Chapter 20
TREATMENT-CONFOUNDER FEEDBACK

The previous chapter identified sequential exchangeability as a key condition to identify the causal effects of time-varying treatments. Suppose that we have a study in which the strongest form of sequential exchangeability holds: the measured time-varying confounders are sufficient to validly estimate the causal effect of any treatment strategy. Then the question is what confounding adjustment method to use. The answer to this question highlights a key problem in causal inference about time-varying treatments: treatment-confounder feedback.

When treatment-confounder feedback exists, using traditional adjustment methods may introduce bias in the effect estimates. That is, even if we had all the information required to validly estimate the average causal effect of any treatment strategy, we would be generally unable to do so. This chapter describes the structure of treatment-confounder feedback and the reasons why traditional adjustment methods fail.

20.1 The elements of treatment-confounder feedback

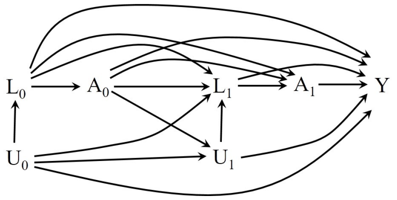

Figure 20.1

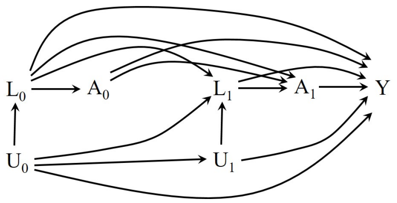

Figure 20.2

Consider again the sequentially randomized trial of individuals with HIV that we discussed in the previous chapter. For every person in the study, we have data on treatment A_k (1: treated, 0: untreated) and covariates L_k at each month of follow-up $k = 0, 1, 2...K$, and on an outcome Y that measures health status at month $K + 1$. The causal diagram in Figure 20.1, which is equal to the one in Figure 19.2, represents the first two months of the study. The time-varying covariates L_k are time-varying confounders. (As in the previous chapter, we are using this example without censoring so that we can focus on confounding.)

Something else is going on in Figure 20.1. Not only is there an arrow from CD4 cell count L_k to treatment A_k, but also there is an arrow from treatment A_{k-1} to future CD4 cell count L_k—because receiving treatment A_{k-1} increases future CD4 cell count L_k. That is, the confounder affects the treatment *and* the treatment affects the confounder. There is *treatment-confounder feedback* (see also Fine Point 20.1).

Note that time-varying confounding can occur without treatment-confounder feedback. The causal diagram in Figure 20.2. is the same as the one in Figure 20.1, except that the arrows from treatment A_{k-1} to future L_k and U_k have been deleted. In a setting represented by this diagram, the time-varying covariates L_k are time-varying confounders, but they are not affected by prior treatment. Therefore, there is time-varying confounding, but there is no treatment-confounder feedback.

Treatment-confounder feedback creates an interesting problem for causal inference. To state the problem in its simplest form, let us simplify the causal diagram in Figure 20.1 a bit more. Figure 20.3 is the smallest subset of Figure 20.1 that illustrates treatment-confounder feedback in a sequentially randomized trial with two time points. When drawing the causal diagram in Figure 20.3, we made four simplifications:

- Because our interest is in the implications of confounding by L_1, we

Fine Point 20.1

Representing feedback cycles with acyclic graphs. Interestingly, an *acyclic* graph—like the one in Figure 20.1—can be used to represent a treatment-confounder feedback loop or *cycle*. The trick to achieve this visual representation is to elaborate the treatment-confounder feedback loop in time. That is, $A_{k-1} \to L_k \to A_k \to L_{k+1}$ and so on.

The representation of feedback cycles with acyclic graphs also requires that time be considered as a discrete variable. That is, we say that treatment and covariates can change during each interval $[k, k+1)$ for $k = 0, 1, ...K$, but we do not specify when exactly during the interval the change takes place. This discretization of time is not a limitation in practice: the length of the intervals can be chosen to be as short as the granularity of the data requires. For example, in a study where individuals see their doctors once per month or less frequently (as in our HIV example), time may be safely discretized into month intervals. In other cases, year intervals or day intervals may be more appropriate. Also, as we said in Chapter 17, time is typically measured in discrete intervals (years, months, days) any way, so the discretization of time is often not even a choice.

did not bother to include a node L_0 for baseline CD4 cell count. Just suppose that treatment A_0 is marginally randomized and treatment A_1 is conditionally randomized given L_1.

