Introduction to Causal Analysis using Mendelian Randomisation

Tayyaba Alvi, Mark Olenik, and Handan Melike Dönertas

Leibniz Institute on Aging – Fritz Lipmann Institute (FLI)

Jena, Germany

Slides, code & data available at:

https://donertas-group.github.io/ismb2025_mr_tutorial/

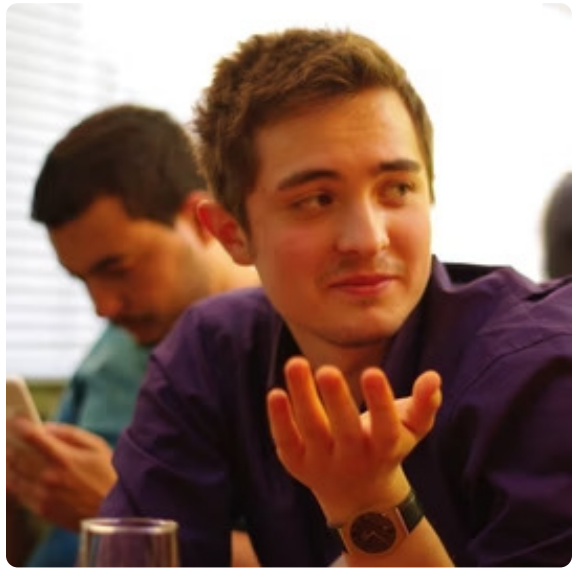
ISMB2025 – July 20, 2025

Who are we?

Tayyaba Alvi

PhD Candidate

Main interests: Microbiome, Bioinformatics



Mark Olenik

Postdoctoral Researcher

Main interests: Machine Learning, Mathematical Biology, Aging

Handan Melike Dönertas

Principal Investigator (PI)

Main interests: Aging, Computational Biology, Microbiome

Welcome!

1. **Part 1: The Problem (~45m)**

   Why Association is Not Causation - The fundamental motivation for MR. **(📌 you are here)**

2. **Part 2: The Theory (~30m)**

   Statistical Foundations (Mark Olenik) - The statistical engine that makes MR work.

3. **Part 3: Technical Introduction (~15m)**

   Delving into the core statistical and biological principles.

4. **Part 4: Simulations to Explain Assumptions (30m)**

   Understanding the underlying principles through interactive simulations.

5. **Part 5: Hands-on Practical (~45m)**

   Applying Mendelian Randomisation methods to real-world datasets.

6. **Part 6: Your Turn! Exercise (~30m)**

   An opportunity to apply what you've learned.

Tell us about yourself

Why are you interested in Mendelian Randomisation?

What sparked your curiosity in this method?

Are you planning to use MR in your work or research?

How do you envision applying it?

What's your background?

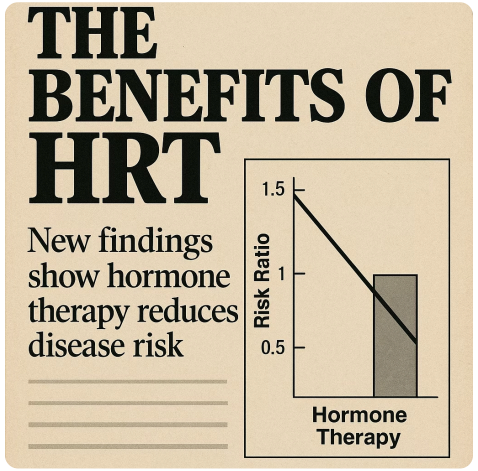
e.g., genetics, statistics, epidemiology, computational biology, other?

A Promise from the 90s

Hormone Replacement Therapy (HRT) showed promising results in early studies:

- Dozens of large observational studies showed a strong, consistent link
- Women on HRT appeared to have a **30-50% lower risk** of coronary heart disease
- Biological rationale seemed plausible
- Estrogen was known to have beneficial effects on lipid profiles

HRT became a standard of care for primary prevention, taken by millions of women.



20.07.25 | ISMB2025

What did RCT say?

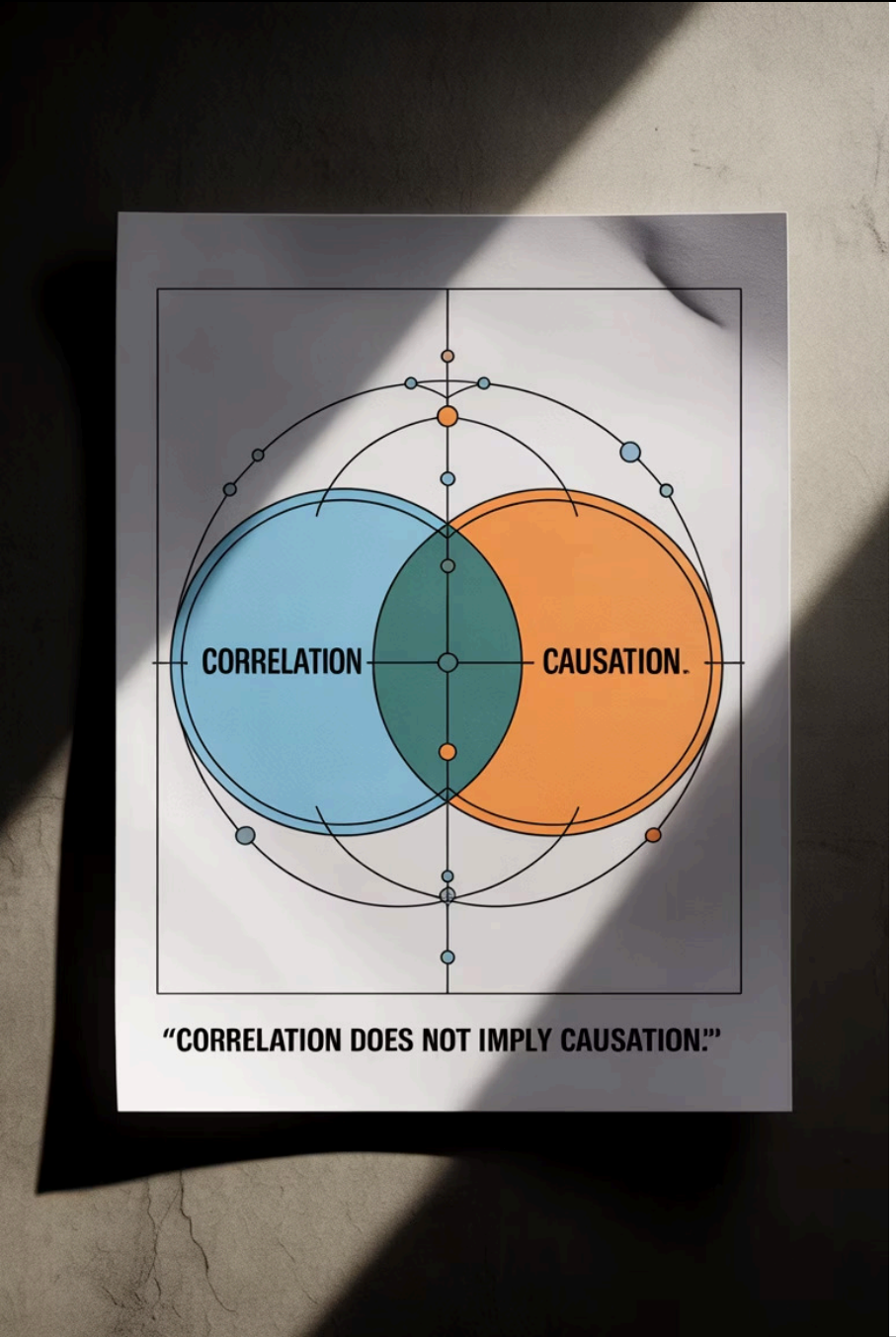
1. Large-scale Randomized Controlled Trials (RCTs) were finally conducted to confirm the benefit.

