

Chapter 19

TIME-VARYING TREATMENTS

So far this book has dealt with fixed treatments which do not vary over time. However, many causal questions involve treatments that vary over time. For example, we may be interested in estimating the causal effects of medical treatments, lifestyle habits, employment status, marital status, occupational exposures, etc. Because these treatments may take different values for a single individual over time, we refer to them as time-varying treatments.

Restricting our attention to time-fixed treatments during Parts I and II of this book helped us introduce basic concepts and methods. It is now time to consider more realistic causal questions that involve the contrast of hypothetical interventions that are played out over time. Part III extends the material in Parts I and II to time-varying treatments. This chapter describes some key terminology and concepts for causal inference with time-varying treatments. Though we have done our best to simplify those concepts (if you don't believe us, check out the causal inference literature), this is still one of the most technical chapters in the book. Unfortunately, further simplification would result in too much loss of rigor. But if you made it this far, you are qualified to understand this chapter.

19.1 The causal effect of time-varying treatments

Consider a time-fixed treatment variable A (1: treated, 0: untreated) at time zero of follow-up and an outcome variable Y measured 60 months later. We have previously defined the average causal effect of A on the outcome Y as the contrast between the mean counterfactual outcome $Y^{a=1}$ under treatment and the mean counterfactual outcome $Y^{a=0}$ under no treatment, that is, $E[Y^{a=1}] - E[Y^{a=0}]$. Because treatment status is determined at a single time (time zero) for everybody, the average causal effect does not need to make reference to the time at which treatment occurs. In contrast, causal contrasts that involve time-varying treatments need to incorporate time explicitly.

For simplicity, we will provisionally assume that no individuals were lost to follow-up or died during this period, and we will also assume that all variables were perfectly measured.

For compatibility with many published papers, we use zero-based indexing for time. That is, the first time of possible treatment is $k = 0$ rather than $k = 1$.

To see this, consider a time-varying dichotomous treatment A_k that may change at every month k of follow-up, where $k = 0, 1, 2 \dots K$ with $K = 59$. For example, in a 5-year follow-up study of individuals infected with the human immunodeficiency virus (HIV), A_k takes value 1 if the individual receives antiretroviral therapy in month k , and 0 otherwise. No individuals received treatment before the start of the study at time 0, i.e., $A_{-1} = 0$ for all individuals.

We use an overbar to denote treatment history, i.e., $\bar{A}_k = (A_0, A_1, \dots, A_k)$ is the history of treatment from time 0 to time k . When we refer to the entire treatment history through K , we often represent \bar{A}_K as \bar{A} without a subscript. In our HIV study, an individual who receives treatment continuously throughout the follow-up has treatment history $\bar{A} = (A_0 = 1, A_1 = 1, \dots, A_{59} = 1) = (1, 1, \dots, 1)$, or $\bar{A} = \bar{1}$. Analogously, an individual who never receives treatment during the follow-up has treatment history $\bar{A} = (0, 0, \dots, 0) = \bar{0}$. Most individuals are treated during part of the follow-up only, and therefore have intermediate treatment histories with some 1s and some 0s—which we cannot

To keep things simple, our example considers an outcome measured at a fixed time. However, the concepts discussed in this chapter also apply to time-varying outcomes and failure time outcomes (see Technical Point 21.8).

Remember that we use lower-case to denote possible realizations of a random variable: a_k is a realization of treatment A_k .

represent as compactly as $\bar{1}$ and $\bar{0}$.

Suppose Y measures health status—with higher values of Y indicating better health—at the end of follow-up at time $K + 1 = 60$. We would like to estimate the average causal effect of the time-varying treatment \bar{A} on the outcome Y . But we can no longer define the average causal effect of a time-varying treatment as a contrast at a single time k , because the contrast $E[Y^{a_k=1}] - E[Y^{a_k=0}]$ quantifies the effect of treatment A_k at a single time k , not the effect of the time-varying treatment A_k at all times k between 0 and 59.

Indeed we will have to define the average causal effect as a contrast between the counterfactual mean outcomes under two treatment strategies that involve treatment at all times between the start ($k = 0$) and the end ($k = K$) of the follow-up. As a consequence, the average causal effect of a time-varying treatment is not uniquely defined. In the next section, we describe many possible definitions of average causal effect for a time-varying treatment.

19.2 Treatment strategies

A general counterfactual theory to compare treatment strategies was first articulated by Robins (1986, 1987, 1997a).

A treatment strategy—also referred to as a plan, policy, protocol, or regime—is a rule to assign treatment at each time k of follow-up. For example, two treatment strategies are “always treat” and “never treat” during the follow-up. The strategy “always treat” is represented by $\bar{a} = (1, 1, \dots, 1) = \bar{1}$, and the strategy “never treat” is represented by $\bar{a} = (0, 0, \dots, 0) = \bar{0}$. We can now define an average causal effect of \bar{A} on the outcome Y as the contrast between the mean counterfactual outcome $Y^{\bar{a}=\bar{1}}$ under the strategy “always treat” and the mean counterfactual outcome $Y^{\bar{a}=\bar{0}}$ under the strategy “never treat”, i.e., $E[Y^{\bar{a}=\bar{1}}] - E[Y^{\bar{a}=\bar{0}}]$.

But there are many other possible causal effects for the time-varying treatment \bar{A} , each of them defined by a contrast of outcomes under two particular treatment strategies. For example, we might be interested in the average causal effect defined by the contrast $E[Y^{\bar{a}}] - E[Y^{\bar{a}'}]$ that compares the strategy “treat at every other month” $\bar{a} = (1, 0, 1, 0, \dots)$ with the strategy “treat at all months except the first one” $\bar{a}' = (0, 1, 1, 1, \dots)$. The number of possible contrasts is very large: we can define at least 2^K treatment strategies because there are 2^K possible combinations of values (a_0, a_1, \dots, a_K) for a dichotomous a_k . In fact, as we next explain, these 2^K such strategies do not exhaust all possible treatment strategies.

