Bayesian Statistics for Genetics

Lecture 7: Mendelian Randomization

July, 2023

In this session

- What is Mendelian randomization (MR)?
- Non-Bayesian and Bayesian methods in MR

Survey:

- Please raise your hand if you have heard about MR before.
- please raise your hand if you have heard about instrumental variables before.

MR: an epidemiological approach to infer causality

Primer | Published: 10 February 2022

Mendelian randomization

Eleanor Sanderson, M. Maria Glymour, Michael V. Holmes, Hyunseung Kang, Jean Morrison, Marcus R. Munafò, Tom Palmer, C. Mary Schooling, Chris Wallace, Qingyuan Zhao & George Davey Smith

Nature Reviews Methods Primers **2**, Article number: 6 (2022) | Cite this article

Abstract

Mendelian randomization (MR) is a term that applies to the use of genetic variation to address causal questions about how modifiable exposures influence different outcomes. The principles of MR are based on Mendel's laws of inheritance and instrumental variable estimation methods, which enable the inference of causal effects in the presence of unobserved confounding. In this Primer, we outline the principles of MR, the instrumental

The basis of MR (Davey Smith and Ebrahim, 2003)

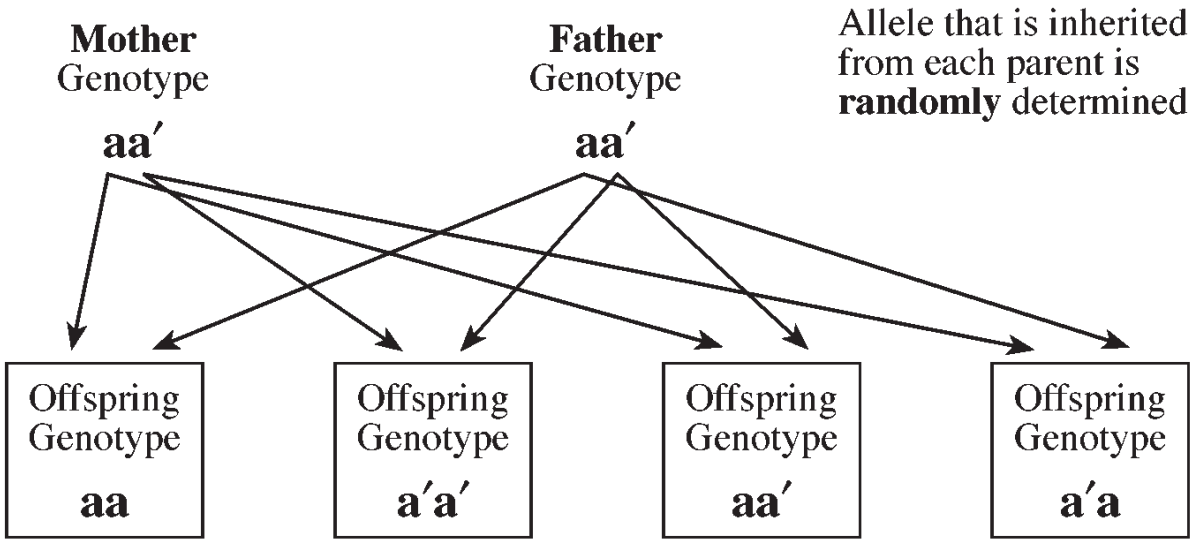


Figure 3 Mendelian randomization in parent–offspring design

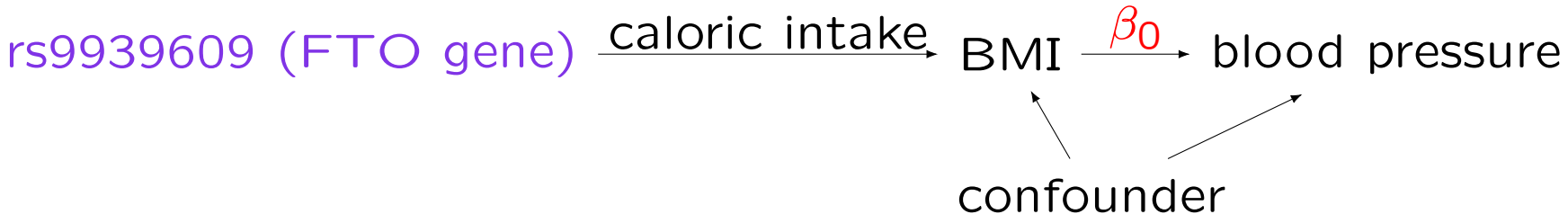
Offspring should have an equal chance of receiving either of the alleles that the parents have at any particular locus

MR in non-familial studies (Davey Smith and Ebrahim, 2003)

> Of course populations share much common ancestry and the genetic make-up of individuals can be traced back through the random segregation of alleles during a sequence of matings, but associating genetic markers with disease risk or phenotype within such populations is not as well protected against potential distorting factors as are parent–offspring comparisons. Thus the Mendelian randomization in genetic association studies is approximate, rather than absolute.

