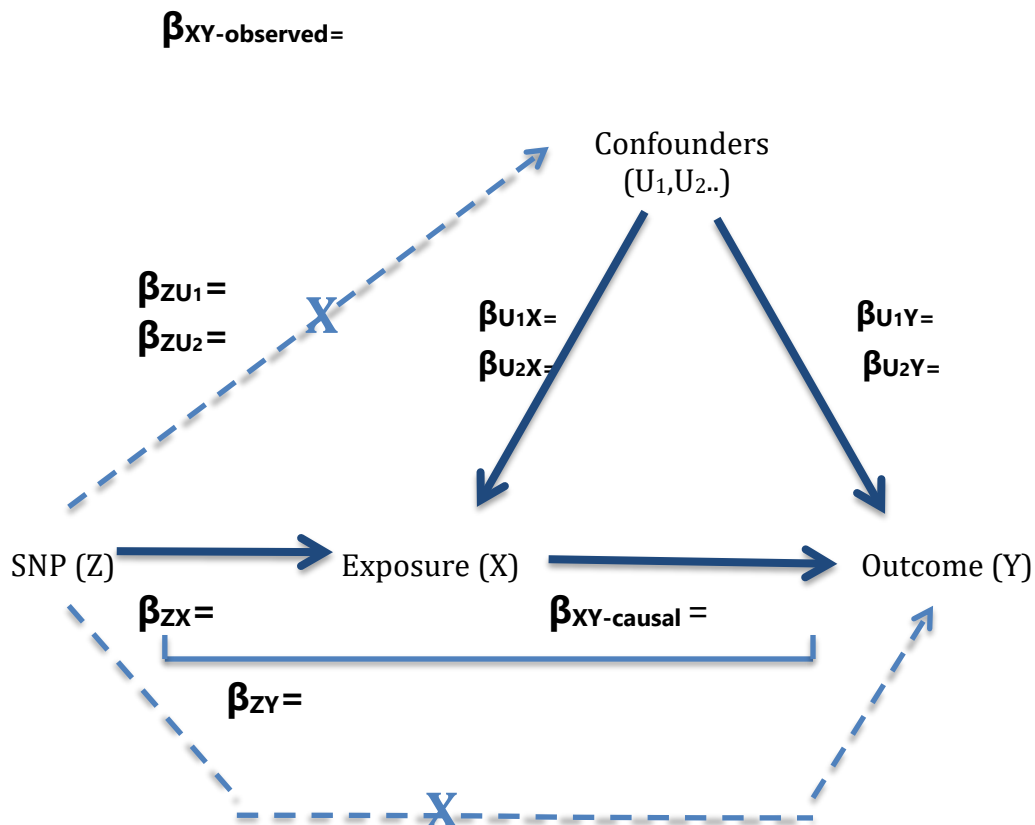


## Mendelian Randomization Practical Exercise 1

Applied Research Question:

Does having higher proinflammatory CRP causally increase your blood pressure?

