

The Use of Polygenic Risk Scores in Mendelian Randomization Studies

Linbo Wang
University of Toronto



IGES PRS Workshop
Oct 12th, 2021

Overview

Classical PRS: Aim to learn the risk of developing a disease Y

$$PRS = \sum_{j=1}^J \hat{\beta}_j G_{ij}$$

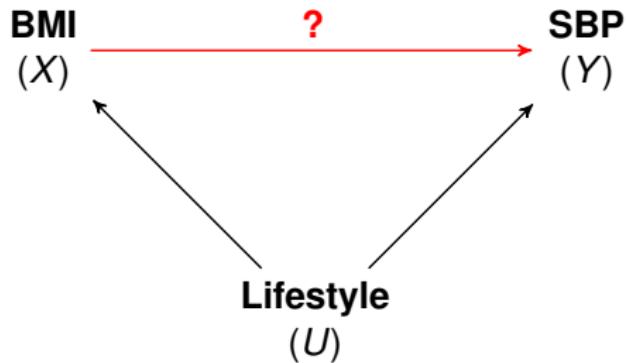
PRS in Mendelian Randomization: Aim to learn the causal effect of a risk factor X on an outcome (e.g. disease) Y

- Construct PRS for X
- NOT aiming for a good prediction of X

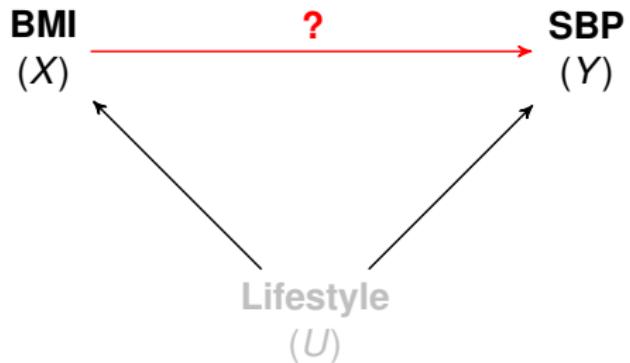
Running example



Running example

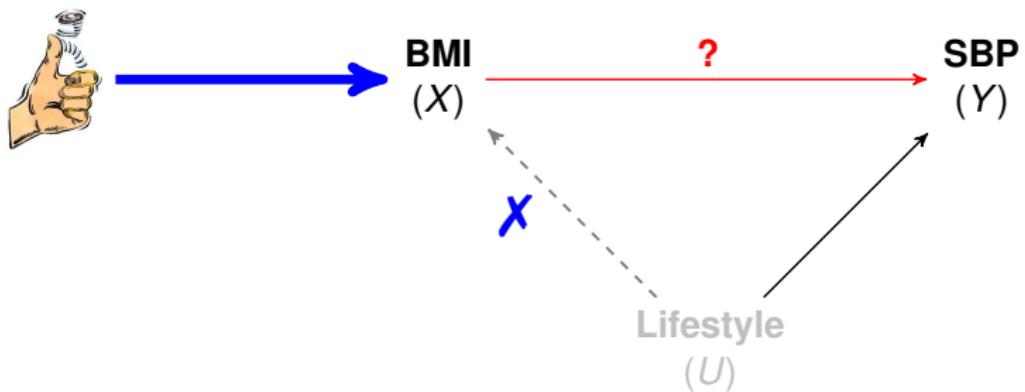


Running example

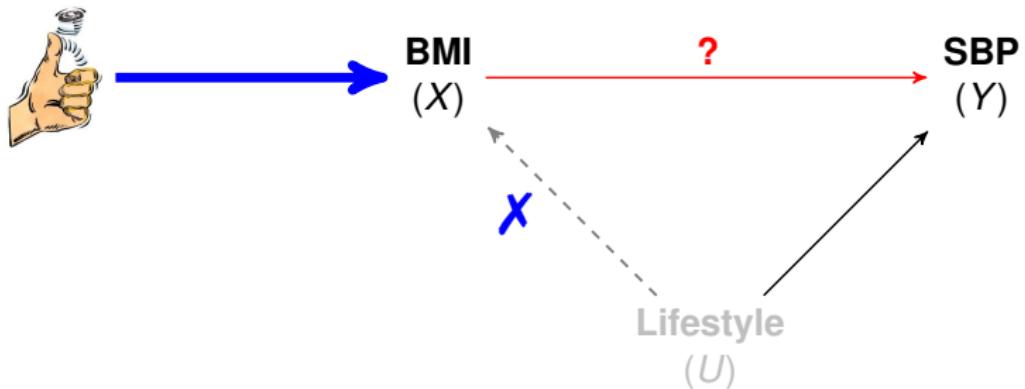


👉 Unmeasured confounding!

Randomized experiment

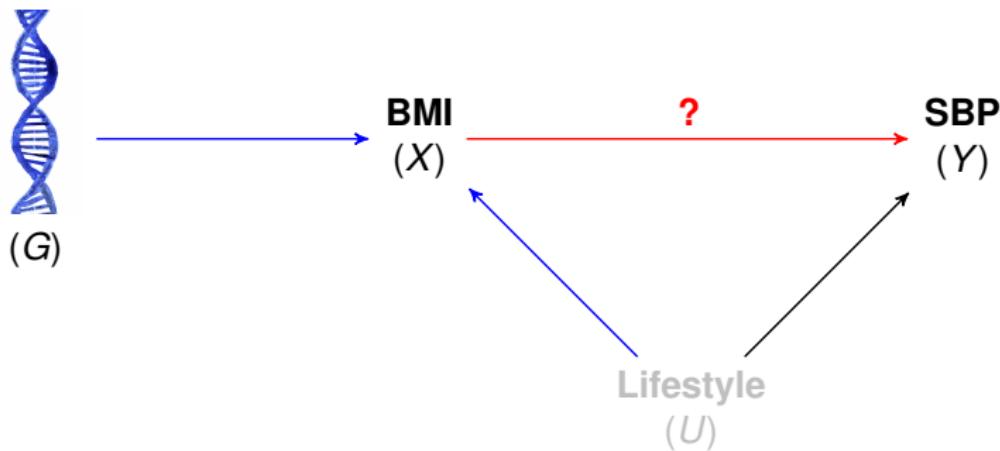


Randomized experiment



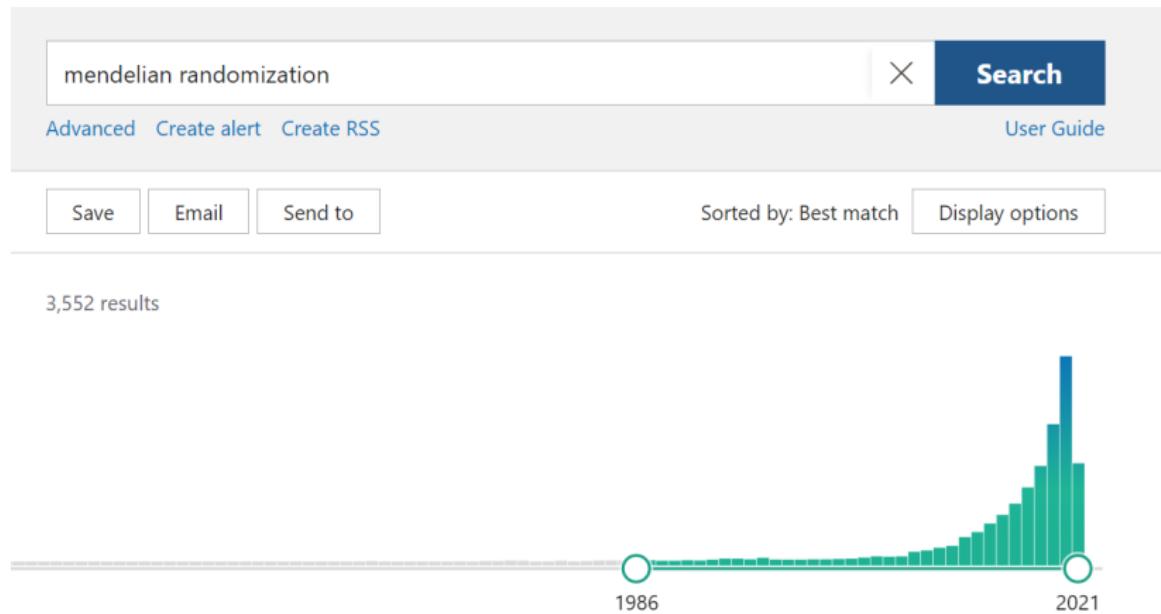
Not so easy...