2. The Women's Health Initiative (WHI) studied two forms: estrogen+progestin and estrogen-alone.

3. **No benefit for heart disease - may even cause a harm.**

4. Later studies suggest timing is very important.

20.07.25 | ISMB2025



What did go wrong?

Correlation IS NOT Causation

20.07.25 | ISMB2025

Objectives

1. Understand the **motivation and logic** behind MR
2. Describe the **three core assumptions** for valid causal inference
3. Critically evaluate **MR results** and recognize **common biases**
4. Access and prepare **GWAS summary statistics** for MR
5. Perform MR using key methods: Inverse-Variance Weighted (IVW) and MR-Egger
6. **Interpret MR estimates** in the context of biomedical research

Why Causal Inference?

The Problem

Observational studies often find correlations that are misleading or non-actionable

The Solution

Causal inference helps identify which associations can be turned into interventions

Correlation IS NOT Causation

20.07.25 | ISMB2025

Bias 1: Confounding

The Problem

A third factor (a confounder) is linked to both the exposure and the outcome, creating a false "back-door" association.

The Culprit

"Healthy User Bias." Women who opted for HRT were systematically different.

The Reality

HRT users were of higher socioeconomic status, more educated, less likely to smoke, more active, and generally more health-conscious.

Conclusion: It was their healthy lifestyle, not the drug, that was truly protective.

Correlation IS NOT Causation

Bias 2: Reverse Causation

The Problem

The causal arrow is pointing in the wrong direction; the outcome actually causes the exposure.

Example 1

Low body weight and mortality. Severe illness (the outcome) often causes unintentional weight loss (the exposure).

Example 2

Diet and disease. A patient diagnosed with heart disease (outcome) is told to eat a low-fat diet (exposure).

An observational study might then falsely associate low-fat diets with heart disease.

Correlation IS NOT Causation

Bias 3: Measurement Error

1. Many exposures are hard to measure precisely (e.g., long-term diet, physical activity, stress).
2. Self-report questionnaires are notoriously "noisy."
3. This **non-differential** (random) error doesn't usually create false associations.
4. Instead, it adds statistical noise that **biases true effects towards the null** (an effect of zero), a phenomenon called "attenuation bias."

It makes real effects harder to see.

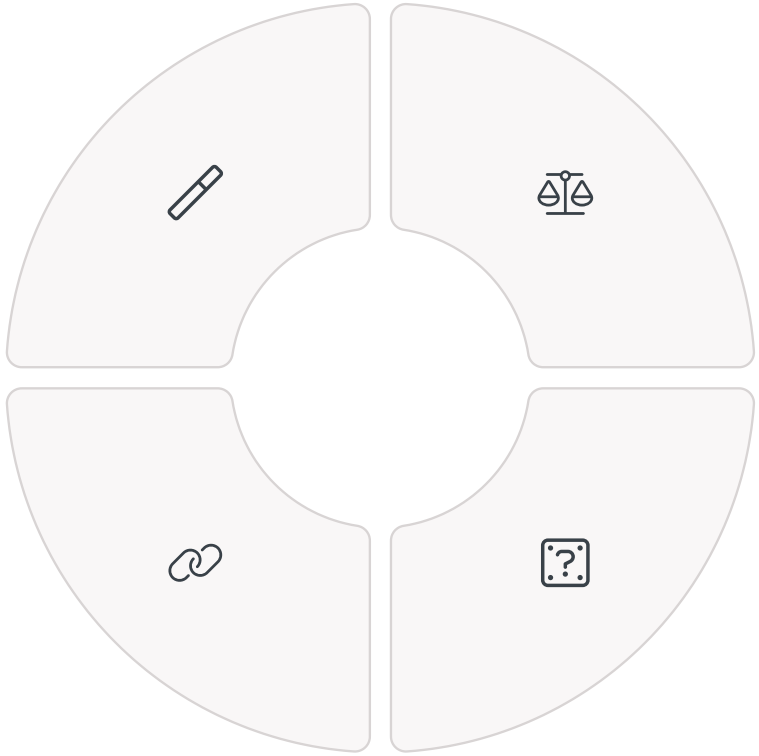
The Gold Standard: RCTs

Randomization Magic

By randomly allocating the intervention, we create two groups that are, on average, identical.

Breaking the Link

It's the only method that can break the link with unmeasured confounders.



Known Confounders

Groups are balanced on factors like age, sex, and other measured variables.

Unknown Confounders

Also balanced on unmeasured factors like genetics, wealth, personality, and "healthy user" habits.

Can we always do an RCT?

Cost

Prohibitively expensive (often hundreds of millions of dollars).

Time

Can take decades for diseases with long latency.

Ethics

We cannot randomize people to exposures we know are harmful (e.g., smoking, pollution).

Generalizability

Trial volunteers are often a select group, limiting how well results apply to the real world.

