

# Assignment 06

PUBH 8878

## Part 1: Instrumental Variable (IV) assumptions in Mendelian randomization (MR) (25 pts)

1.1. Define the IV conditions for MR and explain why only IV1 is directly testable, while IV2–IV3 can be falsified but not proved.

- Relevance (IV1)
- Exchangeability (IV2)
- Exclusion restriction (IV3)

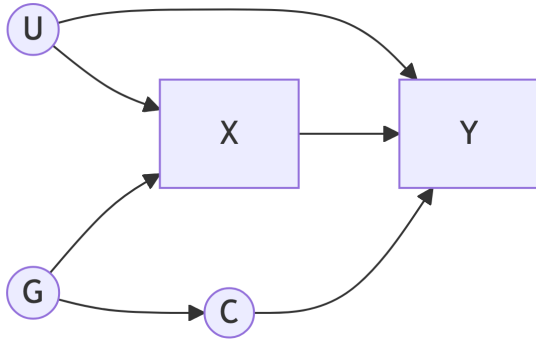
1.2. What additional condition is needed to point-identify a causal effect from MR beyond IV1–IV3? Contrast homogeneity vs monotonicity/LATE, and state the estimand under each.

1.3. Distinguish **horizontal pleiotropy** (biasing vs non-biasing), **vertical pleiotropy**, **LD confounding**, and **correlated pleiotropy**. For each, state whether IV3 (or IV2) is violated and why.

1.4 Define **gene-environment equivalence** and **selection/survivor bias** in MR. Discuss implications for interpretation of MR estimates.

## Part 2: DAGS (25 pts)

2.1. Explain why this DAG violates IV3 and which sensitivity estimators are intended to be robust in this scenario.



**2.2.** Consider the task of estimating the causal effect of alcohol consumption on blood pressure. Look up one GWAS for alcohol consumption. Identify one potential genetic instrument (SNP) from the alcohol GWAS, and draw a DAG including this SNP, alcohol consumption, blood pressure, and at least two potential confounders (measured or unmeasured). Discuss which IV assumptions may be violated and why. Cite your sources.

### Part 3: TSLS (25 pts)

Reproduce and extend the lecture's 2SLS simulation.

#### Data-generating process (DGP)

$$\begin{aligned}
 U &\sim \mathcal{N}(0, 1), \quad Z \sim \text{Bernoulli}(0.5) \\
 X &= \theta Z + \lambda U + \varepsilon_x, \quad \varepsilon_x \sim \mathcal{N}(0, 1) \\
 Y &= \beta X + \alpha U + \varepsilon_y, \quad \varepsilon_y \sim \mathcal{N}(0, 1)
 \end{aligned}$$

Use  $\beta = 1.5$ ,  $\theta = 0.8$ ,  $\lambda = 0.9$ ,  $\alpha = 1.0$ ,  $n = 2,000$ .

**3.1.** Fit OLS  $Y \sim X$  and 2SLS using manual TSLS. Report  $\hat{\beta}$  from OLS and 2SLS and the first-stage  $F$ -statistic\*\* of  $X \sim Z$ . Comment on bias in OLS vs 2SLS, and the strength of the instrument.

**3.2.** Vary instrument strength:  $\theta$  from 0.05 to 1.5. Simulate  $R = 200$  replicates at  $n = 1000$  each, and summarize the mean 2SLS  $\hat{\beta}$  (with 2.5–97.5% quantiles) and mean OLS  $\hat{\beta}$  across  $\theta$ . Plot  $\hat{\beta}_{\text{TSLS}}$  vs  $\theta$  with a ribbon. Interpret: how do weak instruments (small  $\theta$ ) affect bias and variance?

**3.3.** Add a direct violation of IV3 by modifying the DGP to  $Y = \beta X + \alpha U + \delta Z + \varepsilon_y$  with  $\delta$  ranging from 0 to 0.3. Fix  $\theta = 0.8$  and compare 2SLS  $\hat{\beta}$  across  $\delta$  values. Explain why 2SLS is biased when exclusion fails.

#### Part 4: Exchangeability threats and mitigation (25 pts)

**4.1** define population stratification, dynastic effects, and assortative mating. For each, state whether IV2 is threatened and one design/analysis strategy to mitigate the threat.

**4.2** Generate a group indicator  $S \in \{0, 1\}$  with  $\Pr(S = 1) = 0.5$ . Let:

- $G \sim \text{Bernoulli}(0.5 + 0.15 \cdot (S - 0.5))$  (allele frequency shift by S),
- $U \sim \mathcal{N}(0, 1)$ ,
- $X = 0.6G + U + \varepsilon_x$ ,
- $Y = 0.8X + 0.7 \cdot S + U + \varepsilon_y$ .

Compare **2SLS** estimates of  $\beta$  (a) without  $S$ , (b) adjusting for  $S$  in both stages. Summarize the bias reduction when accounting for  $S$ .