

Chapter 23

CAUSAL MEDIATION

In Part III, we have presented approaches to estimate the causal effect of a time-varying treatment on an outcome. Our goal was to quantify changes in the outcome distribution under different treatment strategies that are sustained over time. Our goal was not to determine *how* treatment exerted its effect on the outcome. We now turn our attention to causal mediation: the study of the causal pathways through which the treatment affects the outcome.

The study of causal mediation can be seen as a special case of causal inference with time-varying treatments. Rather than having a single treatment that takes different values over time, in mediation analysis we have two different variables—the treatment of interest and the mediator—at different times. This chapter describes a theoretical framework for causal mediation that, like the rest of the book, is based on hypothetical interventions that can be mapped into a target trial. Unlike other approaches to mediation that are based on pure direct effects and total indirect effects, an interventionist framework for mediation opens the door to empirical verification of the causal estimates, a desirable condition for any scientific endeavor.

23.1 Mediation analysis under attack

Consider a randomized trial in which a random sample of cigarette smokers are assigned to either smoking cessation ($A = 0$) or to continuation of smoking ($A = 1$). Suppose that there was perfect adherence: all individuals assigned to $A = 0$ quit smoking and all individuals assigned to ($A = 1$) continued to smoke. The investigators found a beneficial effect of smoking cessation on the risk of myocardial infarction Y at 1 year, i.e., $E[Y|A = 1] > E[Y|A = 0]$

After the beneficial effect of smoking cessation A on the risk of heart disease Y has been established, the investigators wonder whether the benefit is a consequence of the effect of A on reducing hypertension M . Besides data on treatment A at baseline and outcome Y at one year, they also have data on the presence of hypertension M measured at 6 months. (For simplicity, suppose that no individual experienced the outcome Y in the first 6 months.) The causal diagram in Figure 23.1 depicts these variables under the assumption that interventions on M are sufficiently well defined.

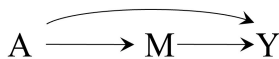


Figure 23.1

Investigators are interested in decomposing the total effect of A on Y into the indirect causal pathways mediated by M and the direct pathways not mediated by M . We then say that investigators are interested in a *causal mediation analysis*.

In the previous chapter we introduced several approaches to formalize the concept of direct effects (Technical Points 22.1 and 22.2) but not the concept of indirect effect. To formalize the decomposition of a total treatment into direct and indirect effects, we can define the pure direct effect and the total indirect effect of A on Y .

The *pure direct effect* of A on Y not through M is the average causal effect of A on Y if, for each individual, the value of M had been set to the value that M would have taken if A had been set to 0, i.e., if M had been set to the value $M^{a=0}$ (which is 1 for some individuals and 0 for others). The *pure direct*

Robins and Greenland (1992) introduced the pure direct effect and the total indirect effect.

Technical Point 23.1

Proof of the mediation formula. In the setting depicted by the causal diagram in Figure 23.1 with treatment A , mediator M , and outcome Y , the counterfactual mean $E[Y^{a=1, M^{a=0}}]$ is equal to

$$\begin{aligned} &= \sum_m E[Y^{a=1, m} | M^{a=0} = m] \Pr[M^{a=0} = m] \text{ by the laws of probability} \\ &= \sum_m E[Y^{a=1, m}] \Pr[M^{a=0} = m] \text{ by the cross-world conditional independence } Y^{a=1, m} \perp\!\!\!\perp M^{a=0} \\ &= \sum_m E[Y | A = 1, M = m] \Pr[M = m | A = 0] \text{ by exchangeability and consistency.} \end{aligned}$$

The cross-world conditional independence used above is assumed when the causal diagram represents an NPSEM-IE, but not when it represents an FFRCISTG model.

effect is then the contrast

$$E[Y^{a=1, M^{a=0}}] - E[Y^{a=0, M^{a=0}}]$$

In our example, the pure direct effect is the effect of smoking cessation A if we could assign to each person the value of hypertension M (1 or 0) that the person would have had if she had quit smoking. This value of hypertension M is known for individuals who actually quit smoking, but it remains unknown for individuals who continued to smoke. We therefore say that $E[Y^{a=1, M^{a=0}}]$, and therefore the pure direct effect, is a cross-world quantity because it involves a counterfactual outcome indexed by two treatment values, $a = 1$ and $a = 0$, that cannot occur simultaneously for the same individual in the same world.

The *total indirect effect*, also a cross-world quantity, is defined as

$$E[Y^{a=1, M^{a=1}}] - E[Y^{a=1, M^{a=0}}]$$

The sum of the pure direct effect and the total indirect effect $E[Y^{a=1, M^{a=1}}] - E[Y^{a=0, M^{a=0}}]$ is, by consistency, the total effect $E[Y^{a=1}] - E[Y^{a=0}]$.

In Figure 23.1, there do not exist unmeasured common causes of the mediator M and the outcome Y . If common causes were present, no direct effects could be identified, including controlled direct effects and the mediation effects, i.e., the pure direct and the total indirect effects. To identify the mediation effects, we need to identify the cross world quantity $E[Y^{a=1, M^{a=0}}]$ under the causal diagram in Figure 23.1. The so-called *mediation formula* does that:

$$\sum_m E[Y | A = 1, M = m] \Pr[M = m | A = 0]$$

The proof is shown in Technical Point 23.1.

The mediation formula seems surprising because it identifies a cross-world counterfactual quantity whose value cannot be empirically confirmed by any experimental intervention on A and M , not even in principle. To achieve this, however, one needs to assume that the causal diagram in Figure 23.1 represents an NPSEM-IE rather than an FFRCISTG, the latter of which is the counterfactual model that we have been using throughout the book (see Technical Point 6.2).

The proof of the mediation formula requires the unverifiable assumption that some counterfactual variables defined in separate worlds are independent. Specifically, the proof requires that the value of the outcome $Y^{a=1, m}$ in a world

Pearl (2001) referred to the pure direct effect and the total indirect effect as the natural direct effect and the natural indirect effect, respectively.

In some very unusual situations, we could identify the value of this cross-world counterfactual quantity by a crossover trial (see Fine Points 2.1 and 3.2).

in which we have jointly intervened by setting the value of treatment to 1 and the value of the mediator to m is independent from the value of the mediator $M^{a=0}$ in a world in which we have intervened to set the value of treatment to 0. Therefore, the variables $Y^{a=1,m}$ and $M^{a=0}$ cannot be simultaneously observed for the same individual in a single world, and thus their independence is not empirically verifiable in any experiment in which A and M are randomly assigned singly or jointly (see Technical Point 7.1).

Under an FFRCISTG model, the pure direct effect and the total indirect effect are not point identified, but sharp bounds can be obtained (Robins and Richardson 2010).

