Chapter 15: Reading Biostatistical Research Part II – Observational Studies

Reading Observational Studies

- Identify the **response variable(s)** being targeted in the study.
  - o Is there just one, or multiple response variable targets?
  - o Is the response variable categorical or numeric or categorical?
    - ▪ **Logistic** Regression Modeling is common when modeling a binary response variable, while **Ordinary least squares** (i.e., **Linear** Regression Modeling) is used for a numeric response variable.
    - ▪ You might occasionally see **Cox Proportional Hazards** Modeling when the researchers are considering time-to-event modeling.
- What **causal factor(s)** are being explored in the study?
  - o Are there one or several variables being assessed as causal factors?
  - o Are the researchers studying an *interaction* effect between these predictors?
    - ▪ For example: how does the fertilizer and sunlight level affect the growth of this plant? Does the effectiveness of the fertilizer increase when sunlight levels are higher?
  - o What other "covariates" are the researchers controlling for? (age, race, biological sex, and medical history are common covariates for researchers to use)
- What *type* of **observational study** was used?
  - o They will likely have used a cross-sectional, cohort, or case-control design
- What **population** is this study trying to generalize to, and what **sample** have they used to represent this population?
  - o Is the population all people with a particular disease/condition? Only those with a mild form or severe form? All ages or select ages?
  - o How large is the sample? What are the sampling units? (e.g., People, Bacteria cultures, Plots of land, etc.)
- What did the researchers **find**, and **what statistical evidence** are they using to make these claims?
  - o For categorical response variables (e.g., having or not having a disease), researchers may report a p-value or confidence intervals for a difference in absolute risk, relative risk, or odds ratios *after controlling for other covariates.*
  - o For numeric response variables researchers will commonly report a p-value for a predictor *after controlling for other covariates.*
  - o Researchers may report a hazard ratio (and Kaplan-meier plot) when comparing time until an adverse event occurs *after controlling for other covariates.*
- Any threats to the **internal validity** (causality argument) being made?
  - o Threats to group selection, setting differences, timing differences, test familiarity, attrition differences, or independence of the units in each group
- Any threats to the **external validity** (generalizability argument) being made?
  - o Threats to participant selection, setting, or historical sustainability?

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."
https://doi.org/10.1016/j.eclinm.2023.102063

1. What is the response outcome in this study?

Cardiovascular disease (CVD) with focus on diagnosis

2. What is the proposed causal factor in this study?

Psychiatric disorders (broken down into many categories for more specific comparisons)

3. Describe the population of interest and describe the sample being examined. In particular, check Figure 1 and note the various group comparisons being made.

Population would be any adult (in Nordic country) and sample would be all registered individuals with psychiatric disorder and no CVD, their non-affected siblings with no CVD, and a large reference population.

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

Cohort study with longitudinal matching. Siblings have a lot in common, but are still different people, where those with psychiatric disorders may be systematically different than siblings in habits

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?

Hazard ratio and confidence interval shows those with psychiatric disorder much more likely to be diagnosed with CVD in first year, and somewhat more likely over 30 year period. These other figures break it down by psychiatric disorder type, where some (like autism) show little correlation with CVD, but others do.

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?

It threatens the causality because their likelihood of diagnosis is higher due to more frequent visits. There is a greater chance for under-diagnosing in the comparison group.

Chapter 15 Additional Practice

The following article was published in the British Medical Journal (BMJ) 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.*" https://www.bmj.com/content/379/bmj-2022-072141

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

Change in Effectiveness of COVID vaccine over time with different waves

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

Pop: All people at risk of covid hospitalization and sample is people hospitalized with covid symptoms (some had covid and some had something else). More patient details in study pop section.

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?*

<u>Keep it simple. Whether or not they had COVID.</u>

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	(<u>Explanatory Variable</u> / Covariate)
Time since Last vaccination	(<u>Explanatory Variable</u> / Covariate)
Age of patient	(Explanatory Variable / <u>Covariate</u>)
Presence of comorbidities	(Explanatory Variable / <u>Covariate</u>)
Dominant COVID strain/wave at time of data collection	(<u>Explanatory Variable</u> / Covariate)
Race/Ethnicity of the patient	(Explanatory Variable / <u>Covariate</u>)

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

<u>One is a comparison of hospital admission with COVID risk and the other is a comparison of ICU/urgent care admission with COVID risk.</u>

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

<u>Vaccine effectiveness for immunocompromised individuals is much poorer at all 2 dose levels (even if recent), but better at 3 dose levels (if recent).</u>

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?

Because wave severity and contagion factors are different in different places and times. Trying to make careful comparison between groups of interested while holding those constant. Strengthens causal claim between explanatory factors and response outcomes

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.

Masking for one. Maybe people more vaxxed also more likely to wear mask or be more careful.

Chapter 15 Learning Goals

After this chapter, you should be able to…

- Read a structured abstract, visualizations/tables, and select passages from a biomedical research paper documenting an observational study and identify the following:
  - The response variable(s) (and variable type(s))
  - The causal factor(s) being examined, and covariates that were controlled for when examining these potential causal effects.
  - The type of observational study used
  - The population being generalized to and details about the sample used
  - The main findings and the statistical evidence supporting those findings
- Based on selected passages from the article…
  - Identify threats to the internal validity (causality argument) *if* the authors are making a causal argument
  - Identify threats to the external validity (generalizability argument)