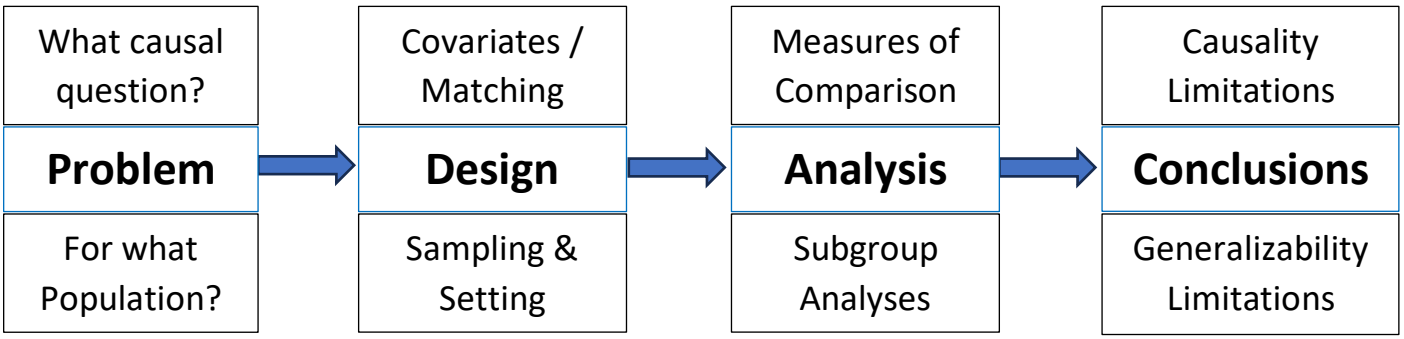
Chapter 15: Reading Biomedical Research – Observational Studies

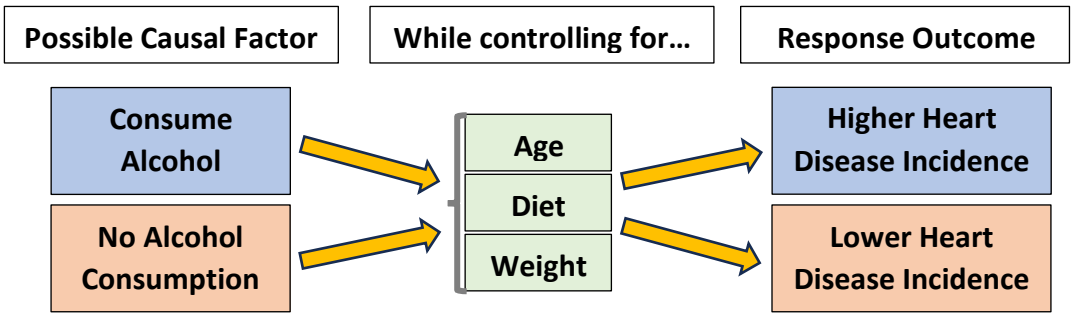
Reading Observational Studies

- With experiments, we examined how groups were sorted and conditions kept as similar as possible as part of our evaluation of the causality argument.
- Now, we'll need to consider the groups that have been compared and what the authors have done to try to create as balanced a comparison as possible.



Causal Inference with Observational Studies

- **Controlling for covariates** involves adding other variables in the model that we believe may correlate with the response and then seeing whether the supposed causal variable still shows predictive power.
    - For example, consider how we might be comparing those who consume and don't consume alcohol to determine a possible causal link to heart disease incidence. To make a good causal argument, we should control for the influences of age, diet, weight, etc. in our model.
    - P-values and confidence intervals for the alcohol variable's coefficient help us determine if our predictor has noticeable correlation with heart disease while holding the covariates constant.



- **Propensity score matching** (or **factor matching** in general) is another method for causal inference where we identify people from each group that match on certain factors and simply use these individuals for the comparison.
    - This is especially useful when we have access to very large databases and have room to exclude cases.

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.

Problem

What are the response and explanatory variables in this study?

What problem / causal relationship are they trying to understand?

What population or setting does this problem or question apply to?