Mendelian randomization (MR)



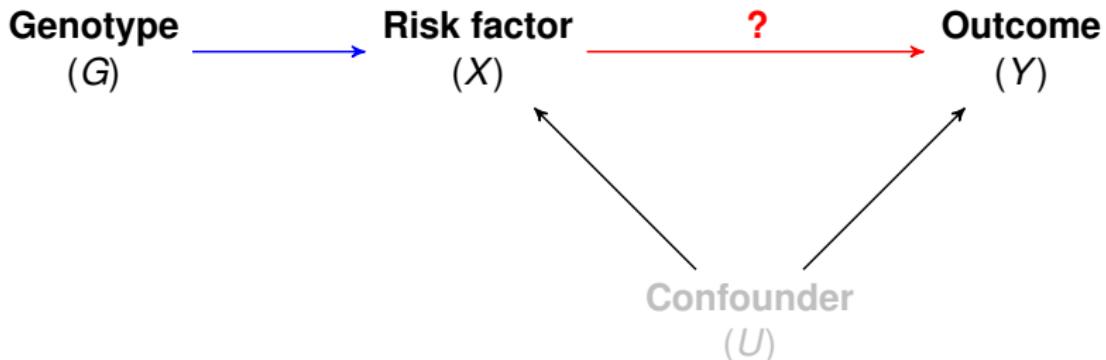
Nature rolls the dice!

History of Mendelian Randomization



Source: Pubmed

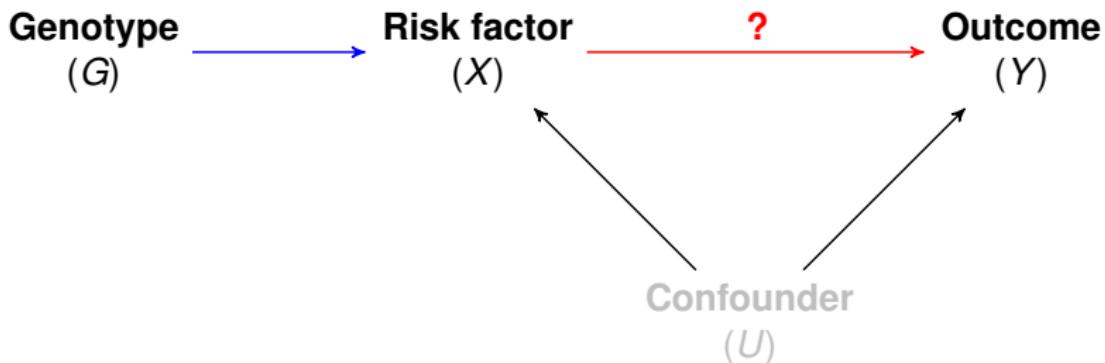
The math behind



Under some conditions,

$$G \xrightarrow{\text{blue}} X \xrightarrow{\text{red}} Y = G \xrightarrow{\text{blue}} X \times X \xrightarrow{\text{red}} Y$$