To define even more treatment strategies in our HIV example, consider the time-varying covariate L_k which denotes CD4 cell count (in cells/ μL) measured at month k in all individuals. The variable L_k takes value 1 when the CD4 cell count is low, which indicates a bad prognosis, and 0 otherwise. At time zero, all individuals have a high CD4 cell count, $L_0 = 0$. We could then consider the strategy “do not treat while $L_k = 0$, start treatment when $L_k = 1$ and treat continuously after that time”. This treatment strategy is different from the ones considered in the previous paragraph because we cannot represent it by a rule $\bar{a} = (a_0, a_1, \dots, a_K)$ under which all individuals get the same treatment a_0 at time $k = 0$, a_1 at time $k = 1$, etc. Now, at each time, some individuals will be treated and others will be untreated, depending on the value of their evolving L_k . This is an example of a *dynamic treatment strategy*, a rule in which the treatment a_k at time k depends on the evolution of an individual’s

Fine Point 19.1

Deterministic and random treatment strategies. A deterministic dynamic treatment strategy is a rule $g = [g_0(\bar{a}_{-1}, l_0), \dots, g_K(\bar{a}_{K-1}, \bar{l}_K)]$, where $g_k(\bar{a}_{k-1}, \bar{l}_k)$ specifies the treatment assigned at k to an individual with past history $(\bar{a}_{k-1}, \bar{l}_k)$. An example in our HIV study: $g_k(\bar{a}_{k-1}, \bar{l}_k)$ is 1 if an individual's CD4 cell count (a function of \bar{l}_k) was low at or before k ; otherwise $g_k(\bar{a}_{k-1}, \bar{l}_k)$ is 0. A deterministic static treatment strategy is a rule $g = [g_0(\bar{a}_{-1}), \dots, g_K(\bar{a}_{K-1})]$, where $g_k(\bar{a}_{k-1})$ does not depend on \bar{l}_k . We will often abbreviate $g_k(\bar{a}_{k-1}, \bar{l}_k)$ as $g(\bar{a}_{k-1}, \bar{l}_k)$.

Most static and dynamic strategies we are interested in comparing are *deterministic treatment strategies*, which assign a particular value of treatment (0 or 1) to each individual at each time. More generally, we could consider *random treatment strategies* that do not assign a particular value of treatment, but rather a probability of receiving a treatment value. Random treatment strategies can be static (e.g., “independently at each month, treat individuals with probability 0.3 and do not treat with probability 0.7”) or dynamic (e.g., “independently at each month, treat individuals whose CD4 cell count is low with probability 0.3, but do not treat individuals with high CD4 cell count”).

We refer to the strategy g for which the mean counterfactual outcome $E[Y^g]$ is maximized (when higher values of outcome are better) as the optimal treatment strategy. For a drug treatment, the optimal strategy will almost always be dynamic because treatment needs to be discontinued when toxicity develops. Also, no random strategy can ever be preferred to the optimal deterministic strategy. However, random strategies (i.e., randomized trials) remain scientifically necessary because, before the trial, it is unknown which deterministic strategy is optimal. See Young et al. (2014) for a taxonomy of treatment strategies. In the text, except if noted otherwise, the letter g will refer only to deterministic treatment strategies.

time-varying covariate(s) \bar{L}_k . Strategies \bar{a} for which treatment does not depend on covariates are non-dynamic or *static treatment strategies*. See Fine Point 19.1 for a formal definition.

Causal inference with time-varying treatments involves the contrast of counterfactual outcomes under two or more treatment strategies. The average causal effect of a time-varying treatment is only well-defined if the treatment strategies of interest are specified. In our HIV example, we can define an average causal effect based on the difference $E[Y^{\bar{a}}] - E[Y^{\bar{a}'}]$ that contrasts strategy \bar{a} (say, “always treat”) versus strategy \bar{a}' (say, “never treat”), or on the difference $E[Y^{\bar{a}}] - E[Y^g]$ that contrasts strategy \bar{a} (“always treat”) versus strategy g (say, “treat only after CD4 cell count is low”). Note we will often use g to represent any—static or dynamic—strategy. When we use it to represent a static strategy, we sometimes write $Y^{g=\bar{a}}$ rather than just Y^g or $Y^{\bar{a}}$.

That is, there is not a single definition of causal effect for time-varying treatments. Even when only two treatment options—treat or do not treat—exist at each time k , we can still define as many causal effects as pairs of treatment strategies exist. In the next section, we describe a study design under which all these causal effects can be validly estimated: the sequentially randomized experiment.

19.3 Sequentially randomized experiments

The causal diagrams in Figures 19.1, 19.2, and 19.3 summarize three situations that can occur in studies with time-varying treatments. In all three diagrams, A_k represents the time-varying treatment, L_k the set of measured variables, Y the outcome, and U_k the set of unmeasured variables at k that are common

Technical Point 19.1

On the definition of dynamic strategies. Each dynamic strategy $g = [g_0(\bar{a}_{-1}, \bar{l}_0), \dots, g_K(\bar{a}_{K-1}, \bar{l}_K)]$ that depends on past treatment and covariate history is associated with a dynamic strategy $g' = [g'_0(\bar{l}_0), \dots, g'_K(\bar{l}_K)]$ that depends only on past covariate history. By consistency (see Technical Point 19.2), an individual will have the same treatment, covariate, and outcome history when following strategy g from time zero as when following strategy g' from time zero. In particular, $Y^g = Y^{g'}$ and $\bar{L}^g(K) = \bar{L}^{g'}(K)$. Specifically, g' is defined in terms of g recursively by $g'_0(l_0) = g_0(\bar{a}_{-1} = 0, l_0)$ (by convention, \bar{a}_{-1} can only take the value zero) and $g'_k(\bar{l}_k) = g_k[g'_k(\bar{l}_{k-1}), \bar{l}_k]$. For any strategy g for which treatment at each k already does not depend on past treatment history, g and g' are the identical set of functions. The above definition of g' in terms of g guarantees that an individual has followed strategy g through time t in the observed data, i.e., $A_k = g_k(\bar{A}_{k-1}, \bar{L}_k)$ for $k \leq t$, if and only if the individual has followed strategy g' through t , i.e., $A_k = g'_k(\bar{L}_k)$ for $k \leq t$.

By definition, a causal graph must always include all common causes of any two variables on the graph.