Example: causal effect of BMI on blood pressure

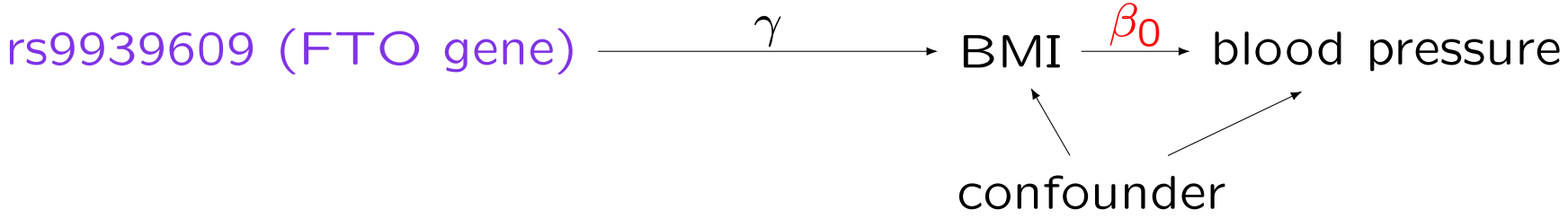
MR: People who inherited certain alleles have naturally higher BMI.



A hypothetical RCT emulated by MR

Example: causal effect of BMI on blood pressure

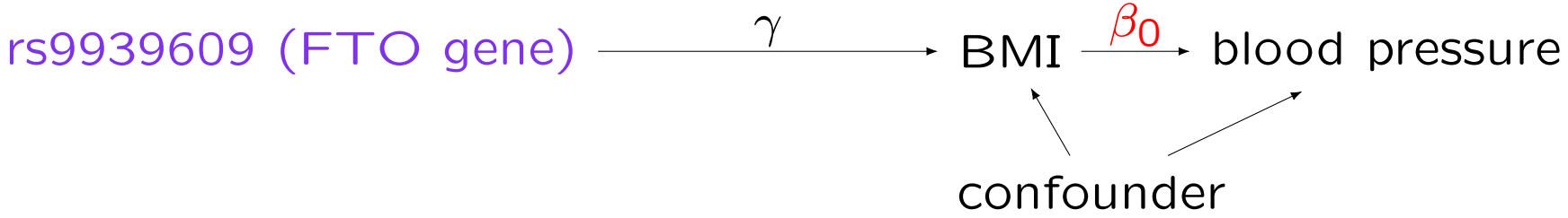
MR: People who inherited certain alleles have naturally higher BMI.



- SNP $\uparrow 1$ unit $\Rightarrow$ BMI $\uparrow \gamma$ units
- BMI $\uparrow 1$ unit $\Rightarrow$ blood pressure $\uparrow \beta_0$ units
- SNP $\uparrow 1$ unit $\Rightarrow$ blood pressure $\uparrow$? units

Example: causal effect of BMI on blood pressure

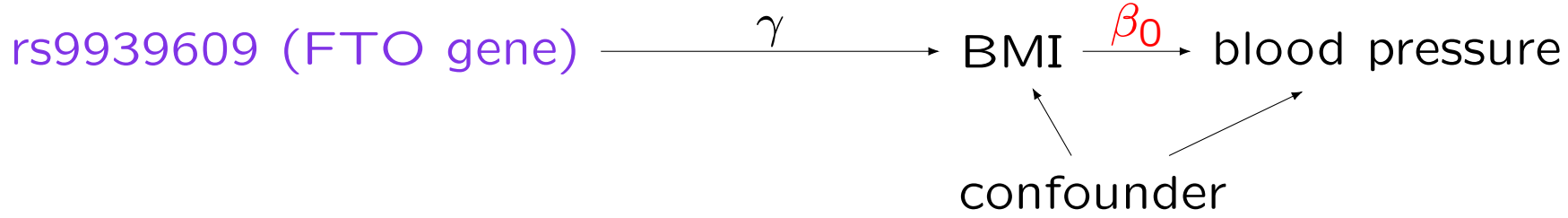
MR: People who inherited certain alleles have naturally higher BMI.