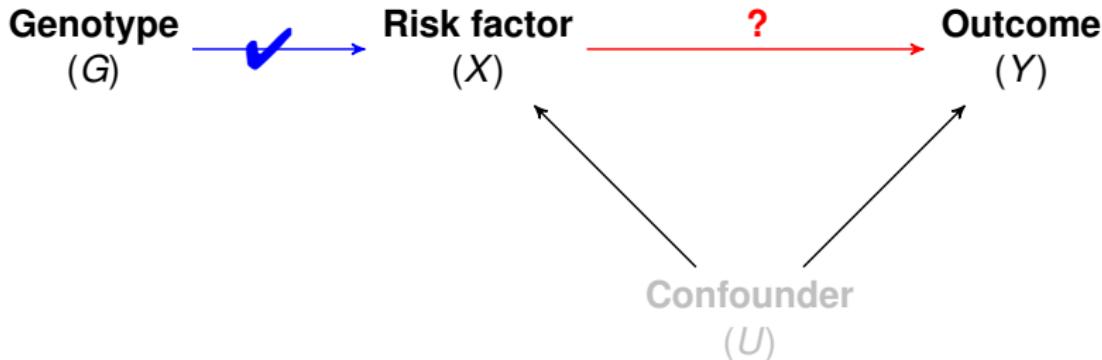
The math behind



Under some conditions,

$$X \rightarrow Y = \frac{G \xrightarrow{\text{blue}} X \xrightarrow{\text{red}} Y}{G \xrightarrow{\text{blue}} X}$$

The math behind

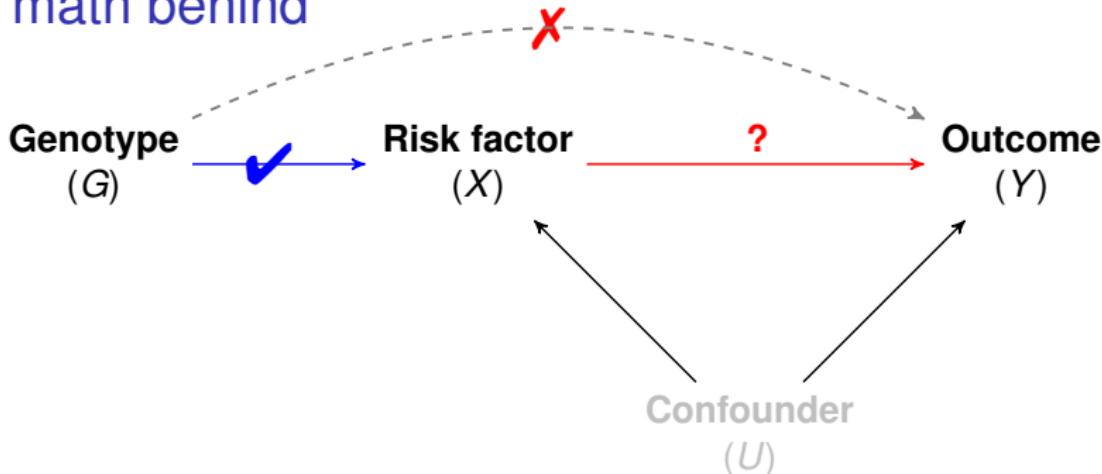


Under some conditions,

$$X \rightarrow Y = \frac{G \xrightarrow{\text{blue}} X \xrightarrow{\text{red}} Y}{G \xrightarrow{\text{blue}} X}$$