These *cross-world independencies* are assumed under an NPSEM-IE but not under an FFRCISTG model. This is one reason for our privileging an FFRCISTG model over an NPSEM-IE: because cross-world independencies cannot be verified by any randomized experiment, we want to use causal inference methods whose results are, in principle, verifiable.

Based on this discussion, some policy makers are unimpressed that the investigators can identify the pure direct effect. As a cross-world parameter, they find the pure direct effect to be without policy or public health importance because it does not correspond to any intervention. To convince these skeptics, advocates of the NPSEM-IE will need to work harder.

23.2 A defense of mediation analysis

The investigators of the smoking cessation trial, as advocates of the NPSEM-IE, prepare the following rationale to convince the skeptics about the practical utility of the pure direct effect.

Pearl (2001) provided a similar argument to justify the use of pure direct effects.

“Suppose that, starting a year from now, nicotine-free cigarettes will be available and that your policy goal is learning the benefits of nicotine-free cigarettes as soon as possible. To do so, we use the already collected data from the above smoking cessation trial to estimate the two-year risk of heart disease Y under an intervention requiring all smokers to change to nicotine-free cigarettes (when they become available). Suppose that strong experimental evidence exists that (i) the entire causal effect of nicotine on the outcome Y (heart disease) is through its effect on the mediator M (hypertension), and that (ii) the non-nicotine components in cigarettes have no direct causal effect on the mediator M (hypertension). Under assumptions (i) and (ii), the hypertensive status M of a smoker of nicotine-free cigarettes will equal her hypertensive status under non-exposure to cigarettes and hence $E[Y^{a=1,M^{a=0}}]$ will be the risk of heart disease in smokers were all smokers to change to nicotine-free cigarettes a year from now. That is exactly your public health effect of interest and this is what the mediation formula would compute when applied to the data of our study”.

This is interesting. To argue for the substantive importance of the parameter $E[Y^{a=1,M^{a=0}}]$, the NPSEM-IE advocates tell a story about the effect of an intervention on the nicotine content of cigarettes—an intervention that makes no reference to the mediator M at all. Also, they make the “no direct effect” assumptions (i) and (ii) about the absence of direct effects of variables that were not even included on the causal diagram in Figure 23.1.

The investigators’ story states that the variable A can be decomposed into two separable components N and O . The separable component N directly

Technical Point 23.2

When the mediation formula is the g-formula. According to the causal diagram in Figure 23.3, exchangeability holds for the separable components N and O and therefore the mean $E[Y^{n=0,o=1}]$ may be identified by the g-formula. However, in the smoking cessation trial described in the main text, no individual has data ($N = 0, O = 1$). Therefore, positivity does not hold and it appears that the g-formula cannot identify $E[Y^{n=0,o=1}]$. But, given exchangeability and consistency, positivity is a sufficient but not necessary condition for identification by the g-formula; identification only requires that the g-formula is a function of the observed data distribution. We now show that the deterministic bold arrows in Figure 23.2 and the no direct effect assumptions (i) and (ii) together imply that the g-formula is indeed a function of the observed data distribution of (A, M, Y) —that function being the mediation formula.

Under an FFRCISTG represented by Figure 23.2, if data on N and O were available, the mean $E[Y^{n=0,o=1}]$ would be identified by the g-formula

$$\sum_m E[Y | O = 1, M = m] \Pr(M = m | N = 0)$$

since, in the g-formula $N = 0$ needs not be included in the first conditioning event and $O = 1$ needs not be included in the second conditioning event because N is not a parent of Y and O is not a parent of M in Figure 23.2. Furthermore, even though positivity does not hold, the g-formula is a function of the observed data distribution only because $O = 1$ if and only if $A = 1$ and $N = 0$ if and only if $A = 0$. Thus, we can rewrite the g-formula for $E[Y^{n=0,o=1}]$ as

$$\sum_m E[Y | A = 1, M = m] \Pr(M = m | A = 0)$$

which is exactly the mediation formula. This derivation, based on the g-formula, is somewhat heuristic because of the presence of null sets arising from determinism between N, O and L . Robins et al. (2022) provide an alternative, rigorous proof using the SWIG Markov property described in Technical Point 21.12.

Not only does the g-formula equal the mediation formula under the expanded causal diagram in Figure 23.2, but the g-formula also equals the front door formula under the expanded causal diagram in Figure 23.7, as shown in Technical Point 23.3.

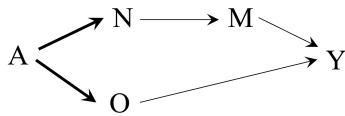


Figure 23.2

These are “no controlled direct effect” assumptions. Assumption (i) says that there is no direct effect of N on Y when controlling for (setting) the value of M .

affects M but not Y . The separable component O directly affects Y but not M . Also, according to the story, each separable component can in principle be intervened on separately. For example, $Y^{n=0,o=1}$ represents the counterfactual outcome under an intervention that removes (only) the nicotine component of cigarettes.

The most direct representation of this causal story is provided by an FFRCISTG model represented by the causal DAG in Figure 23.2 where N is a binary variable representing nicotine exposure, and O is a binary variable representing exposure to the other non-nicotine components of a cigarette. The bold arrows from A to N and O indicate deterministic relationships. This is because, in the actual data from the trial, either one continues to smoke normal cigarettes so $A = N = O = 1$, or quits smoking cigarettes so $A = N = O = 0$. The absence of an arrow from N to Y encodes assumption (i) that N does not have a direct effect on Y . The absence of an arrow from O to M encodes assumption (ii) that O does not have an effect on M . Figure 23.3 shows the associated SWIG. If O is not a cause of M for every individual we can additionally write the random variable $M^{n,o}$ as M^n .

Under this characterization of the problem, the g-formula that identifies $E[Y^{n=0,o=1}]$ is exactly the mediation formula (see proof in Technical Point 23.2). Let us review the lessons we learned.

For a researcher that initially accepted an NPSEM-IE associated with the

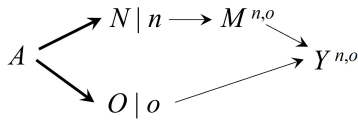


Figure 23.3

DAG in Figure 22.1, the mean $E[Y^{a=1, M^{a=0}}]$ was already identified by the mediation formula so the story of separable effects of N and O did not contribute to identification. Rather, it served only to show that the parameter $E[Y^{a=1, M^{a=0}}]$ —and thus the pure direct effect and the total indirect effect—encodes a parameter $E[Y^{n=0, o=1}]$ of public health interest. For a researcher that takes the FFRCISTG point of view, the story not only provides an interventional interpretation of the mean $E[Y^{a=1, M^{a=0}}]$ as the mean $E[Y^{n=0, o=1}]$, but in addition makes both means identifiable via the g-formula or, equivalently, the mediation formula. In fact, when we interpret absence of an arrow as an absence of a direct causal effect for all individuals, $Y^{a=1, M^{a=0}}$ and $Y^{n=0, o=1}$ are equal for every individual.