### Graphical representation of a Mendelian Randomization IV analysis



### Formulas for Wald Estimator

Where Z=SNP instrument, X=Exposure, Y=Outcome

$$\text{Causal } \beta_{IV} = \frac{\beta_{ZY}}{\beta_{ZX}}$$

$$SE_{IV} = \frac{SE_{ZY}}{\beta_{ZX}}$$

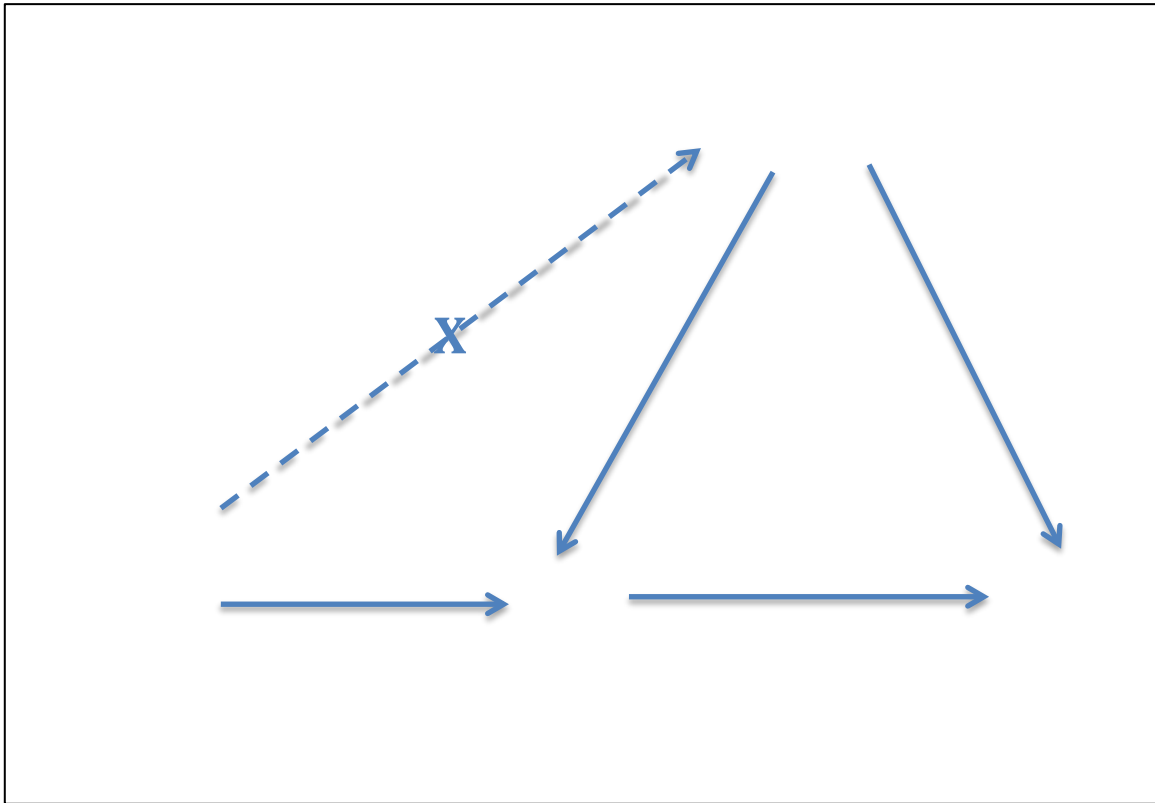
$$95\% \text{ CI} = \beta_{IV} \pm 1.96^* SE_{IV}$$



## CRP AND BLOOD PRESSURE

### Graphical Representation

**Q1.** As you're running the commands below, fill in the graphical representation of the IV analysis with the appropriate variables and beta-coefficients:



```
# Read in the dataset
setwd("/***YOUR DIRECTORY ***/")
example <- read.table("data.txt", header=T)
attach(example)
```

```
# Look at the data
head(example)
summary(example)
```

```
# Units: SNP (0,1,2), CRP mmol/L, SBP mmHg, Income per $10,000, HDL mmol/L
```

### Observational analyses

**Q2.** What does the observational linear regression of SBP on CRP show?

**A.**

```
# Run observational OLS regression (ordinary least squares) for BP & CRP
summary(lm(SBP~CRP))
```

```
# Plot the observational association between BP and CRP
plot(CRP,SBP)
abline(lm(SBP~CRP),col="red")
```

**Q3.** What does the OLS regression of the CRP SNP rs3091244 on CRP show?

**A.**

```
# Observational OLS regression of CRP on CRP SNP
summary(lm(CRP~rs3091244))
```

```
# Plot the relationship between CRP and rs3091244
plot(rs3091244, CRP)
abline(lm(CRP~rs3091244),col="red")
```

**Q4.** What do the OLS regressions of potential confounders (income, HDL) show?

**A.**

```
# Confounders
summary(lm(SBP~INCOME))
summary(lm(CRP~INCOME))
summary(lm(INCOME~rs3091244))
```

```
summary(lm(SBP~HDL))
summary(lm(CRP~HDL))
summary(lm(HDL~rs3091244))
```

**\*CHECK\***

Add the observational-based association variables and parameters to your graphical representation.