- SNP $\uparrow 1$ unit $\Rightarrow$ BMI $\uparrow \gamma$ units
- BMI $\uparrow 1$ unit $\Rightarrow$ blood pressure $\uparrow \beta_0$ units
- SNP $\uparrow 1$ unit $\Rightarrow$ blood pressure $\uparrow \gamma \cdot \beta_0$ units

Example: causal effect of BMI on blood pressure

MR: People who inherited certain alleles have naturally higher BMI.

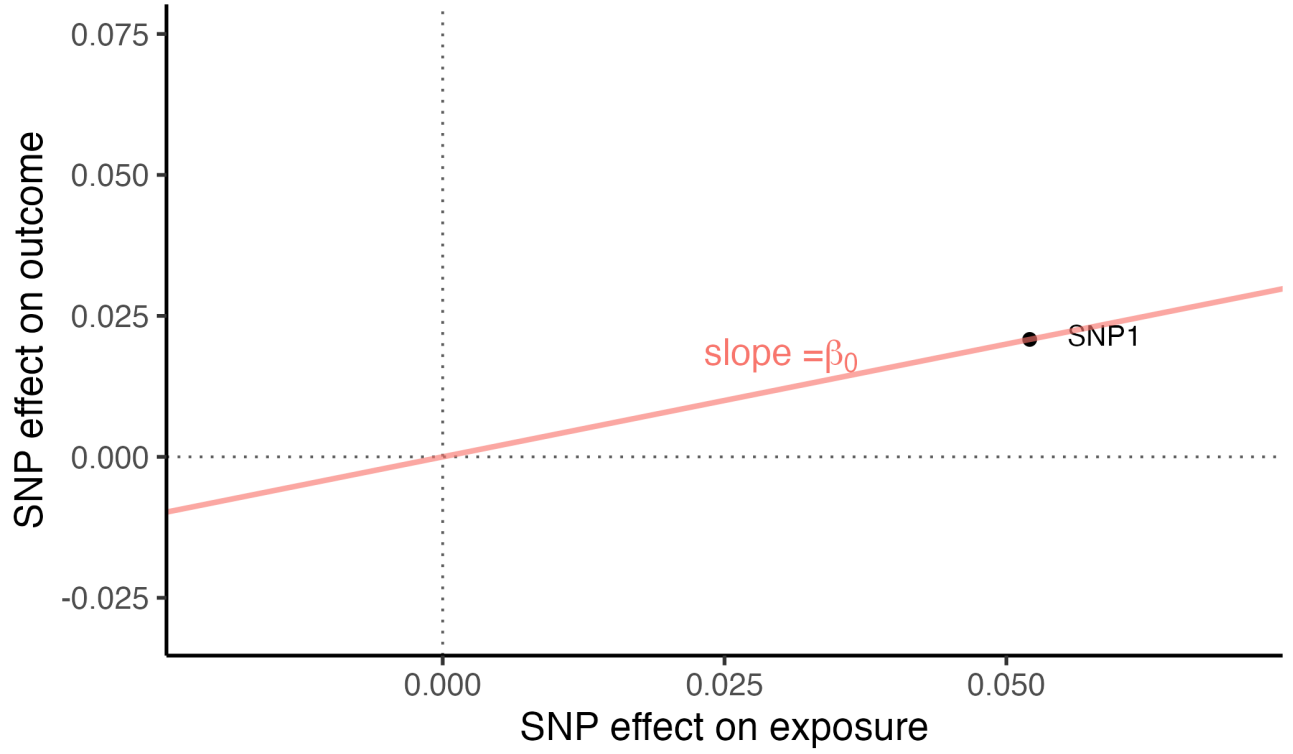


- SNP $\uparrow 1$ unit $\Rightarrow$ BMI $\uparrow \gamma$ units
- BMI $\uparrow 1$ unit $\Rightarrow$ blood pressure $\uparrow \beta_0$ units
- SNP $\uparrow 1$ unit $\Rightarrow$ blood pressure $\uparrow \gamma \cdot \beta_0$ units

Hence, a **linear relationship**: SNP's effect on blood pressure($\gamma \cdot \beta_0$)
= SNP's effect on BMI(γ) $\times$ Causal effect of BMI on blood pressure(β_0).