- The unmeasured variable U_0 is not included.

- There is no arrow from A_0 to A_1, which implies that treatment is assigned using information on L_1 only.

- There are no arrows from A_0, L_1 and A_1 to Y, which would be the case if treatment has no causal effect on the outcome Y of any individual, i.e., the sharp null hypothesis holds.

None of these simplifications affect the arguments below. A more complicated causal diagram would not add any conceptual insights to the discussion in this chapter; it would just be harder to read.

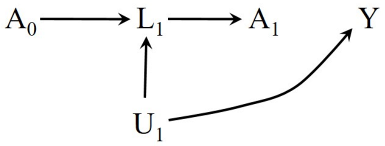

Figure 20.3

Now suppose that treatment has no effect on any individual's Y, which implies the causal diagram in Figure 20.3 is the correct one, but the investigators do not know it. Also suppose that we have data on treatment A_0 in month 0 and A_1 in month 1, on the confounder CD4 cell count L_1 at the start of month 1, and on the outcome Y at the end of follow-up. We wish to use these data to estimate the average causal effect of the static treatment strategy "always treat", $(a_0 = 1, a_1 = 1)$, compared with the static treatment strategy "never treat", $(a_0 = 0, a_1 = 0)$ on the outcome Y, i.e., $\mathrm{E}\left[Y^{a_0=1,a_1=1}\right] - \mathrm{E}\left[Y^{a_0=0,a_1=0}\right]$. According to Figure 20.3, the true, but unknown to the investigator, average causal effect is 0 because there are no forward-directed paths from either treatment variable to the outcome. That is, one cannot start at either A_0 or A_1 and, following the direction of the arrows, arrive at Y.

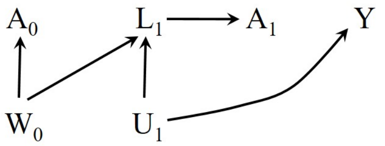

Figure 20.4

Figure 20.3 can depict a sequentially randomized trial because there are no direct arrows from the unmeasured U into the treatment variables. Therefore, as we discussed in the previous chapter, we should be able to use the observed data on A_0, L_1, A_1, and Y to conclude that $\mathrm{E}\left[Y^{a_0=1,a_1=1}\right] - \mathrm{E}\left[Y^{a_0=0,a_1=0}\right]$ is equal to 0. However, as we explain in the next section, we will not generally be able to correctly estimate the causal effect when we adjust for L_1 using traditional methods, like stratification, outcome regression, and matching. That is, in this example, an attempt to adjust for the confounder L_1 using these methods will generally result in an effect estimate that is different from 0, and thus invalid.

Figure 20.3 represents either a sequentially randomized trial or an observational study with no unmeasured confounding; Figure 20.4 represents an observational study.

In other words, when there are time-varying confounders and treatment-confounder feedback, traditional methods cannot be used to correctly adjust for those confounders. Even if we had sufficient longitudinal data to ensure sequential exchangeability, traditional methods would not generally provide a valid estimate of the causal effect of any treatment strategies. In contrast, g-methods appropriately adjust for the time-varying confounders even in the presence of treatment-confounder feedback.

This limitation of traditional methods applies to settings in which the time-varying confounders are affected by prior treatment as in Figure 20.3, but also to settings in which the time-varying confounders share causes W with prior treatment as in Figure 20.4, which is a subset of Figure 19.4. We refer to both Figures 20.3 and 20.4 (and Figures 19.2 and 19.4) as examples of treatment-confounder feedback. The next section explains why traditional methods cannot adequately handle treatment-confounder feedback.

20.2 The bias of traditional methods

This is an ideal trial with full adherence to the assigned treatment strategy and no losses to follow-up.

