Contents

1 Tutorial: Understanding Backdoor Paths in Causal Inference

1.1 Overview

This tutorial will teach you how to identify and block backdoor paths in causal graphs. By the end, you'll understand when you can replace an intervention with conditioning and how to identify valid adjustment sets.

Learning Objectives:

- Understand what paths and backdoor paths are
- Learn the rules for blocking paths
- Apply the backdoor criterion to real examples
- Practice diagnosing causal graphs

1.2 1. The Core Problem

1.2.1 What We Want to Know

In causal inference, we often want to estimate the causal effect of treatment C on outcome X:

$$p(X \mid do(C))$$

But we can only observe conditional probabilities:

$$p(X \mid C, Z)$$

Key Question: When can we replace intervention with conditioning?

$$p(X \mid do(C)) \stackrel{?}{=} \sum_{z} p(X \mid C, z) \cdot p(Z)$$

Answer: When Z blocks all backdoor paths from C to X.

1.3 2. Foundations: What is a Path?

1.3.1 Definition

A **path** is any sequence of variables connected by edges, **ignoring arrow direction**.

1.3.2 Example Paths Between C and X

Path 1: $C \rightarrow X$ (direct causal path)

Path 2: $C \leftarrow U \rightarrow X$ (backdoor path through confounder)

Path 3: $C \leftarrow U \rightarrow Z \rightarrow X$ (longer backdoor path)

Path 4: $C \rightarrow M \leftarrow U \rightarrow X$ (path through collider)

Key Point: Direction doesn't matter for *being* a path. Direction matters for *classifying* and *blocking* paths.

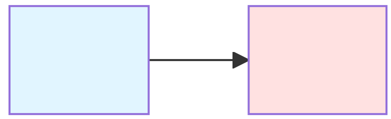

Figure 1: Mermaid diagram

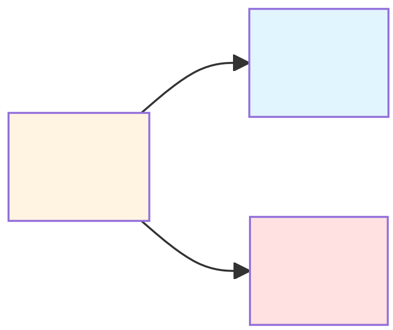

Figure 2: Mermaid diagram

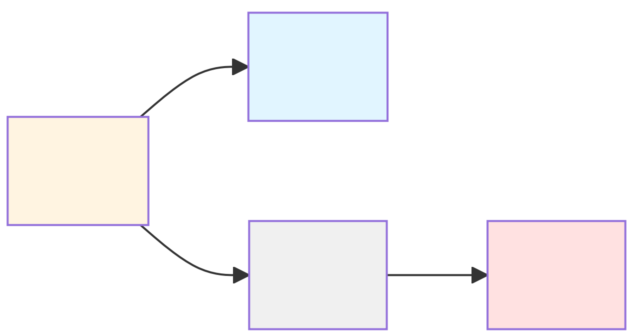

Figure 3: Mermaid diagram

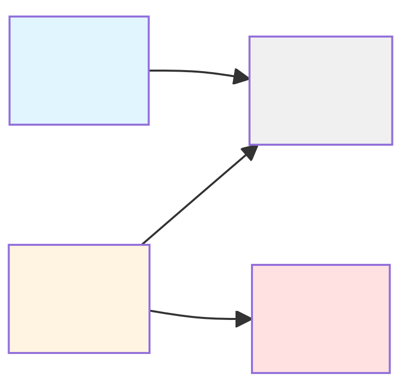

Figure 4: Mermaid diagram

1.4 3. What Makes a Path a Backdoor Path?

1.4.1 Precise Definition

> A **backdoor path** from C to X is any path that **starts with an arrow pointing INTO** C.

That's it. Nothing more complex.