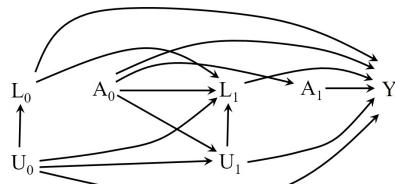


Figure 19.1

causes of at least two other variables on the causal graph. Because the covariates U_k are not measured, their values are unknown and therefore unavailable for the analysis. In our HIV study, the time-varying CD4 cell count L_k is a consequence of the true, but unmeasured, chronic damage to the immune system U_k . The greater an individual's immune damage U_k , the lower her CD4 cell count L_k and her health status Y . For simplicity, the causal diagrams show only the first two times of follow-up $k = 0$ and $k = 1$, and we assume that all participants adhered to the assigned treatment.

The causal diagram in Figure 19.1 lacks arrows from either the measured covariates \bar{L}_k or the unmeasured covariates \bar{U}_k into treatment A_k . The causal diagram in Figure 19.2 has arrows from the measured covariates \bar{L}_k , but not from the unmeasured covariates \bar{U}_k , into treatment A_k . The causal diagram in Figure 19.3 has arrows from both the measured covariates \bar{L}_k and the unmeasured covariates \bar{U}_k into treatment A_k .

Figure 19.1 could represent a randomized experiment in which treatment A_k at each time k is randomly assigned with a probability that depends only on prior treatment history (for simplicity, we will assume perfect adherence throughout). Our HIV study would be represented by Figure 19.1 if, e.g., an individual's treatment value at each month k were randomly assigned with probability 0.5 for individuals who did not receive treatment during the previous month ($A_{k-1} = 0$), and with probability 1 for individuals who did receive treatment during the previous month k ($A_{k-1} = 1$). When interested in the contrast of static treatment strategies, Figure 19.1 is the proper generalization of no confounding by measured or unmeasured variables for time-varying treatments. Under this causal diagram, the counterfactual outcome mean $E[Y^{\bar{a}}]$ if everybody had followed the static treatment strategy \bar{a} is simply the mean outcome $E[Y|\bar{A} = \bar{a}]$ among those who followed the strategy \bar{a} . (Interestingly, the same is not true for dynamic strategies. The counterfactual mean $E[Y^g]$ under a dynamic strategy g that depends on the variables L is only the mean outcome among those who followed the strategy g if the probability of receiving treatment $A_k = 1$ is exactly 0.5 at all times k at which treatment A_k depends on \bar{L}_k . Otherwise, identifying $E[Y^g]$ requires the application of g-methods to data on \bar{L} , \bar{A} , and Y under either Figure 19.1 or Figure 19.2.)

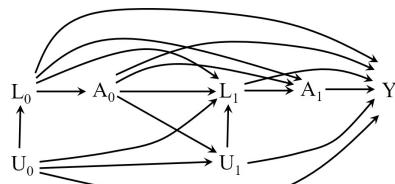


Figure 19.2

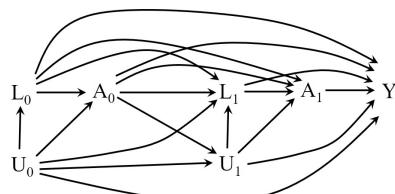


Figure 19.3

Figure 19.2 could represent a randomized experiment in which treatment A_k at each time k is randomly assigned by the investigators with a probability that depends on prior treatment *and* measured covariate history. Our study would be represented by Figure 19.2 if, e.g., an individual's treatment value

at each month k were randomly assigned with probability 0.4 for untreated individuals with high CD4 cell count ($A_{k-1} = 0, L_k = 1$), 0.8 for untreated individuals with low CD4 cell count ($A_{k-1} = 0, L_k = 0$), and 0.5 for previously treated individuals, regardless of their CD4 cell count ($A_{k-1} = 1$). In Figure 19.2, there is confounding by measured, but not unmeasured, variables for the time-varying treatment.

An experiment in which treatment is randomly assigned at each time k to each individual is referred to as a *sequentially randomized experiment*. Therefore Figures 19.1 and 19.2 could represent sequentially randomized experiments. On the other hand, Figure 19.3 cannot represent a randomized experiment: the value of treatment A_k at each time k depends partly on unmeasured variables U which are causes of L_k and Y , but unmeasured variables obviously cannot be used by investigators to assign treatment. That is, a sequentially randomized experiment can be represented by a causal diagram with many time points $k = 0, 1 \dots K$ and with no direct arrows from the unmeasured prognostic factors U into treatment A_k at any time k .

In observational studies, decisions about treatment often depend on outcome predictors such as prognostic factors. Therefore, observational studies will be typically represented by either Figure 19.2 or Figure 19.3 rather than Figure 19.1. For example, suppose our HIV follow-up study were an observational study (not an experiment) in which the lower the CD4 cell count L_k , the more likely a patient is to be treated. Then our study would be represented by Figure 19.2 if, at each month k , treatment decisions in the real world were made based on the values of prior treatment and CD4 cell count history (\bar{A}_{k-1}, \bar{L}_k), but not on the values of any unmeasured variables \bar{U}_k . Thus, an observational study represented by Figure 19.2 would differ from a sequentially randomized experiment only in that the assignment probabilities are unknown (but could be estimated from the data). Unfortunately, it is impossible to show empirically whether an observational study is represented by the causal diagram in either Figure 19.2 or Figure 19.3. Observational studies represented by Figure 19.3 have unmeasured confounding, as we describe later.

Sequentially randomized experiments are not frequently used in practice. However, the concept of sequentially randomized experiment is helpful to understand some key conditions for valid estimation of causal effects of time-varying treatments. The next section presents these conditions formally.

19.4 Sequential exchangeability

As described in Parts I and II, valid causal inferences about time-fixed treatments typically require conditional exchangeability $Y^a \perp\!\!\!\perp A|L$. When exchangeability $Y^a \perp\!\!\!\perp A|L$ holds, we can obtain unbiased estimates of the causal effect of treatment A on the outcome Y if we appropriately adjust for the variables in L via standardization, IP weighting, g-estimation, or other methods. We expect conditional exchangeability to hold in conditionally randomized experiments—a trial in which individuals are assigned treatment with a probability that depends on the values of the covariates L . Conditional exchangeability holds in observational studies if the probability of receiving treatment depends on the measured covariates L and, conditional on L , does not further depend on any unmeasured, common causes of treatment and outcome.