1. **(Relevant)** $G \rightarrow X \neq 0$

The math behind

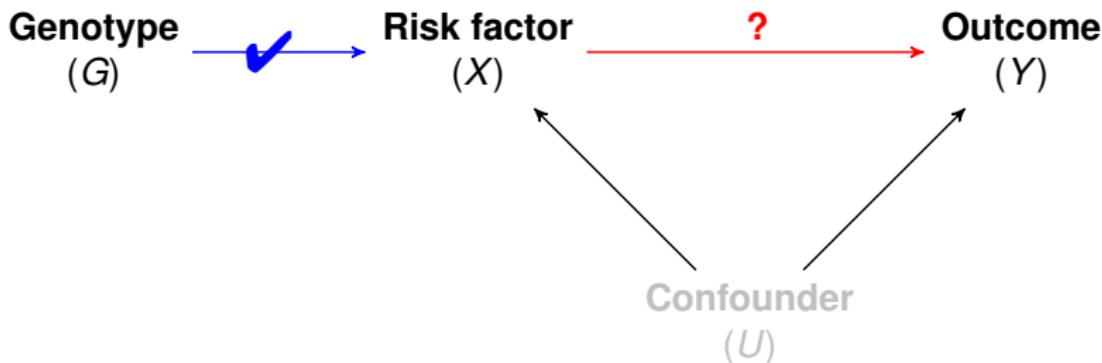


Under some conditions,

$$X \rightarrow Y = \frac{G \rightarrow X \rightarrow Y}{G \rightarrow X}$$

1. **(Relevant)** $G \rightarrow X \neq 0$
2. **(Valid)** $G \rightarrow Y = 0$

The math behind



Under some conditions,

$$X \rightarrow Y = \frac{G \rightarrow X \rightarrow Y}{G \rightarrow X}$$

1. **(Relevant)** $G \rightarrow X \neq 0$
2. **(Valid)** $G \rightarrow Y = 0$

The genotype G is known as an **instrumental variable (IV)**

Natural experiments are imperfect!

Problem: Most of the SNPs are weak IV

$$X \rightarrow Y = \frac{G \rightarrow X \rightarrow Y}{G \rightarrow X}$$

☞: Large variance and/or amplified bias

PRS come to rescue!

Two stage regression:

Stage I :

$$D \sim G$$

- $\hat{D} = \sum_{j=1}^J \hat{\beta}_j G_{ij}$ is a PRS

Stage II :