This characterization of the problem also allows us to define and identify the *separable direct effect* of the component N on the outcome Y , $E[Y^{n=1, o=1}] - E[Y^{n=0, o=1}]$, which is a controlled direct effect for the separable component N . This effect is also equal to the total indirect effect $E[Y^{a=1}] - E[Y^{a=1, M^{a=0}}]$. Analogously, $E[Y^{n=1, o=1}] - E[Y^{n=1, o=0}]$ is the pure direct effect $E[Y^{a=1, M^{a=0}}] - E[Y^{a=0}]$.

23.3 Empirically verifiable mediation

The interventional interpretation of $E[Y^{a=1, M^{a=0}}]$ is only valid if the story about the separable effects of N and O is correct and Figure 23.3 represents an FFRCISTG. An advantage of the interventional interpretation is that the validity of this story can be, in principle, empirically refuted via a randomized trial, as we now describe.

When nicotine-free cigarettes become available in a year, we conduct a new randomized trial in which a random sample of cigarette smokers are randomly assigned to one of three groups: smoking cessation ($A = N = O = 0$), continuation of smoking of standard cigarettes ($A = N = O = 1$), or continuation of smoking of nicotine-free cigarettes ($N = 0, O = 1$). That is, the new trial has the same two arms as the original trial plus an additional arm of nicotine-free cigarettes. In the absence of temporal trends, the mean outcomes for the first two arms should be the same in both trials. We assume that the sample size of both trials is large enough to ignore sampling variability.

By randomization, the mean outcome $E[Y|N = 0, O = 1]$ in the new trial is expected to equal $E[Y^{n=0, o=1}]$. Therefore, if our story is correct, we expect that $E[Y|N = 0, O = 1]$ will equal the mediation formula from the earlier trial. Otherwise, the N and O story has been empirically refuted, which implies that one or more of the following assumptions is incorrect: (i) no direct effect of nicotine on the outcome, (ii) no direct effect of other, non-nicotine components on the mediator, (iii) no unmeasured common cause U of M and Y on the causal diagram in Figure 23.2.

If $E[Y|N = 0, O = 1]$ differs from the mediation formula, we can use the data from this new trial to investigate which of the assumptions (i)-(iii) are false. To do so, we start by checking whether O and M are associated among those with $N = 0$, i.e.,

$$E[M|N = 0, O = 1] - E[M|N = 0, O = 0] \neq 0$$

Fine Point 23.1

Empirical falsification of the assumptions for separable effects. Suppose we had conducted the three-arm trial with interventions on N and O , and found that N and Y are associated within joint levels of M and O . Then, as stated in the main text, we conclude that either assumption (i) or assumption (iii), or both, are false.

To determine which of the two assumptions are false, we would need another randomized trial with, say, 8 arms in which we intervene on M as well as on N and O . If N has no direct effect on Y except through M , N will be independent of Y given M and O in the eight-arm trial. If M and Y do not share an unmeasured common cause, the conditional distribution of Y given M within levels of N and O should be the same in the three-arm trial and in the eight-arm trial.

One might wonder how Figure 23.2 can have an unmeasured common cause of M and Y left off the graph if we are correct in assuming that Figure 23.1 represents an FFRCISTG model with no common causes of M and Y . The explanation is that common causes may only be present (i.e., active) under interventions that assign different values of N and O to each person. The FFRCISTG model associated with Figure 23.1, unlike that associated with Figure 23.2, only considers interventions $A = N = O = 1$ and $A = N = O = 0$ that assign the same value to each person. Interestingly, if a common cause U of Y and M is the only reason $E[Y|N = 0, O = 1]$ in the new trial differs from the mediation formula, it is still the case that $Y^{n=0,o=1} = Y^{a=1,M^{a=0}}$ for every individual. However, $E[Y^{n=0,o=1}] = E[Y^{a=1,M^{a=0}}]$ is not identified by the mediation formula. Robins et al. (2022) describe a hypothetical but realistic study of treatment for river blindness under which Figure 23.1 is an FFRCISTG but Figure 23.2 is not because a common cause U of Y and M is missing from the latter graph.

If that is the case, then assumption (ii) is refuted and the arrow $O \rightarrow M$ should be added to Figure 23.2 (and the corresponding arrow to Figure 22.3).

We next check whether N and Y are associated within joint levels of M and O , i.e., whether

$$E[Y|N = 1, O = 1, M = m] - E[Y|N = 0, O = 1, M = m] \neq 0$$

for some values of m . If that is the case, then assumption (i) is refuted (which implies that an arrow $N \rightarrow Y$ should be added to the causal diagrams) or an unmeasured common cause U of M and Y exists (and should be added to the causal diagrams). See Fine Point 23.1.

Now suppose that in two years, when the results of the new trial become available, the mean outcome estimates are equal in the arms that correspond to the same intervention in both trials, but the estimate of the mean outcome in the third arm of the new trial, $E[Y|N = 0, O = 1]$, differs from the mediation formula in the first trial. How would different people react?

Both a person who had assumed that Figure 23.1 was an NPSEM-IE and a person who only assumed that it was an FFRCISTG would agree that the story of separable effects of O and N was incorrect, and that $E[Y^{n=0,o=1}]$ is correctly estimated by the mean outcome in the third arm of the new trial and not by the mediation formula in the earlier trial.

Those who had assumed an NPSEM-IE can continue to believe that the mean $E[Y^{a=1,M^{a=0}}]$ is still equal to the mediation formula, but they cannot justify the policy interest of the pure direct effect as the effect of taking nicotine-free cigarettes. Those who assumed an FFRCISTG may have little interest in the original mediation formula. Instead they will be interested in what they have learned about the effects of nicotine-free cigarettes on the outcome from the three-arm randomized trial. They will also be interested in learning that one or more of the following assumptions were false: (i) no direct effect of

If $E[Y^{a=1, M^{a=0}}]$ were always point identified by the mediation formula, the sharp bounds in Robins and Richardson (2010) would be violated.

nicotine on the outcome, (ii) no direct effect of other, non-nicotine components on the mediator, (iii) no unmeasured common cause U of M and Y .

Continue to assume that Figure 23.1 represents an FFRCISTG model and that $E[Y^{n=0, o=1}]$ in the future nicotine-free cigarette trial differs from the mediation formula. An interesting question is whether treatment A can always be decomposed into other (possibly unknown) intervenable components N' and O' such that Figure 23.2 is an FFRCISTG and the no direct effect assumptions (i) and (ii) hold, in which case $E[Y^{n'=0, o'=1}]$ would equal $E[Y^{a=1, M^{a=0}}]$ and both would equal the mediation formula? The answer is no, since otherwise $E[Y^{a=1, M^{a=0}}]$ would always be point identified by the mediation formula, which is not the case.

