# Chapter 7: Confounding

## Contents



In Chapter 3, we introduced exchangeability as a key identifiability condition. In Chapter 6, we learned to represent causal relationships using DAGs and introduced the backdoor criterion for identifying confounding. This chapter provides a detailed examination of **confounding**—the most common threat to validity in observational studies.

This chapter is based on Hernán and Robins (2020, chap. 7, pp. 77-92).

# 1 7.1 The Structure of Confounding (pp. 77-80)

---

Confounding occurs when a common cause of treatment and outcome creates a non-causal association between them.

**Definition 1.1** (Confounding Structure)**.** A variable $L$ is a **confounder** of the effect of $A$ on $Y$ if:

1. $L$ causes $A$ (or shares a common cause with $A$)
2. $L$ causes $Y$ (or shares a common cause with $Y$)
3. $L$ is not affected by $A$ (not a consequence of treatment)

**Causal diagram representation**:

```
L → A → Y
L → Y
```

The path $A \leftarrow L \rightarrow Y$ is a **backdoor path** that creates non-causal association.

**Why confounding creates bias**:

Without confounding:

$$E[Y|A=1] - E[Y|A=0] = E[Y^{a=1}] - E[Y^{a=0}]$$

(association equals causation)

With confounding:

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

(association does not equal causation)

The observed association includes both:

- The causal effect of $A$ on $Y$
- The confounding bias due to the common cause $L$

### 1.1 Common Confounding Scenarios

**Example 1: Healthy worker bias**

- Healthier individuals are more likely to be employed (employed → more likely to be exposed at work)
- Healthier individuals have better outcomes
- Comparing employed vs. unemployed introduces confounding by health status

**Example 2: Confounding by indication**

- Sicker patients receive more aggressive treatment
- Sicker patients have worse outcomes
- Treatment appears harmful when in fact it may be beneficial

**Confounding can go in either direction**:

- **Positive confounding**: Makes treatment appear more beneficial (or more harmful) than it truly is
- **Negative confounding**: Makes treatment appear less beneficial (or less harmful) than it truly is
- **Zero confounding**: The magnitude depends on the strength and direction of the $L \rightarrow A$ and $L \rightarrow Y$ relationships

## 2 7.2 Confounding and Exchangeability (pp. 80-82)

---

Confounding is equivalent to lack of (conditional) exchangeability.

### 2.1 No Confounding = Exchangeability

**No confounding** means:

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

This is **marginal exchangeability**: the counterfactual outcomes are independent of treatment.

**Confounding** means exchangeability does not hold:

$$Y^a \not\perp\!\!\!\perp A$$

The treated and untreated differ with respect to their potential outcomes.

**Example 2.1** (Confounding and Exchangeability)**.** Suppose exercise ($A$) affects heart disease ($Y$), and both are affected by age ($L$):

**Without confounding**:

- Young and old people equally likely to exercise
- $E[Y^{a=1}|A=1] = E[Y^{a=1}|A=0]$ (exchangeable)

**With confounding**:

- Younger people more likely to exercise

- Younger people have lower baseline risk
- $E[Y^{a=1}|A=1] \neq E[Y^{a=1}|A=0]$ (not exchangeable)
- Those who exercise would have had better outcomes even without exercising

## 2.2 Conditional Exchangeability

Even when marginal exchangeability fails, we may achieve **conditional exchangeability** by adjusting for confounders:

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

Within levels of $L$, the treated and untreated are exchangeable.

**Key insight**: Confounding can be eliminated by conditioning on (adjusting for) the confounders.

**Requirements**: 1. We must **identify** all confounders based on subject-matter knowledge 2. We must **measure** them accurately 3. We must **adjust** for them appropriately in analysis

If these requirements are met, we can estimate causal effects from observational data.

# 3 7.3 Confounding and the Backdoor Criterion (pp. 82-85)

---

The backdoor criterion (Chapter 6) provides a graphical method for identifying confounding.