Similarly, causal inference with time-varying treatments requires adjusting for the time-varying covariates \bar{L}_k to achieve conditional exchangeability at

For those with treatment history $[A_0 = g(L_0), A_1 = g(A_0, L_0, L_1)]$ equal to (i.e., compatible with) the treatment they would have received under strategy g through the end of follow-up, the counterfactual outcome Y^g is equal (by consistency) to the observed outcome Y and therefore also to the counterfactual outcome under the static strategy (a_0, a_1) with $a_0 = A_0, a_1 = A_1$.

In Figure 19.1, sequential unconditional exchangeability for Y holds, i.e., for all static strategies \bar{a} , $Y^{\bar{a}} \perp\!\!\!\perp A_k | \bar{A}_{k-1} = \bar{a}_{k-1}$. Unconditional exchangeability implies that association is causation, i.e., $E[Y^{\bar{a}}] = E[Y | \bar{A} = \bar{a}]$.

Whenever we talk about identification of causal effects, the identifying formula will be the g-formula (see Chapter 21). In rare cases not relevant to our discussion, effects can be identified by formulas that are related to, but not equal to, the g-formula (e.g., Technical Point 7.3).

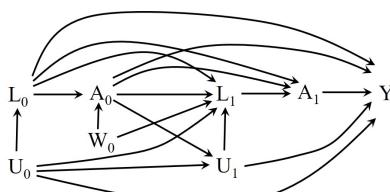


Figure 19.4

each time point, i.e., sequential conditional exchangeability. For example, in a study with two time points, sequential conditional exchangeability is the combination of conditional exchangeability at both the first time and the second time of the study. That is, $Y^g \perp\!\!\!\perp A_0 | L_0$ and $Y^g \perp\!\!\!\perp A_1 | A_0 = g(L_0), L_0, L_1$. (For brevity, in this book we drop the word “conditional” and simply say sequential exchangeability.) We will refer to this set of conditional independences as *sequential exchangeability for Y^g* under any—static or dynamic—strategy g that involves interventions on both components of the time-varying treatment (A_0, A_1) .

A sequentially randomized experiment—an experiment in which treatment A_k at each time k is randomly assigned with a probability that depends only on the values of their prior covariate history \bar{L}_k and treatment history \bar{A}_{k-1} —implies sequential exchangeability for Y^g . That is, for any strategy g , the treated and the untreated at each time k are exchangeable for Y^g conditional on prior covariate history \bar{L}_k and any observed treatment history $\bar{A}_{k-1} = g(\bar{A}_{k-2}, \bar{L}_{k-1})$ compatible with strategy g . Formally, sequential exchangeability for Y^g is defined as

$$Y^g \perp\!\!\!\perp A_k | \bar{A}_{k-1} = g(\bar{A}_{k-2}, \bar{L}_{k-1}), \bar{L}_k \text{ for all strategies } g \text{ and } k = 0, 1, \dots, K$$

This form of sequential exchangeability (there are others, as we will see) always holds in any causal graph which, like Figure 19.2, has no arrows from the unmeasured variables U into the treatment variables A . Therefore sequential exchangeability for Y^g holds in sequentially randomized experiments and observational studies in which the probability of receiving treatment at each time depends on their treatment and measured covariate history $(\bar{A}_{k-1}, \bar{L}_k)$ and, conditional on this history, does not depend on any unmeasured causes of the outcome.

That is, in observational studies represented by Figure 19.2 the mean of the counterfactual outcome $E[Y^g]$ under all strategies g is identified, whereas in observational studies represented by Figure 19.3 no mean counterfactual outcome $E[Y^g]$ is identified. In observational studies represented by other causal diagrams, the mean counterfactual outcome $E[Y^g]$ under some but not all strategies g is identified.

For example, consider an observational study represented by the causal diagram in Figure 19.4, which includes an unmeasured variable W_0 . In our HIV example, W_0 could be an indicator for a scheduled clinic visit at time 0 that was not recorded in our database. In that case W_0 would be a cause shared by treatment A_0 and the measured (with some error) CD4 cell count L_1 , with U_1 representing the underlying but unknown true value of CD4 cell count. Even though W_0 is unmeasured, the mean counterfactual outcome is still identified under any static strategy $g = \bar{a}$; however, the mean counterfactual outcome $E[Y^g]$ is not identified under any dynamic strategy g with treatment assignment depending on L_1 . To illustrate why identification is possible under some but not all strategies, we will use SWIGs in the next section.

In addition to some form of sequential exchangeability, causal inference involving time-varying treatments also requires a sequential version of the conditions of positivity and consistency. In a sequentially randomized experiment, both sequential positivity and consistency are expected to hold (see Technical Point 19.2). Below we will assume that sequential positivity and consistency hold. Under the three identifiability conditions, we can identify the mean counterfactual outcome $E[Y^g]$ under a strategy of interest g as long as we use methods that appropriately adjust for treatment and covariate history $(\bar{A}_{k-1}, \bar{L}_k)$, such as the g-formula (standardization), IP weighting, and g-estimation.

Technical Point 19.2

Positivity and consistency for time-varying treatments. The positivity condition needs to be generalized from the fixed version “if $f_L(l) \neq 0$, $f_{A|L}(a|l) > 0$ for all a and l ” to the sequential version

$$\text{If } f_{\bar{A}_{k-1}, \bar{L}_k}(\bar{a}_{k-1}, \bar{l}_k) \neq 0, \text{ then } f_{A_k|\bar{A}_{k-1}, \bar{L}_k}(a_k|\bar{a}_{k-1}, \bar{l}_k) > 0 \text{ for all } (\bar{a}_k, \bar{l}_k)$$

In a sequentially randomized experiment, positivity will hold if the randomization probabilities at each time k are never either 0 nor 1, no matter the past treatment and covariate history. If we are interested in a particular strategy g , the above positivity condition needs to only hold for treatment histories compatible with g , i.e., for each k , $a_k = g(\bar{a}_{k-1}, \bar{l}_k)$.