23.4 An interventionist theory of mediation

This theory of interventionist mediation was presented by Robins and Richardson (2010) and extended by Robins, Richardson, and Shpitser (2022). The theory has been extended to survival analysis (Didelez 2019, Aalen et al. 2020), competing risks (Stensrud et al. 2020, 2021), and settings with interference (Shpitser et al. 2021).

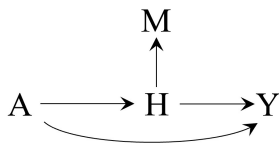


Figure 23.4

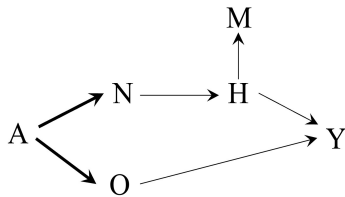


Figure 23.5

Stensrud et.al. (2021) emphasized the need for the above equalities of mean outcomes across arms of the six-arm trial.

In this chapter we have introduced an interventionist theory of causal mediation that differs from the standard approach based on pure direct and total indirect effects. Because the interventionist theory was developed in response to perceived deficiencies in the standard theory, we first described the standard theory and its problematic aspects resting on the use of cross-world (nested) counterfactuals. However, the interventionist theory can be viewed as autonomous, providing a self-contained framework for discussing mediation without reference to cross-world nested counterfactuals.

To do so, we used a (simplified) randomized trial with a time-fixed treatment as an intervention. The key idea was reframing the mediation question as a question about the effects of interventions on substantively meaningful, separable components (N and O) of treatment A on Y . If N and O are separable effects of A , then, in a future randomized trial with six arms:

1. treat with $a = 1$
2. treat with $a = 0$
3. treat with $n = 1, o = 1$
4. treat with $n = 0, o = 0$
5. treat with $n = 0, o = 1$
6. treat with $n = 1, o = 0$

then the following two statements are true. First, the mean of the outcome Y in the first arm equals that in the third and the mean outcome in the second arm equals that in the fourth. Second, $E[Y^{n=0, o=1}]$ is identified as the mean outcome in the fifth arm and $E[Y^{n=1, o=0}]$ as the mean outcome in the sixth arm of the future trial, regardless of whether $E[Y^{n=0, o=1}]$ is not identified from the observed data.

We can use currently available data on A , M , and Y to identify the separable effects of N and O under the assumptions of (i) no unmeasured common cause of the mediator M and the outcome Y and (ii) no direct effects of the component O on the mediator M and of component N on outcome Y . In particular, $E[Y^{n=0, o=1}]$ equals the mediation formula.

Fine Point 23.2

Separable effects with a surrogate mediator. In the following we adopt the theory of causal diagrams introduced in Section 9.5. Since the arrow from M to Y is only causally interpretable when there exist well-defined interventions for the effect of M on Y , the DAG in Figure 23.1 is not causal when those interventions do not exist. That is, M is not a mediator but rather a surrogate for an unknown variable H that is the true mediator for which well-defined interventions exist. The causal DAG in Figure 23.4 represents this scenario.

Consider the causal diagram in Figure 23.5, which is an extension of Figure 23.4 with separable components N and O . In contrast to Figure 23.2, N and Y are not d-separated given M and O in Figure 23.5. Suppose, surprisingly, we observe that N and Y are independent given M and O in our three-arm trial data. How should we interpret this observation? There are 4 possibilities: a) we were mistaken: Figure 23.1 rather than 23.4 is the true causal diagram and M is the true causal mediator, b) M is a one-to-one deterministic function of the true causal mediator H , c) there is a non-deterministic faithfulness violation in Figure 23.4, and d) N and O are not conditionally independent but the three-arm trial was too small to detect their dependence. Advocates of causal discovery would tend to adopt interpretation a) if the sample size in the three-arm trial was large (see Technical Point 10.7).

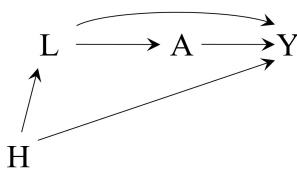


Figure 23.6

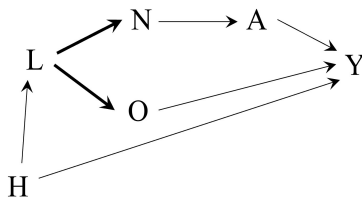


Figure 23.7

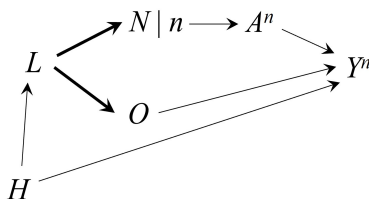


Figure 23.8

We now discuss one further reason to prefer an interventionist approach over earlier approaches to causal mediation. Specifically, it is often the case that

interventions on the putative mediator M are not well-defined and counterfactuals like $Y^{a,m}$ are then not meaningful. Then neither the pure direct effect nor the controlled direct effects (described in Technical Point 22.1) based on M exist. In contrast, the interventionist theory is concerned with effects that exist, even if interventions on M are not well-defined, as long as substantively meaningful separable components N and O exist and can be intervened on as discussed above.

To summarize, an interventionist approach to mediation includes the following components.

- the hypothesis that treatment A can be decomposed into multiple, substantially meaningful, separable components, each of which contributes to the overall effect of treatment and which can, in principle, be independently intervened on
- the assumptions needed to identify the effects of the separable components are, in principle, empirically verifiable in future randomized trials in which these components are actually intervened on
- well-defined interventions on a purported mediator need not exist
- enhanced communication with subject-matter experts owing to the substantive specificity of the separable components
- when the identifying assumptions hold, the identifying g-formula is identical to the mediation formula; however, the identified causal effects refer to interventions on the separable components.

For simplicity, we considered an example with only two separable components, but the interventionist framework can accommodate multiple separable components of treatment, including those that vary over time.

In this chapter we have reviewed theoretical concepts of mediation, but we have not emphasized the practical aspects of data analyses to estimate pure

Technical Point 23.3

Path-specific effects and the front door formula. Under Figure 23.6, which is the modified version of Figure 9.14, the total effect of L on Y is not identified. This is because the contribution of the pathway $L \rightarrow Y$ to the total effect cannot be separated from the confounding effect of H . In contrast, under the unmodified causal diagram in Figure 9.14 lacking the $L \rightarrow Y$ edge, we showed in Fine Point 9.5 that the total effect of L on Y was identified by the front door formula, as $L \rightarrow A \rightarrow Y$ was the sole causal pathway.

