

# Chapter 20: Treatment-Confounder Feedback

## Contents

<b>1</b>	<b>20.1 The Elements of Treatment-Confounder Feedback (p. 267)</b>	<b>1</b>
1.1	The Role of Unmeasured Common Causes . . . . .	2
<b>2</b>	<b>20.2 The Bias of Traditional Methods (p. 269)</b>	<b>2</b>
2.1	A Numerical Example . . . . .	2
2.2	Direction of the Bias . . . . .	3
<b>3</b>	<b>20.3 Why Traditional Methods Fail (p. 271)</b>	<b>3</b>
3.1	A DAG-Based Explanation . . . . .	3
<b>4</b>	<b>20.4 Why Traditional Methods Cannot Be Fixed (p. 273)</b>	<b>4</b>
4.1	Attempts to Fix Traditional Regression . . . . .	4
4.2	What Would a Correct Method Need to Do? . . . . .	4
<b>5</b>	<b>20.5 Adjusting for Past Treatment (p. 274)</b>	<b>5</b>
5.1	When Adjusting for Past Treatment Is Appropriate . . . . .	5
5.2	The Way Forward . . . . .	5
<b>6</b>	<b>Summary</b>	<b>6</b>
<b>7</b>	<b>References</b>	<b>6</b>

Chapter 19 introduced the concept of time-varying confounders: covariates that change over time, are causally influenced by past treatment, and simultaneously predict future treatment and the outcome. This chapter shows exactly **why** this creates a problem for standard analytic methods — and why that problem cannot be fixed by any simple modification of those methods. Understanding this failure is the intellectual motivation for the g-methods presented in Chapter 21.

This chapter is based on Hernán and Robins (2020, chap. 20, pp. 267–275).

**Key insight:** When treatment-confounder feedback is present, any analytic approach that conditions on the time-varying confounder (e.g., standard regression, stratification, propensity score adjustment) simultaneously over-adjusts for the causal effect of past treatment and under-controls for confounding of future treatment. There is no allocation of weight to these two errors that corrects both simultaneously.

## 1 20.1 The Elements of Treatment-Confounder Feedback (p. 267)

---

**Treatment-confounder feedback** arises whenever a time-varying covariate  $L_k$  satisfies all of the following:

1.  $L_k$  is a **confounder**:  $L_k$  is an independent predictor of both future treatment  $A_k$  and the outcome  $Y$ , creating a backdoor path  $A_k \leftarrow L_k \rightarrow Y$ .
2.  $L_k$  is **affected by prior treatment**: past treatment  $A_{k-1}$  (or any  $A_j$  with  $j < k$ ) is a cause of  $L_k$ , so  $A_{k-1} \rightarrow L_k$  is a causal arrow in the DAG.

3. **Feedback continues:** the adjusted value of  $L_k$  can influence the next treatment decision  $A_k$ , which in turn affects  $L_{k+1}$ , and so on.

The causal diagram capturing two time points of feedback is:

$$A_0 \rightarrow L_1 \rightarrow A_1 \rightarrow Y, \quad A_0 \rightarrow Y, \quad L_0 \rightarrow A_0, \quad L_0 \rightarrow Y, \quad L_1 \rightarrow Y.$$

The arrow  $A_0 \rightarrow L_1$  is the feedback arrow: prior treatment changes the confounder.

**Example 1.1** (HIV Treatment and CD4 Counts). In an HIV cohort study, let  $A_k$  be antiretroviral therapy (ART) at time  $k$ ,  $L_k$  be the CD4 T-cell count at time  $k$ , and  $Y$  be death within five years.

- Physicians initiate or intensify ART when CD4 drops ( $L_k$  predicts  $A_k$ ).
- ART partially restores immune function, raising future CD4 counts ( $A_k$  causes  $L_{k+1}$ ).
- CD4 count is an independent predictor of mortality ( $L_k$  causes  $Y$ ).

All three conditions above are satisfied:  $L_k$  is a time-varying confounder with treatment feedback.

## 1.1 The Role of Unmeasured Common Causes

To appreciate why conditioning on  $L_k$  is problematic, it helps to introduce an unmeasured common cause  $U$  of  $L_k$  and  $Y$ . For example,  $U$  might be an unmeasured aspect of immune function that affects both CD4 count (and hence  $L_k$ ) and mortality (and hence  $Y$ ).