To illustrate the bias of traditional methods, let us consider a (hypothetical) sequentially randomized trial with 32,000 individuals with HIV and two time points $k = 0$ and $k = 1$. Treatment $A_0 = 1$ is randomly assigned at baseline with probability 0.5. Treatment A_1 is randomly assigned in month 1 with a probability that depends only on the value of CD4 cell count L_1 at the start of month 1—0.4 if $L_1 = 0$ (high), 0.8 if $L_1 = 1$ (low). The outcome Y, which is measured at the end of follow-up, is a function of CD4 cell count, concentration of virus in the serum, and other clinical measures, with higher values of Y signifying better health.

Table 20.1

N	A_0	L_1	A_1	Mean Y
2400	0	0	0	84
1600	0	0	1	84
2400	0	1	0	52
9600	0	1	1	52
4800	1	0	0	76
3200	1	0	1	76
1600	1	1	0	44
6400	1	1	1	44

Table 20.1 shows the data from this trial. To save space, the table displays one row per combination of values of A_0, L_1, and A_1, rather than one row per individual. For each of the eight combinations, the table provides the number of subjects N and the mean value of the outcome $\mathrm{E}[Y|A_0, L_1, A_1]$. Thus, row 1 shows that the mean of the 2400 individuals with $(A_0 = 0, L_1 = 0, A_1 = 0)$ was $\mathrm{E}[Y|A_0 = 0, L_1 = 0, A_1 = 0] = 84$. In this sequentially randomized trial, the identifiability conditions—sequential exchangeability, positivity, consistency—hold. By design, there are no confounders for the effect of A_0 on Y, and L_1 is the only confounder for the effect of A_1 on Y so (conditional on L_1) sequential exchangeability holds. By inspection of Table 20.1, we can conclude that the positivity condition is satisfied, because otherwise one or more of the eight rows would have zero individuals.

If there were additional times k at which treatment A_k were affected by L_k, then L_k would be a time-varying confounder

The causal diagram in Figure 20.3 depicts this sequentially randomized experiment when the sharp null hypothesis holds. To check whether the data in Table 20.1 are consistent with the causal diagram in Figure 20.3, we can separately estimate the average causal effects of each of the time-fixed treatments A_0 and A_1 within levels of past covariates and treatment, which should all be null. In the calculations below, we will ignore random variability.

Figure 20.3 represents the null because there is no arrow from L_1 to Y. Otherwise, A_0 would have an effect on Y through L_1

A quick inspection of the table shows that the average causal effect of treatment A_1 is indeed zero in all four strata defined by A_0 and L_1. Consider the effect of A_1 in the 4000 individuals with $A_0 = 0$ and $L_1 = 0$, whose data are shown in rows 1 and 2 of Table 20.1. The mean outcome among those who did not receive treatment at time 1, $\mathrm{E}[Y|A_0 = 0, L_1 = 0, A_1 = 0]$, is 84, and the mean outcome among those who did receive treatment at time 1,

Technical Point 20.1

G-null test. Suppose the sharp null hypothesis is true. Then any counterfactual outcome Y^g is the observed outcome Y. In this setting, sequential exchangeability for all Y^g can be written as $Y \perp\!\!\!\perp A_0 | L_0$ and $Y \perp\!\!\!\perp A_1 | A_0, L_0, L_1$ in a study with two time points. (We have used the fact that, for any values of a_0 and l_0, there exist strategies g such that $a_0 = g(l_0)$.) Therefore, under sequential exchangeability, a test of these conditional independencies is a test of the sharp null. This is the g-null test (Robins 1986). Note the first independence implies no causal effect of A_0 in any strata defined by L_0, and the second independence implies no causal effect of A_1 in any strata defined by L_1 and A_0.

More generally, the g-null theorem of Robins (1986) says that, under sequential randomization for all g, the above two independencies hold if and only if the distribution of Y^g and therefore the mean $\mathrm{E}[Y^g]$ is the same for all g, and also equal to the distribution and mean of the observed Y.