The Epidemiologist's Dilemma

Observational Studies

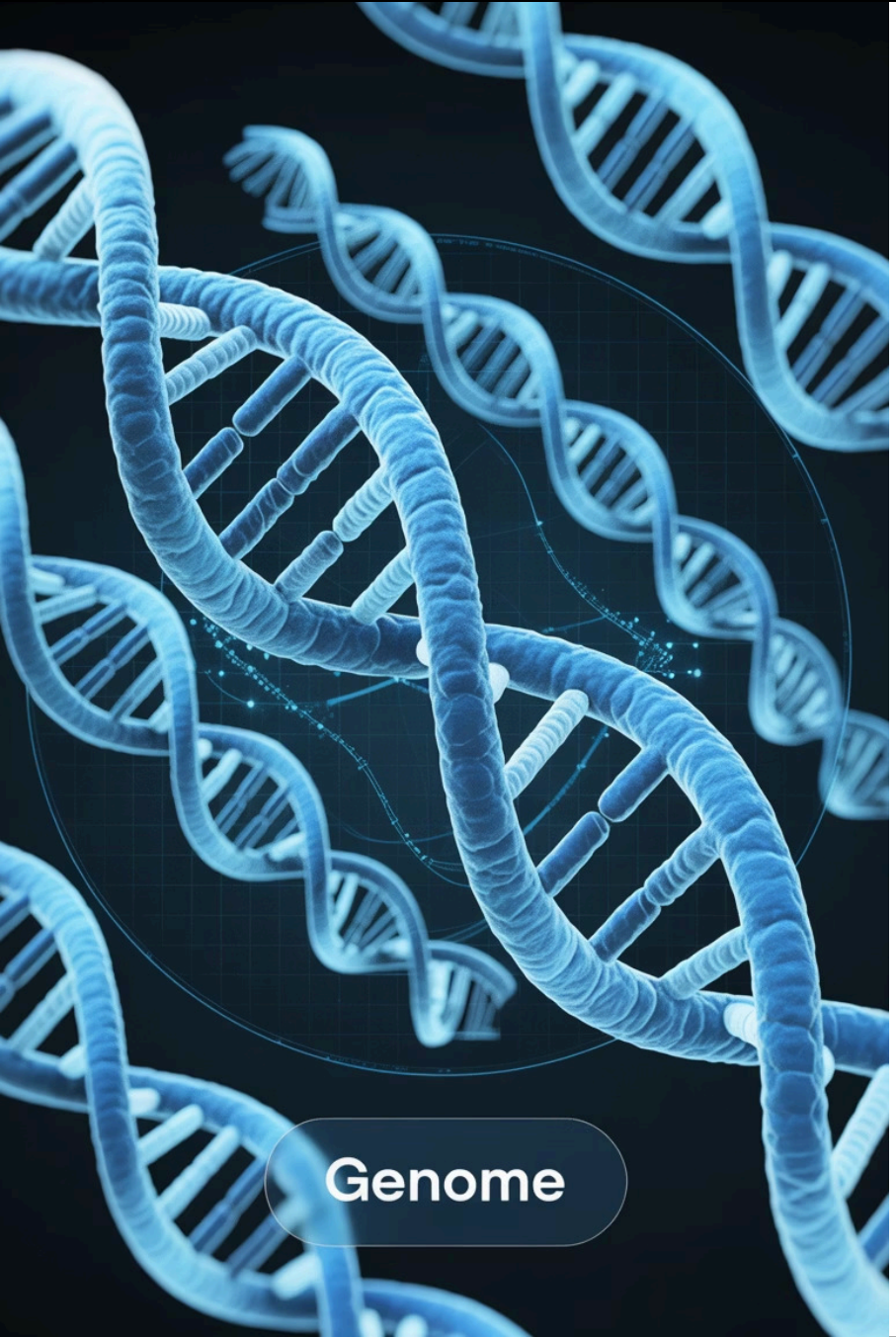
- Cheap and fast
- Can study any exposure
- Prone to intractable bias

RCTs

- Rigorous and causal
- Often impossible due to cost
- Time or ethical constraints

The Key Question: How can we get the causal rigor and confounding control of an RCT, but using the cheap, readily available data from observational studies?

20.07.25 | ISMB2025



The Core Solution

Nature's RCT – Mendelian Randomization

20.07.25 | ISMB2025

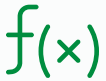
MR – Biological Basis

Central Dogma

The flow of information is DNA -> RNA -> Protein -> Biological Function.

SNPs

Single Nucleotide Polymorphisms are common, single "letter" changes in the DNA code.



Functional Changes

A SNP can alter a protein's function or change how much of it is made (gene expression).

Phenotype Impact

This is the fundamental mechanism through which genes influence our observable traits.

Nature's RCT: MR

The Concept

Genetic variants are randomly assigned at conception, creating natural experiment groups.

- Not related to confounders
- Practical to study
- Ethical to analyze

20.07.25 | ISMB2025

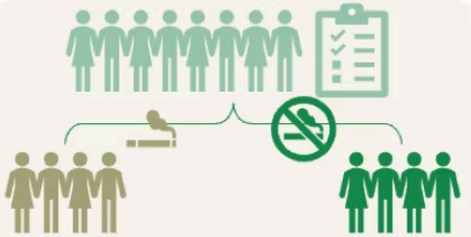
Example Pathway

Genetic Variants → Smoking (Exposure) → Lifespan (Outcome)

Using genetic variants associated with smoking behavior as instruments to study the causal effect on lifespan.

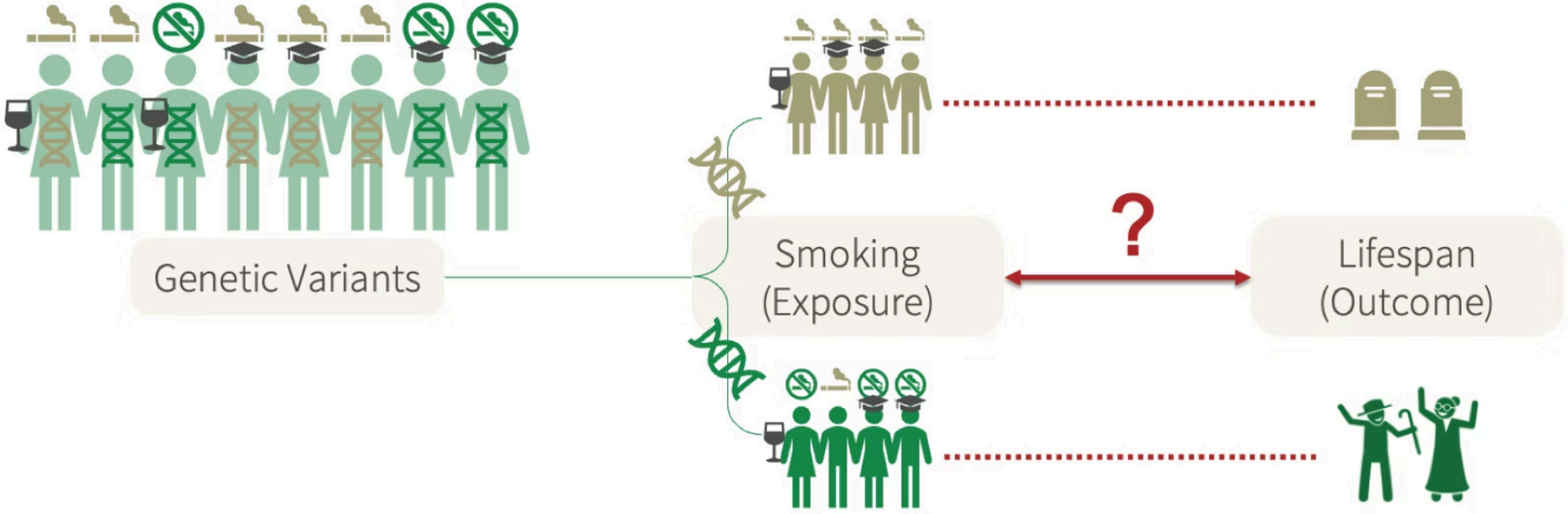
A hypothetical experiment could aim to investigate the **causal effect of smoking on lifespan**.

Smokers tend to have other factors associated with lifespan, which may act as **confounding factors**.



A randomized controlled trial (RCT) **assigning smoking and non-smoking** status would account for confounders but would be **difficult** and **unethical**.

Alternatively, **genetic variants associated with smoking** can be used for causal inference.