The extended DAG then contains the path

$$A_{k-1} \rightarrow L_k \leftarrow U \rightarrow Y.$$

Here,  $L_k$  is a **collider** on the path  $A_{k-1} \rightarrow L_k \leftarrow U \rightarrow Y$ . As discussed in Chapter 6, conditioning on a collider opens a previously blocked path, inducing a spurious association between  $A_{k-1}$  and  $Y$  through  $U$ .

**Why the unmeasured  $U$  matters even if we don't know it exists:**

Every causal diagram should be understood as containing all relevant common causes, measured or not. When we say “condition on  $L_k$ ”, we open the collider path through  $U$  regardless of whether  $U$  is in our dataset. The collider bias induced by conditioning on  $L_k$  is a structural feature of the problem, not a statistical artifact that can be resolved by collecting more data (unless we measure  $U$  itself).

## 2 20.2 The Bias of Traditional Methods (p. 269)

**Traditional methods** for confounding adjustment include:

- Stratification (comparing treated and untreated within levels of  $\bar{L}_k$ ).
- Multivariable regression of  $Y$  on  $\bar{A}_K$  and  $\bar{L}_K$ .
- Propensity score adjustment conditioning on  $\bar{L}_k$ .

All of these approaches **condition on the time-varying confounder  $L_k$** . We now show that this conditioning introduces bias in the presence of treatment-confounder feedback.

### 2.1 A Numerical Example

Consider a two-time-point study ( $k = 0, 1$ ) with binary treatment and binary confounder, and suppose the true causal effect of the treatment strategy “always treat” ( $\bar{a} = (1, 1)$ ) versus “never treat” ( $\bar{a} = (0, 0)$ ) on  $Y$  is zero.

If we stratify on  $L_1$  (the time-varying confounder), we find a non-zero association between  $A_0$  and  $Y$  within strata of  $L_1$ . This spurious association arises because conditioning on  $L_1$  opens the backdoor path  $A_0 \rightarrow L_1 \leftarrow U \rightarrow Y$ .

Conversely, if we do *not* stratify on  $L_1$ , we fail to control for the confounding path  $A_1 \leftarrow L_1 \rightarrow Y$ , also introducing bias. There is no strategy within the traditional regression paradigm that avoids both forms of bias simultaneously.

### Why both errors occur simultaneously:

Let us be precise. Traditional adjustment for  $\bar{L}_K = (L_0, L_1)$  causes:

1. **Collider bias from conditioning on  $L_1$ :** Since  $A_0 \rightarrow L_1$  and  $L_1 \leftarrow U \rightarrow Y$ , conditioning on  $L_1$  opens the noncausal path  $A_0 \rightarrow L_1 \leftarrow U \rightarrow Y$ , creating a spurious association between  $A_0$  and  $Y$ .
2. **Incomplete confounding control from not conditioning on  $L_1$  for  $A_1$ :** If we do not condition on  $L_1$ , the backdoor path  $A_1 \leftarrow L_1 \rightarrow Y$  remains open, creating confounding of the  $A_1 \rightarrow Y$  effect.

A traditional analyst might try to fix (2) while ignoring (1) by conditioning on  $L_1$ , thereby simultaneously creating the problem described in (1). Or they might try to avoid (1) by not conditioning on  $L_1$ , thereby failing to fix (2). The two problems are structurally inseparable.

## 2.2 Direction of the Bias

The **direction** of the bias introduced by traditional methods depends on the signs of the associations in the feedback loop. Without additional information, the bias can be in either direction — traditional methods can either underestimate or overestimate the treatment effect.

This unpredictability is particularly concerning: a naive analyst who uses traditional regression may not only miss the true effect but may estimate an effect of the wrong sign.

## 3 20.3 Why Traditional Methods Fail (p. 271)

The fundamental reason traditional methods fail is that **they cannot simultaneously adjust for time-varying confounding and preserve the causal effect of prior treatment.**

To estimate the total effect of the strategy “always treat” versus “never treat”, we need to include the causal path  $A_0 \rightarrow L_1 \rightarrow A_1 \rightarrow Y$  — that is, the effect of  $A_0$  working through  $L_1$  on  $A_1$  and then on  $Y$ . But to control confounding of the  $A_1 \rightarrow Y$  relationship, we feel compelled to condition on  $L_1$ , which blocks the path  $A_0 \rightarrow L_1 \rightarrow \dots \rightarrow Y$  and introduces collider bias.