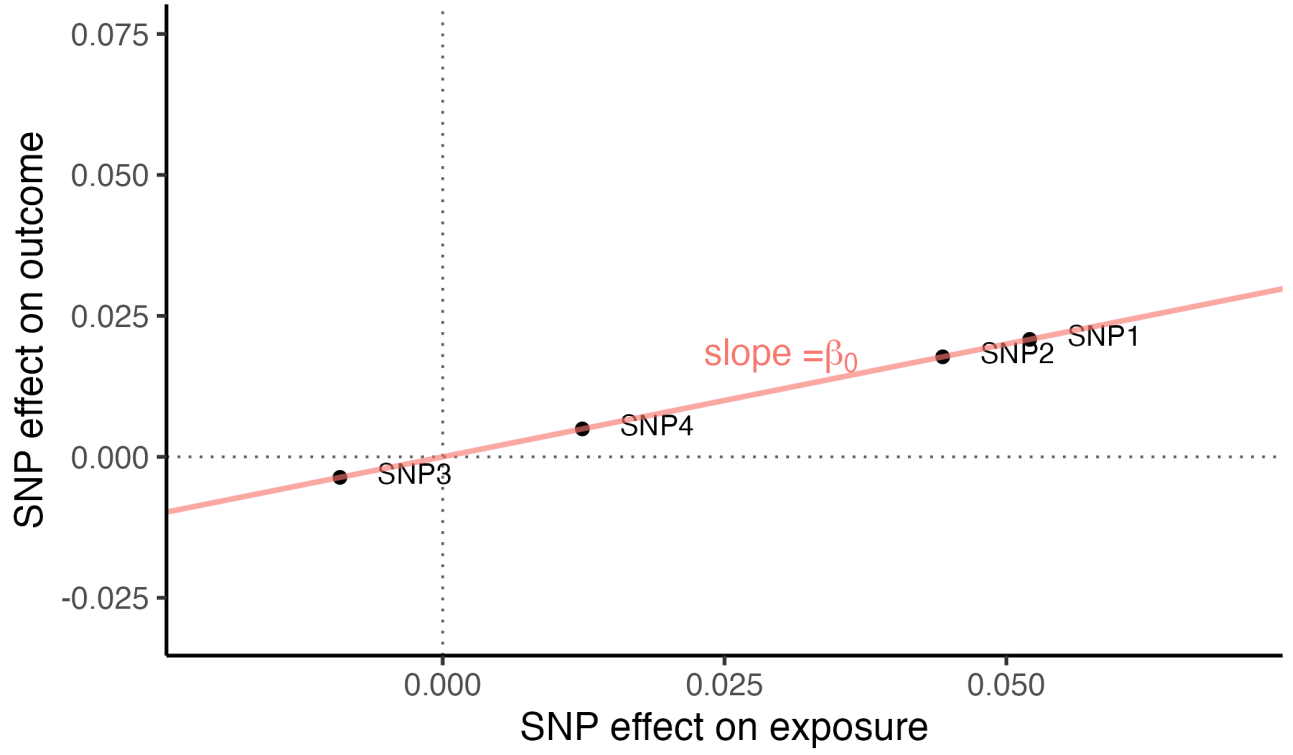
Example: causal effect of BMI on blood pressure

Key: a simple linear relationship that passes (0,0)