Why does this work?

1. **Mendel's Law of Independent Assortment:** Alleles for different genes are passed from parents to offspring independently of one another.

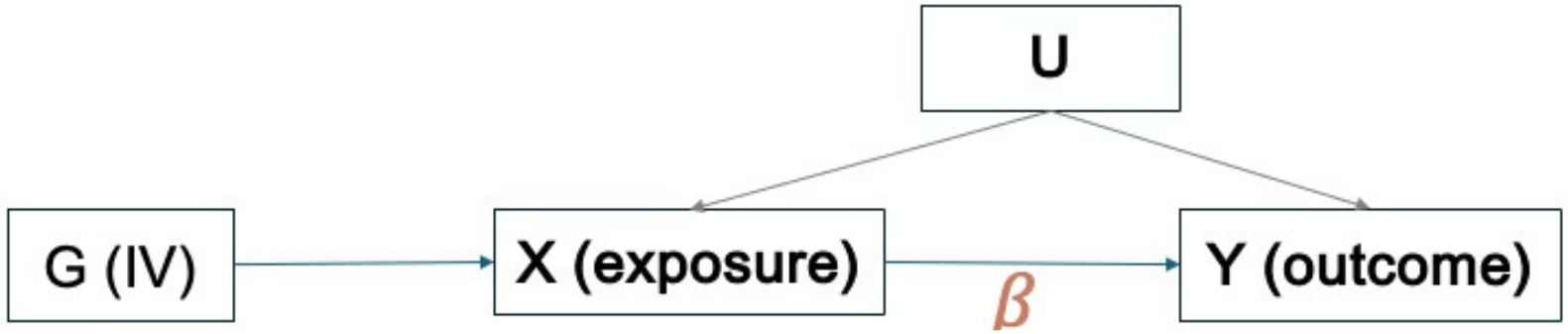
2. This means the set of genes you get is **randomly allocated** at conception.

3. It's a **"Natural Lottery"** that happens before birth, creating naturally randomized groups.

4. This process mimics the active randomization in an RCT.

20.07.25 | ISMB2025

Instrumental Variable (IV) Analysis



Genetic Instrument

Mendelian Randomization is a specific application of Instrumental Variable analysis.

Clean Proxy

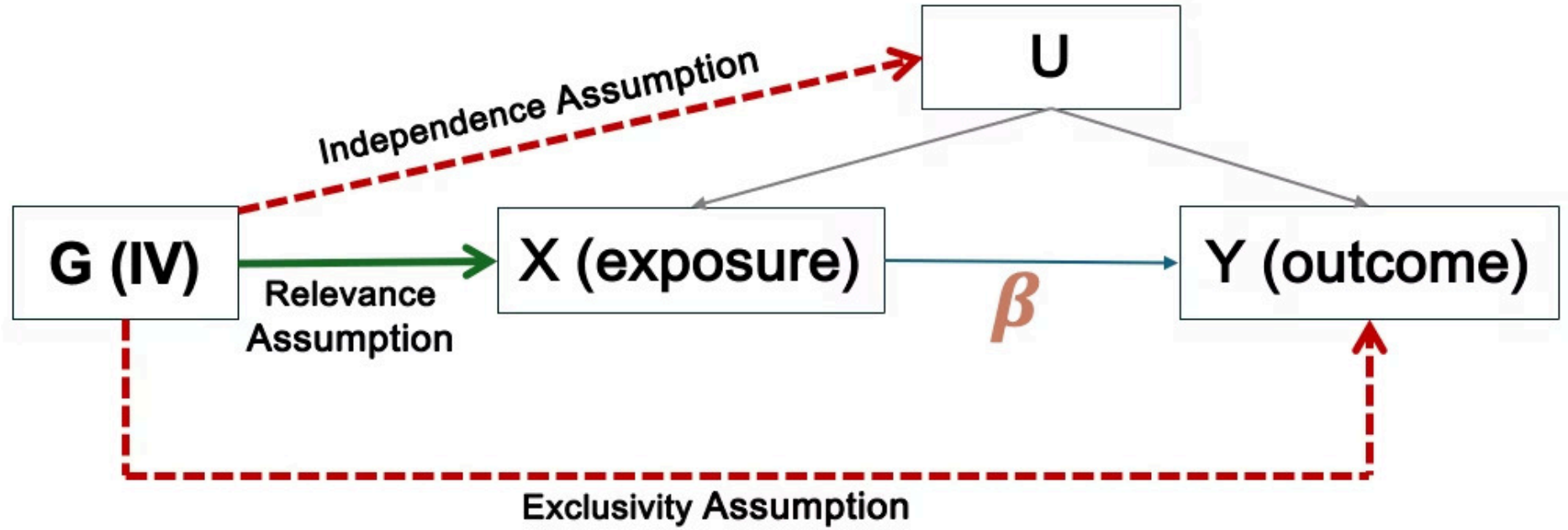
We use a genetic variant (G) as an "instrument" — a clean proxy for the modifiable exposure (X).

Causal Testing

We use this instrument to test the causal effect of X on the outcome Y, bypassing confounders (U).

20.07.25 | ISMB2025

Mendelian Randomization Assumptions



Relevance Assumption

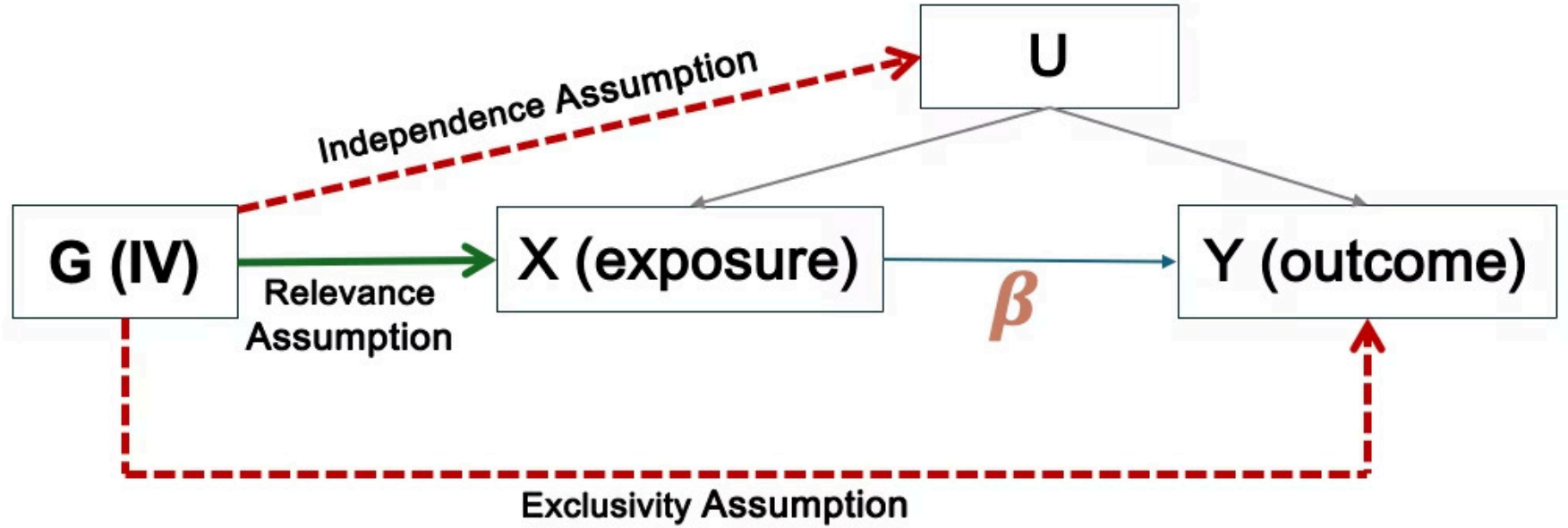
Genetic variants must strongly associate with the exposure.