$\mathrm{E}[Y | A_0 = 0, L_1 = 0, A_1 = 1]$, is also 84. Therefore the difference

$$\mathrm{E}[Y | A_0 = 0, L_1 = 0, A_1 = 1] - \mathrm{E}[Y | A_0 = 0, L_1 = 0, A_1 = 0]$$

is zero. Because the identifiability conditions hold, this associational difference validly estimates the average causal effect

$$\mathrm{E}\left[Y^{a_1=1} | A_0 = 0, L_1 = 0\right] - \mathrm{E}\left[Y^{a_1=0} | A_0 = 0, L_1 = 0\right]$$

in the stratum $(A_0 = 0, L_1 = 0)$. Similarly, it is easy to check that the average causal effect of treatment A_1 on Y is zero in the remaining three strata $(A_0 = 0, L_1 = 1)$, $(A_0 = 1, L_1 = 0)$, $(A_0 = 1, L_1 = 1)$, by comparing the mean outcome between rows 3 and 4, rows 5 and 6, and rows 7 and 8, respectively.

We can now show that the average causal effect of A_0 is also zero. To do so, we need to compute the associational difference $\mathrm{E}[Y | A_0 = 1] - \mathrm{E}[Y | A_0 = 0]$ which, because of randomization, is a valid estimator of the causal contrast $\mathrm{E}\left[Y^{a_0=1}\right] - \mathrm{E}\left[Y^{a_0=0}\right]$. The mean outcome $\mathrm{E}[Y | A_0 = 0]$ among the 16,000 individuals treated at time 0 is the weighted average of the mean outcomes in rows 1, 2, 3 and 4, which is 60. And $\mathrm{E}[Y | A_0 = 1]$, computed analogously, is also 60. Therefore, the average causal effect of A_0 is zero.

The weighted average is $\frac{2400}{16000} \times 84 + \frac{1600}{16000} \times 84 + \frac{2400}{16000} \times 52 + \frac{9600}{16000} \times 52 = 60$

We have confirmed that the causal effects of A_0 and A_1 (conditional on the past) are zero when we treat A_0 and A_1 separately as time-fixed treatments. What if we now treat the joint treatment (A_0, A_1) as a time-varying treatment and compare two treatment strategies? For example, let us say that we want to compare the strategies "always treat" versus "never treat", that is $(a_0 = 1, a_1 = 1)$ versus $(a_0 = 0, a_1 = 0)$. Because the identifiability conditions hold, the data in Table 20.1 should suffice to validly estimate this effect.

Because the effect for each of the individuals components of the strategy, a_0 and a_1, is zero, it follows from the g-null theorem (see Technical Point 20.1) that the average causal effect $\mathrm{E}\left[Y^{a_0=1,a_1=1}\right] - \mathrm{E}\left[Y^{a_0=0,a_1=0}\right]$ is zero. But is this what we conclude from the data if we use conventional analytic methods? To answer this question, let us conduct two data analyses. In the first one, we do not adjust for the confounder L_1, which should give us an incorrect effect estimate. In the second one, we do adjust for the confounder L_1 via stratification.

1. We compare the mean outcome in the 9600 individuals who were treated at both times (rows 6 and 8 of Table 20.1) with that in the 4800 individuals who were untreated at both times (rows 1 and 3). The respective averages are $\mathrm{E}[Y | A_0 = 1, A_1 = 1] = 54.7$, and $\mathrm{E}[Y | A_0 = 0, A_1 = 0] =$

$\mathrm{E}[Y|A_0 = 1, A_1 = 1]$
$\frac{3200}{9600} \times 76 + \frac{6400}{9600} \times 44 = 54.7$
$\mathrm{E}[Y|A_0 = 0, A_1 = 0]$
$\frac{2400}{4800} \times 84 + \frac{2400}{4800} \times 52 = 68.0$

68. The associational difference is $54.7 - 68 = -13.3$ which, if interpreted causally, would mean that not being treated at either time is better than being treated at both times. This analysis gives the wrong answer—a non-null difference—because $\mathrm{E}[Y|A_0 = a_0, A_1 = a_1]$ is not a valid estimator of $\mathrm{E}[Y^{a_0,a_1}]$. Adjustment for the confounder L_1 is needed.

