Chapter 3: Observational Studies

Contents

In Chapter 1, we defined causal effects using counterfactual outcomes. In Chapter 2, we showed how randomized experiments allow us to estimate causal effects because randomization ensures exchangeability. But most research questions cannot be answered with randomized experiments—either because it would be unethical, impractical, or impossible to randomize treatment. This chapter discusses observational studies, which do not involve randomization of treatment.

The key question is: under what conditions can we validly estimate causal effects from observational data? This chapter introduces the fundamental identifiability conditions necessary for causal inference in observational studies.

This chapter is based on Hernán and Robins (2020, chap. 3, pp. 25-36).

1 3.1 Identifiability Conditions (pp. 25-27)

In an observational study, treatment is not randomly assigned by the investigator. Instead, individuals receive treatment based on their characteristics, preferences, physician recommendations, or other factors. As a result, treated and untreated individuals may differ systematically in ways that affect the outcome.

Can we still estimate causal effects from observational data? The answer is yes—but only under certain conditions called **identifiability conditions**.

Definition 1.1 (Identifiability)**.** A causal quantity (such as the average treatment effect) is **identifiable** if it can be computed from the observed data distribution under a given set of assumptions.

Three key identifiability conditions are required for causal inference from observational data:

1. **Exchangeability** (also called "no unmeasured confounding")
2. **Positivity** (also called "experimental treatment assignment")
3. **Consistency** (linking counterfactual outcomes to observed outcomes)

These three conditions are sometimes called the "causal inference triad" or "fundamental assumptions of causal inference." Each is necessary for identifying causal effects from observational data. Throughout this book, we will repeatedly return to these conditions and explore their implications.

Note that randomized experiments automatically satisfy exchangeability (by design) and typically satisfy positivity (by protocol). However, even randomized experiments require the consistency assumption.

1.1 Why Identifiability Matters

Without these conditions, association does not equal causation. Specifically:

- If **exchangeability** fails, observed associations may reflect confounding rather than causal effects
- If **positivity** fails, we cannot learn about causal effects for all individuals
- If **consistency** fails, the causal question itself may be ill-defined

In practice, we can never be certain that all identifiability conditions hold. Instead, we must:

1. Think carefully about whether these assumptions are plausible given subject-matter knowledge
2. Conduct sensitivity analyses to assess robustness to violations
3. Be transparent about the limitations of our causal conclusions

The remainder of this chapter explores each condition in detail.

2 3.2 Exchangeability (pp. 27-29)

The most critical identifiability condition is **exchangeability**. Informally, exchangeability means that the treated and untreated are comparable with respect to their potential outcomes.

Definition 2.1 (Conditional Exchangeability)**.** The treated and untreated are **exchangeable** conditional on covariates L when:

$$Y^a \perp\!\!\!\perp A \mid L \quad \text{for all } a$$

This means the potential outcome Y^a is independent of treatment A within levels of L.

Under conditional exchangeability:

$$Pr[Y^a = 1|A = 1, L] = Pr[Y^a = 1|A = 0, L] = Pr[Y^a = 1|L]$$

for all values of a and L.

Conditional exchangeability means the treated and untreated are exchangeable within strata defined by L. This is weaker than **marginal** exchangeability (which requires $Y^a \perp\!\!\!\perp A$ without conditioning on anything).

In randomized experiments, we have marginal exchangeability by design. In observational studies, we typically only hope for conditional exchangeability after adjusting for measured confounders.

The notation $\perp\!\!\!\perp$ denotes statistical independence.

2.1 Exchangeability and Confounding

Confounding is the absence of exchangeability. When exchangeability fails, comparing treated and untreated groups yields a biased estimate of the causal effect.

Example 2.1 (Confounding Example)**.** Suppose we want to estimate the causal effect of smoking on lung cancer using observational data. If smokers differ from non-smokers in other ways that affect lung cancer risk (e.g., occupational exposures, genetic factors), then:

$$Pr[Y^{a=1} = 1|A = 1] \neq Pr[Y^{a=1} = 1|A = 0]$$

The potential outcome under smoking is not independent of actual smoking status. The treated (smokers) are not exchangeable with the untreated (non-smokers).

2.2 Achieving Conditional Exchangeability

In observational studies, we try to achieve conditional exchangeability by:

1. **Identifying** potential confounders L based on subject-matter knowledge
2. **Measuring** these confounders accurately
3. **Adjusting** for confounders using appropriate statistical methods

Critical limitation: Conditional exchangeability is an **untestable assumption**. We can never verify from data alone that we have measured and adjusted for all confounders.

This is why subject-matter knowledge is essential in causal inference. We must use:

- Prior research
- Biological mechanisms

- Causal diagrams (introduced in Chapter 6)
- Expert opinion

to identify the set of confounders L that, if measured and adjusted for, would render the treated and untreated exchangeable.