Independence Assumption

Genetic variants must not be related to confounders.

Exclusion Restriction

Genetic variants must affect outcome only through the exposure.



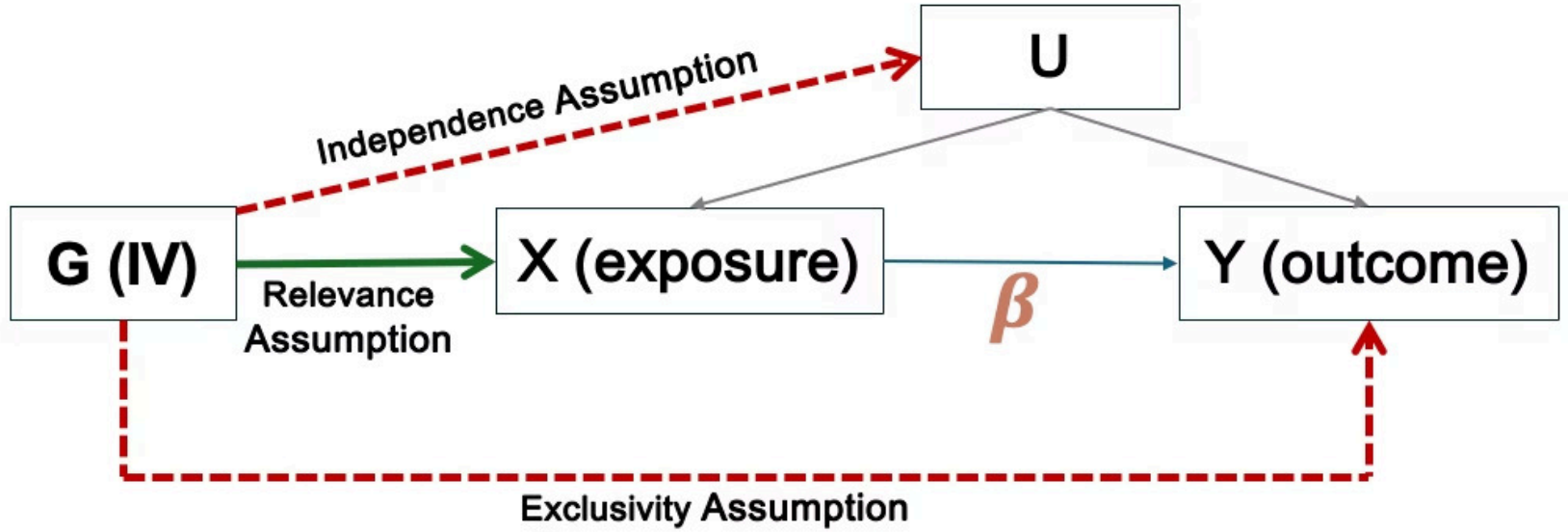
Assumption 1: Relevance

1 The instrument (G) must be **strongly and reliably associated** with the exposure (X).

2 **How we check:** We select SNPs with a genome-wide significant p-value ($p<5\times10-8$).

3 We then quantify their strength using the **F-statistic**. Conventionally, **F > 10** indicates a sufficiently strong instrument.

4 **Why it matters:** A weak instrument can **bias the MR estimate back towards the confounded observational estimate**.



Assumption 2: Independence

The Assumption

The instrument (G) must be **independent of all confounders** (U) of the exposure-outcome relationship.

Why It's Plausible

This is the "natural RCT" assumption. Your genotype is determined at conception and is fixed.

It cannot be influenced by later-life choices (smoking, diet) or circumstances (socioeconomic status) that act as confounders.

Pitfall: Population Stratification

The Problem

This assumption can be violated in genetically diverse populations.

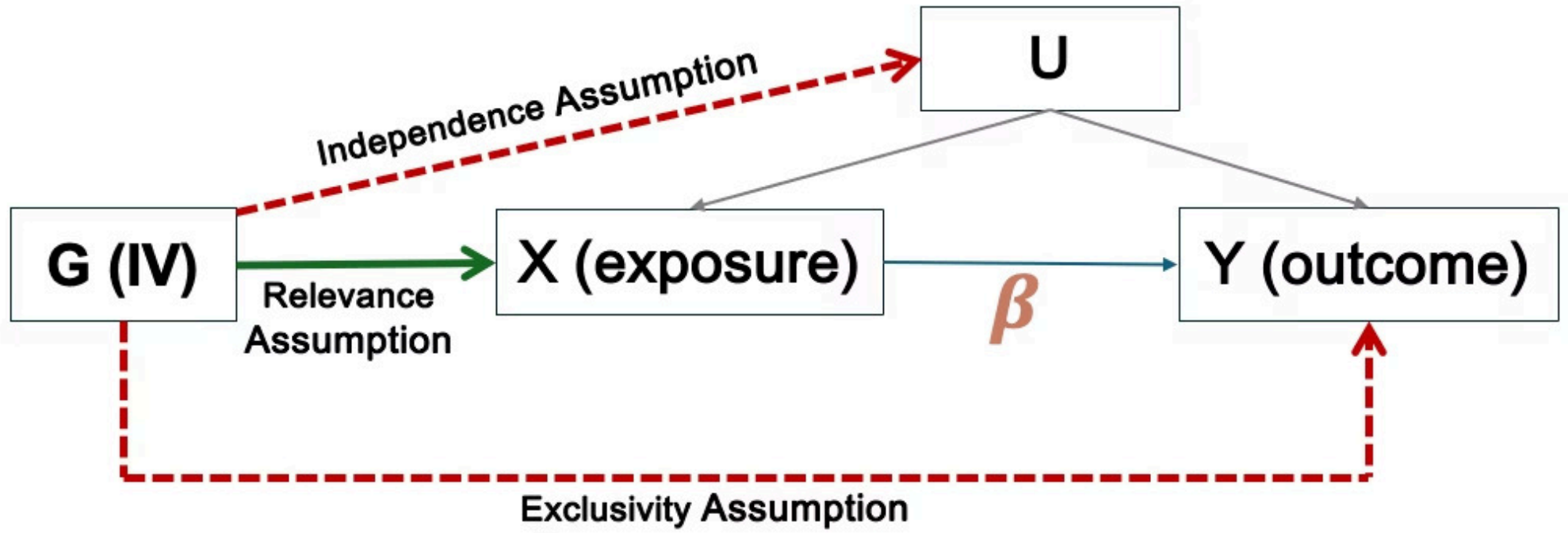
Allele frequencies AND outcome risk can both differ systematically by ancestry.