We now consider whether the effect of L along the pathway $L \rightarrow A \rightarrow Y$ is also identified, and by the same front door formula when the $L \rightarrow Y$ edge is present. In the spirit of this chapter, we shall use the substantive story in Fine Point 9.5 to provide an interventionist formulation of this question. Figure 23.7 is an expanded graph where N records the value of BMI L reported to the physician responsible for prescribing drug A , and O records the value of L used to determine referral to physical therapy and diet counseling. The bold arrows indicate a deterministic relationship, since, in the observed data, we always have $L = N = O$. Figure 23.8 is a SWIG representing an (unethical) intervention in which a value n of BMI different from the truth $L = N$ was reported to the physician, but the true value $L = O$ was used for the referrals. In contrast with earlier in this chapter, only N has been intervened on. Further, as was our goal, the effect on Y of the intervention on n is restricted to the path through A . Since $L \equiv O$ even in the intervened world, we have no need of O . We therefore remove O from Figures 23.7 and 23.8 and again have Y as a child of L . Then noting trivially that $Y^n \perp\!\!\!\perp N|L$ on Figure 23.8, $E[Y^n]$ should be given by the g-formula based on Figure 23.7 which equals

$$\sum_{l,a} E[Y|A=a, L=l] \Pr[L=l] \Pr[A=a|N=n] = \sum_a \left\{ \sum_l E[Y|A=a, L=l] \Pr[L=l] \right\} \Pr[A=a|L=n]$$

where the last equality is by the determinism $L \equiv N$ in the data. The right hand side is indeed the front door formula.

However, this derivation is somewhat heuristic due to the presence of null sets arising from determinism. A rigorous proof is readily obtained by combining determinism with the approach used in Technical Point 21.12 to prove the front door formula. Note also that, substantively speaking, one generally would not consider N and O as separable components of L . But that is irrelevant. What matters is that our substantive causal story implies an expanded graph 23.7 containing the new intervention variable N that is deterministically related to L in the actual world (Stensrud et al., 2023). Wen et al. (2023) have independently and concurrently obtained results essentially equivalent to those of this technical point. Fulcher et al. (2020) had earlier shown that the above front door formula identifies the cross world counterfactual quantity $E[Y^{L,A^{l=n}}]$ under the NPSEM-IE model associated with Figure 23.6. Both Wen et al. and this Technical Point can be seen as interventionist reformulations of Fulcher et al's result.

direct effects and separable direct effects. In the previous chapter, we adopted a similar approach when introducing controlled direct effects and principal stratum direct effects. We hope that our presentation has clarified the scientific advantages of an interventionist approach to the identification of direct effects.

In the absence of randomized trials with actual interventions on either the mediator (for controlled direct effects) or components of treatment (for separable direct effects), all these methodologies rely on observational data and therefore on the exchangeability assumptions that we have discussed throughout the book. Hence, valid mediation analyses require adjustment for the confounders for the effect of treatment and for the effect of the mediator in addition to other assumptions that we discussed in this chapter. Our causal diagram in Figure 23.1 was intended as a teaching device to explore theoretical issues related to mediation in a simplified setting, not as a realistic representation of most studies of causal mediation. In practice, causal mediation analyses are observational analyses that rely on more heroic assumptions than non-mediation analyses.

For separable effects, additional issues arise if there is a measured cause of M and Y that is a child of N and O (Robins and Richardson 2010; Robins et al. 2022).

References

- Aalen OO, Stensrud MJ, Didelez V, Daniel R, Røysland K, Strohmaier S (2020). Time-dependent mediators in survival analysis: Modeling direct and indirect effects with the additive hazards model. *Biometrical Journal* 62(3):532–549
- Abadie A (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies* 72(1): 1-19.
- Abadie A, Chingos MM, West MR (2018). Endogenous stratification in randomized experiments. *The Review of Economics and Statistics* 100(4): 567–580.
- Abadie A, Imbens GW (2006). Large sample properties of matched estimates for average treatment effects. *Econometrica* 74:235-267.
- Amrhein V, Greenland S, McShane B (2019). Scientists rise up against statistical significance. *Nature* 567:305-307.
- Angrist JD, Imbens GW (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association* 90:431-442.
- Angrist JD, Imbens GW, Rubin DB (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91:444-455.
- Angrist JD, Krueger AB (1999). Empirical strategies in labor economics. In: Ashenfelter O, Card D, eds. *Handbook of Labor Economics* 3A, 1277-1366. Elsevier.
- Angrist JD, Pischke J-S (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- Avagyan V, Vansteelandt S (2021). Stable inverse probability weighting estimation for longitudinal studies. *Scandinavian Journal of Statistics* 48(3), 1046– 1067.
- Baiocchi M, Small D, Lorch S, Rosenbaum P (2010). Building a stronger instrument in an observational study of perinatal care for premature infants. *Journal of the American Statistical Association* 105(492): 1285-1296.
- Baiocchi M, Cheng J, Small D (2014). Instrumental variable methods for causal inference. *Statistics in Medicine* 33: 2297-2340.
- Baker SG, Lindeman KS (1994). The paired availability design, a proposal for evaluating epidural analgesia during labor. *Statistics in Medicine* 13:2269-2278.
- Baker SG, Kramer BS, Lindeman KL (2016). Latent class instrumental variables. A clinical and biostatistical perspective. *Statistics in Medicine* 35:147-160 (errata in *Statistics in Medicine* 2019; 38: 901).
- Balke A, Pearl J (1994). Probabilistic evaluation of counterfactual queries. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, Volume I, pp. 230-237.

- Balke A, Pearl J (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association* 92(439):1171-1176.
- Bang H, Robins JM (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* 61: 962-972 (errata in *Biometrics* 2008;64:650).
- Barnow SN, Cain GG, Goldberger AS (1980). Issues in the analysis of selectivity bias. In: *Evaluation Studies* (Vol. 5). Stromsdorfer EW, Farkas G, eds. Beverly Hills, CA: Sage Publication.
- Berkson J (1946). Limitations of the application of fourfold table analysis to hospital data. *Biometrics* 2: 47-53.
- Berkson J (1955). The statistical study of association between smoking and lung cancer. *Proceedings of the Staff Meetings of the Mayo Clinic* 30: 319-348.
- Blot WJ, Day NE (1979). Synergism and interaction: are they equivalent? [letter] *American Journal of Epidemiology* 110:99-100.
- Blyth CR (1972). On Simpson's paradox and the sure-thing principle. *Journal of the American Statistical Association* 67:364-66.
- Bonet B (2001). Instrumentality tests revisited. In: *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*. San Francisco, CA: Morgan Kaufmann, pp. 48-55.
- Bound J, Jaeger D, Baker R (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak. *Journal of the American Statistical Association* 90:443-450.
- Brown LD, Cai T, DasGupta A (2001). Interval estimation for a binomial proportion (with discussion). *Statistical Science* 16:101-133.
- Brookhart MA, Schneeweiss S (2007). Preference-based instrumental variable methods for the estimation of treatment effects: assessing validity and interpreting results. *International Journal of Biostatistics* 3:14.
- Brookhart MA, Rassen J, Schneeweiss S (2010). Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiology and Drug Safety* 19:537-554.
- Buehler RJ (1982). Some ancillary statistics and their properties. Rejoinder. *Journal of the American Statistical Association* 77:593-594.
- Cain LE, Robins JM, Lanoy E, Logan R, Costagliola D, Hernán MA. When to start treatment? A systematic approach to the comparison of dynamic regimes using observational data. *International Journal of Biostatistics* 2010; 6(2) Article 18.
- Card D (1990). The Impact of the Mariel Boatlift on the Miami Labor Market. *Industrial and Labor Relations Review* 43(2):245-257.