2. We adjust for L_1 via stratification. That is, we compare the mean outcome in individuals who were treated with that in individuals who were untreated at both times, within levels of L_1. For example, take the stratum $L_1 = 0$. The mean outcome in the treated at both times, $\mathrm{E}[Y|A_0 = 1, L_1 = 0, A_1 = 1]$, is 76 (row 6). The mean outcome in the untreated at both times, $\mathrm{E}[Y|A_0 = 0, L_1 = 0, A_1 = 0]$, is 84 (row 1). The associational difference is $76 - 84 = -8$ which, if interpreted causally, would mean that, in the stratum $L_1 = 0$, not being treated at either time is better than being treated at both times. Similarly, the difference $\mathrm{E}[Y|A_0 = 1, L_1 = 1, A_1 = 1] - \mathrm{E}[Y|A_0 = 0, L_1 = 1, A_1 = 0]$ in the stratum $L_1 = 1$ is also -8.

Note that, because the effect is -8 in both strata of L_1, it is not possible that a weighted average of the stratum-specific effects will yield the correct value 0.

What? We said that the effect estimate should be 0, not -8. How is it possible that the analysis adjusted for the confounder also gives a wrong answer? This estimate reflects the bias of traditional methods to adjust for confounding when there is treatment-confounder feedback. The next section explains why the bias arises.

20.3 Why traditional methods fail

Table 20.1 shows data from a sequentially randomized trial with treatment-confounder feedback, as represented by the causal diagram in Figure 20.3. Even though no data on the unmeasured variable U_1 (immunosuppression level) is available, all three identifiability conditions hold: U_1 is not needed if we have data on the confounder L_1. Therefore, as discussed in Chapter 19, we should be able to correctly estimate causal effects involving any static or dynamic treatment strategies. And yet our analyses in the previous section did not yield the correct answer, whether or not we adjusted for L_1.

The problem was that we did not use the correct method to adjust for confounding. Stratification is a commonly used method to adjust for confounding, but it cannot handle treatment-confounder feedback. Stratification means estimating the association between treatment and outcome in subsets—strata—of the study population defined by the confounders—L_1 in our example. Because the variable L_1 can take only two values—1 if the CD4 cell count is low, and 0 otherwise—there are two such strata in our example. To estimate the causal effect in those with $L_1 = l$, we selected (i.e., conditioned or stratified on) the subset of the population with value $L_1 = l$.

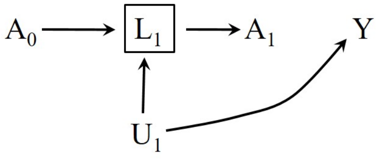

Figure 20.5

But stratification can have unintended effects when the association measure is computed within levels of a variable L_1 that is caused by prior treatment A_0. Indeed Figure 20.5 shows that conditioning on L_1—a collider—opens the path $A_0 \longrightarrow L_1 \longleftarrow U_1 \longrightarrow Y$. That is, stratification induces a noncausal association between the treatment A_0 at time 0 and the unmeasured variable U_1, and therefore between A_0 and the outcome Y, within levels of L_1. Among those with low CD4 count ($L_1 = 1$), being on treatment ($A_0 = 1$) becomes a marker for severe immunosuppression (high value of U_1); among those with a high level

Fine Point 20.2

Confounders on the causal pathway. Conditioning on confounders L_1 which are affected by previous treatment can create selection bias even if the confounder is not on a causal pathway between treatment and outcome. In fact, no such causal pathway exists in Figures 20.5 and 20.6.

On the other hand, in Figure 20.7 the confounder L_1 for subsequent treatment A_1 lies on a causal pathway from earlier treatment A_0 to outcome Y, i.e., the path $A_0 \longrightarrow L_1 \longrightarrow Y$. If U_1 were not a common cause of L_1 and Y in Figure 20.7 (i.e., if there were no selection bias), the A-Y associations within strata of L_1 would be an unbiased estimate of the direct effects of A_0 on Y not through L_1, but still would not be an unbiased estimate of the overall effect of $\bar{A}$ on Y, because the effect of A_0 mediated through L_1 is not included.