1.4.2 Example 1: Direct Causal Path (NOT a Backdoor)

Figure 5: Mermaid diagram

Path: $C \to X$

- Starts with arrow **out of** C
- **Front-door (causal) path**
- **Not a backdoor path**

1.4.3 Example 2: Classic Confounder (IS a Backdoor)

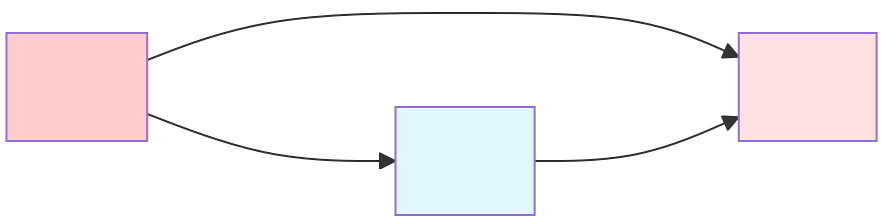

Figure 6: Mermaid diagram

Backdoor Path: $C \leftarrow U \to X$

- Starts with arrow **into** C (i.e., $C \leftarrow U$)
- **Backdoor path**
- This path carries **spurious association**

Why it's dangerous: U causes both C and X, creating a non-causal association between them.

1.4.4 Example 3: Longer Backdoor Path

Backdoor Path: $C \leftarrow U \to Z \to X$

- Still starts with $C \leftarrow U$
- **Backdoor path** (even though it's longer)

1.5 4. Why Backdoor Paths Are Dangerous

Backdoor paths allow **information to flow from** C **to** X without C causing X.

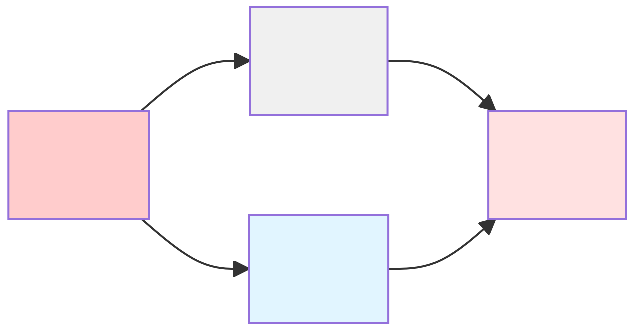

Figure 7: Mermaid diagram

1.5.1 The Algebra

With backdoor path (observational):

$$p(X \mid C) = \sum_u p(X \mid C, U) \cdot p(U \mid C)$$

Without backdoor path (interventional):

$$p(X \mid do(C)) = \sum_u p(X \mid C, U) \cdot p(U)$$

The Problem: The backdoor path makes $p(U \mid C) \neq p(U)$, introducing bias.

1.5.2 Intuition

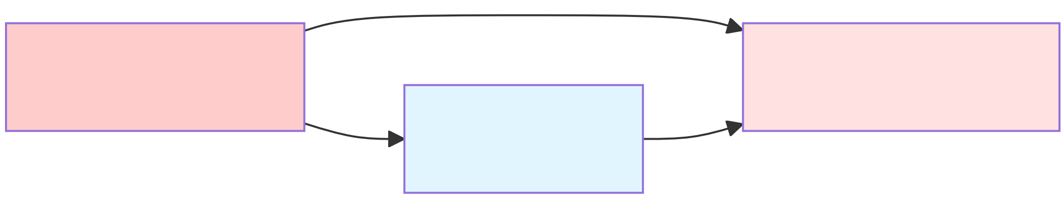

Figure 8: Mermaid diagram

People with high genetic risk might exercise more (or less). This creates a spurious association between exercise and heart disease that isn't purely causal.

1.6 5. Blocking Paths: The Rules

A path is **blocked** when information cannot flow through it. Here are the rules:

1.6.1 Rule 1: Conditioning on a Non-Collider Blocks the Path

Path: $C \to U \to X$

If we condition on U: The path is **blocked**

Why: Conditioning on U "holds it constant," preventing information flow.

Figure 9: Mermaid diagram

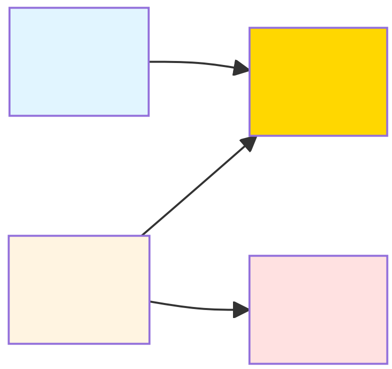

Figure 10: Mermaid diagram

1.6.2 Rule 2: A Collider Blocks the Path by Default

Path: $C \to Z \leftarrow U \to X$

Collider: Z (two arrows point into it)

Status: Path is **blocked by default**

Key Point: Do NOT condition on colliders!