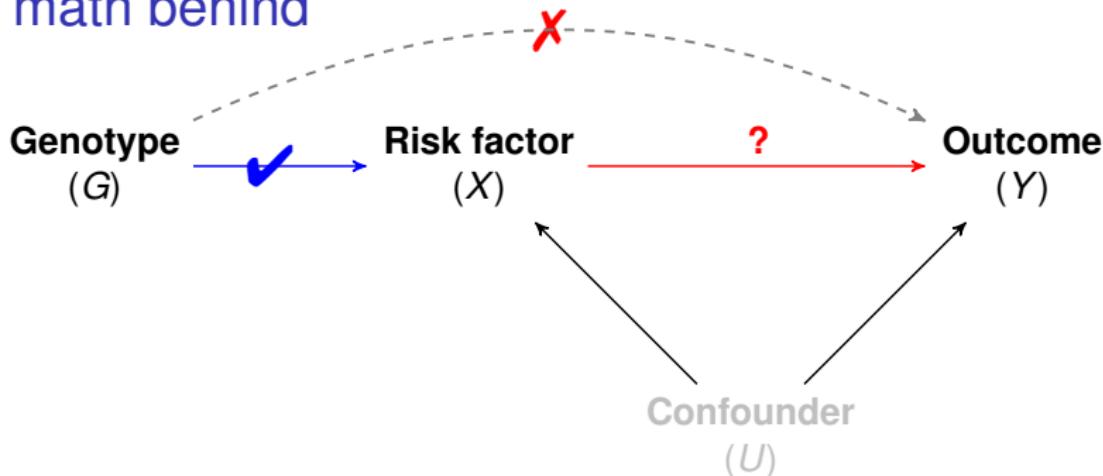
$$Y \sim \hat{D}$$

Insight: Combine the strength of multiple SNPs for a stronger IV

Challenges

1. What if some SNPs are invalid IVs (a.k.a. pleiotropy)?

The math behind



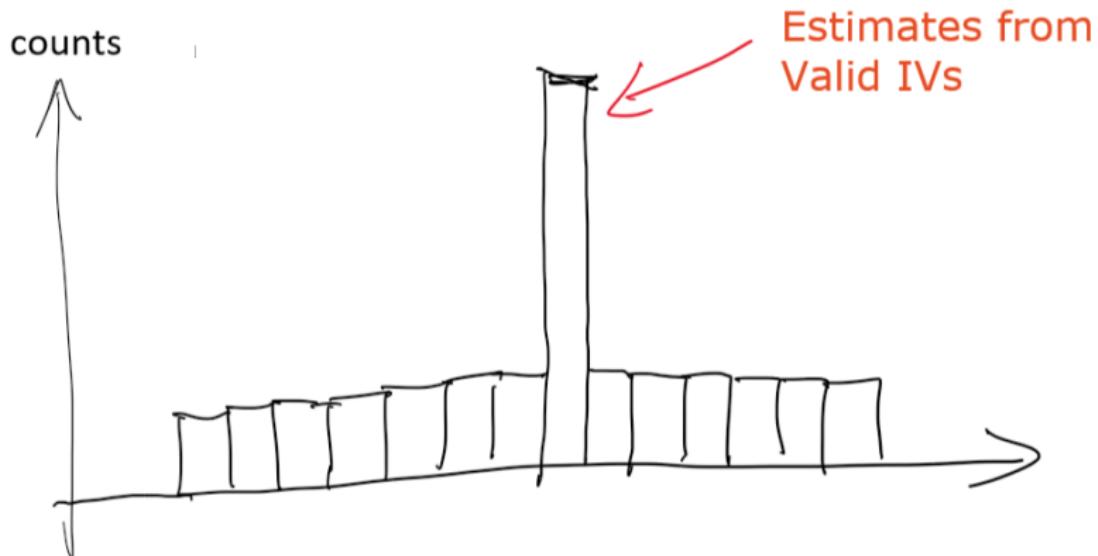
Under some conditions,

$$X \rightarrow Y = \frac{G \rightarrow X \rightarrow Y}{G \rightarrow X}$$

1. **(Relevant)** $G \rightarrow X \neq 0$
2. **(Valid)** $G \rightarrow Y = 0$

Valid causal inference with some invalid instruments

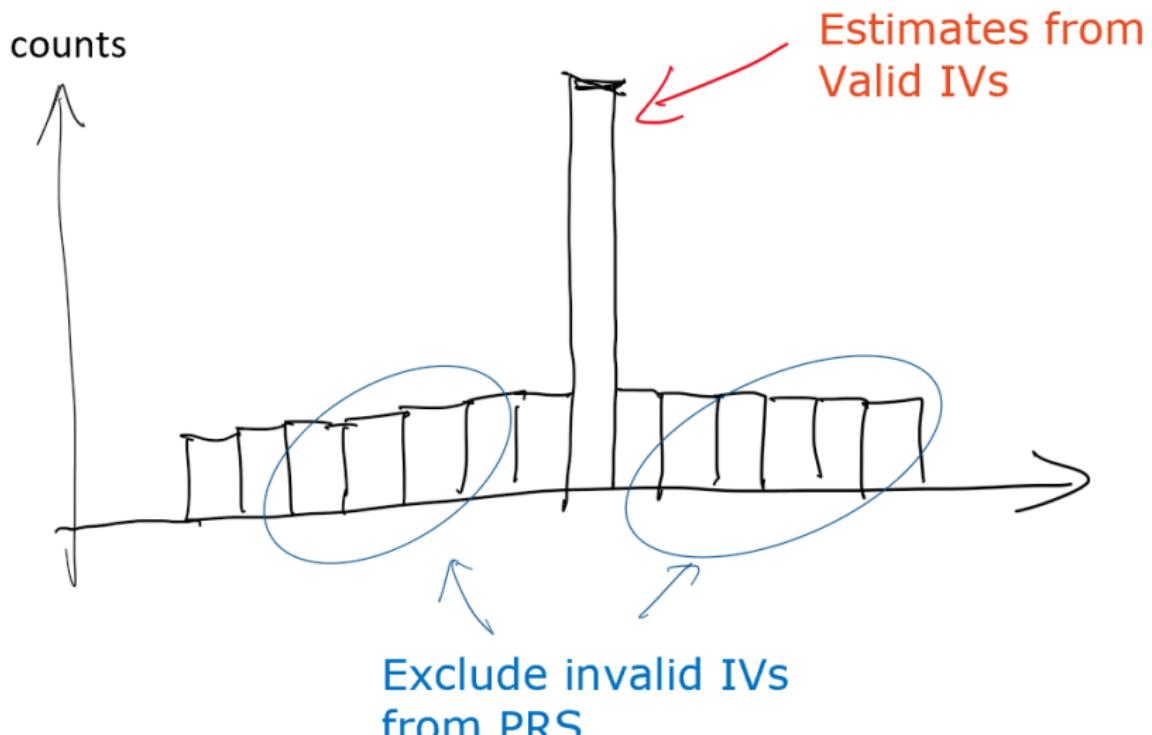
All happy instruments are alike; each unhappy instrument is unhappy in its own way



Histogram of $X \rightarrow Y$ effect estimates

Valid causal inference with some invalid instruments

All happy instruments are alike; each unhappy instrument is unhappy in its own way



Lesson: Exclude invalid IVs from the PRS even if they are predictive of the risk factor

- The best predictor of the risk factor may not be the best PRS for causal effect estimation!