It is sometimes said that variables on a causal pathway between treatment and outcome cannot be considered as confounders, because adjusting for those variables will result in a biased effect estimate. However, this characterization of confounders is inaccurate for time-varying treatments. Figure 20.7 shows that a confounder for subsequent treatment A_1 can be on a causal pathway between past treatment A_0 and the outcome. As for whether adjustment for confounders on a causal pathway induces bias for the effect of a treatment strategy, that depends on the choice of adjustment method. Stratification will indeed induce bias; g-methods will not.

of CD4 ($L_1 = 0$), being off treatment ($A_0 = 0$) becomes a marker for milder immunosuppression (low value of U_1). Thus, the side effect of stratification is to induce an association between treatment A_0 and outcome Y.

In other words, stratification eliminates confounding for A_1 at the cost of introducing selection bias for A_0. The associational differences

$$\mathrm{E}\left[Y|A_0 = 1, L_1 = l, A_1 = 1\right] - \mathrm{E}\left[Y|A_0 = 0, L_1 = l, A_1 = 0\right]$$

may be different from 0 even if, as in our example, treatment has no effect on the outcome of any individuals at any time. This bias arises from choosing a subset of the study population by selecting on a variable L_1 affected by (a component A_0 of) the time-varying treatment. The net bias depends on the relative magnitude of the confounding that is eliminated and the selection bias that is created.

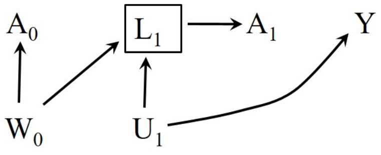

Figure 20.6

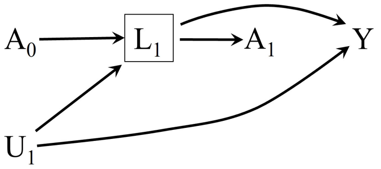

Figure 20.7

Technically speaking, the bias of traditional methods will occur not only when the confounders are affected by prior treatment (in randomized experiments or observational studies), but also when the confounders share an unmeasured cause W_0 with prior treatment (in observational studies). In the observational study depicted in Figure 20.6, conditioning on the collider L_1 opens the path $A_0 \longleftarrow W_0 \longrightarrow L_1 \longleftarrow U_1 \longrightarrow Y$. For this reason, we referred to both settings in Figures 20.3 and 20.4—which cannot be distinguished using the observed data—as examples of treatment-confounder feedback.

The causal diagrams that we have considered to describe the bias of traditional methods are all very simple. They only represent settings in which treatment does not have a causal effect on the outcome. However, conditioning on a confounder in the presence of treatment-confounder feedback also induces bias when treatment has a non-null effect, as in Figure 20.7. The presence of arrows from A_0, A_1, or L_1 to Y does not change the fact that conditioning on L_1 creates an association between A_0 and Y that does not have a causal interpretation (see also Fine Point 20.2). Also, our causal diagrams had only two time points and a limited number of nodes, but the bias of traditional methods will also arise from high-dimensional data with multiple time points and variables. In fact, the presence of time-varying confounders affected by previous treatment at multiple times increases the possibility of a large bias.

In general, valid estimation of the effect of treatment strategies is only possible when the joint effect of the treatment components A_k can be estimated simultaneously and without bias. As we have just seen, this may be impossible to achieve using stratification, even when data on all time-varying confounders are available.

20.4 Why traditional methods cannot be fixed

We showed that stratification cannot be used as a confounding adjustment method when there is treatment-confounder feedback. But what about other traditional methods? For example, we could have used parametric outcome regression, rather than nonparametric stratification, to adjust for confounding. Would outcome regression succeed where plain stratification failed?

This question is particularly important for settings with high-dimensional data, because in high-dimensional settings we will be unable to conduct a simple stratified analysis like we did in the previous section. Consider data generated under Figure 20.5. Treatment A_k occurs at two months $k = 0, 1$, which means that there are only $2^2 = 4$ static treatment strategies $\bar{a}$. But when the treatment A_k occurs at multiple points $k = 0, 1...K$, we will not be able to present a table with all the combinations of treatment values. If, as is not infrequent in practice, K is of the order of 100, then there are 2^{100} static treatment strategies $\bar{a}$, a staggering number that far exceeds the sample size of any study. The total number of treatment strategies is much greater when we consider dynamic strategies as well.

The number of data combinations is even greater because there are multiple confounders L_k measured at each time point k.