1.6.3 Rule 3: Conditioning on a Collider OPENS the Path (The Trap!)

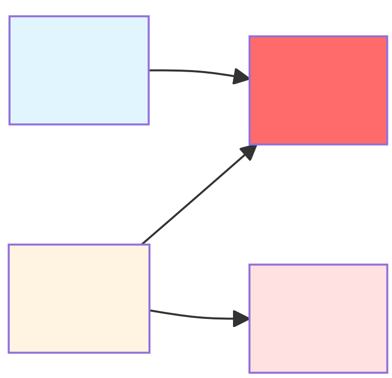

Figure 11: Mermaid diagram

If we condition on Z: The path **opens**

Why: Conditioning on a collider creates a spurious association between its parents.

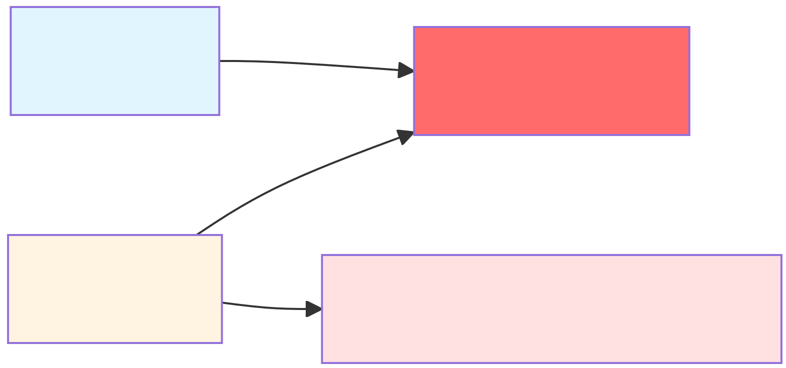

Figure 12: Mermaid diagram

1.6.4 Example: Collider Bias

Among NBA players (conditioning on S), talent and height appear negatively correlated—but this is spurious!

1.7 6. The Backdoor Criterion (Formal Statement)

A set of variables Z satisfies the **backdoor criterion** relative to (C, X) if:

1.7.1 Condition 1: No Descendants of Treatment

No variable in Z is a descendant of C

Reason: You don't adjust for consequences of treatment (that would block the causal effect!)

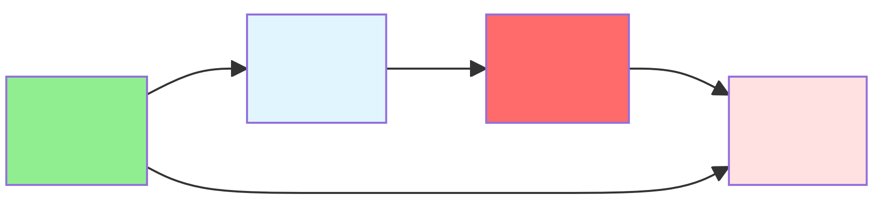

Figure 13: Mermaid diagram

- Adjust for U (confounder)
- Do NOT adjust for M (mediator/descendant)

1.7.2 Condition 2: Block All Backdoor Paths

Z blocks every backdoor path from C to X

Every path starting with an arrow into C must be blocked.

1.7.3 The Payoff

If both conditions hold:

$$p(X \mid do(C)) = \sum_z p(X \mid C, z) \cdot p(z)$$

You can estimate causal effects from observational data!

1.8 7. Worked Examples

1.8.1 Example 1: Simple Confounder

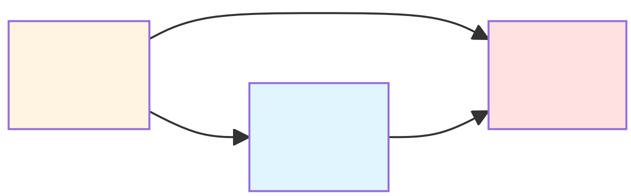

Figure 14: Mermaid diagram

Question: Should we adjust for U?

Analysis:

- **Backdoor path:** $C \leftarrow U \rightarrow X$
- **Is U a descendant of C?** No
- **Does U block the backdoor path?** Yes

Answer: Adjust for U

1.8.2 Example 2: Collider Bias

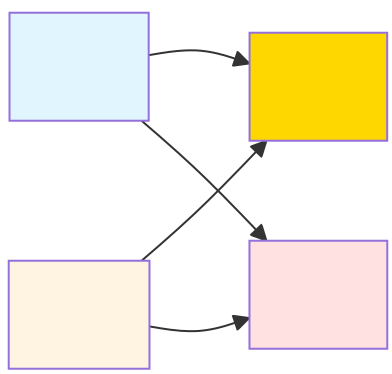

Figure 15: Mermaid diagram

Question: Should we adjust for Z?