The Result

A spurious, non-causal association between the gene and the outcome driven entirely by ancestry.

Solutions

- Restrict analysis to a single, homogeneous ancestral group
- Statistically adjust for genetic ancestry using principal components analysis (PCA)



Assumption 3: Exclusion Restriction

1. The genetic instrument (G) can affect the outcome (Y) **ONLY through its effect on the exposure of interest (X)**.
2. This is an assumption of **no alternative causal pathways**.
3. This is the most challenging assumption and cannot be definitively proven, only tested for violations.

Pitfall: Horizontal Pleiotropy

Pleiotropy: The phenomenon of one gene affecting multiple, distinct traits.

Vertical Pleiotropy (Acceptable)

The gene affects traits that are *on the causal pathway*.

Example: A gene for BMI (X) influences blood pressure (Z), which is a step in how BMI causes a stroke (Y).

Horizontal Pleiotropy (Problem)

The gene affects the outcome through a *separate, independent pathway*. This violates the exclusion restriction.

Example: A gene for alcohol intake (X) also influences smoking behavior (Z), which affects lung cancer risk (Y).

GWAS Summary Statistics

1. A hypothesis-free approach that scans millions of SNPs across the entire genome.

2. Involves massive international consortia (e.g., UK Biobank, GIANT, CARDIoGRAM) with millions of participants.

3. **Goal:** Pinpoint which specific SNPs are statistically associated with a trait of interest.

4. The publicly available summary statistics (e.g. in GWAS Catalog) from these studies are the raw material for MR.

Study Designs in MR

One Sample MR

- Exposure and outcome measured in the same individuals
- Risk of overfitting
- Weak instrument bias

Two Sample MR

- Exposure and outcome GWAS conducted in separate cohorts
- Avoids within-sample bias
- Requires careful alignment (harmonization) of SNP data

Why Use Multiple Instruments?

1. Increased Statistical Power

The power to detect a causal effect scales with the amount of variance in the exposure explained by the instruments.

More instruments = more variance explained = more power.

2. Assumption-Checking

Multiple instruments unlock a whole toolkit of sensitivity analyses to formally test for violations of the MR assumptions, especially pleiotropy.

LD & Clumping

The Problem

Genes located close together on a chromosome are often co-inherited in correlated "blocks." This is called Linkage Disequilibrium (LD), measured by r^2.

Correlated SNPs are not independent pieces of evidence. Including them all would be like pseudo-replication.

The Solution: Clumping

This is a crucial pre-analysis step. In each genetic region, we keep only the single most significant SNP (the "lead SNP") and prune away the rest that are in high LD (e.g., $r^2 > 0.001$) with it.

20.07.25 | ISMB2025

Inverse-Variance Weighted (IVW)

1. The default and most statistically powerful method for combining multiple instruments.

2. It's essentially a meta-analysis that combines the individual Wald Ratios from all SNPs.

3. Each SNP's contribution is "weighted" by its precision (the inverse of its variance for the outcome association).

4. SNPs with stronger, more certain effects on the outcome get more "say" in the final combined estimate.

The IVW Assumption

1. IVW's power comes at a cost. It relies on a crucial assumption: any horizontal pleiotropy present must be **"balanced."**
2. This means the pleiotropic "side effects" are random and their net effect on the outcome averages out to zero across all instruments.
3. This is a strong, untestable assumption.
4. **This is why sensitivity analyses are not optional, they are mandatory.**

20.07.25 | ISMB2025

The Sensitivity Toolkit

Multiple Methods Approach

Principle: We never trust just one number. A robust MR study is a story told by multiple methods.

Triangulation

We use a toolkit of methods, each with different underlying assumptions about pleiotropy, to check for bias and robustness.

Interpretation

Agreement = Confidence.

Disagreement = Red Flag for Pleiotropy.

20.07.25 | ISMB2025

Test 1: Heterogeneity (Cochran's Q)

1. **The Question:** Are the causal estimates from each of our individual SNPs more scattered or different from one another than we would expect by random chance alone?

2. **The Null Hypothesis:** All SNPs are estimating the same true causal effect.

3. A significant Q-test (low p-value) rejects this null, indicating high heterogeneity.

4. This is a **major red flag for pleiotropy or other model violations.**

Test 2: MR-Egger Regression

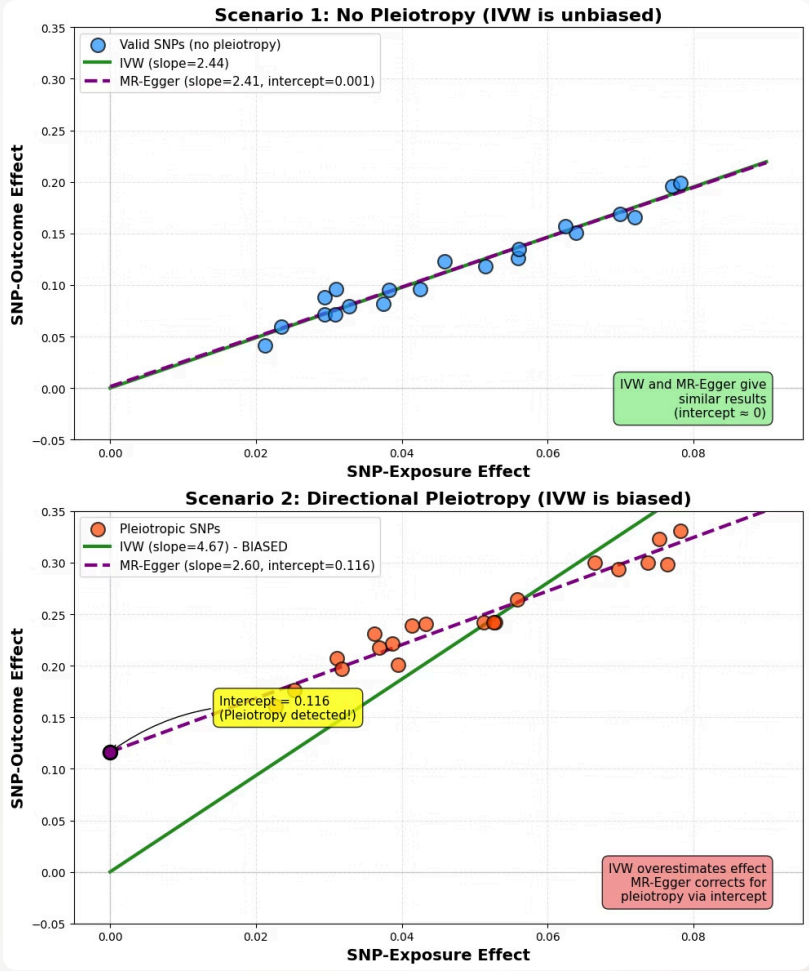
1. A method that formally tests for **directional pleiotropy**.

