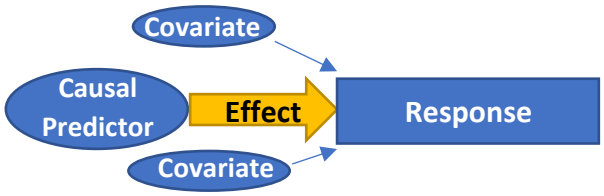
# Chapter 15: Reading Biostatistical Research Part II – Observational Studies

## Introduction to Causal Inference

- It is more _____ for researchers to make strong causal arguments with observational studies than it is with well-designed experiments.
- To build a causal argument, the researchers will need to build a model that controls for any legitimate confounders to see whether the supposed causal variable still shows predictive power. In statistics, this approach is known as **"Causal Inference."**
- If studying whether one specific predictor is actually causing changes in the response, other potential confounders should be included in the model as _____ (i.e., other predictors whose exact contribution is not important to us).
    - This allows us to stratify by other variables, and then observe if there is still a relationship between the supposed causal predictor and the response.
    - Our interest is on the _____, which reveals the probability that that predictor is contributing only random chance correlation to the model.
    - Researchers may also commonly provide a _____ for the slope coefficient, which can also be used to estimate how much influence that predictor has on the response while controlling for other predictors.




## Model Types

- Readers should also be aware that different response variables will require different types of models.
    - In this course, we only learned about **Linear Regression Modeling** (i.e., Ordinary least squares modeling). This is a common choice when working with a numeric response variable. In these cases, we're interested in testing whether a slope coefficient is different from 0.
    - **Logistic** Regression Modeling is common when modeling a binary response variable. Most often, researchers will examine model coefficients as relative risks or odds ratios, in which case, we're testing whether they are different from __.
    - You might occasionally see **Cox Proportional Hazards** Modeling when the response variable is time-to-event data. These will typically test whether hazard ratios are different from 1.
- Be on the lookout for _____ coefficients.
    - These would be model coefficients for our primary predictors while controlling for the other covariates in the model.
    - When we say controlling for covariates, we're asking whether changes to our predictor correlate with changes to the response, even if the covariates are held constant.

**Practice:** The following article was published in *The Lancet* in June 2023:
"Psychiatric disorders and subsequent risk of cardiovascular disease: a longitudinal matched cohort study across three countries."


**Abstract**

**Background**

Several psychiatric disorders have been associated with increased risk of cardiovascular disease (CVD), however, the role of familial factors and the main disease trajectories remain unknown.

**Methods**

In this longitudinal cohort study, we identified a cohort of 900,240 patients newly diagnosed with psychiatric disorders during January 1, 1987 and December 31, 2016, their 1,002,888 unaffected full siblings, and 1:10 age- and sex-matched reference population from nationwide medical records in Sweden, who had no prior diagnosis of CVD at enrolment. We used flexible parametric models to determine the time-varying association between first-onset psychiatric disorders and incident CVD and CVD death, comparing rates of CVD among patients with psychiatric disorders to the rates of unaffected siblings and matched reference population. We also used disease trajectory analysis to identify main disease trajectories linking psychiatric disorders to CVD. Identified associations and disease trajectories of the Swedish cohort were validated in a similar cohort from nationwide medical records in Denmark (N = 875,634 patients, same criteria during January 1, 1969 and December 31, 2016) and in Estonian cohorts from the Estonian Biobank (N = 30,656 patients, same criteria during January 1, 2006 and December 31, 2020), respectively.

**Findings**

During up to 30 years of follow-up of the Swedish cohort, the crude incidence rate of CVD was 9.7, 7.4 and 7.0 per 1000 person-years among patients with psychiatric disorders, their unaffected siblings, and the matched reference population. Compared with their siblings, patients with psychiatric disorders experienced higher rates of CVD during the first year after diagnosis (hazard ratio [HR], 1.88; 95% confidence interval [CI], 1.79–1.98) and thereafter (1.37; 95% CI, 1.34–1.39). Similar rate increases were noted when comparing with the matched reference population. These results were replicated in the Danish cohort. We identified several disease trajectories linking psychiatric disorders to CVD in the Swedish cohort, with or without mediating medical conditions, including a direct link between psychiatric disorders and hypertensive disorder, ischemic heart disease, venous thromboembolism, angina pectoris, and stroke. These trajectories were validated in the Estonian Biobank cohort.

**Interpretation**

Independent of familial factors, patients with psychiatric disorders are at an elevated risk of subsequent CVD, particularly during first year after diagnosis. Increased surveillance and treatment of CVDs and CVD risk factors should be considered as an integral part of clinical management, in order to reduce risk of CVD among patients with psychiatric disorders.

1. What is the response outcome in this study?

2. What is the proposed causal factor in this study? Could this have been studied as an experiment?

3. Describe the three groups being examined in this study.

4. What type of observational study this is? What about this particular design might help *limit* (but not completely eliminate) concerns of confounding differences between groups?

5. What are the *primary* findings of this study (focusing on the outcome more generally)? How do Figures 3 and 4 add more detail to the findings?

6. Consider the following statement in the limitations (second to last paragraph): "patients with psychiatric disorders may have more frequent health care consultations and therefore be more likely to be diagnosed with CVD." Why is this a limitation? What type of limitation/threat is it?

## Chapter 15 Additional Practice (if you need it!)

This article from the British Medical Journal (BMJ) was published in September 2022: "*Waning of vaccine effectiveness against moderate and severe covid-19 among adults in the US from the VISION network: test negative, case-control study.*"

1. What is the purpose of this study? What are the researchers trying to measure?


2. Describe the population of interest and describe the sample being examined. Also note what type of observational study this is.

3. What response outcome are the researchers using to help them answer their question? *What information is being collected from each person as the target outcome?*

4. Take a look at Figures 1 and 2. Which of the following are being framed as primary explanatory factors and which are simply serving as covariates (confounders) being controlled for in the model?

| | |
|---|---|
| Vaccination Status | (Explanatory Variable / Covariate) |
| Time since Last vaccination | (Explanatory Variable / Covariate) |
| Age of patient | (Explanatory Variable / Covariate) |
| Presence of comorbidities | (Explanatory Variable / Covariate) |
| Dominant COVID strain/wave at time of data collection | (Explanatory Variable / Covariate) |
| Race/Ethnicity of the patient | (Explanatory Variable / Covariate) |

5. What is the *difference* in information being displayed in Figures 1 and 2?

6. Take a look at Figure 4. What story are the researchers telling with this figure?

7. Take a look at the strengths and limitations section of the paper. Why did the researchers control so carefully for geography and calendar time?

8. Take a look at the strengths and limitations section of the paper, second paragraph. Name a factor that this study did not control for the researchers propose as a possible confounder here that could partly explain differences in infection rates between unvaccinated and vaccinated/boosted individuals.