Analysis:

- **Backdoor path:** $C \rightarrow Z \leftarrow U \rightarrow X$
- **Is Z a collider?** Yes
- **Is the path blocked by default?** Yes
- **What happens if we condition on Z?** Opens the path

Answer: Do NOT adjust for Z

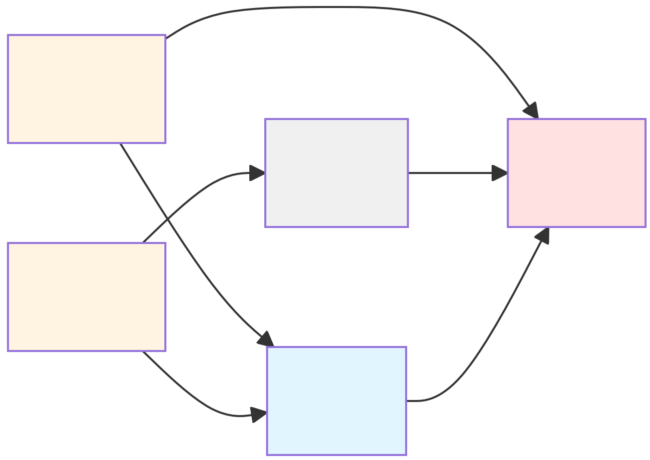

Figure 16: Mermaid diagram

1.8.3 Example 3: Multiple Paths

Question: What should we adjust for?

Analysis:

- **Backdoor path 1:** $C \leftarrow U_1 \rightarrow X$
- **Backdoor path 2:** $C \leftarrow U_2 \rightarrow M \rightarrow X$
- **Adjustment set:** $\{U_1, U_2\}$
- **Are they descendants of C?** No
- **Do they block all backdoor paths?** Yes

Answer: Adjust for $\{U_1, U_2\}$

1.8.4 Example 4: Mediator Trap

Figure 17: Mermaid diagram

Question: Should we adjust for M?

Analysis:

- **No backdoor paths** (no arrows into C)
- **Is M a descendant of C?** Yes
- **What happens if we condition on M?** Blocks the causal path!

Answer: Do NOT adjust for M (it's a mediator)

1.9 8. Practice Problems

Test your understanding with these diagnostic exercises.

1.9.1 Problem 1

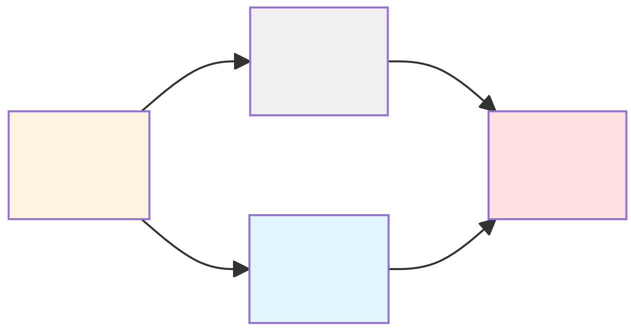

Figure 18: Mermaid diagram

Questions:

1. Identify all paths from C to X
2. Which paths are backdoor paths?
3. What should you adjust for?

Click to see answer

Paths:

1. $C \to X$ (causal path)
2. $C \leftarrow U \to Z \to X$ (backdoor path)

Backdoor paths: Path 2 (starts with $C \leftarrow U$)

Adjustment: Adjust for either U or Z (both block the backdoor path)

- U is preferred (blocks closer to the source)

1.9.2 Problem 2

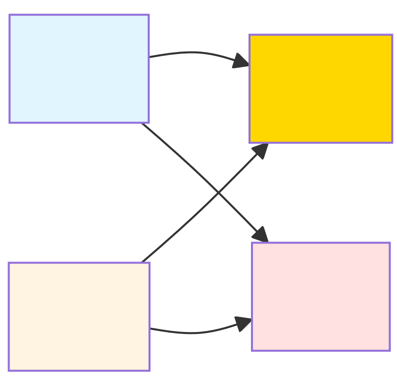

Figure 19: Mermaid diagram

Questions:

1. Is there a backdoor path from C to X?
2. Should you adjust for M?
3. Should you adjust for U?

Click to see answer

Backdoor path: $C \to M \leftarrow U \to X$

Wait! This path starts with $C \to M$, not an arrow into C. So it's NOT a backdoor path.