The consistency condition also needs to be generalized from the fixed version “If $A = a$ for a given individual, then $Y^a = Y$ for that individual” to the sequential version

$$Y^{\bar{a}} = Y^{\bar{a}^*} \text{ if } \bar{a}^* = \bar{a}; Y^{\bar{a}} = Y \text{ if } \bar{A} = \bar{a}; \bar{L}_k^{\bar{a}} = \bar{L}_k^{\bar{a}^*} \text{ if } \bar{a}^* = \bar{a}_{k-1}; \bar{L}_k^{\bar{a}} = \bar{L}_k \text{ if } \bar{A}_{k-1} = \bar{a}_{k-1}$$

where $\bar{L}_k^{\bar{a}}$ is the counterfactual L -history through time k under strategy \bar{a} . Technically, the identification of effects of time-varying treatments on Y requires weaker consistency conditions: “If $\bar{A} = \bar{a}$ for a given individual, then $Y^{\bar{a}} = Y$ for that individual” is sufficient for static strategies, and “For any strategy g , if $A_k = g_k(\bar{A}_{k-1}, \bar{L}_k)$ at each time k for a given individual, then $Y^g = Y$ ” is sufficient for dynamic strategies. However, the stronger sequential consistency is a natural condition that we will always accept.

Note that, if we expect that the interventions “treat in month k ” corresponding to $A_k = 1$ and “do not treat in month k ” corresponding to $A_k = 0$ are sufficiently well defined at all times k , then all static and dynamic strategies involving A_k will be similarly well defined.

19.5 Identifiability under some but not all treatment strategies

Pearl and Robins (1995) proposed a generalized backdoor criterion for static strategies. Robins (1997a) extended the procedure to dynamic strategies.

In Chapter 7, we presented a graphical rule—the backdoor criterion—to assess whether exchangeability holds for a time-fixed treatment under a particular causal diagram. The backdoor criterion can be generalized for time-varying treatments. For example, for static strategies, a sufficient condition for identification of the causal effect of treatment strategies is that, at each time k , all backdoor paths into A_k that do not go through any future treatment are blocked.

However, the *generalized backdoor criterion* does not directly show the connection between blocking backdoor paths and sequential exchangeability, because the procedure is based on causal directed acyclic graphs that do not include counterfactual outcomes. An alternative graphical check for identifiability of causal effects is based on SWIGs, also discussed in Chapter 7. SWIGs are especially helpful for time-varying treatments.

Consider the causal diagrams in Figures 19.5 and 19.6, which are simplified versions of those in Figures 19.2 and 19.4. We have omitted the nodes U_0 and L_0 and the arrow from A_0 to U_1 . In addition, the arrow from L_1 to Y is absent so L_1 is no longer a direct cause of Y . Figures 19.5 and 19.6 (like Figures 19.2 and 19.4) differ in whether A_k and subsequent covariates L_t for $t > k$ share a cause W_k .

As discussed in Part I of this book, a SWIG represents a counterfactual world under a particular intervention. The SWIG in Figure 19.7 represents the world in Figure 19.5 if all individuals had received the static strategy (a_0, a_1) , where a_0 and a_1 can take values 0 or 1. For example, Figure 19.7 can be used to represent the world under the strategy “always treat” ($a_0 = 1, a_1 = 1$) or under the strategy “never treat” ($a_0 = 0, a_1 = 0$). To construct this SWIG, we

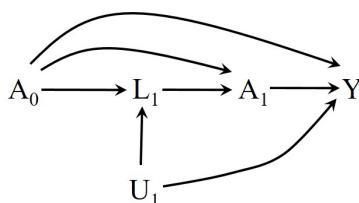


Figure 19.5

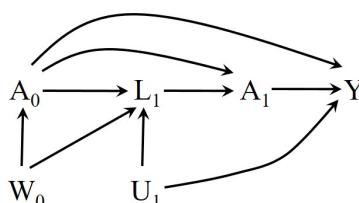


Figure 19.6

Technical Point 19.3

The many forms of sequential exchangeability. Consider a sequentially randomized experiment of a time-varying treatment A_k with multiple time points $k = 0, 1, \dots, K$. The SWIG that represents this experiment is just a longer version of Figure 19.7. The following conditional independence can be directly read from the SWIG:

$$(Y^{\bar{a}}, \underline{L}_{k+1}^{\bar{a}}) \perp\!\!\!\perp A_k | \bar{A}_{k-1}^{\bar{a}_{k-1}}, \bar{L}_k^{\bar{a}_{k-1}}$$

where $\underline{L}_{k+1}^{\bar{a}}$ is the counterfactual covariate history from time $k+1$ through the end of follow-up. The above conditional independence implies $(Y^{\bar{a}}, \underline{L}_{k+1}^{\bar{a}}) \perp\!\!\!\perp A_k | \bar{A}_{k-1}^{\bar{a}_{k-1}} = \bar{a}_{k-1}, \bar{L}_k^{\bar{a}_{k-1}}$ for the particular instance $\bar{A}_{k-1}^{\bar{a}_{k-1}} = \bar{a}_{k-1}$, with \bar{a}_{k-1} being a component of strategy \bar{a} . Because of consistency, the last conditional independence statement equals

$$(Y^{\bar{a}}, \underline{L}_{k+1}^{\bar{a}}) \perp\!\!\!\perp A_k | \bar{A}_{k-1} = \bar{a}_{k-1}, \bar{L}_k$$

When this statement holds for all \bar{a} , we say that there is *sequential exchangeability*. Interestingly, even though this sequential exchangeability condition only refers to static strategies $g = \bar{a}$, it is equivalent to the seemingly stronger

$$(Y^g, \underline{L}_{k+1}^g) \perp\!\!\!\perp A_k | \bar{A}_{k-1} = g(\bar{A}_{k-1}, \bar{L}_k), \bar{L}_k \text{ for all } g,$$

and, if positivity holds, is therefore sufficient to identify the outcome and covariate distribution under any static and dynamic strategies g (Robins 1986). This identification results from the joint conditional independence between $(Y^{\bar{a}}, \underline{L}_{k+1}^{\bar{a}})$ and A_k . Note that, for dynamic strategies, sequential exchangeability does not follow from the separate independences $Y^{\bar{a}} \perp\!\!\!\perp A_k | \bar{A}_{k-1} = \bar{a}_{k-1}, \bar{L}_k$ and $\underline{L}_{k+1}^{\bar{a}} \perp\!\!\!\perp A_k | \bar{A}_{k-1} = \bar{a}_{k-1}, \bar{L}_k$.

Stronger conditional independences are expected to hold in a sequentially randomized experiment, but they (i) cannot be read from SWIGs and (ii) are not necessary for identification of the causal effects of treatment strategies in the population. For example, a sequentially randomized trial implies the stronger joint independence $\{Y^{\bar{a}}, \underline{L}_{k+1}^{\bar{a}}; \text{all } \bar{a}\} \perp\!\!\!\perp A_k | \bar{A}_{k-1}, \bar{L}_k$.