- Card D (1995). Using geographic variation in college proximity to estimate the return to schooling. In: Christofides LN, Grant EK, Swidinsky R, eds. *Aspects of labor market behavior: Essays in honour of John Vanderkamp*. Toronto, Canada: University of Toronto Press.
- Casella G, Berger RL (2002). *Statistical Inference*, 2nd ed. Pacific Grove, CA: Duxbury Press.
- Cochran WG (1972). Observational studies. In: Bancroft TA, ed. *Statistical Papers in Honor of George W Snedecor*. Iowa State University Press, pp. 77-90
- Cole SR, Hernán MA (2002). Fallibility in estimating direct effects. *International Journal of Epidemiology* 31(1):163-165.
- Cornfield J, Haenszel W, Hammond EC, Lilienfeld AM, Shimkin MB, Wynder EL (1959). Smoking and lung cancer: Recent evidence and a discussion of some questions. *Journal of the National Cancer Institute* 22:173-203. Reprinted in *International Journal of Epidemiology* 2009; 38(5): 1175-1191.
- Cox DR (1958). *Planning of Experiments*. New York, NY: John Wiley and Sons.
- Cox DR, Wermuth N (1999). Likelihood factorizations for mixed discrete and continuous variables. *Scandinavian Journal of Statistics* 26(2): 209-220.
- Cui Y, Pu H, Shi X, Miao W, Tchetgen Tchetgen EJ (2024). Semiparametric proximal causal inference. *Journal of the American Statistical Association* 119(546): 1348-1359.
- Dahabreh IJ, Hernán (2019). Extending inferences from a randomized trial to a target population. *European Journal of Epidemiology* 34(8): 719-722.
- Dahabreh IJ, Petito LC, Robertson SE, Hernán MA, Steingrimsson JA (2020a). Toward causally interpretable meta-analysis: transporting inferences from multiple randomized trials to a new target population. *Epidemiology* 31(3):334-344.
- Dahabreh IJ, Robertson SE, Steingrimsson JA, Stuart EA, Hernán MA (2020b). Extending inferences from a randomized trial to a new target population. *Statistics in Medicine* 39(14):1999-2014.
- Danaei G, Robins JM, Hu FB, Manson J, Hernán MA (2016). Effect of weight loss on coronary heart disease and mortality in middle-aged or older women: sensitivity analysis for unmeasured confounding by undiagnosed disease. *Epidemiology* 27(2): 302-310.
- Davey Smith G, Ebrahim S (2004). Mendelian randomization: prospects, potentials, and limitations. *International Journal of Epidemiology* 33: 30-42.
- Dawid AP (1979). Conditional independence in statistical theory (with discussion). *Journal of the Royal Statistical Society B* 41:1-31.
- Dawid AP (2000). Causal inference without counterfactuals (with discussion). *Journal of American Statistical Association* 95: 407-424.

- Dawid AP (2002). Influence diagrams for causal modelling and inference. *International Statistical Review* 70: 161-189.
- Dawid, A. P. (2003). Causal inference using influence diagrams: The problem of partial compliance (with discussion). In: Green PJ, Hjort NL, Richardson S, eds. *Highly Structured Stochastic Systems*. New York, NY: Oxford University Press, 45-65.
- de Finetti (1972). *Probability, Induction, and Statistics*. John Wiley & Sons.
- Deaton A (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature* 48 424-455.
- Detels R, Muñoz A, McFarlane G, Kingsley LA, Margolick JB, Giorgi J, Schrager L, Phair J, for the Multicenter AIDS Cohorts Study (1998). *JAMA* 280(17): 1497-1503.
- Diaz I, Williams N, Hoffman KL, Schenck EJ (2021). Nonparametric causal effects based on longitudinal modified treatment policies. *Journal of the American Statistical Association* 118(542):846-857.
- Didelez V (2019). Defining causal mediation with a longitudinal mediator and a survival outcome. *Lifetime Data Analysis* 25(4):593-610.
- Didelez V, Sheehan N (2007). Mendelian randomization as an instrumental variable approach to causal inference. *Statistical Methods in Medical Research* 16: 309-330.
- Ding P, VanderWeele TJ, Robins JM (2017). Instrumental variables as bias amplifiers with general outcome and confounding. *Biometrika* 104(2): 291-302.
- Dorn HF (1953). Philosophy of inferences for retrospective studies. *American Journal of Public Health* 43: 677-83.
- Dosemeci M, Wacholder S, Lubin JH (1990). Does nondifferential misclassification of exposure always bias a true effect toward the null value? *American Journal of Epidemiology* 132:746-748.
- Earle CC, Tsai JS, Gelber RD, Weinstein MC, Neumann PJ, Weeks JC (2001). Cancer in the elderly: instrumental variable and propensity analysis. *Journal of Clinical Oncology* 19(4): 1064-1070.
- Efron B, Hinkley DV (1978). Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information. *Biometrika* 65: 657-687.
- Feinstein AR (1971). Clinical biostatistics. XI. Sources of 'chronology bias' in cohort statistics. *Clinical Pharmacology and Therapeutics* 12(5): 864-79.
- Fisher RA (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A*; 222 (594-604): 309-368.
- Flanders WD (2006). On the relation of sufficient component cause models with potential (counterfactual) models. *European Journal of Epidemiology* 21: 847-853.