But there IS a backdoor path: None! (No arrows point into C)

Adjustment for M: No! M is a collider. Conditioning opens a path.

Adjustment for U: No! There's no backdoor path to block.

Correct answer: No adjustment needed (or adjust for nothing)

1.9.3 Problem 3

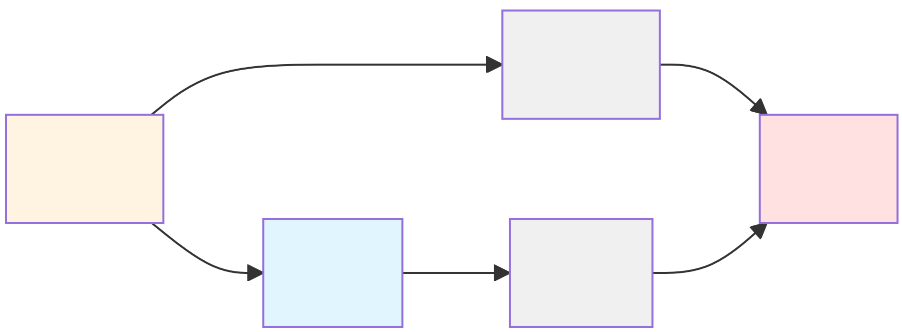

Figure 20: Mermaid diagram

Questions:

1. Identify all backdoor paths
2. What are valid adjustment sets?
3. Can you adjust for M?

Click to see answer

Backdoor path: $C \leftarrow U_1 \to U_2 \to X$

Valid adjustment sets:

- $\{U_1\}$
- $\{U_2\}$
- $\{U_1, U_2\}$

Adjust for M? No! M is a descendant of C (mediator)

1.9.4 Problem 4: Complex Graph

Questions:

1. How many backdoor paths are there?
2. What is the minimal sufficient adjustment set?
3. Is $\{U_2, Z\}$ a valid adjustment set?

Click to see answer

Backdoor paths:

1. $C \leftarrow U_1 \to U_2 \to X$
2. $C \leftarrow U_3 \to Z \to X$

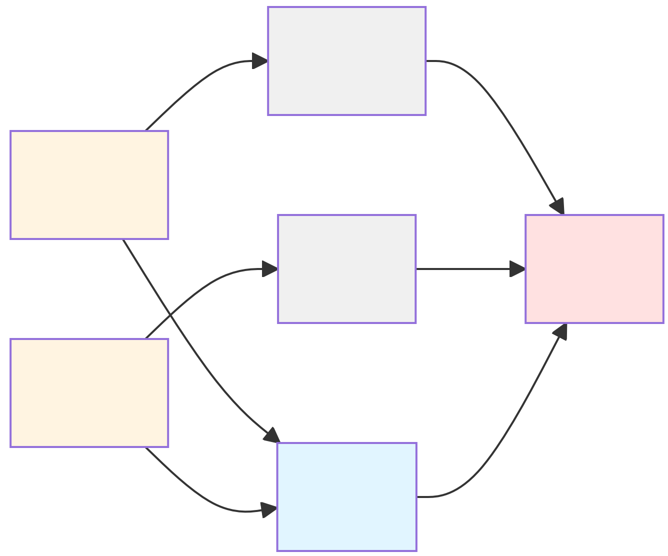

Figure 21: Mermaid diagram

Minimal sufficient adjustment sets:

- $\{U_1, U_3\}$ (blocks at the source)
- $\{U_1, Z\}$
- $\{U_2, U_3\}$
- $\{U_2, Z\}$

Is $\{U_2, Z\}$ valid? Yes! It blocks both backdoor paths.