Design

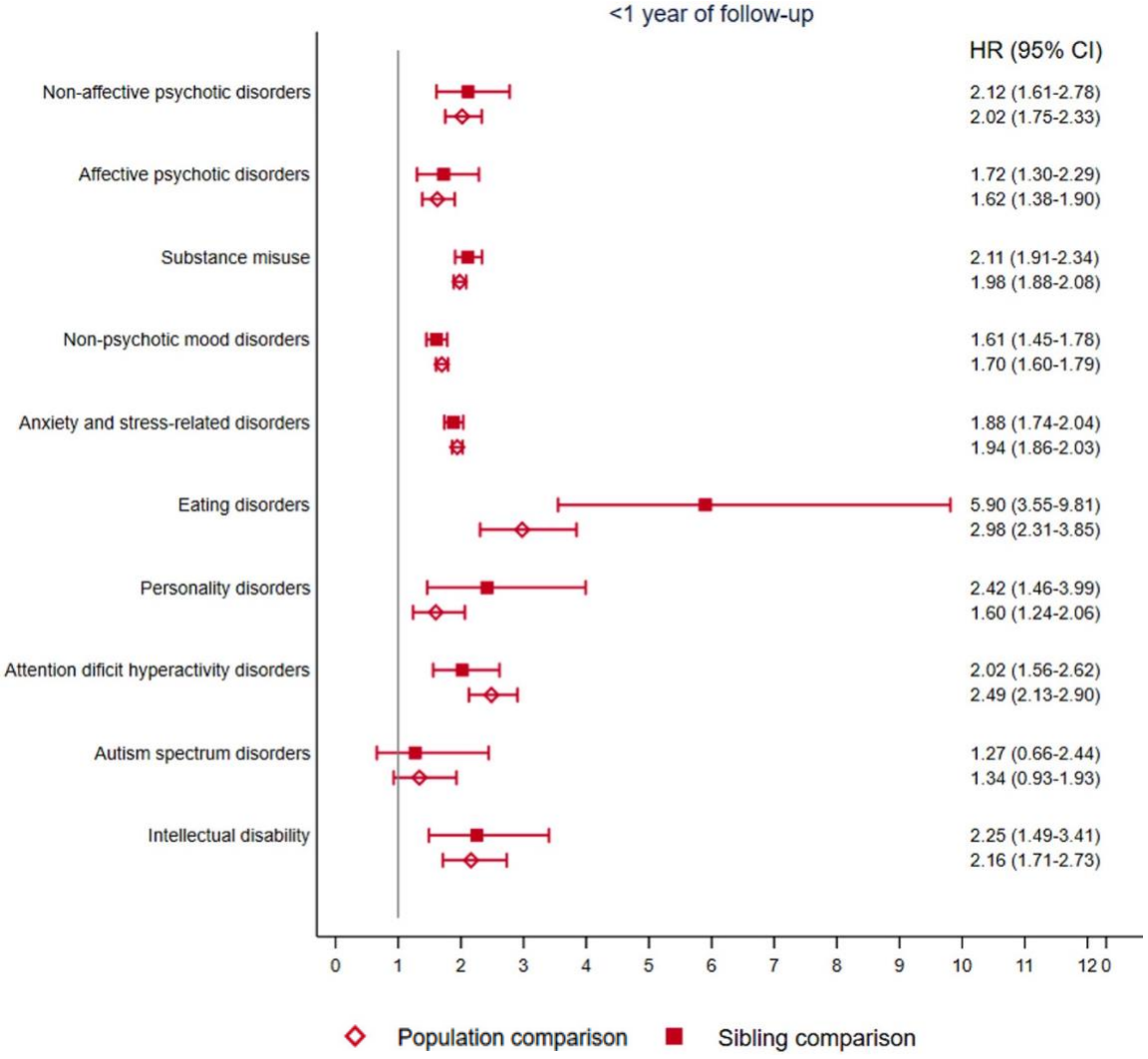
What kind of observational study is this? What steps have been taken to create a balanced comparison?

Who comprises the sample? Any representation questions at this point?

Analysis

What statistical measure(s) was/were reported? What do we learn from the confidence intervals or p-values in determining evidence for a difference more generally?

Figure 3: Comparing different categories of CVD among patients with psychiatric disorders to the various comparison groups within 1 year. Cox models further controlled for age, sex, education level, family income, cohabitation status, history of somatic disease, and family history of CVD in population comparison.



Analysis Continued

What subgroups were examined, and do the effects vary across different populations or conditions?

Conclusions

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?

Model Types

- 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 **1**.
    - 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 **adjusted** 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.

Reflection Questions

15.1. Why is it often harder to build a causality argument when working with observational study data?

15.2. What does it mean to control for covariates in a model? Why might this strengthen an observational study's causality argument?

15.3. What does it mean to do propensity score matching / factor matching? Why might this strengthen an observational study's causality argument?