As we have been arguing since Chapter 11, we will need modeling to estimate average causal effects involving $\mathrm{E}\left[Y^{\bar{a}}\right]$ when there are many possible treatment strategies $\bar{a}$. To do so, we will need to hypothesize a dose-response function for the effect of treatment history $\bar{a}$ on the mean outcome Y. One possibility would be to assume that the effect of treatment strategies $\bar{a}$ increases linearly as a function of the cumulative treatment under each strategy. Under this assumption, all strategies that assign treatment for exactly three months have the same effect, regardless of the period when those three months of treatment occur the first 3 months of follow-up, the last 3 months of follow-up, etc. The price paid for modeling is yet another threat to the validity of our estimates due to possible model misspecification of the dose-response function.

And yet paying this price does not buy any protection against the failure of traditional methods. In the presence of treatment-confounder feedback, regression modeling cannot possibly remove the bias of conventional stratification-based methods because regression is a conventional stratification-based method itself. For example, suppose that we have data generated under Figure 20.5. Let us define cumulative treatment $cum\left(\bar{A}\right) = A_0 + A_1$, which can take 3 values: 0 (if the individuals remains untreated at both times), 1 (if the subject is treated at time 1 only or at time 2 only), and 2 (if the subject is treated at both times). The treatment strategies of interest can then be expressed as "always treat" $cum\left(\bar{a}\right) = 2$, and "never treat" $cum\left(\bar{a}\right) = 0$, and the average causal effect as $\mathrm{E}\left[Y^{cum(\bar{a})=2}\right] - \mathrm{E}\left[Y^{cum(\bar{a})=0}\right]$. Again, any valid method should estimate that the value of this difference is 0.

Under the assumption that the mean outcome $\mathrm{E}\left[Y|\bar{A}, L_1\right]$ depends linearly on the covariate $cum\left(\bar{A}\right)$, we could fit the outcome regression model

$$\mathrm{E}\left[Y|\bar{A}, L_1\right] = \theta_0 + \theta_1 cum\left(\bar{A}\right) + \theta_2 L_1$$

The associational difference $\mathrm{E}\left[Y|cum\left(\bar{A}\right)=2,L_1\right]-\mathrm{E}\left[Y|cum\left(\bar{A}\right)=0,L_1\right]$ is equal to $\theta_1\times 2$. (The model correctly assumes that the difference is the same in the strata $L_1=1$ and $L_1=0$.) Therefore some might want to interpret $\theta_1\times 2$ as the average causal effect of "always treat" versus "never treat" within levels of the covariate L_1. But such causal interpretation is unwarranted because, as Figure 20.5 shows, conditioning on L_1 induces an association between A_0, a component of treatment $cum\left(\bar{A}\right)$, and the outcome Y. This implies that θ_1—and therefore the associational difference of means—is non-zero even if the true causal effect is zero and the regression model for $\mathrm{E}\left[Y|\bar{A},L_1\right]$ is correct. A similar argument can be applied to matching. G-methods are needed to appropriately adjust for time-varying confounders in the presence of treatment-confounder feedback.

20.5 Adjusting for past treatment

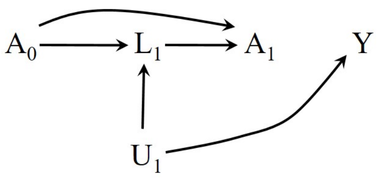

Figure 20.8

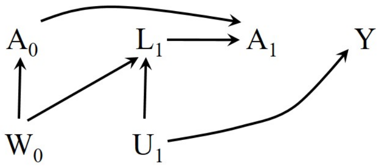

Figure 20.9

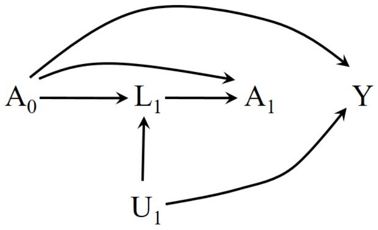

Figure 20.10

One more thing before we discuss g-methods. For simplicity, we have so far described treatment-confounder feedback under simplified causal diagrams in which past treatment does not directly affect subsequent treatment. That is, the causal diagrams in Figures 20.3 and 20.4 did not include an arrow from A_0 to A_1. We now consider the more general case in which past treatment may directly affect subsequent treatment.