**Proposition 3.1** (Traditional Methods Fail under Treatment-Confounder Feedback). *Let the causal DAG contain the path  $A_0 \rightarrow L_1 \leftarrow U \rightarrow Y$  and the confounding path  $A_1 \leftarrow L_1 \rightarrow Y$ . Then no estimator that conditions on  $L_1$  as a standard covariate (in a regression or stratification) can consistently estimate  $E[Y^{\bar{a}=1}] - E[Y^{\bar{a}=0}]$ .*

### 3.1 A DAG-Based Explanation

The failure can be visualized on the causal DAG. Suppose the full DAG (including unmeasured  $U$ ) is:

$$L_0 \rightarrow A_0 \rightarrow L_1 \leftarrow U \rightarrow Y, \quad L_1 \rightarrow A_1 \rightarrow Y, \quad A_0 \rightarrow Y, \quad L_0 \rightarrow Y.$$

The treatment effect of interest is mediated through multiple paths:  $A_0 \rightarrow Y$  (direct) and  $A_0 \rightarrow A_1 \rightarrow Y$  (through its effect on  $A_1$  via  $L_1$ ).

Now consider what happens when we condition on  $L_1$  in a regression:

- We block the path  $A_0 \rightarrow L_1 \rightarrow A_1 \rightarrow Y$  (over-adjustment).
- We open the path  $A_0 \rightarrow L_1 \leftarrow U \rightarrow Y$  (collider bias).

Both effects distort the estimated association between  $A_0$  and  $Y$ . The g-methods of Chapter 21 avoid this by not conditioning on  $L_1$  as a standard covariate, but instead using it in a more structured way that respects the feedback.

**Historical note:** The problem of time-varying confounders was first formally analyzed by James Robins in the 1980s. He showed that the standard methods produce inconsistent estimates and developed the g-computation formula (g-formula) as a solution. The term “g” stands for “generalized” — the g-formula generalizes the standard adjustment formula to the longitudinal setting with feedback.

## 4 20.4 Why Traditional Methods Cannot Be Fixed (p. 273)

One might hope that a more sophisticated version of traditional regression — perhaps including interaction terms, polynomial terms, or machine learning — could solve the time-varying confounding problem. This hope is misguided.

The failure of traditional methods is not due to model misspecification or insufficient flexibility. It is a **structural** failure: the data cannot be analyzed correctly using any approach that conditions on  $L_k$  as a standard covariate.

### 4.1 Attempts to Fix Traditional Regression

Several “fixes” have been proposed and shown to be inadequate:

**Adding lagged treatments:** Including  $A_0, A_1$  and  $L_0, L_1$  in the same regression of  $Y$  does not solve the problem because it still conditions on  $L_1$  simultaneously with  $A_0$  and  $A_1$ , blocking the causal path and opening the collider.

**Marginal models without adjustment:** Fitting a marginal regression of  $Y$  on  $A_0$  and  $A_1$  without adjusting for  $L_k$  fails to control confounding, since  $L_k$  is a confounder for  $A_k$ .

**Fixed-effects panel models:** These control for time-invariant unmeasured confounders but do not address time-varying confounders that are affected by past treatment.

**Instrumental variable methods:** Require instruments that are hard to find and do not directly address the structural issue with treatment-confounder feedback.

#### **i** Fine Point 20.1: The Structural Problem Is Not About Sample Size

A common misconception is that the bias of traditional methods in the presence of treatment-confounder feedback is a finite-sample problem that disappears with large enough datasets. This is incorrect. The bias is present even in infinite samples because it arises from the structural relationship between variables in the DAG, not from estimation error. Doubling the sample size doubles our precision in estimating the wrong quantity. The only remedy is to use an estimation strategy that is structurally correct — the g-methods of Chapter 21.

### 4.2 What Would a Correct Method Need to Do?

A method that correctly handles treatment-confounder feedback must:

1. **Use  $L_k$  to adjust for confounding of  $A_k$**  — so  $L_k$  cannot simply be ignored.
2. **Not condition on  $L_k$  as a standard covariate in estimating the effect of  $A_{k-1}$**  — to avoid blocking the causal path and opening the collider.
3. **Account for the entire treatment and covariate history** — through a model that respects the sequential, recursive nature of the data-generating process.

