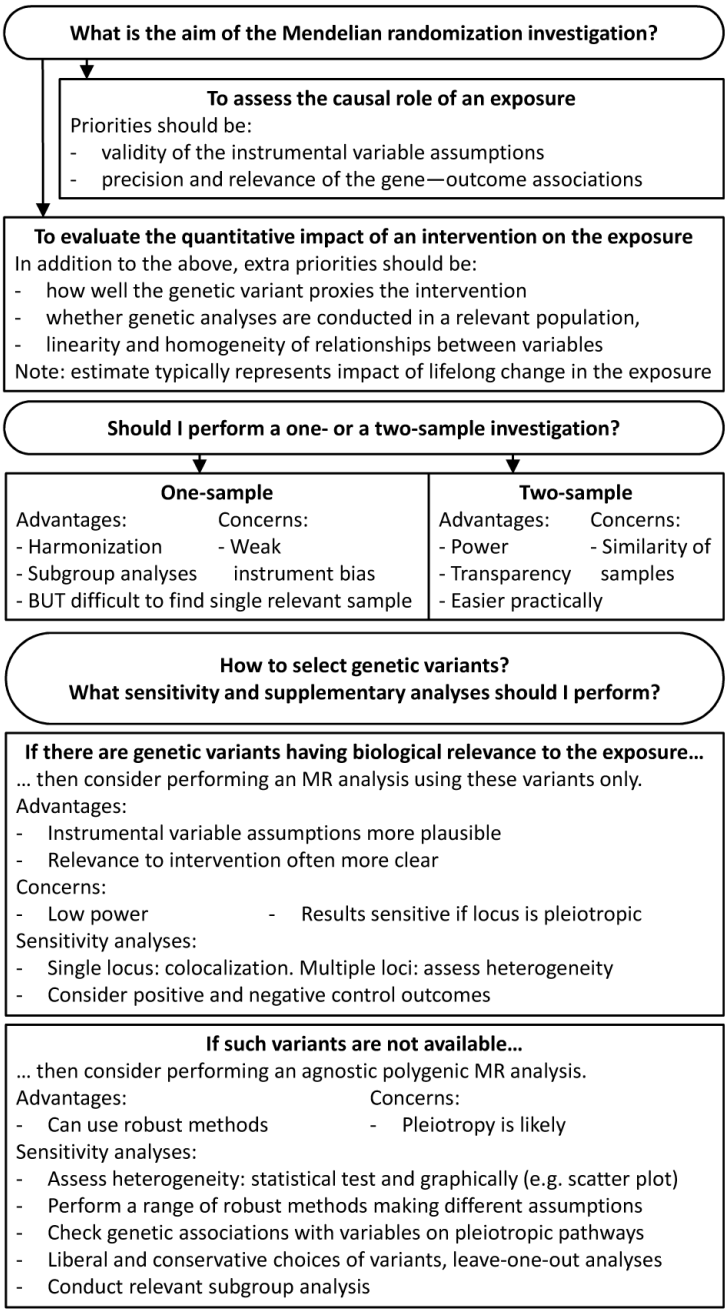


Figure 1. Flowchart highlighting some of the key analytic choices in performing a Mendelian randomization (MR) analysis.

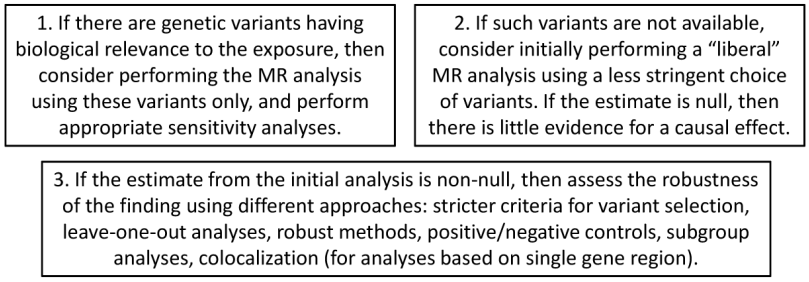


Figure 2. Generic analytic pipeline for Mendelian randomization (MR).

Checklist for reviewing Mendelian randomization investigations

1. What is the primary hypothesis of interest? What is the motivation for using Mendelian randomization? What is the scope of the investigation? What and how many primary analyses?

---

2. Data sources

   What type of Mendelian randomization investigation is this? One-sample or two-sample? Sample overlap? Summarized data or individual-level data? Drawn from same population? Relevance to applied research?

---

3. Selection of genetic variants – how were the genetic variants chosen? Single or multiple gene regions?
   a. Biological rationale?
   b. GWAS analysis? If so, what dataset? What was the p-value threshold? Clumping?
   c. Were genetic variants excluded from the analysis? Associations with pleiotropic pathways?
   d. How else was the validity of genetic variants as instrumental variables assessed?

---

4. Variant harmonization (for two-sample analyses)

   Was it checked that genetic variants were appropriately orientated across the datasets?

---

5. Primary analysis

   What was the primary analysis? What was the statistical method? How implemented? Multiple testing?

---

6 and 7. Supplementary and sensitivity analyses

   What analyses were performed to support and assess the validity of the primary analysis?
   For example: stricter criteria for variant selection, assess heterogeneity, robust methods, subgroup analysis, positive/negative controls, 'leave-one-out' analyses, colocalization (single gene region)

---

8. Extensions and additional analyses

   What additional analyses were performed to better understand the nature of the causal effect?

---

9. Data presentation

   How are the data and results presented to allow readers to assess the analysis and assumptions?
   For example: scatter plot, forest/funnel/radial plot, R^2/F statistics, comparison of methods, power

---

10. Interpretation

    How have results been interpreted, particularly any numerical estimates?

Figure 3. Checklist of questions to consider when reviewing a Mendelian randomization investigation.