If there exist **unmeasured confounders**—variables that affect both treatment and outcome but are not in L—then conditional exchangeability fails, and our causal effect estimates will be biased.

3 3.3 Positivity (pp. 29-30)

The second identifiability condition is **positivity**, also called the experimental treatment assignment assumption.

Definition 3.1 (Positivity)**. Positivity** requires that, for every combination of values of L for which $Pr[L] > 0$:

$$Pr[A = a|L] > 0 \quad \text{for all } a$$

In words: every individual has a non-zero probability of receiving every level of treatment, conditional on their measured covariates.

3.1 Why Positivity Matters

Without positivity, we cannot estimate causal effects for all individuals. If certain individuals (with specific values of L) have zero probability of receiving treatment, we cannot learn about their counterfactual outcome under treatment from the data.

Example 3.1 (Positivity Violation Example)**.** Suppose we want to estimate the effect of a treatment on 90-year-old individuals, but in our data, no 90-year-old person received the treatment. Then:

$$Pr[A = 1|\text{Age} = 90] = 0$$

We cannot estimate $\text{E}[Y^{a=1}|\text{Age} = 90]$ from the data because we have no treated 90-year-olds to observe.

Positivity violations can occur due to:

1. **Structural violations**: Some individuals cannot receive treatment (e.g., males cannot receive a pregnancy intervention)
2. **Random violations**: By chance, no individuals with certain covariate values received treatment in our sample
3. **Practical violations**: Treatment is theoretically possible but extremely rare for certain subgroups

When positivity is violated, we must either:

- Restrict our causal question to the subpopulation where positivity holds
- Make additional modeling assumptions (with their own limitations)
- Accept that we cannot answer the original causal question with the available data

In practice, **near-violations** of positivity (where $Pr[A = a|L]$ is very close to 0 or 1) can also cause problems, leading to unstable estimates with large variance.

4 3.4 Consistency: First, Define the Counterfactual Outcome (pp. 30-32)

The third identifiability condition is **consistency**. Unlike exchangeability and positivity, consistency is not about the relationship between treatment and outcome. Instead, it concerns the definition of the counterfactual outcome itself.

Definition 4.1 (Consistency)**.** The **consistency assumption** states that:

$$Y = Y^A$$

In words: the observed outcome Y for an individual equals their counterfactual outcome Y^a under the treatment level a that they actually received.

4.1 Well-Defined Interventions

For consistency to hold, the treatment must be **well-defined**. This means we must be able to precisely specify what it means to receive treatment level a.

Example 4.1 (Ill-Defined Treatment Example)**.** Consider "exercise" as a treatment. What does $A = 1$ (receives exercise) mean?

- Running 3 miles?
- Swimming for 30 minutes?
- Walking 10,000 steps?
- Weight training?

If different individuals in the $A = 1$ group received different forms of exercise, then $Y^{a=1}$ is not well-defined. The potential outcome under "exercise" depends on which specific form of exercise.

When treatment is not well-defined, we have **treatment variation irrelevance**. Different versions of "treatment" may have different effects on the outcome. In this case:

$$Y = Y^{A,K}$$

where K represents the specific version of treatment received.

For consistency to hold, we need either: 1. **No treatment variation**: All individuals who receive $A = a$ receive exactly the same intervention 2. **Treatment variation irrelevance**: Different versions of treatment have the same effect on the outcome (which is often implausible)

This is sometimes called the **Stable Unit Treatment Value Assumption (SUTVA)**.

5 3.5 Consistency: Second, Link Counterfactuals to Observed Data (pp. 32-34)

The consistency assumption also requires that there is no **interference** between individuals.

5.1 No Interference

Interference occurs when one individual's treatment affects another individual's outcome. For consistency to hold, we need:

$$Y_i = Y_i^{A_i}$$

The outcome for individual i depends only on individual i's treatment, not on other individuals' treatments.

> **Example 5.1** (Interference Example)**.** Consider a vaccine study. If vaccinating person A reduces person B's risk of disease (through herd immunity), then:
>
> $$Y_B \neq Y_B^{A_B}$$
>
> Person B's outcome depends not just on A_B (whether B was vaccinated), but also on A_A and the vaccination status of others in the population.

When interference exists, the full counterfactual notation must specify everyone's treatment:

$$Y_i^{\mathbf{a}} = Y_i^{a_1,a_2,\ldots,a_n}$$

where $\mathbf{a}$ is a vector of everyone's treatment assignments.

Interference is common in:

- Infectious disease studies (herd immunity)
- Social network studies (peer effects)
- Environmental exposures (shared exposures)
- Cluster randomized trials

Special methods are needed to handle interference, which is beyond the scope of the standard framework introduced in this chapter. For now, we assume no interference.