Example: causal effect of BMI on blood pressure

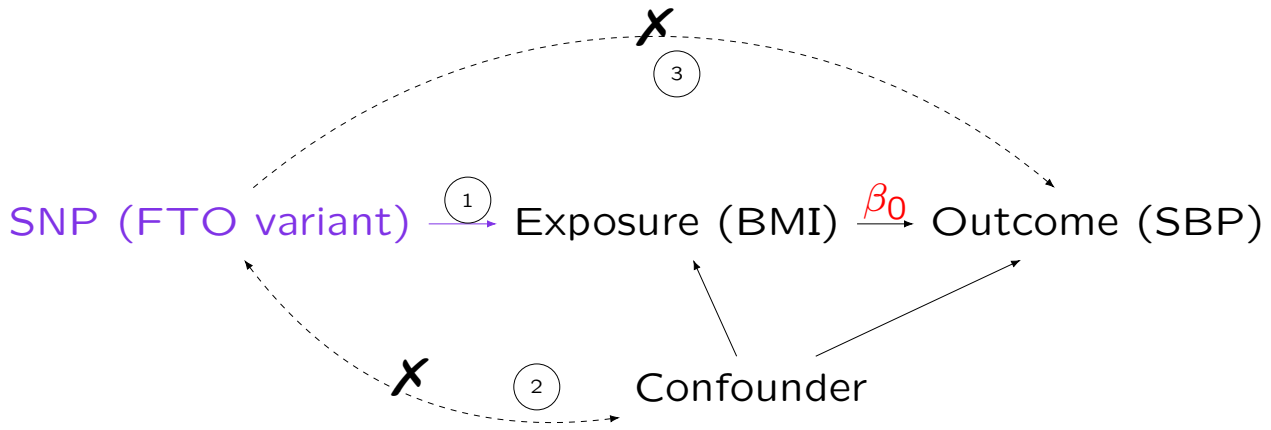
Key: a simple linear relationship that passes (0,0)



The principle of MR

MR uses single-nucleotide polymorphisms (SNPs) as instrumental variables (IVs), if they satisfy three core IV assumptions:

1. Relevance: the SNP is related with the exposure
2. Randomness: the SNP is independent of the confounder
3. Exclusion restriction: the SNP has no direct effect on the outcome (i.e. no horizontal pleiotropy)



Two-sample summary-data MR

Combine publically available summary data* on **SNP-exposure** and **SNP-outcome** associations from **two separate samples**.

Example: estimate the effect of BMI on SBP using 160 independent SNPs.

- **Exposure dataset**: A GWAS for BMI by the GIANT consortium, $\text{lm}(X \sim Z_j) \Rightarrow \hat{\gamma}_j, \sigma_{Xj},\ j = 1, \ldots, p.$
- **Outcome dataset**: A GWAS for SBP in the UK BioBank, $\text{lm}(Y \sim Z_j) \Rightarrow \hat{\Gamma}_j, \sigma_{Yj},\ j = 1, \ldots, p.$
- Assume that $\hat{\gamma}_j \sim N(\gamma_j, \sigma^2_{Xj})$ and $\hat{\Gamma}_j \sim N(\Gamma_j, \sigma^2_{Yj})$, where $\Gamma_j = \gamma_j \beta_0$ for $j = 1, \ldots, p$, and $\sigma^2_{Xj}, \sigma^2_{Yj}$'s are known. Here, p can be very large (e.g., thousands).

*Many GWAS summary data are publically available at, e.g., IEU GWAS database and GWAS Catalog.

The IVW estimator in MR

- In practice, we can fit a weighted linear regression (with the intercept at 0)