**Q5.** What are the implications for these income and HDL associations for the observational CRP-SBP association?

**A.**

**Q6.** Compare the unadjusted and covariate-adjusted OLS observational regressions. What do they show?

**A.**

```
# Run a covariate-adjusted model for the association between CRP & BP
summary(lm(SBP~CRP))
summary(lm(SBP~CRP+INCOME+HDL))
```

**Q7.** What could explain this?

**A.**

### **MR/IV Analyses: Wald Estimator**

Obtain the required estimates to compute the causal effect using the Wald Estimator from your dataset. Note, however, that an advantage of the Wald estimator is that you do not need individual level datasets to do the MR analysis. Reported SNP effects from published GWAS is sufficient and you can take the SNP effect on the exposure from one GWAS, and the SNP effect on outcome from a different GWAS sample ("Two sample MR")

**Q8.** Run the necessary OLS regressions to compute a Wald estimator

```
# OLS regression of CRP on CRP SNP
```

```
summary(lm(CRP~rs3091244))
```

```
# OLS regression of BP on CRP SNP
```

```
summary(lm(SBP~rs3091244))
```

**Q9.** From the above output, compute the causal effect using the Wald estimator, as well as it's SE and 95% CI. What do the results show and what do they mean?

**A.**

**Wald estimator causal Beta =**

**SE =**

**95% CI =**

**Q10.** Rerun the observational OLS of CRP and SBP and compare with the results from the Wald estimator. What do you notice about the Beta and SEs?

**A.**

```
# Observational OLS regression
```

```
summary(lm(SBP~CRP))
```

### **MR/IV Analyses: TSLS**

Two-stage least squares (TSLS) MR requires individual level data, and the exposure, SNP and outcome in the one sample ("Single sample MR").

```
# Call the AER library to run TSLS (if the AER package has been installed)
```

```
library(AER)
```

```
# If AER has not been installed, run the command below first:
```

```
install.packages("AER")
```

```
# General format for TSLS command:
```

```
# summary(ivreg(Outcome~Exposure | Instrument))
```

**Q11.** What do the TSLS results show and did it differ to the Wald estimator?

**A.**

```
# TSLS regression
```

```
summary(ivreg(SBP~CRP | rs3091244))
```

### Manual TSLS

To better understand what two-stage least squares regression is doing, let's perform it manually.

```
# Regress the exposure (CRP) on the instrument (rs3091244)
First_Stage <- lm(CRP~rs3091244)
```

Create a new variable (Pred\_CRP), which is the predicted values of the exposure (CRP) from the first-stage regression with the instrument. You can think of this as the expected value of exposure (CRP) given we know the particular individual's genotype (for *CRP* SNP rs3091244).

```
# Create predicted CRP values, from first-stage regression
Pred_CRP <- predict(First_Stage)
```

Have a quick look at these values

```
table(Pred_CRP)
plot(Pred_CRP, CRP)
abline(lm(CRP~ Pred_CRP), col="red")
```

Now regress the outcome variable (BP) on the predicted values of CRP, from the first-stage regression.

```
# Second stage regression
Second_Stage <- lm(SBP~Pred_CRP)
```

```
# Look at the results:
summary(Second_Stage)
```

**Q12.** Are they the same as 'ivreg' TSLS function?

**A.**

**\*CHECK\***

Are all the variables and parameters now complete in your graphical representation?

### **Weak instruments bias**

Assessing instrument strength with the F-stat (looking for  $\geq 10$ ).

For Single SNP MR, the F-statistic is calculated as:

$$F_{\text{stat}} = \frac{R^2 * (N-1)}{(1-R^2)}$$

where  $R^2$  is the variance explained in exposure by the SNP, and  $N$  is number of individuals in the study. This statistic is available in the output for OLS and TSLS

**Q13.** Looking at the F-statistic, determine if weak instruments may be an issue  
**A.**

```
#Look at F-stat from the first-stage linear regression  
summary(lm(CRP~rs3091244))
```

```
#Look at F-stat from 'diagnostics' by AER package  
summary(ivreg(SBP~CRP | rs3091244), diagnostics=T)
```

Discuss:

**Q14.** How would having weak instruments change the causal estimate of CRP on SBP, in this study (single sample)?