
CAUSAL INFERENCE SERIES · 01

Mendelian Randomization

A Practical Guide to Causal Inference

ENGLISH EDITION

For Graduate Researchers in
Epidemiology, Biostatistics & Genetics

Statistical Foundations · Study Design · Advanced Methods · R in Practice

Updated for 2026: estimands, identification, collider bias, omics MR

Su Hyun Lee, PhD

PhD in Public Health, Yonsei University · Genetic Epidemiology & Causal Inference

2026 GRADUATE RESEARCHER EDITION

FREE SAMPLE · 2026 ENGLISH EDITION

Mendelian Randomization: A Practical Guide to Causal Inference

2026 Graduate Researcher Edition

© 2026 Su Hyun Lee, PhD. All rights reserved. This ebook was written for research methods education. It is intended for the purchaser's personal study and research reference; resale or redistribution of all or a substantial part without the author's permission is prohibited. It does not replace original papers or scholarly sources, and researchers conducting empirical analyses should consult the original publications and current software documentation.

Names such as Dataset-A, Dataset-B, and K-Health Genetic Cohort, and all example figures used in this book, are pseudonyms or hypothetical examples for teaching purposes. They do not refer to any specific institution or real research data.

Disclaimer

This book is a practical guide for research methods education. It does not replace statistical consultation, clinical decision-making, or the original methodological papers. Software defaults and public GWAS resources can change, so consult current documentation before conducting an empirical analysis.

How to Use This Book

This practical guide is designed for graduate students and researchers seeking to apply Mendelian randomization (MR) in empirical research. It goes beyond conceptual foundations, providing a step-by-step progression from study design and diagnostic assessment to interpretation and manuscript preparation.

Structure

PART I	Foundations	Why observational studies fall short; what MR assumes
PART II	Statistical Foundations	Wald ratio, 2SLS, F-statistic, IVW, estimands, ALICE
PART III	Design and Estimation	One- and two-sample designs, instruments, harmonization, family designs
PART IV	Diagnostics and Advanced MR	Pleiotropy, heterogeneity, collider bias, MVMR, cis-MR, MR-SPI
PART V	Applied Analysis	TwoSampleMR workflow, R code, case study, QC, omics MR
PART VI	Interpretation and Writing	Interpretation, STROBE-MR, templates, reviewer responses

Reading order

If MR is new to you, read Parts I, III, and V in that order and consult Part IV as your analysis progresses. If you already have MR experience, you can start with

Parts II and IV. If you are writing a paper, keep Part VI and the checklists in the appendices open first.

Notation

Key terms follow standard usage in the international methods literature. Each abbreviation is spelled out at first use in a chapter, after which the abbreviation is used.

Abbreviation	Full term
MR	Mendelian randomization
IV	Instrumental variable
GWAS	Genome-wide association study
SNP	Single-nucleotide polymorphism
LD	Linkage disequilibrium
IVW	Inverse-variance weighted method
2SLS	Two-stage least squares
MVMR	Multivariable Mendelian randomization
EAF	Effect allele frequency
OR	Odds ratio

Contents

PART I Foundations of Causal Inference and MR	7
01 Why MR? From Correlation to Causal Effect	8
02 Questions Observational Studies Alone Cannot Answer	10
03 Minimal Genetics: SNPs, Alleles, LD, and GWAS	13
04 The Three Core Assumptions of MR	14
05 Understanding MR through DAGs	16
06 Association, Causality, and Triangulation	17
PART II Statistical Foundations	19
07 Core Statistics in MR	20
08 The Statistical Core of Instrumental Variable Analysis	25
09 The Ratio Method, 2SLS, and Confidence Intervals	27
10 Weak Instruments and Finite-Sample Bias	29
11 Multiple Instruments, Allele Scores, and Power	33
12 Causal Estimands: What Exactly Does MR Estimate?	35
13 The ALICE Model: A Common Language for Summary-Data MR	39
PART III Design and Estimation	43
14 One-Sample MR: Starting from Individual-Level Data	44
15 Designing Two-Sample MR	45
16 SNP Selection and LD Clumping	46
17 Harmonization: Aligning Effect Directions	47
18 The Wald Ratio and IVW	48
19 Summary-Data MR and Evidence Synthesis	49
20 Generalizability of MR Estimates	51
21 Binary Outcomes and Non-Linear Causal Questions	52
22 Population-Based versus Family-Based MR	53
PART IV Diagnostics and Advanced MR	56
23 MR-Egger, Weighted Median, and Weighted Mode	57
24 Diagnosing Heterogeneity and Pleiotropy	58
25 MR-PRESSO and Outliers	59
26 Leave-One-Out, Single-SNP, and Funnel Plots	60
27 Directionality: Steiger and Bidirectional MR	61
28 Multivariable MR and Mediation	62
29 Sample Overlap, Winner's Curse, and Selection Bias	64