As an example, suppose doctors in our HIV study use information on past treatment history $\bar{A}_{k-1}$ when making a decision about whether to prescribe treatment A_k at time k. To represent this situation, we add an arrow from A_0 to A_1 to the causal diagrams in Figures 20.3 and 20.4, as depicted in Figures 20.8 and 20.9.

The causal diagrams in Figures 20.8 and 20.9 show that, in the presence of treatment-confounder feedback, conditioning on L_1 is insufficient to block all backdoor paths between treatment A_1 and outcome Y. Indeed conditioning on L_1 opens the path $A_1 \leftarrow A_0 \rightarrow L_1 \leftarrow U_1 \rightarrow Y$ in Figure 20.8, and the path $A_1 \leftarrow A_0 \leftarrow W_0 \rightarrow L_1 \leftarrow U_1 \rightarrow Y$ in Figure 20.9. Of course, regardless of whether treatment-confounder feedback exists, conditioning on past treatment history is always required when past treatment has a non-null effect on the outcome, as in the causal diagram of Figure 20.10. Under this diagram, treatment A_0 is a confounder of the effect of treatment A_1.

Therefore, sequential exchangeability at time k generally requires conditioning on treatment history $\bar{A}_{k-1}$ before k; conditioning only on the covariates L is not enough. That is why, in this and in the previous chapter, all the conditional independence statements representing sequential exchangeability were conditional on treatment history.

Past treatment plays an important role in the estimation of effects of time-fixed treatments too. Suppose we are interested in estimating the effect of the time-fixed treatment A_1—as opposed to the effect of a treatment strategy involving both A_0 and A_1—on Y. (Sometimes the effect of A_1 is referred to as the short-term effect of the time-varying treatment $\bar{A}$.) Then lack of adjustment for past treatment A_0 will generally result in selection bias if there is treatment-confounder feedback, and in confounding if past treatment A_0 directly affects the outcome Y. In other words, the difference $\mathrm{E}\left[Y|A_1=1,L_1\right]-\mathrm{E}\left[Y|A_1=0,L_1\right]$ would not be zero even if treatment A_1 had no effect on any individual's outcome Y, as in Figures 20.8-20.10. In practice, when making causal inferences about time-fixed treatments, bias may arise in analyses that compare current users ($A_1 = 1$) versus nonusers ($A_1 = 0$) of treatment. To avoid the bias, one can adjust for prior treatment history or restrict the analysis to individuals with a particular treatment history. This is the idea behind "new-user designs" for time-fixed treatments: restrict the analysis to individuals who had not used treatment in the past.

If one could correctly adjust for past treatment, the analysis would not need to be restricted to new users.

The requirement to adjust for past treatment has additional bias implications when past treatment is mismeasured. As discussed in Section 9.3, a mismeasured confounder may result in effect estimates that are biased, either upwards or downwards. In our HIV example, suppose investigators did not have access to the study participants' medical records. Rather, to ascertain prior treatment, investigators had to ask participants via a questionnaire. Since not all participants provided an accurate recollection of their treatment history, treatment A_0 was measured with error. Investigators had data on the mismeasured variable A_0^* rather than on the variable A_0. To depict this setting in Figures 20.8-20.10, we add an arrow from the true treatment A_0 to the mismeasured treatment A_0^*, which shows that conditioning on A_0^* cannot block the biasing paths between A_1 and Y that go through A_0. Investigators will then conclude that there is an association between A_1 to Y, even after adjusting for A_0^* and L_1, despite the lack of an effect of A_1 on Y.

Therefore, when treatment is time-varying, we find that, contrary to a widespread belief, mismeasurement of treatment—even if the measurement error is independent and non-differential—may cause bias under the null. This bias arises because past treatment is a confounder for the effect of subsequent treatment, even if past treatment has no causal effect on the outcome. Furthermore, under the alternative, this imperfect bias adjustment may result in an exaggerated estimate of the effect.

Robins (1987) showed that randomly mismeasured treatment may lead to bias away from the null.