Challenges

- 2 How do we find the candidate IVs in the first place?
 - Traditional wisdom: turn to experts

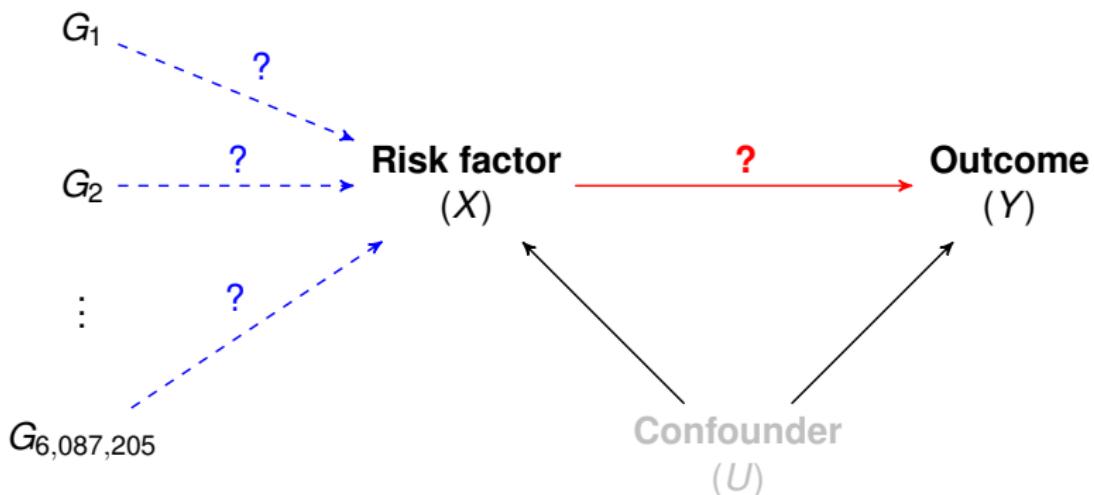
Challenges

- 2 How do we find the candidate IVs in the first place?
 - Traditional wisdom: turn to experts

Challenges

- 2 How do we find the candidate IVs in the first place?
 - Traditional wisdom: turn to experts

Modern solution (Step 1): GWAS



Challenges

- 2 How do we find the candidate IVs in the first place?
 - Traditional wisdom: turn to experts

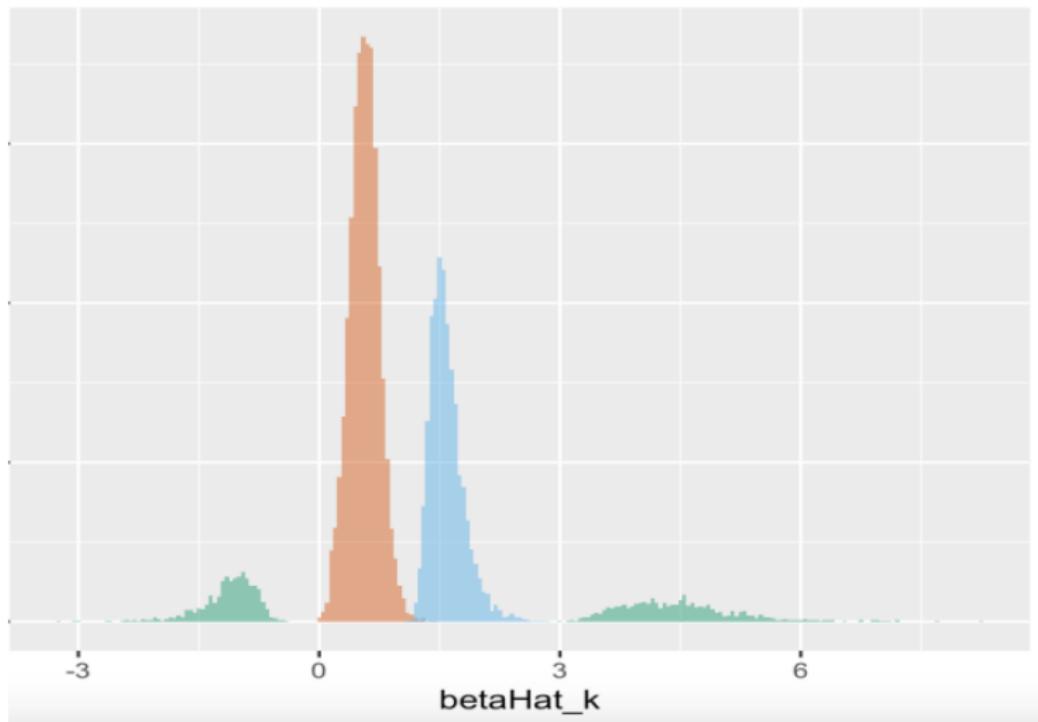
Modern solution (Step 2): Combine multiple genetic variants via polygenic risk score (PRS)