- Flanders WD, Klein M, Darrow LA, Strickland MJ, Sarnat SE, Sarnat JA, Waller LA, Winquist A, Tolbert PE (2011). A Method for Detection of Residual Confounding in Time-series and Other Observational Studies. *Epidemiology* 22(1): 59-67.
- Fleming TR, Harrington DP (2005). *Counting Processes and Survival Analysis*, 2nd ed. New York: Wiley.
- Frangakis CE, Rubin DB (2002). Principal stratification in causal inference. *Biometrics* 58(1); 21-29.
- Fulcher IR, Shpitser I, Marealle S, Tchetgen Tchetgen EJ (2020). Robust inference on population indirect causal effects: the generalized front door criterion. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 82(1); 199-214.
- Glymour MM, Spiegelman D (2016). Evaluating public health interventions: 5. Causal inference in public health research-Do sex, race, and biological factors cause health outcomes? *American Journal of Public Health* 107(1): 81-85.
- Glymour MM, Tchetgen Tchetgen EJ, Robins JM (2012). Credible Mendelian randomization studies: approaches for evaluating the instrumental variable assumptions. *American Journal of Epidemiology* 175: 332-339.
- Goodfellow I, Bengio Y, Courville A (2016). *Deep Learning*. Cambridge, MA: MIT Press.
- Greenland S (1977). Response and follow-up bias in cohort studies. *American Journal of Epidemiology* 106(3):184-187.
- Greenland S (1980). The effect of misclassification in the presence of covariates. *American Journal of Epidemiology* 112(4):564-569.
- Greenland S (1987). Interpretation and choice of effect measures in epidemiologic analyses. *American Journal of Epidemiology* 125(5):761-768.
- Greenland S (1991). On the logical justification of conditional tests for two-by-two contingency tables. *The American Statistician* 45(3):248-251.
- Greenland S (1996a). Basic methods for sensitivity analysis of bias. *International Journal of Epidemiology* 25(6):1107-1116.
- Greenland S (1996b). Absence of confounding does not correspond to collapsibility of the rate ratio or rate difference. *Epidemiology* 7(5):498-501.
- Greenland S (2000). An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology* 29(4):722-729.
- Greenland S (2003). Quantifying biases in causal models: classical confounding versus collider-stratification bias. *Epidemiology* 14:300-306.
- Greenland S (2009a). Bayesian perspectives for epidemiologic research. III. Bias analysis via missing-data methods. *International Journal of Epidemiology* 38:1662-1673.
- Greenland S (2009b). Relaxation penalties and priors for plausible modeling of nonidentified bias sources. *Statistical Science* 24:195-210.

- Greenland S (2019). Some misleading criticisms of P-values and their resolution with S-values. *The American Statistician* 73(supplement 1): 106-114.
- Greenland S, Brumback B (2002). An overview of relations among causal modelling methods. *International Journal of Epidemiology* 31:1030-1037.
- Greenland S, Lash TL (2008). Bias analysis. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology*, 3rd edition. Philadelphia, PA: Lippincott Williams & Wilkins, pp. 345-380.
- Greenland S, Lash TL, Rothman KJ (2008). Concepts of interaction. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology*, 3rd edition. Philadelphia, PA: Lippincott Williams & Wilkins, pp. 71-83.
- Greenland S, Poole C (1988). Invariants and noninvariants in the concept of interdependent effects. *Scandinavian Journal of Work, Environment & Health* 14(2):125-129.
- Greenland S, Robins JM (1986). Identifiability, exchangeability, and epidemiological confounding. *International Journal of Epidemiology* 15:413-19.
- Greenland S, Robins JM (1988). Conceptual problems in the definition and interpretation of attributable fractions. *American Journal of Epidemiology* 128:1185-1197.
- Greenland S, Robins JM (2009). Identifiability, exchangeability, and confounding revisited. *Epidemiologic Perspectives & Innovations* 6:4.
- Greenland S, Pearl J, Robins JM (1999). Causal diagrams for epidemiologic research. *Epidemiology* 10:37-48.
- Greenland S, Robins JM, Pearl J (1999). Confounding and collapsibility in causal inference. *Statistical Science* 14:29-46.
- Greenland S, Rothman KJ (2008). Introduction to stratified analysis. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology*, 3rd edition. Philadelphia, PA: Lippincott Williams & Wilkins, pp. 258-282.
- Grieve AP (2003). The number needed to treat: a useful clinical measure or a case of the Emperor's new clothes? *Pharmaceutical Statistics* 2:87-102.
- Hajek J (1971). Comment on "An essay on the logical foundations of survey sampling by D. Basu". In: Godambe VP, Sprott DA, eds. *Foundations of Statistical Inference*. New York City, NY: Holt, Rinehart, and Winston; 1971 (p. 236).
- Halloran ME, Struchiner CJ (1995). Causal inference in infectious diseases. *Epidemiology* 6: 142-151.
- Hahn J, Todd P, van der Klaauw W (2001). Identification and estimation of treatment effects with a regression discontinuity design. *Econometrica* 69: 201-209
- Hansen BB (2008). The prognostic analogue of the propensity score. *Biometrika* 95(2): 481-488.

- Harrell FE (2015). *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. New York: Springer.
- Hastie TJ, Tibshirani RJ (1990). *Generalized Additive Models*. London: Chapman & Hall.
- Hastie TJ, Tibshirani RJ, Friedman J (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York: Springer.
- Heckman JJ, Vytlačil EJ (1999). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences USA* 96:4730-4734.
- Hernán MA (2010). The hazards of hazard ratios. *Epidemiology* 21(1):13-15.
- Hernán MA (2016). Does water kill? A call for less casual causal inferences. *Annals of Epidemiology* 26: 674-680.
- Hernán MA (2017). Selection bias without colliders. *American Journal of Epidemiology* 21(1):13-15.
- Hernán MA, Brumback B, Robins JM (2001). Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association* 96(454):440-448.
- Hernán MA, Clayton D, Keiding N (2011). The Simpson's paradox unraveled. *International Journal of Epidemiology* 40:780-785.
- Hernán MA, Cole SR, Margolick JB, Cohen MH, Robins JM (2005). Structural accelerated failure time models for survival analysis in studies with time-varying treatments. *Pharmacoepidemiology and Drug Safety* 14(7):477-491.
- Hernán MA, Hernández-Díaz S, Werler MM, Mitchell AA (2002). Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *American Journal of Epidemiology* 155:176-184.
- Hernán MA, Hernández-Díaz S, Robins JM (2004). A structural approach to selection bias. *Epidemiology* 15:615-625.
- Hernán MA, Hernández-Díaz S (2012). Beyond the intention to treat in comparative effectiveness research. *Clinical Trials*; 9(1):48-55.
- Hernán MA, Hsu J, Healy B (2019). A second chance to get causal inference right: a classification of data science tasks. *Chance* 32(1):42-49.
- Hernán MA, Robins JM (2006a). Estimating causal effects from epidemiological data. *Journal of Epidemiology and Community Health* 60: 578-586.
- Hernán MA, Robins JM (2006b). Instruments for causal inference: An epidemiologist's dream? *Epidemiology* 17(4): 360-372.
- Hernán MA, Alonso A, Logan R, Grodstein F, Michels KB, Willett WC, Manson JE, Robins JM (2008). Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease (with discussion). *Epidemiology* 19(6):766-779.