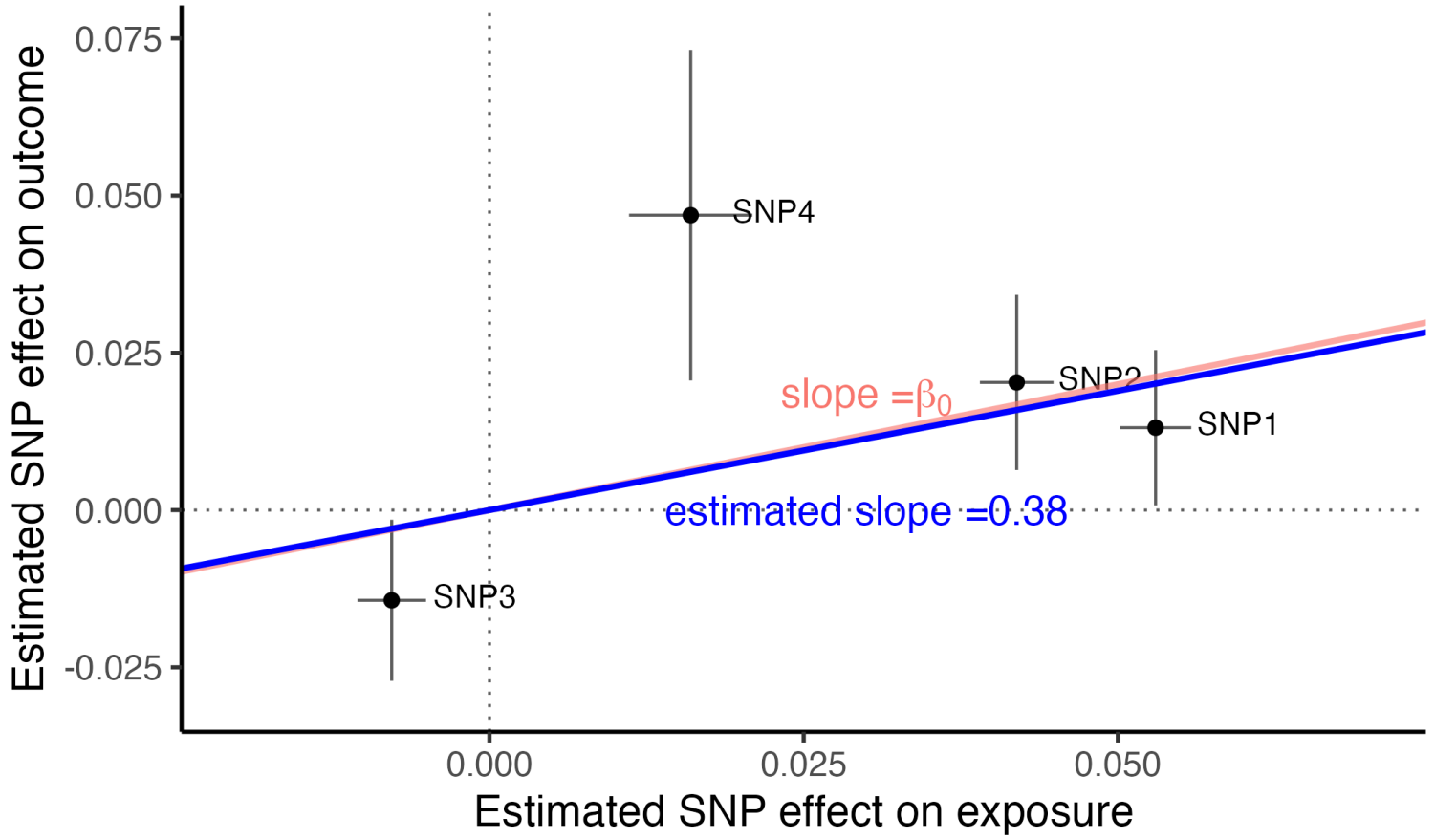
```
> lm(beta.outcome~0+beta.exposure,weights = 1/se.outcome^2,data=df_summary)
# fix the intercept at 0
# the estimated slope is the estimated causal effect

Coefficients:
beta.exposure
         0.38
```

- We can fit linear regression as $(\widehat{\gamma}_j, \widehat{\Gamma}_j, j = 1, \ldots, p)$ are approximately mutually independent if SNPs are independent (i.e., not in linkage disequilibrium).
- This is the popular inverse-variance weighted (IVW) estimator, which is a meta-analysis (Session 8) of $\widehat{\Gamma}_j/\widehat{\gamma}_j$ across all SNPs.
- The IVW estimator can also be understood as a Beyasian estimator of the common ratio β_0 with a flat prior on β_0 (see Session 4), when ignoring the measurement error of $\widehat{\gamma}_j$'s.

The IVW estimator in MR

- However, because $\widehat{\gamma}_j$'s have measurement errors, directly using the standard error from the linear model output can underestimate the uncertainty. If γ_j's are close to zero (i.e., weak IVs), IVW will be biased toward zero (weak IV bias) (Ye, Shao, Kang, 2021).
- What happens when $\alpha_j := \Gamma_j - \beta_0\gamma_j \neq 0$? If $\alpha_j \mid \gamma_j \sim_{iid} N(0, \tau^2)$ (a random effect), the IVW estimator is still unbiased but has larger variability. Otherwise, the IVW can be biased.
- We can check heterogeneity across all SNPs (i.e. all SNPs do not agree on a common causal effect) using Cochran's Q statistic; see more in Session 8.

The IVW estimator in MR



Likelihood

Recall that we assume $\hat{\gamma}_j \sim N(\gamma_j, \sigma^2_{Xj})$ and $\hat{\Gamma}_j \sim N(\Gamma_j, \sigma^2_{Yj})$, for $j = 1, \ldots, p$, which are mutually independent, and $\sigma^2_{Xj}, \sigma^2_{Yj}$'s are known.