5.2 Temporal Consistency

Consistency also requires that the **timing** of treatment and outcome measurement is well-defined.

If we ask "What is the effect of treatment on the outcome?", we must specify:

- When is treatment received?
- When is the outcome measured?

For example, the effect of exercise on weight at 6 months may differ from the effect on weight at 1 year. We need to be precise about the follow-up time.

6 3.6 The Target Trial (pp. 34-36)

A useful framework for thinking about identifiability conditions in observational studies is the **target trial**.

> **Definition 6.1** (Target Trial)**.** The **target trial** is the (hypothetical) randomized experiment we would conduct if we could. By specifying the target trial, we clarify:
> 1. The causal question we are trying to answer
> 2. The treatment comparison of interest
> 3. The outcome and follow-up time
> 4. The eligibility criteria
> 5. The estimand (parameter to be estimated)

6.1 Emulating the Target Trial

When conducting an observational study, we should ask: "How closely can we emulate the target trial with the available data?"

The identifiability conditions can be understood as requirements for successfully emulating a randomized experiment:

- **Exchangeability**: Can we adjust for confounders to simulate randomization?
- **Positivity**: Do we have both treated and untreated individuals in all relevant subgroups?
- **Consistency**: Is the treatment well-defined, matching what would be implemented in the trial?

The target trial framework is especially useful for:

1. **Clarifying the causal question**: By specifying exactly what randomized trial we would run, we force ourselves to be precise about what causal effect we want to estimate
2. **Identifying potential biases**: Differences between the target trial and our observational study reveal potential sources of bias
3. **Guiding analysis**: The analysis plan should mimic what we would do in the target trial (e.g., intention-to-treat analysis, per-protocol analysis)

This framework will be revisited throughout the book, particularly in Part III when we discuss time-varying treatments and complex longitudinal data.

6.2 Example: Target Trial for Statins and Cardiovascular Disease

> **Example 6.1** (Target Trial Example)**. Research question**: Does taking statins reduce the risk of cardiovascular disease (CVD)?
> **Target trial specification**:
> - **Eligibility**: Adults aged 40-75 with LDL cholesterol > 190 mg/dL, no history of CVD
> - **Treatment**: Daily statin (yes/no)

- **Assignment**: Random assignment to statin or no statin
- **Outcome**: CVD event within 5 years
- **Estimand**: Risk difference $Pr[Y^{a=1} = 1] - Pr[Y^{a=0} = 1]$

Observational study emulation:

- Identify individuals meeting eligibility criteria in electronic health records
- Compare those who initiated statins vs. those who did not
- Adjust for baseline confounders L (age, sex, blood pressure, smoking, etc.)
- Follow individuals for 5 years, censoring if they move or die from non-CVD causes

In this example, the key challenges are:

1. **Exchangeability**: Did we measure all confounders? Are there unmeasured factors (e.g., health behaviors, genetics) that affect both statin use and CVD risk?
2. **Positivity**: Are there individuals with certain characteristics (e.g., very high cholesterol) who always receive statins? If so, we cannot estimate the effect for them.
3. **Consistency**: Is "taking statins" well-defined? Do adherence levels vary? Should we analyze intention-to-treat or per-protocol?

The target trial framework helps us think systematically about these issues.

7 Summary

This chapter introduced the fundamental **identifiability conditions** for causal inference in observational studies:

1. **Exchangeability**: $Y^a \perp\!\!\!\perp A \mid L$ (no unmeasured confounding)
2. **Positivity**: $Pr[A = a|L] > 0$ for all a and L (every subgroup can receive every treatment level)
3. **Consistency**: $Y = Y^A$ (well-defined interventions, no interference)

Under these three conditions, we can identify causal effects from observational data by adjusting for measured confounders L.

The **target trial framework** provides a useful way to think about observational studies: we should try to emulate the randomized experiment we would have conducted if we could. The identifiability conditions tell us what is required for this emulation to succeed.

Looking ahead:

- Chapter 4 introduces effect modification, where causal effects differ across subgroups
- Chapter 5 discusses interaction between treatments
- Chapter 6 introduces causal diagrams, a powerful tool for understanding confounding
- Chapters 7-9 explore specific types of bias: confounding, selection bias, and measurement bias
- Chapter 10 introduces time-varying treatments and confounding

The three identifiability conditions introduced in this chapter will appear repeatedly throughout the book. They are the foundation of all causal inference from observational data.

8 References

Hernán, Miguel A, and James M Robins. 2020. *Causal Inference: What If*. Chapman & Hall/CRC. https://miguelhernan.org/whatifbook.