30 Collider Bias: Spurious Associations from Selection, Survival, and Conditioning	65
31 MR-TRYX: Using Pleiotropy as Information	71
32 Drug-Target MR, cis-MR, and Colocalization	72
33 Recent Methods: MR-RAPS, CAUSE, LCV, and Bayesian MR	74
34 Three Identifying Assumptions that Tolerate Invalid Instruments	77
35 Weak-Instrument Bias: Modern Remedies	80
36 Instrument Selection and Post-Selection Inference: MR-SPI and Uniformly Valid Intervals	83
PART V Applied Analysis	86
37 A Practical TwoSampleMR Workflow	87
38 R Code Templates and Reproducibility	88
39 Case Study: A Metabolic Exposure and a Cardiovascular Outcome	90
40 Harmonization in Practice: Common Errors and Quality Checks	92
41 Building Publication-Quality Results Tables	93
42 Reading Figures: Scatter, Forest, Funnel, and Leave-One-Out	94
43 Worked Examples by Hand	95
44 A Minimal Reproducible Project	97
45 Quality-Control Checklist for MR Analyses	99
46 Omics MR (xMR): Finding Causal Targets in the Proteome	100
PART VI Interpretation and Writing	103
47 Interpreting MR Estimates Beyond Statistical Significance	104
48 Writing with STROBE-MR	105
49 Sentence Templates for Methods, Results, and Discussion	106
50 Responding to Reviewers	109
51 Reading MR Papers Critically	111
52 Pre-Submission Checklist	113
53 From the 2015 Classic to Modern MR: A Learning Map	115
APPENDIX Appendices	117
A Glossary of MR Terms	118
B Ten Self-Test Questions	120
C Exercises for Graduate Researchers	121
D References	122

PART I

Foundations of Causal Inference and MR

Starting with questions that observational studies alone cannot resolve, we examine what MR does differently and the assumptions on which it relies.

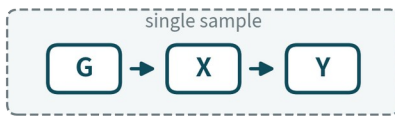
CHAPTER 01

Why MR? From Correlation to Causal Effect

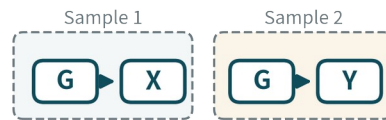
Observational studies are powerful for identifying candidate causes of disease, but the fact that an exposure and an outcome move together cannot by itself establish a causal effect. Unmeasured confounding, measurement error, and reverse causation can all produce associations in the same direction. Mendelian randomization (MR) is a causal inference design that uses genetic variants as instrumental variables to sidestep some of these problems.

The central question is not “Do genes cause disease?” but “Are genetically predicted differences in an exposure linked to differences in the outcome?” MR is therefore best understood as a tool for evaluating the potential causal effects of modifiable risk factors, biomarkers, behaviors, or drug targets.

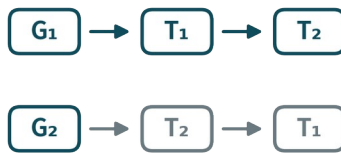
RCTs and MR are not the same. An RCT evaluates the average effect of an actual intervention, whereas MR usually exploits differences in exposure induced by lifelong genetic variation. An MR estimate should not be equated directly with “the effect of giving a drug.” The strength of MR lies in addressing questions where randomization is impractical; its weakness lies in its dependence on instrumental variable assumptions and on the underlying genetic architecture.