This gives the likelihood

$$
\begin{aligned}
& p(\hat{\Gamma}_j, \hat{\gamma}_j, j = 1, \ldots, p \mid \Gamma_j, \gamma_j, \sigma^2_{Xj}, \sigma^2_{Yj}, j = 1, \ldots, p) \\
&= \prod_{j=1}^{p} p(\hat{\gamma}_j \mid \gamma_j, \sigma^2_{Xj}) p(\hat{\Gamma}_j \mid \Gamma_j, \sigma^2_{Yj}) \\
&\propto \exp\left(- \sum_{j=1}^{p} \frac{(\hat{\gamma}_j - \gamma_j)^2}{2\sigma^2_{Xj}} - \sum_{j=1}^{p} \frac{(\hat{\Gamma}_j - \Gamma_j)^2}{2\sigma^2_{Yj}} \right).
\end{aligned}
$$

The causal effect of interest β_0 is defined by the relationship of γ_j, Γ_j.

Causal assumptions and some choices of priors

- **Valid IVs**[†]: $\Gamma_j = \gamma_j \beta_0$. Priors: $\gamma_j \sim_{iid} N(0, \psi^2)$ and $\beta_0 \sim N(0, \nu^2)$.
- **Balanced horizontal pleiotropy**: $\Gamma_j = \gamma_j \beta_0 + \alpha_j$. Priors: $\gamma_j \sim_{iid} N(0, \psi^2)$ and $\beta_0 \sim N(0, \nu^2)$, and $\alpha_j \sim_{iid} N(0, \tau^2)$.

... for some ψ^2, ν^2, τ^2 which may in turn have hyperpriors (More in Session 8).

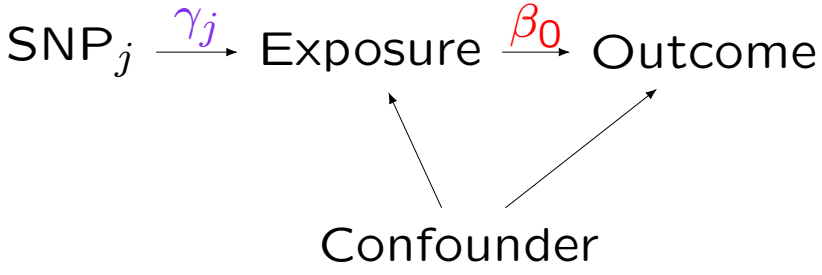


Fig 1. Valid IVs

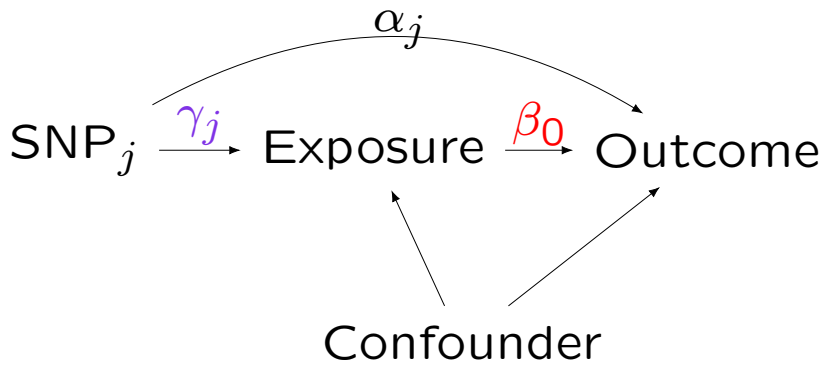


Fig 2. Balanced horizontal pleiotropy

[†]Another good choice of prior for γ_j is the spike-and-slab Gaussian mixture prior (Zhao et al. 2019)

Causal assumptions and some choices of priors

- **Uncorrelated pleiotropy or InSIDE**: $\Gamma_j = \gamma_j\beta_0 + \alpha_j$. Priors: $\gamma_j \sim_{iid} N(0, \psi^2)$ and $\beta_0 \sim N(0, \nu^2)$, $\alpha_j \sim_{iid} N(\alpha, \tau^2)$.
- **Correlated pleiotropy**[‡]: $\Gamma_j = \gamma_j\beta_0 + \gamma_j\eta + \alpha_j$. Priors: $\gamma_j \sim_{iid} N(0, \psi^2)$ and $\beta_0 \sim N(0, \nu^2)$, $\eta \sim N(0, \nu^2)$, $\alpha_j \sim_{iid} N(0, \tau^2)$.