1.9.5 Problem 5: The Butterfly

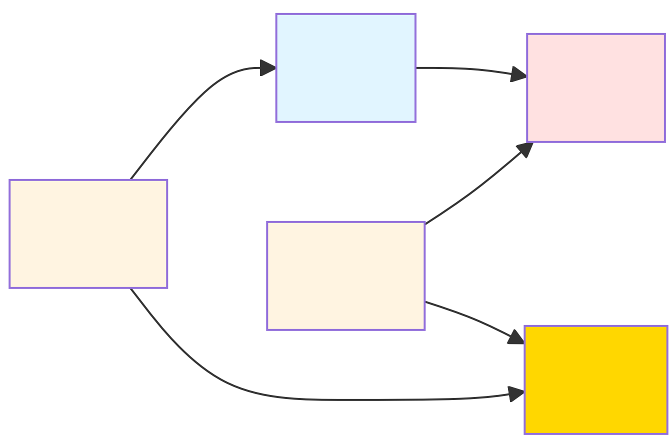

Figure 22: Mermaid diagram

Questions:

1. Is there a backdoor path?
2. Should you adjust for M?
3. What should you adjust for?

Click to see answer

Backdoor path: $C \leftarrow U_1 \rightarrow M \leftarrow U_2 \rightarrow X$

Is it blocked by default? Yes! M is a collider.

Adjust for M? No! That would OPEN the path.

Correct adjustment: No adjustment needed (the path is already blocked)

Alternative: Could adjust for U_1 or U_2 (but not necessary)

1.10 9. Common Pitfalls

1.10.1 Pitfall 1: Adjusting for Mediators

Figure 23: Mermaid diagram

Wrong: Adjust for M

Why wrong: Blocks the causal path you want to estimate!

1.10.2 Pitfall 2: Adjusting for Colliders

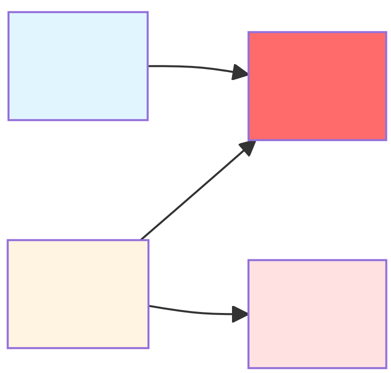

Figure 24: Mermaid diagram

Wrong: Adjust for Z

Why wrong: Opens a backdoor path!

1.10.3 Pitfall 3: Adjusting for Descendants of Treatment

Wrong: Adjust for D

Why wrong: D is a descendant of C (post-treatment variable)

Right: Adjust for U only

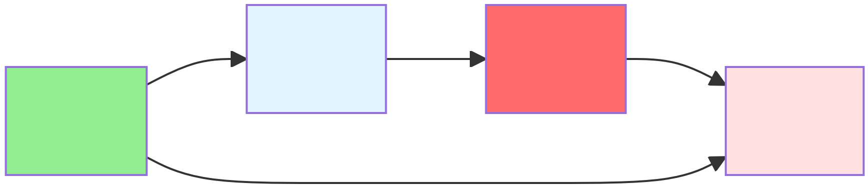

Figure 25: Mermaid diagram

1.11 10. Key Takeaways

1.11.1 The One-Sentence Summary

> **A backdoor path is any non-causal route by which observing C tells you something about other causes of X.**

Blocking backdoor paths is how we make *seeing* C behave like *doing* C.

1.11.2 The Rules (Quick Reference)

1. **Backdoor path:** Any path starting with an arrow INTO the treatment
2. **Blocking:** Condition on non-colliders, avoid colliders
3. **Backdoor criterion:**
   - Block all backdoor paths
   - Don't adjust for descendants of treatment

1.11.3 Decision Tree

1.12 11. Further Practice

1.12.1 Challenge Problem 1: The M-Graph

Identify all backdoor paths and determine valid adjustment sets.

1.12.2 Challenge Problem 2: Selection Bias

What happens if you condition on S (e.g., selecting only certain samples)?

1.12.3 Challenge Problem 3: Instrumental Variable Setup

Can you estimate the causal effect without measuring U? (Hint: This is an IV setup)

1.13 12. Next Steps

Now that you understand backdoor paths, you can explore:

1. **Front-door criterion:** What if you can't block all backdoor paths?
2. **Instrumental variables:** Using variables that affect treatment but not outcome directly
3. **Do-calculus:** The complete rules for causal identification
4. **Sensitivity analysis:** What if there are unmeasured confounders?