2. Unlike IVW (which forces its line through the origin), MR-Egger fits a regression line and allows for an **intercept**.

3. **The Intercept Test:** If the intercept is statistically different from zero, it provides evidence of unbalanced, directional pleiotropy biasing the result.

4. **Its own assumption:** Relies on the "InSIDE" assumption (Instrument Strength is Independent of the Direct Effect).

20.07.25 | ISMB2025



Test 3: Weighted Median

1. A robust method that works on a different principle.

2. It calculates the causal estimate for every SNP, weights them, and then simply takes the **median** of this weighted distribution.

3. **Key strength:** The median is resistant to outliers.

4. This method provides a valid estimate as long as **at least 50% of the *weight* in the analysis comes from valid, non-pleiotropic instruments.**

A Real-World MR Example

(Case study in the Hands-on Session)

The Question

Is the observational link between the inflammatory marker CRP and CHD causal?

The MR Finding

The Conclusion

The MR Publication Explosion

Publication Trend

The accessibility of GWAS data has led to an explosion in published MR studies.

Quality Concerns

However, the quality, rigor, and interpretation of these studies can vary widely.

Knowing how to be a critical consumer of this literature is an essential skill for any modern scientist or clinician.

20.07.25 | ISMB2025

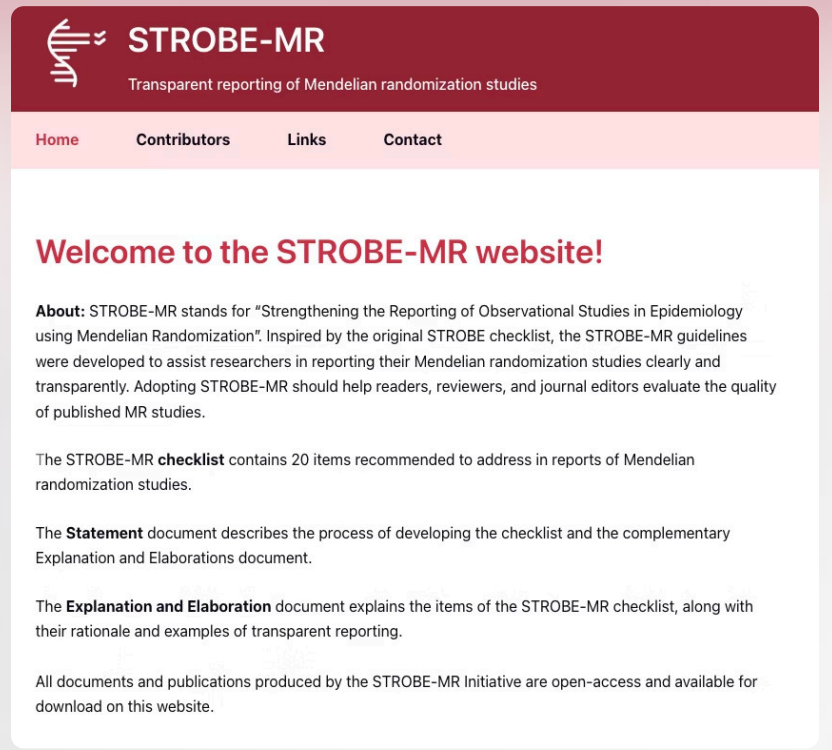
The STROBE-MR Checklist

1. **Instruments:** Are they strong (F>10)? Independent (clumped appropriately)? Are the GWAS sources large, well-powered, and of an appropriate ancestry?

2. **Assumptions:** Did the authors treat pleiotropy as a serious threat? Look for the full "constellation of evidence" (IVW, MR-Egger, Weighted Median).

3. **Interpretation:** Do they acknowledge limitations? Is the causal language appropriately cautious, or do they overstate their claims as "proof"?

20.07.25 | ISMB2025



Gene-Environment Equivalence: Understanding Lifelong Exposure

1 Lifelong Exposure

Genetic variants act as natural experiments in dose and time, providing insight into lifelong exposure effects from birth.

2 Biological Adaptation

The body can adapt over a lifetime to chronic exposures (e.g., genetic predisposition to high cholesterol leading to compensatory clearance mechanisms).

3 Cumulative Impact

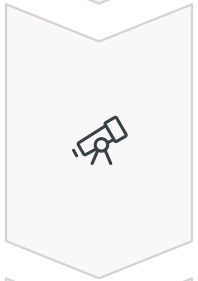
Mendelian Randomization estimates reflect the integrated effect of an exposure over an entire lifespan, not just short-term or acute effects.

The MR effect size may not perfectly match the effect of a short-term drug intervention started in adulthood.

Molecular Data & Drug Discovery

New Frontier

Moving beyond traditional risk factors to molecular data at a massive scale.



Omics MR

Proteomics-MR and **Metabolomics-MR** use genetic instruments for thousands of circulating proteins and metabolites.

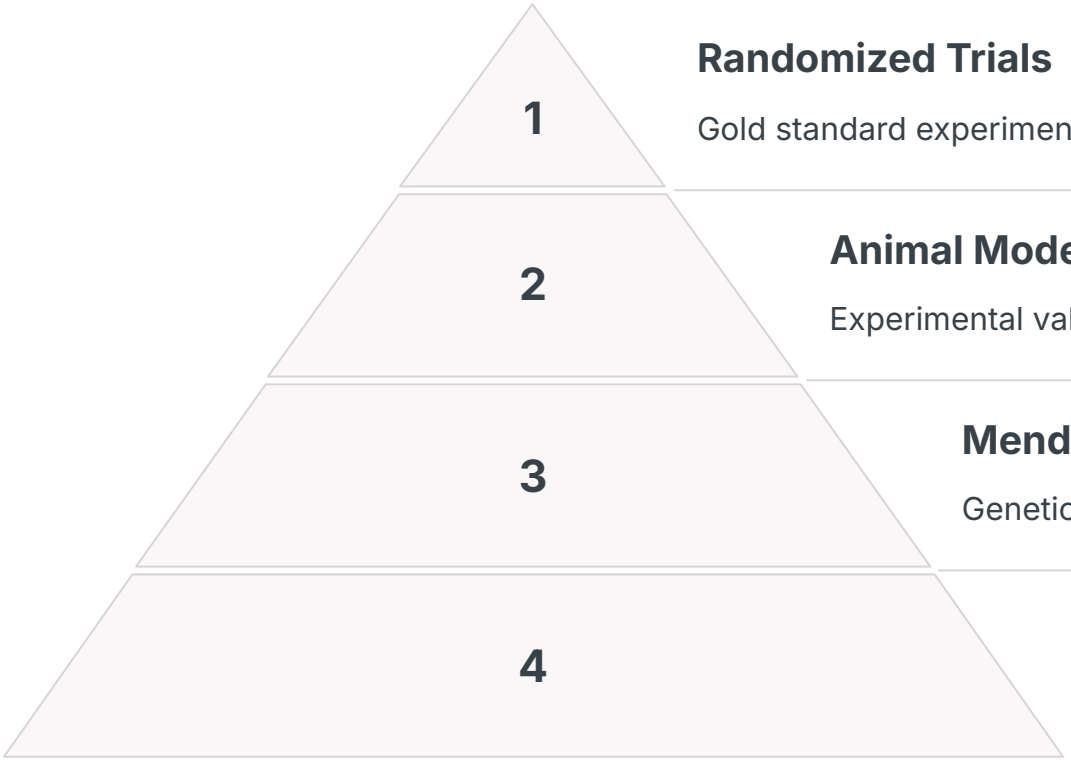
Drug Targets

This is **revolutionizing drug discovery** by directly identifying which specific proteins are causal for disease.