## 3.1 Backdoor Paths and Confounding

A **backdoor path** from $A$ to $Y$:

- Starts with an arrow into $A$ (i.e., $\cdot \to A$)
- Connects $A$ to $Y$ through any sequence of arrows

**Confounding exists if backdoor paths are open (unblocked)**.

**Example 3.1** (Identifying Confounders with the Backdoor Criterion)**.** **Diagram 1**:

```
L → A → Y
L → Y
```

**Backdoor path**: $A \leftarrow L \rightarrow Y$ **Confounders**: $L$ **Solution**: Adjust for $L$

**Diagram 2**:

```
U → L → A → Y
      L → Y
```

**Backdoor paths**: $A \leftarrow L \rightarrow Y$, $A \leftarrow L \leftarrow U \rightarrow Y$ (if U causes Y) **Confounders**: $L$ (and $U$ if it affects $Y$) **Solution**: Adjust for $L$ (and $U$ if measured)

**Diagram 3**:

```
A → M → Y
L → A
L → Y
```

**Backdoor path**: $A \leftarrow L \rightarrow Y$ **Confounders**: $L$ **Do NOT adjust for** $M$: $M$ is a mediator (on the causal path), not a confounder

**Common mistakes**:

1. **Adjusting for mediators**: Variables on the causal path from $A$ to $Y$ should NOT be adjusted for, as this blocks the causal effect we're trying to estimate
2. **Adjusting for colliders**: Variables caused by both $A$ and $Y$ should NOT be adjusted for, as this induces bias
3. **Failing to adjust for all confounders**: If even one confounder is unmeasured or unadjusted, bias remains
4. **Overadjustment**: Including unnecessary variables (especially descendants of treatment) can introduce bias

# 4 7.4 Confounding and Confounders (pp. 85-87)

---

The traditional definition of "confounder" in epidemiology differs slightly from the causal DAG perspective.

## 4.1 Traditional Confounder Definition

Traditionally, a variable $L$ is considered a confounder if: 1. $L$ is associated with treatment $A$ 2. $L$ is associated with outcome $Y$ (among the untreated) 3. $L$ is not affected by treatment $A$

## 4.2 DAG-Based Definition

From the DAG perspective, $L$ is a confounder if:

- $L$ opens a backdoor path from $A$ to $Y$

**Differences**:

The traditional definition is based on **associations** (statistical relationships). The DAG definition is based on **causal structure** (graphical relationships).

**Why this matters**:

1. **A variable can be associated with both $A$ and $Y$ without being a confounder**
   - Example: A collider caused by both $A$ and $Y$
   - Adjusting for it would induce bias, not remove it
2. **A variable can be a confounder without being associated with both $A$ and $Y$ in the data**
   - Example: A confounder whose effects cancel out, leaving no association
   - Failing to adjust for it would leave bias

**Best practice**: Use DAGs to identify confounders based on causal structure, not associations alone.

# 5 7.5 Single-World Intervention Graphs (pp. 87-89)

---

Single-World Intervention Graphs (SWIGs) are an extension of DAGs that explicitly represent interventions and counterfactual outcomes.

## 5.1 SWIGs vs. DAGs

- **Standard DAGs**: Represent relationships among observed variables
- **SWIGs**: Represent relationships among counterfactual variables under specified interventions

**SWIG notation**:

- $Y_a$: Counterfactual outcome under intervention $do(A = a)$
- $A_a$: Treatment value set to $a$ by intervention
- Edges represent causal effects in the counterfactual world where $A = a$

**Example SWIG**: For the causal effect of $A$ on $Y$ with confounder $L$:

```
L → A_a → Y_a
L → Y_a
```

SWIGs make explicit:

- Which variables are set by intervention
- Which variables remain as observed
- Which counterfactual outcome we're interested in

**Advantages**:

- Clearer representation of counterfactuals
- Explicit about the intervention
- Useful for complex scenarios (time-varying treatments, mediation)

**Disadvantages**:

- More complex notation
- Require more assumptions to be specified
- Not yet as widely used as standard DAGs

For most purposes in this book, standard DAGs suffice. SWIGs are mentioned for completeness and for readers interested in advanced topics.