An even stronger condition that is expected to hold in sequentially randomized experiments is

$$(Y^{\bar{A}}, \bar{L}^{\bar{A}}) \perp\!\!\!\perp A_k | \bar{A}_{k-1}, \bar{L}_k$$

where, for a dichotomous treatment A_k , \bar{A} denotes the set of all 2^K static strategies \bar{a} , $Y^{\bar{A}}$ denotes the set of all counterfactual outcomes $Y^{\bar{a}}$, and $\bar{L}^{\bar{A}}$ denotes the set of all counterfactual covariate histories. Using a terminology analogous to that of Technical Point 2.1, we refer to this joint independence condition as *full sequential exchangeability*.

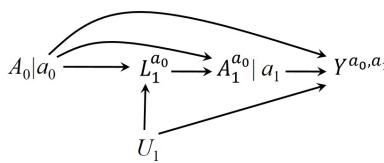


Figure 19.7

first split the treatment nodes A_0 and A_1 . The right side of the split treatments represents the value of treatment under the intervention. The left side represents the value of treatment that would have been observed when intervening on all previous treatments. Therefore, the left side of A_0 is precisely A_0 because there are no previous treatments to intervene on, and the left side of A_1 is the counterfactual treatment $A_1^{a_0}$ that would be observed after setting A_0 to the value a_0 . All arrows into a given treatment in the original causal diagram now point into the left side, and all arrows out of a given treatment now originate from the right side. The outcome variable is the counterfactual outcome Y^{a_0,a_1} and the covariates L are replaced by their corresponding counterfactual variables. Note that we write the counterfactual variable corresponding to L_1 under strategy (a_0, a_1) as $L_1^{a_0}$, rather than $L_1^{a_0,a_1}$, because a future intervention on A_1 cannot affect the value of earlier L_1 .

Unlike the directed acyclic graph in Figure 19.5, the SWIG in Figure 19.7 does include the counterfactual outcome, which means that we can visually check for exchangeability using d-separation.

$Y^{a_0,a_1} \perp\!\!\!\perp A_1^{a_0}|A_0 = a_0, L_1^{a_0}$ equals $Y^{a_0,a_1} \perp\!\!\!\perp A_1|A_0 = a_0, L_1$ because, by consistency, $L_1^{a_0} = L_1$ and $A_1^{a_0} = A_1$ when $A_0 = a_0$.

In Figure 19.7, by d-separation, both $Y^{a_0,a_1} \perp\!\!\!\perp A_0$ and $Y^{a_0,a_1} \perp\!\!\!\perp A_1^{a_0}|A_0, L_1^{a_0}$ hold for any static strategy (a_0, a_1) . This second conditional independence holds even though there seems to be an open path $A_1^{a_0} \leftarrow a_0 \rightarrow L_1^{a_0} \leftarrow U_1 \rightarrow Y^{a_0,a_1}$. However, this path is actually blocked for the following reason. In the counterfactual world, a_0 is a constant and in probability statements constants are always implicitly conditioned on even though, by convention, they are not shown in the conditioning event. See Fine Point 19.2 for details.

The second conditional independence $Y^{a_0,a_1} \perp\!\!\!\perp A_1^{a_0}|A_0, L_1^{a_0}$ implies, by definition, $Y^{a_0,a_1} \perp\!\!\!\perp A_1^{a_0}|A_0 = a_0, L_1^{a_0}$ in the subset of individuals who received treatment $A_0 = a_0$. Therefore, by consistency, we conclude that $Y^{a_0,a_1} \perp\!\!\!\perp A_0$ and $Y^{a_0,a_1} \perp\!\!\!\perp A_1|A_0 = a_0, L_1$ hold under the causal diagram in Figure 19.5, which corresponds to the SWIG in Figure 19.7 where we can actually check for exchangeability. If there were multiple time points, we would say that

$$Y^{\bar{a}} \perp\!\!\!\perp A_k | \bar{A}_{k-1} = \bar{a}_{k-1}, \bar{L}_k \text{ for } k = 0, 1, \dots, K$$

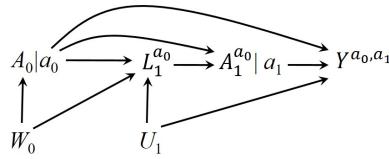


Figure 19.8

We refer to the above condition as *static sequential exchangeability for $Y^{\bar{a}}$* , which is weaker than sequential exchangeability for Y^g , because it only requires conditional independence between counterfactual outcomes $Y^{\bar{a}}$ indexed by static strategies $g = \bar{a}$ and treatment A_k . Static sequential exchangeability is sufficient to identify the mean counterfactual outcome under any static strategy $g = \bar{a}$. See also Technical Point 19.3.

Static sequential exchangeability also holds under the causal diagram in Figure 19.6, as can be checked by applying d-separation to its corresponding SWIG in Figure 19.8. Thus, in an observational study represented by Figure 19.6, we can identify the mean counterfactual outcome under any static strategy (a_0, a_1) . Let us return to Figure 19.5. Let us now assume that the arrow from L_1 to A_1 were missing. In that case, the arrow from $L_1^{a_0}$ to $A_1^{a_0}$ would also be missing from the SWIG in Figure 19.7. It would then follow by d-separation that sequential unconditional exchangeability holds, and therefore that the mean counterfactual outcome under any static strategy could be identified without data on L_1 . Now let us assume that, in Figure 19.5, there was an arrow from U_1 to A_1 . Then the SWIG in Figure 19.7 would include an arrow from U_1 to $A_1^{a_0}$, and so no form of sequential exchangeability would hold. The counterfactual mean would not be identified under any strategy.