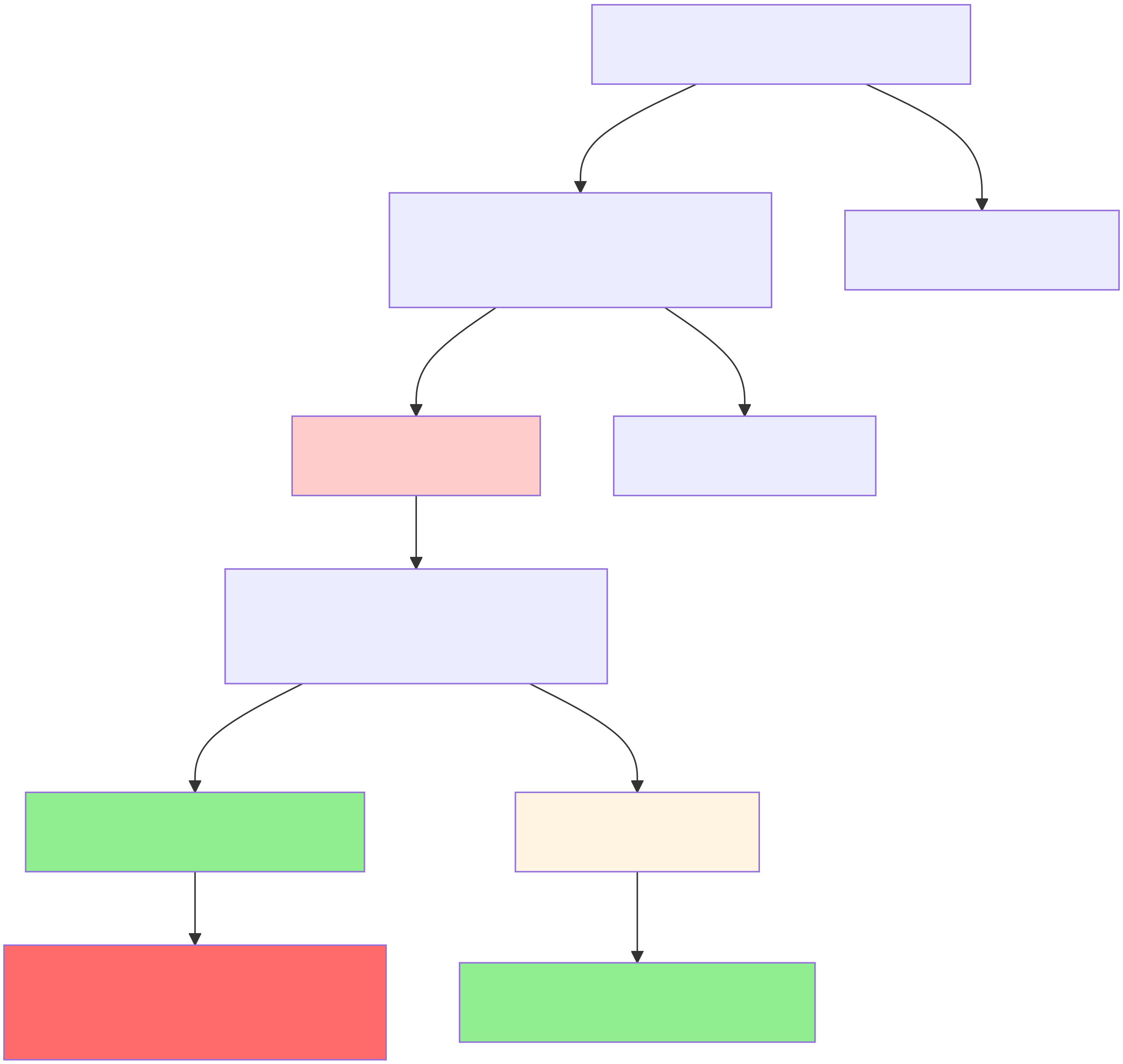

Figure 26: Mermaid diagram

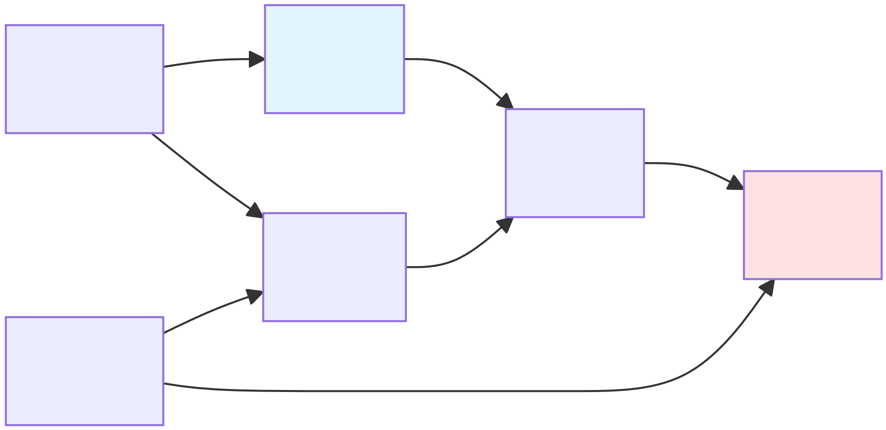

Figure 27: Mermaid diagram

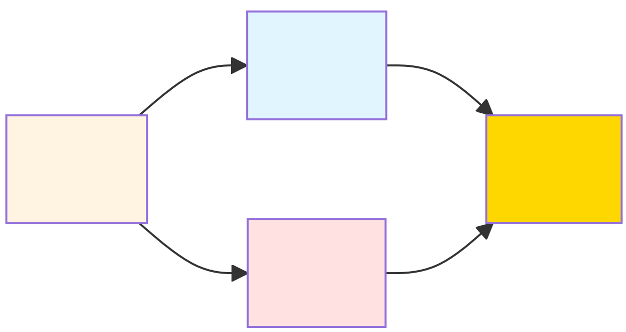

Figure 28: Mermaid diagram

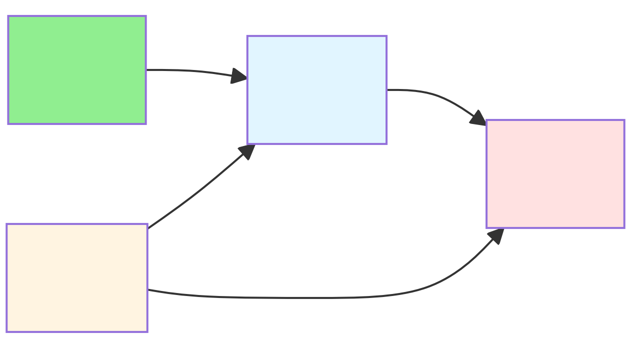

Figure 29: Mermaid diagram

1.14 References and Further Reading

- Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*
- Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal Inference in Statistics: A Primer*
- Hernán, M. A., & Robins, J. M. (2020). *Causal Inference: What If*
- Causal Inference: The Mixtape by Scott Cunningham

1.15 Appendix: Why Descendants Are Forbidden

This is subtle and important. Let's walk through it step-by-step.

1.15.1 Setup

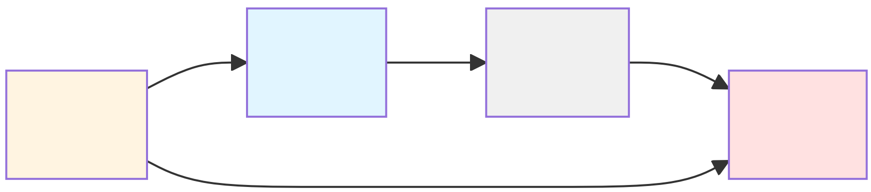

Figure 30: Mermaid diagram

1.15.2 The Question

Why can't we adjust for M (a descendant of C)?

1.15.3 The Answer

Backdoor path: $C \leftarrow U \rightarrow X$

Option 1: Adjust for U only

$$p(X \mid do(C)) = \sum_u p(X \mid C, u) \cdot p(u)$$

Correct!

Option 2: Adjust for M (wrong!)

- M is caused by C
- Conditioning on M means we're looking at different "types" of C based on its effect
- This blocks part of the causal effect we want to estimate!

1.15.4 Concrete Example

- C: Exercise program
- M: Weight loss (mediator)
- X: Blood pressure
- U: Baseline health

If we condition on weight loss (M), we're asking: "Among people who lost the same amount of weight, what's the effect of the exercise program?"

This **blocks the causal pathway** $C \rightarrow M \rightarrow X$ and underestimates the total effect!

1.15.5 The Rule

Never adjust for variables on the causal path from treatment to outcome. This includes mediators, descendants, and any post-treatment variables.