The g-formula, IP weighting with marginal structural models, and g-estimation each accomplish these goals through different mechanisms, as described in Chapter 21.

## 5 20.5 Adjusting for Past Treatment (p. 274)

A specific manifestation of the problem is the question of whether and how to include **past treatment** in a regression model for the outcome.

Consider a regression of  $Y$  on  $(A_0, A_1, L_0, L_1)$ . This model adjusts for both time points of treatment and both time points of the covariate. Is this a valid approach?

**i** Remark (Adjusting for Past Treatment in Regression)

*Remark 5.1* (Adjusting for Past Treatment in Regression). Including  $A_0$  in a regression model for  $Y$  that also includes  $L_1$  creates the following problem:

- $L_1$  is a descendant of  $A_0$ .
- Conditioning on  $L_1$  therefore partitions the effect of  $A_0$  into a component that acts “through  $L_1$ ” and a component that acts “not through  $L_1$ .”
- The regression coefficient on  $A_0$  estimates a **controlled direct effect** of  $A_0$ , not the total causal effect.
- This controlled direct effect is not the estimand of interest when the goal is to compare treatment strategies.

### 5.1 When Adjusting for Past Treatment Is Appropriate

Adjusting for past treatment *is* appropriate in specific circumstances:

- When the research question is about the **incremental effect** of a single time point’s treatment, holding all other time points constant. For example, “what is the effect of adding treatment at time 1 among individuals who received treatment at time 0?” This is a different (and more limited) causal question than comparing strategies.
- When all time-varying covariates are independent of past treatment (i.e., there is no feedback), a standard regression with  $(\bar{A}_{K-1}, \bar{L}_K)$  can yield valid estimates of direct effects. But this assumption is rarely plausible.

### 5.2 The Way Forward

The fundamental lesson of Chapter 20 is that standard analytic tools, designed for cross-sectional or time-fixed settings, are not equipped to handle the recursive causal structure of longitudinal data with treatment-confounder feedback.

The solution — g-methods — are designed from the ground up with the sequential nature of the data in mind. They correctly separate the roles of  $L_k$ : using it to remove confounding of  $A_k$  while preserving the causal path  $A_{k-1} \rightarrow L_k \rightarrow Y$ .

**Summary of why each traditional approach fails:**

Method	Why it fails
Stratification on $\bar{L}_K$	Opens collider paths via $A_{k-1} \rightarrow L_k \leftarrow U$
Regression on $(\bar{A}_K, \bar{L}_K)$	Same as stratification; estimates controlled direct effects, not total effects
Marginal regression (no $L$ adjustment)	Does not control confounding of $A_k$ by $L_k$
Fixed-effects models	Address time-invariant confounders only
Lagged-variable models	Still condition on descendants of past treatment

**The g-methods** break out of this table by using a different conceptual approach: the g-formula uses standardization that respects the temporal ordering; IP weighting creates a pseudo-population where treatment is independent of confounders; g-estimation models the treatment-free counterfactual directly. All three are discussed in Chapter 21.

## 6 Summary

---

- **Treatment-confounder feedback** exists when  $L_k$  is simultaneously a cause of  $A_k$  (confounding) and a descendant of  $A_{k-1}$  (feedback).
- When  $L_k$  is also affected by unmeasured  $U$ , conditioning on  $L_k$  opens the collider path  $A_{k-1} \rightarrow L_k \leftarrow U \rightarrow Y$ , introducing bias.
- **Traditional methods** (stratification, regression, propensity score adjustment) all condition on  $\bar{L}_K$  as a standard covariate, making them structurally unable to produce consistent estimates of strategy effects.
- This failure is **structural**, not statistical: it persists in large samples and cannot be fixed by adding flexibility or interactions to a traditional regression model.
- **Adjusting for past treatment** in a regression that also adjusts for future covariates estimates a controlled direct effect, not the total causal effect of a strategy.
- The solution is to use **g-methods** (Chapter 21), which are specifically designed to respect the recursive causal structure of longitudinal data with feedback.

## 7 References

---

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