We now discuss SWIGs under dynamic treatment strategies. Figure 19.9 represents the world of Figure 19.5 under a dynamic strategy $g = [g_0, g_1(L_1)]$ in which treatment A_0 is assigned a fixed value g_0 (either 0 or 1), and treatment A_1 at time $k = 1$ is assigned a value $g_1(L_1^g)$ that depends on the value of L_1^g that was observed after having assigned treatment value g_0 at time $k = 0$. For example, g may be the strategy “do not treat at time 0, treat at time 1 only if CD4 cell count is low, i.e., if $L_1^g = 1$ ”. Under this strategy $g_0 = 0$ for everybody, and $g_1(L_1^g) = 1$ when $L_1^g = 1$ and $g_1(L_1^g) = 0$ when $L_1^g = 0$. Therefore the SWIG includes an arrow from L_1^g to $g_1(L_1^g)$. This arrow was not part of the original causal graph; it exists only in the counterfactual world associated with this dynamic strategy. We therefore draw this arrow differently from the others, even though we need to treat it as any other arrow when evaluating d-separation. The outcome in the SWIG is the counterfactual outcome Y^g under the dynamic strategy g which uses L_1 to assign treatment A_1 .

By applying d-separation to the SWIG in Figure 19.9, we find that both $Y^g \perp\!\!\!\perp A_0$ and $Y^g \perp\!\!\!\perp A_1^g|A_0 = g_0, L_1^g$ hold for any strategy g . That is, sequential exchangeability for Y^g holds, which means that we can identify the mean counterfactual outcome under all strategies g (see also Fine Point 19.2). This result, however, does not hold for the causal diagram in Figure 19.6.

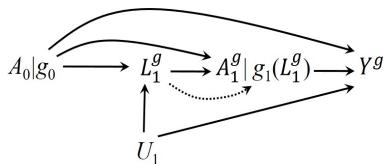


Figure 19.9

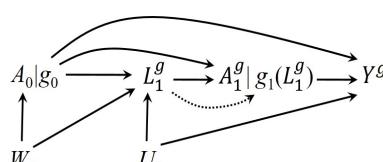


Figure 19.10

Fine Point 19.2

Arrows from intervention nodes in SWIGs When drawing SWIGs, we include arrows from the node a into future variables even though, in a single intervention world, a is a constant and thus cannot affect other variables. One reason we do this is that is a convenient device to keep track of the variables directly affected by A in the original DAG. A second reason is described in Technical Point 21.10.

To illustrate why this is important, consider two causal diagrams: 1) the causal DAG in Figure 7.14 and 2) a causal DAG equal to Figure 7.14 except that it also includes a direct arrow from A to Y . The SWIGs corresponding to these two DAGs would be identical if we did not include arrows leaving from a . As a result, we could not use the SWIG to infer that the causal effect of A on Y is identified in DAG 1 (using the front door formula) but not in DAG 2. Therefore, when using d-separation on a SWIG, we need to remember that all paths that include any intervention node a are blocked even if we do not explicitly condition on a in the notation.

The same logic applies for any intervention node under deterministic strategies. Suppose that we include a baseline confounder L_0 in Figure 19.9 and we consider a deterministic dynamic strategy with g_0 replaced by $g_0(L_0)$. Then, when checking $Y^g \perp\!\!\!\perp A_1^g | A_0, L_0, L_1^g$, we do not need to explicitly condition also on $g_0(L_0)$ because $g_0(L_0)$ becomes a constant conditional on L_0 . However, we need to remember that, when conditioning on L_0 , paths through $g_0(L_0)$ are blocked. When we instantiate A_0 at $g(L_0)$ and use consistency, the statistical independence $Y^g \perp\!\!\!\perp A_1^g | A_0, L_0, L_1^g$ becomes the exchangeability condition $Y^g \perp\!\!\!\perp A_1 | A_0 = g(L_0), L_0, L_1$ described above.

On the other hand, under a random strategy g that assigns a random treatment value $A_0^{+,g}$ to each individual from a distribution that possibly depends on L_0 , we would explicitly include $A_0^{+,g}$ on the SWIG (replacing $g_0(L_0)$) and also in the conditioning event when we check for d-separation, because $A_0^{+,g}$ is no longer perfectly determined by L_0 . Richardson and Robins (2013) showed that a necessary condition for identifiability by the g-formula for such a random strategy is that $Y^g \perp\!\!\!\perp A_1^g | A_0, A_0^{+,g}, L_0, L_1^g$ holds. They also considered strategies that depend on the natural value of treatment (Robins et al. 2004), i.e., strategies that assign treatment $A_t^{+,g}$ at time t based on \bar{A}_t^g , for which they provided exchangeability conditions that license identification by the extended g-formula. Strategies that depend on the natural value of treatment have recently been referred to as “modified treatment policies” (Diaz et al. 2021).

What we read from the SWIG is $Y^g \perp\!\!\!\perp A_1^g | A_0, L_1^g$ which, by consistency, implies $Y^g \perp\!\!\!\perp A_1 | A_0 = g_0, L_1$

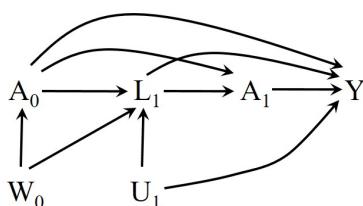


Figure 19.11

The SWIG in Figure 19.10 represents the world of Figure 19.6 under a dynamic treatment strategy $g = [g_0, g_1(L_1)]$. By applying d-separation to the SWIG in Figure 19.10, we find that $Y^g \perp\!\!\!\perp A_0$ does not hold because of the open path $A_0 \leftarrow W_0 \rightarrow L_1^g \rightarrow g_1(L_1^g) \rightarrow Y^g$. That is, sequential exchangeability for Y^g does not hold, which means that we cannot identify the mean counterfactual outcome for any strategy g .

In summary, in observational studies (or sequentially randomized trials) represented by Figure 19.5, sequential exchangeability for Y^g holds, and therefore the data can be used to validly estimate causal effects involving static and dynamic strategies. On the other hand, in observational studies represented by Figure 19.6, only the weaker condition for static strategies holds, and therefore the data can be used to validly estimate causal effects involving static strategies, but not dynamic strategies. Another way to think about this is that in the counterfactual world represented by the SWIG in Figure 19.10, the distribution of Y^g depends on the distribution of $g_1(L_1^g)$ and thus of L_1^g . However, the distribution of L_1^g is not identifiable due to the path $A_0 \leftarrow W_0 \rightarrow L_1^g$.

One last example. Consider Figure 19.11 which is equal to Figure 19.6 except for the presence of an arrow from L_1 to Y , and its corresponding SWIG under a static strategy in Figure 19.12. We can use d-separation to show that neither sequential exchangeability for Y^g nor static sequential exchangeability for $Y^{\bar{a}}$ hold. Therefore, in observational study represented by Figure 19.11, we cannot use the data to validly estimate causal effects involving any strategies.