The validation of PCSK9 for LDL and CHD is a prime example of this success.

20.07.25 | ISMB2025

Triangulation of Evidence



Randomized Trials

Gold standard experimental evidence

Animal Models

Experimental validation in controlled systems

Mendelian Randomization

Genetic approaches to causality

Observational Studies

Traditional epidemiological approaches

MR is not definitive proof. The strongest possible causal inference comes from **triangulation**: combining evidence from multiple study designs with different key sources of bias.

Summary - Limitations of MR

1. Violations of assumptions → misleading results
2. Prone to false negatives – weak instruments → attenuation toward the null
3. Prone to false positives – pleiotropy, measurement errors, poor harmonization introduces bias
4. Other concerns: population stratification, sample overlap

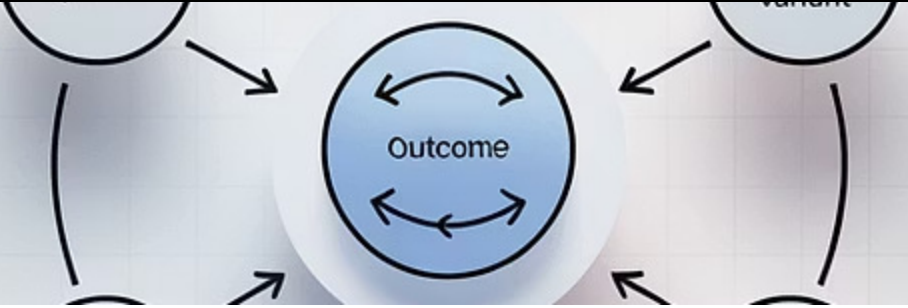
MR is not a replacement for randomized trials or functional validation.

Triangulation is key: integrate evidence across study designs.

20.07.25 | ISMB2025

Randomization:

A Literature Review



References & Further Learning

Key Papers

- Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003
- Burgess S, et al. Guidelines for performing Mendelian randomization investigations. Wellcome Open Res. 2020

Software Tools

- TwoSampleMR (R package)
- MendelianRandomization (R package)
- MR-Base web application

Online Resources

- MR Dictionary (University of Bristol) (**https://mr-dictionary.mrcieu.ac.uk/**)
- STROBE-MR guidelines (**https://www.strobe-mr.org/**)

20.07.25 | ISMB2025

Questions & Discussion

Contact Information

Tayyaba Alvi: tayyaba.alvi@leibniz-fli.de

Mark Olenik: mark.olenik@leibniz-fli.de

Handan Melike Dönertas: melike.donertas@leibniz-fli.de

Resources

All slides, code, and datasets from today's tutorial are available at:

donertas-group.github.io/ismb2025_mr_tutorial

Thank you for your attention! We welcome your questions.

20.07.25 | ISMB2025

Share your Feedback!

Scan the QR code to let us know your thoughts on this tutorial. Thank you!

Other Tools

MR-PRESSO

An outlier-robust method that formally identifies which specific SNPs are pleiotropic outliers, removes them, and provides a corrected causal estimate.

Leave-One-Out Plots

A crucial visualization where you re-calculate the main IVW result repeatedly, each time leaving one instrument out.

This helps you visually inspect if your entire conclusion is being driven by a single, highly influential SNP.

Two-Step MR (Mediation)

Purpose

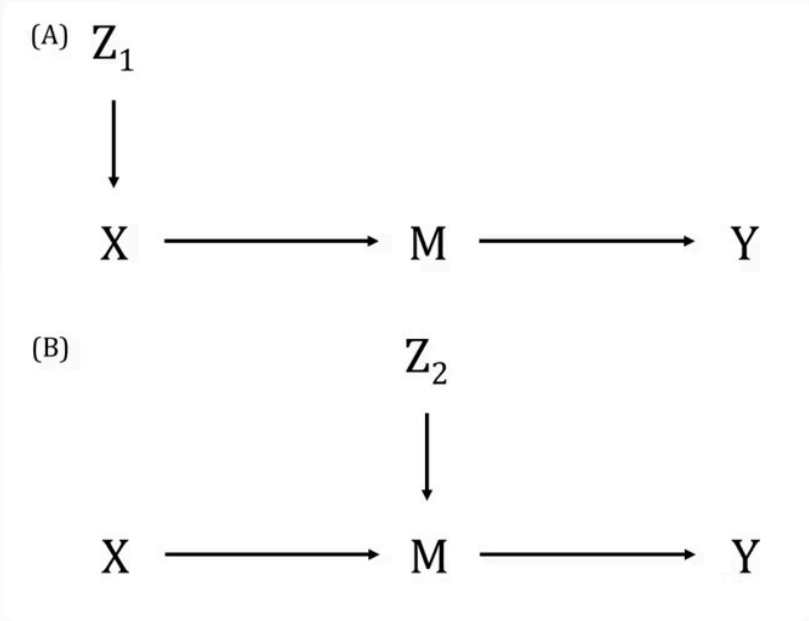
Used to dissect causal pathways and test for **mediation**.

Step 1

Use genes for the IL-6 receptor to test the causal effect on CRP.

Step 2

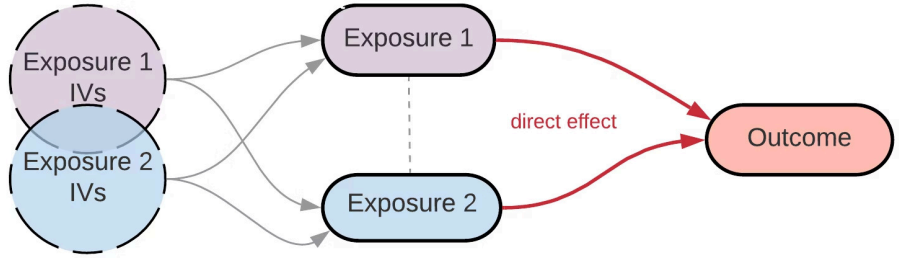
Use the same genes to test the causal effect on CHD.



This approach validated the IL-6 pathway as causal for heart disease, identifying it as a drug target.

20.07.25 | ISMB2025

Multivariable MR (MVMR)



1. Used when multiple exposures are closely correlated and may confound each other.

2. **Classic Example:** LDL cholesterol, HDL cholesterol, and Triglycerides.

3. MVMR simultaneously includes instruments for all exposures in one model.

4. This allows it to estimate the **independent, direct causal effect** of each one, effectively adjusting for the genetic effects of the others.