# 6 7.6 Confounding Adjustment (pp. 89-92)

---

Once confounders are identified, several methods can adjust for them.

## 6.1 Methods for Confounding Adjustment

1. **Stratification**: Estimate effects within strata of $L$, then combine (standardization)
2. **Regression adjustment**: Include $L$ as covariates in a regression model
3. **Inverse probability weighting**: Weight by $1/Pr[A|L]$ to create a pseudo-population where $A$ and $L$ are independent (Chapter 12)
4. **Matching**: Match treated and untreated individuals on $L$

**Example 6.1** (Comparing Adjustment Methods)**. Data**: Effect of smoking ($A$) on lung cancer ($Y$), adjusting for age ($L$)

**Stratification**:

- Estimate effect separately for age = 40, 50, 60, 70
- Combine using weighted average

**Regression**:

```
glm(Y ~ A + L, family = binomial())
```

**IP weighting** (Chapter 12):

```r
weight <- 1 / predict(glm(A ~ L, family = binomial()), type = "response")
glm(Y ~ A, weights = weight, family = binomial())
```

**Matching**:

- For each smoker, find non-smoker of same age
- Compare outcomes

**Choosing an adjustment method**:

**Stratification**:

- Transparent, easy to understand
- Allows checking for effect modification
- Limited to discrete confounders
- Requires large sample sizes for fine strata

**Regression**:

- Handles continuous confounders
- Efficient (uses all data)
- Relies on model assumptions
- Can obscure effect modification

**IP weighting**:

- Flexible, can handle complex confounding
- Estimates marginal (population-average) effects
- Can be unstable with extreme weights
- More complex to implement

**Matching**:

- Intuitive, creates comparable groups
- Discards unmatched data
- Largely superseded by other methods

**Modern recommendation**: Use IP weighting or doubly robust methods (combine regression and weighting) for flexibility and robustness.

# 7 Summary

---

This chapter provided a detailed examination of **confounding**.

**Key concepts**:

1. **Confounding structure**: Common causes of treatment and outcome create backdoor paths

2. **Exchangeability**: Confounding = lack of exchangeability; conditional exchangeability can be achieved by adjusting for confounders

3. **Backdoor criterion**: Provides a graphical method to identify which variables to adjust for

4. **DAG vs. traditional definitions**: DAG-based confounding identification is preferred over association-based criteria

5. **Adjustment methods**: Stratification, regression, IP weighting, and matching can all adjust for confounding

6. **Critical assumptions**:

   - All confounders must be identified (no unmeasured confounding)
   - All confounders must be measured accurately
   - Adjustment must be done correctly

**Practical guidelines for dealing with confounding**:

1. **Draw a DAG** based on subject-matter knowledge before analyzing data
2. **Identify confounders** using the backdoor criterion
3. **Measure confounders** as accurately as possible
4. **Choose an adjustment method** appropriate for your data and confounders
5. **Check assumptions**:
   - Positivity: Do all $(A, L)$ combinations occur?
   - Model fit: Are regression model assumptions met?
   - Balance: After adjustment, are confounders balanced?
6. **Conduct sensitivity analyses**: How robust are findings to unmeasured confounding?
7. **Be transparent**: Report the DAG, adjustment set, and method clearly

**Limitations**:

Even with perfect adjustment for measured confounders, bias can remain if:

- Important confounders are unmeasured (Chapter 19 covers sensitivity analysis)
- Confounders are measured with error (Chapter 9)
- Adjustment methods are applied incorrectly
- Positivity is violated

Confounding control is necessary but not sufficient for valid causal inference.

**Looking ahead**:

- **Chapter 8**: Selection bias
- **Chapter 9**: Measurement bias
- **Chapters 12-15**: Advanced methods for confounding adjustment

# 8 References

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