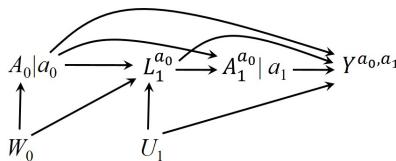


Figure 19.12

19.6 Time-varying confounding and time-varying confounders

No form of sequential exchangeability is guaranteed to hold in observational studies. Achieving approximate exchangeability requires expert knowledge, which will guide investigators in the design of their studies to measure as many of the relevant variables \bar{L}_k as possible. For example, in an HIV study, experts would agree that time-varying variables like CD4 cell count, viral load, and symptoms need to be appropriately measured and adjusted for.

But the question “Are the measured covariates sufficient to ensure sequential exchangeability?” can never be answered with certainty. Yet we can use our expert knowledge to organize our beliefs about exchangeability and represent them in a causal diagram. Figures 19.1 to 19.4 are examples of causal diagrams that summarize different scenarios. Note that we drew these causal diagrams in the absence of selection (e.g., censoring by loss to follow-up) so that we can concentrate on confounding here.

Consider Figure 19.5. Like in Part I of this book, suppose that we are interested in the effect of the time-fixed treatment A_1 on the outcome Y . We say that there is confounding for the effect of A_1 on Y because A_1 and Y share the cause U , i.e., because there is an open backdoor path between A_1 and Y through U . To estimate this effect without bias, we need to adjust for confounders of the effect of the treatment A_1 on the outcome Y , as explained in Chapter 7. In other words, we need to be able to block all open backdoor paths between A_1 and Y . This backdoor path $A_1 \leftarrow L_1 \leftarrow U \rightarrow Y$ cannot be blocked by conditioning on the common cause U because U is unmeasured and therefore unavailable to the investigators. However, this backdoor path can be blocked by conditioning on L_1 , which is measured. Thus, if the investigators collected data on L_1 for all individuals, there would be no unmeasured confounding for the effect of A_1 . We then say that L_1 is a confounder for the effect of A_1 , even though the actual common cause of A_1 and Y was the unmeasured U (re-read Section 7.3 if you need to refresh your memory about confounding and causal diagrams).

As discussed in Chapter 7, the confounders do not have to be direct causes of the outcome. In Figure 19.5, the arrow from the confounder L_1 to the outcome Y does not exist. Then the source of the confounding (i.e., the causal confounder) is the unmeasured common cause U . Nonetheless, because data on L_1 suffice to block the backdoor paths from A_1 to Y and thus to control confounding, we refer to L_1 as a confounder for the effect of A_1 on Y .

Now imagine the very long causal diagram that contains all time points $k = 0, 1, 2, \dots$, and in which L_k affects subsequent treatments A_k, A_{k+1}, \dots and shares unmeasured causes U_k with the outcome Y . Suppose that we want to estimate the causal effects on the outcome Y of treatment strategies defined by interventions on A_0, A_1, A_2 . Then, at each time k , the covariate history \bar{L}_k will be needed, together with the treatment history \bar{A}_{k-1} , to block the backdoor paths between treatment A_k and the outcome Y . Thus, no unmeasured confounding for the effect of \bar{A} requires that the investigators collected data on \bar{L}_k for all individuals. We then say that the time-varying covariates in \bar{L}_k are *time-varying confounders* for the effect of the time-varying treatment \bar{A} on Y at several (or, in our example, all) times k in the study. See Fine Point 19.3 for a more precise definition of time-varying confounding.

Unfortunately, we cannot empirically confirm that all confounders, whether time-fixed or time-varying, are measured. That is, we cannot empirically differentiate between Figure 19.2 with no unmeasured confounding and Figure 19.3 with unmeasured confounding. Interestingly, even if all confounders were

A second backdoor path gets open after conditioning on collider L_1 :
 $A_1 \leftarrow A_0 \rightarrow L_1 \leftarrow U \rightarrow Y$
This second backdoor path can be safely blocked by conditioning on prior treatment A_0 , assuming it is available to investigators.

Time-varying confounders are sometimes referred to as time-dependent confounders.

Fine Point 19.3

A definition of time-varying confounding. In the absence of selection bias, we say there is confounding for causal effects involving $E[Y^{\bar{a}}]$ if $E[Y^{\bar{a}}] \neq E[Y|A = \bar{a}]$, that is, if the mean outcome had, contrary to fact, all individuals in the study followed strategy \bar{a} differs from the mean outcome among the subset of individuals who followed strategy \bar{a} in the actual study.

We say the confounding is solely time-fixed (i.e., wholly attributable to baseline covariates) if $E[Y^{\bar{a}}|L_0] = E[Y|A = \bar{a}, L_0]$, as would be the case if the only arrows pointing into A_1 in Figure 19.2 were from A_0 and L_0 . In contrast, if the identifiability conditions hold, but $E[Y^{\bar{a}}|L_0] \neq E[Y|A = \bar{a}, L_0]$, we say that time-varying confounding is present. If the identifiability conditions do not hold, as in Figure 19.3, we say that there is unmeasured confounding.

A sufficient condition for no time-varying confounding is unconditional sequential exchangeability for $Y^{\bar{a}}$, i.e., $Y^{\bar{a}} \perp\!\!\!\perp A_k | \bar{A}_{k-1} = \bar{a}_{k-1}$. This condition holds in sequentially randomized experiments, like the one represented in Figure 19.1, in which treatment A_k at each time k is randomly assigned with a probability that depends only on the values of prior treatment history \bar{A}_{k-1} . In fact, the causal diagram in Figure 19.1 can be greatly simplified. To do so, first note that L_1 is not a common cause of any two nodes in the graph so it can be omitted from the graph. Once L_1 is gone, then both L_0 and U_1 can be omitted too because they cease to be common causes of two nodes in the graph. In the graph without L_0 , L_1 , and U_1 , the node U_0 can be omitted too. That is, the causal diagram in Figure 19.1 can be simplified to include only the nodes A_0 , A_1 and Y .

correctly measured and modeled, most adjustment methods may still result in biased estimates when comparing treatment strategies. The next chapter explains why g-methods are the appropriate approach to adjust for time-varying confounders.