$$PRS = \sum_{j=1}^J \hat{\beta}_j G_j$$

- $\hat{\beta}_j$: estimated effect size of SNP j on X

More challenges: Selection bias

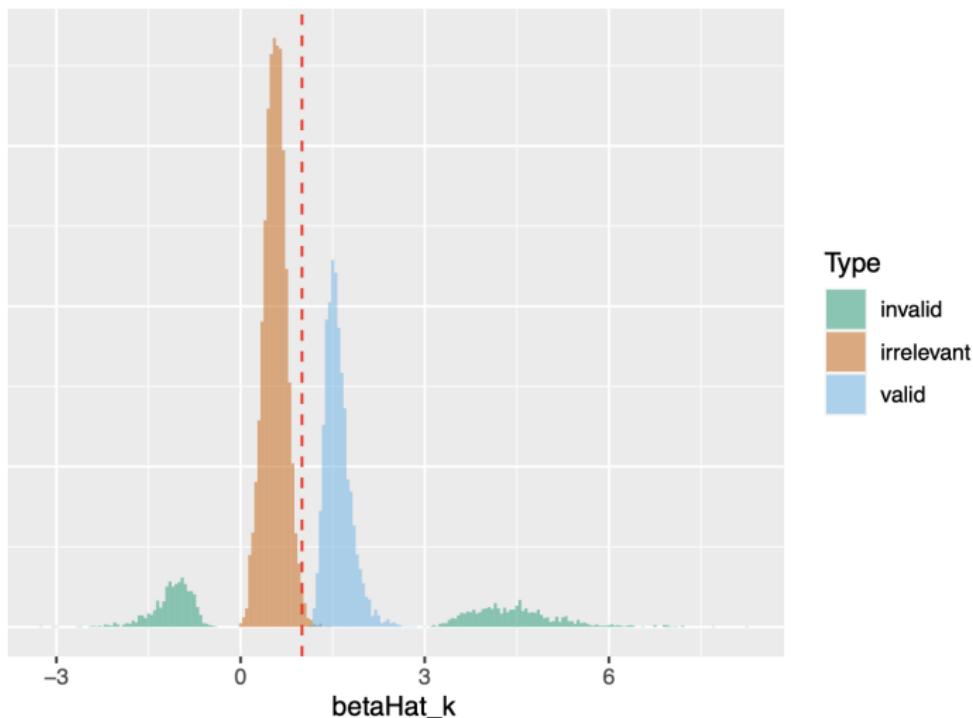
Simulation: true value is 1



Histogram of causal effect estimates

More challenges: Selection Bias

Simulation: true value is 1



Histogram of causal effect estimates

Summary

- Mendelian randomization is a powerful tool for causal effect estimation
- PRS allows one to combine multiple weak IVs to create a strong one
- Not the only challenge...
 - Invalid IVs
 - How to find candidate IVs
 - Selection bias