(a) One-sample MR

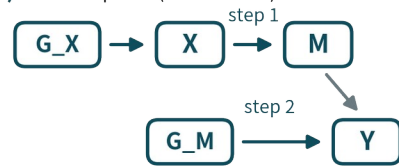
2SLS · every step estimated in one individual-level dataset

(b) Two-sample MR

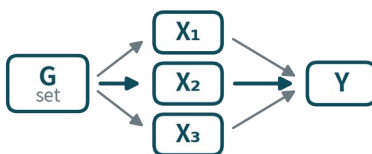
β_{GX} and β_{GY} from two non-overlapping GWAS

(c) Bidirectional MR

forward and reverse MR with separate instruments

(d) Two-step MR (mediation)

$X \rightarrow M$ and $M \rightarrow Y$ estimated separately

(e) Multivariable MR

direct effect of X_2 conditional on X_1 and X_3

(f) Drug-target / cis-MR

colocalization checks for a shared causal variant

Figure 1. Main types of MR study design. (a) One-sample MR, (b) two-sample MR, (c) bidirectional MR, (d) two-step MR, (e) multivariable MR, (f) drug-target/cis-MR. The data structures and assumptions each design requires are covered in Parts III and IV.

CHAPTER 02

Questions Observational Studies Alone Cannot Answer

Association Is Not Causation

Why an Observed Association Can Differ from a Causal Effect

Epidemiological research begins by observing the association between an exposure X and an outcome Y . Case-control, cross-sectional, and cohort studies all explore associations, but confounding, reverse causation, selection bias, information bias, and unmeasured confounding make it difficult to interpret an association directly as causation. This is where MR can provide complementary evidence.

A causal effect is closer to a counterfactual question: “If we intervened to change X , how would Y have differed in the same individual?” An association, by contrast, summarizes the joint distribution of X and Y that was actually observed. A strong association, a small p -value, or a dose-response pattern does not, on its own, establish a causal effect.

Bradford Hill Viewpoints and Modern Causal Inference

The Bradford Hill viewpoints—strength, consistency, temporality, biological gradient, plausibility, and others—offer perspectives to consider when interpreting an association causally. In modern causal inference, however, they are not a mechanical checklist for declaring causality. Temporality is close to a necessary condition, but neither a strong association nor biological plausibility can remove unmeasured confounding. MR adds evidence with a different bias structure by using a genetic instrument.

When Observational and MR Evidence Diverge

Case 1. HDL-C and Myocardial Infarction

Higher high-density lipoprotein cholesterol (HDL-C) was long strongly associated with lower myocardial infarction risk. In the MR study by Voight et al. (Lancet, 2012), however, genetic mechanisms that raise HDL-C did not lower myocardial infarction risk as expected. This finding was an important turning point: raising HDL-C concentration per se may not be the same causal intervention as the protective association seen in observational studies.

Case 2. CRP and Coronary Heart Disease

C-reactive protein (CRP) has shown strong observational associations with coronary heart disease. An individual-participant-data MR meta-analysis of 47 studies and 194,418 participants used variants that raise CRP genetically to evaluate its causal role, revealing a discrepancy between the observational association and the genetic evidence. Cases like this show how useful MR can be for distinguishing whether a biomarker is a risk marker or a causal factor.

How Large Is the MR Literature?

The MR literature has grown very rapidly. A 2025 review reported that a PubMed search on 24 February 2025 for “Mendelian randomization” OR “Mendelian randomisation” already returned more than 15,000 publications. A separate 2025 bibliometric study identified 7,801 publications from 2003–2023 in the Web of Science Core Collection. Because counts vary by database and search string, this book reports search dates and search strings rather than a single fixed number.

REPRODUCIBLE SEARCH

PubMed search string: "Mendelian randomization" OR "Mendelian randomisation". Result

counts grow daily, so always record the search date in publications. As of 2025-02-24, more than 15,000 PubMed records were reported.

Continue with the Complete Guide

This sample introduces the foundations and motivation for Mendelian randomization. The complete 127-page edition continues from core assumptions and statistical foundations to study design, diagnostics, advanced MR, R workflows, interpretation, and manuscript preparation.

The complete 2026 English Edition includes:

- IVW, MR-Egger, weighted median and weighted mode
- Weak instruments, F-statistics, heterogeneity and pleiotropy
- MR-PRESSO, Steiger, bidirectional MR and MVMR
- Collider bias, sample overlap, winner's curse and selection bias
- MR-RAPS, CAUSE, LCV, Bayesian MR and MR-SPI
- Drug-target MR, cis-MR, colocalization and omics MR
- TwoSampleMR workflows, harmonization and quality control
- STROBE-MR, writing templates and reviewer-response guidance

Complete 127-page guide · \$89

PDF + EPUB included · Available on Gumroad

Su Hyun Lee, PhD

Genetic Epidemiology & Causal Inference · 2026 Graduate Researcher Edition