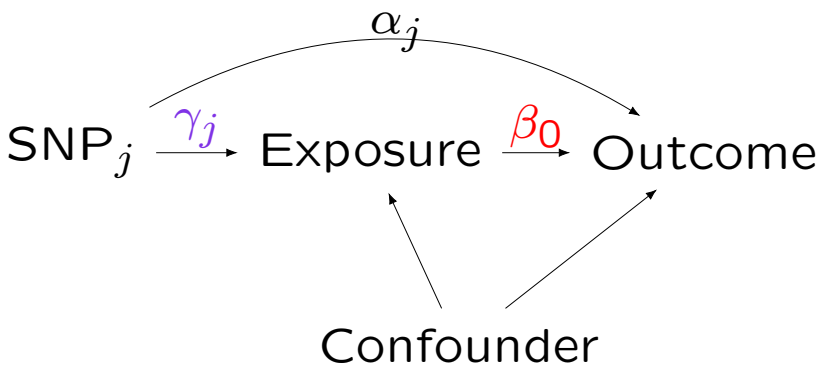


Fig 3. Uncorrelated pleiotropy

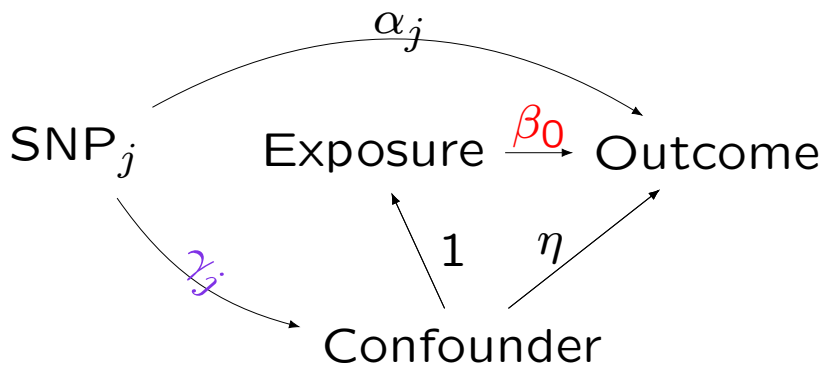


Fig 4. Correlated pleiotropy

[‡]In Morrison et al, 2020, they considered a mixture of different pleiotropy types.

Not so Bayes: Profile likelihood (*)

Previously, we use the full likelihood of the summary statistics, which involves a large number of parameters (including $\gamma_1, \ldots, \gamma_p$). However, γ_j's are nuisance parameters while β_0 is the parameter of interest.

We can profile out $\gamma_1, \ldots, \gamma_p$ and get the profile likelihood. Take the valid IV case (i.e. $\Gamma_j = \beta_0 \gamma_j$) as an example (Zhao et al. 2020). The profile likelihood is

$$\max_{\gamma_1,\ldots,\gamma_p} p(\hat{\Gamma}_j, \hat{\gamma}_j, j = 1, \ldots, p \mid \Gamma_j, \gamma_j, \sigma_{Xj}^2, \sigma_{Yj}^2, j = 1, \ldots, p)$$
$$\propto \exp\left(- \sum_{j=1}^{p} \frac{(\hat{\Gamma}_j - \beta_0 \hat{\gamma}_j)^2}{2(\sigma_{Yj}^2 + \beta_0^2 \sigma_{Xj}^2)} \right),$$

which only has one unknown parameter β_0!

- A non-Bayesian approach: directly maximize the profile likelihood.
- A Bayesian approach: put a prior on β_0, e.g. a normal prior with zero mean.

Strengths and challenges of MR

Strengths:

- Less susceptible to conventional unmeasured confounding (Mendel's laws)
- Less susceptible to reverse causation (genetics are fixed at conception)
- Has a summary-data and a two-sample option

Challenges:

- Weak IV bias
- Genetic-outcome confounding
- Widespread horizontal pleiotropy can cause bias (multi-functions genes)
- Low power
- Assumes constant treatment effect
- Based on gene-environment equivalence (Sanderson et al. 2022)
- Only applicable to heritable exposures

Summary

- MR leverages genetic variation to address causal questions
  - Emulates a RCT
  - Triangulation across multiple sources of evidence for causal inference
  - MR has strengths and challenges
- Non-Bayesian and Bayesian methods in MR
  - A very good tutorial here