- Hernán MA, Robins JM (2016). Using big data to emulate a target trial when a randomized trial is not available. *American Journal of Epidemiology* 183(8):758-764.
- Hernán MA, Robins JM (2017). Per-protocol analyses of pragmatic trials. *New England Journal of Medicine* 377(14): 1391-1398.
- Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I (2016). Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *Journal of Clinical Epidemiology* 79: 70-75.
- Hernán MA, Taubman SL (2008). Does obesity shorten life? The importance of well-defined interventions to answer causal questions. *International Journal of Obesity* 32: S8-S14.
- Hernán MA, VanderWeele TJ (2011). Compound treatments and transportability of causal inference. *Epidemiology* 22:368-377.
- Holland PW (1986). Statistics and causal inference (with discussion). *Journal of the American Statistical Association* 81:945-961.
- Hoover DR, Muñoz A, He Y, Taylor JMG, Kingsley L, Chmiel JS, Saah A (1994). The effectiveness of interventions on incubation of AIDS as measured by secular increases within the population. *Statistics in Medicine* 13:2127-2139.
- Horvitz DG, Thompson DJ (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* 47:663-685.
- Huang Y, Valtorta M (2006). Pearl's calculus of intervention is complete. In: *Proceedings of the 22nd Conference in Uncertainty in Artificial Intelligence*. Cambridge, MA: AUAI Press, pp. 217-224.
- Hosmer DW, Lemeshow S, May S (2008). *Applied Survival Analysis: Regression Modelling of Time to Event Data*. Hoboken, NJ: Wiley.
- Hudgens MG, Halloran ME (2009). Towards causal inference with interference. *Journal of the American Statistical Association* 103:832-842.
- Hume D (1748). *An Enquiry Concerning Human Understanding*. Reprinted and edited 1993, Indianapolis/Cambridge: Hackett.
- Imai K, Ratkovic M (2015). Robust estimation of inverse probability weights for marginal structural models. *Journal of the American Statistical Association* 110, 1013-1023.
- Imbens GW (2004). Nonparametric estimation of average treatment effects under exogeneity: a review. *The Review of Economics and Statistics* 86 (1): 4-29.
- Imbens GW, Angrist JD (1994). Identification and estimation of local average treatment effects. *Econometrica* 62:467-475.
- Imbens GW, Rubin DB (1997). Estimating outcome distributions for compliers in instrumental variables models. *Review of Economic Studies* 64:555-574.

- Imbens G, Lemieux T (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics* 142 (2): 615–635.
- Kalbfleisch and Prentice (2002). *The Statistical Analysis of Failure Time Data*. Hoboken, NJ: Wiley.
- Kallus N, Santacatterina M (2018). Optimal balancing of time-dependent confounders for marginal structural models. arXiv preprint arXiv:1806.01083.
- Katan MB (1986). Apolipoprotein E isoforms, serum cholesterol, and cancer. *Lancet*, 1:507-508.
- Kleiner A, Talwalkar A, Sarkar P, Jordan MI (2014). A scalable bootstrap for massive data. *Journal of the Royal Statistical Society B*, 76 (Part 4):795-816.
- Kosinski S, Stillwell D, Graepel T (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences* 110(15): 5802-5805.
- Kurth T, Walker AM, Glynn RJ, Chan KA, Gaziano JM, Berger K, Robins JM (2006). Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. *American Journal of Epidemiology* 163(3): 262-270.
- Korn EL, Baumrind S (1998). Clinician preferences and the estimation of causal treatment differences. *Statistical Science* 13:209-235
- Laupacis A, Sackett DL, Roberts RS (1988). An assessment of clinically useful measures of the consequences of treatment. *New England Journal of Medicine* 318:1728-1733.
- Lauritzen SL, Dawid AP, Larsen BN, Leimer H-G (1990). Independence properties of directed Markov fields. *Networks* 20:491-505.
- Lash TL, Fox MP, Fink AK (2009). *Applying Quantitative Bias Analysis to Epidemiologic Data*. New York: Springer.
- Lewis D (1973). *Counterfactuals*. Oxford: Blackwell.
- Liang K-Y, Zeger SL (1986). Longitudinal data analysis using generalized linear models. *Biometrika* 73(1):13-22.
- Lin V, McGrath S, Zhang Z, Logan RW, Petito LC, Young JG, Hernán MA (2019). gfoRmula: Parametric G-Formula. R package version 0.2.1. <https://CRAN.R-project.org/package=gfoRmula>.
- Lin L, Mukherjee R, Robins JM (2020). On nearly assumption-free tests of nominal confidence interval coverage for causal parameters estimated by machine learning. *Statistical Science* 35(3): 518-539.
- Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology* 21(3):383-8 (erratum in *Epidemiology* 2010;21(4):589).
- Little RJ, D’Agostino R, Cohen ML, et al (2012). The prevention and treatment of missing data in clinical trials. *New England Journal of Medicine* 367(14): 1355-60.

- Mackie JL (1965). Causes and conditions. *American Philosophical Quarterly* 2:245-264.
- McClellan M, McNeil BJ, Newhouse JP (2004). Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. *JAMA* 272(11):859-866.
- Manski CF (1990). Nonparametric bounds on treatment effects. *American Economic Review* 80(2):319-323.
- Marini MM, Olsen AR, Rubin DB (1980). Maximum-likelihood estimation in panel studies with missing data. *Sociological Methodology* 11:314-357.
- Martens E, Pestman W, de Boer A, Belitser S, Klungel OH (2006). Instrumental variables: applications and limitations. *Epidemiology* 17(4):260-267.
- McCullagh P, Nelder JA (1989). *Generalized Linear Models*, 2nd ed. London: Chapman & Hall.
- McCulloch CE, Searle SE, Neuhaus JM (2008). *Generalized, Linear, and Mixed Models*, 2nd ed. New York, NY: Wiley.
- McGrath S, Young JG, Hernán MA (2022). Revisiting the g-null paradox. *Epidemiology* 33(1):114-120.
- Meyer BD (1995). Natural and quasi-experiments in economics. *Journal of Business & Economic Statistics* 13(2):151-161.
- Miao W, Geng Z, Tchetgen Tchetgen EJ (2018). Identifying causal effects With proxy variables of an unmeasured confounder. *Biometrika* 105(4):987-993.
- Miettinen OS (1972). Standardization of risk ratios. *American Journal of Epidemiology* 96:383-388.
- Miettinen OS (1982). Causal and preventive interdependence: elementary principles. *Scandinavian Journal of Work, Environment & Health* 8:159-168.
- Miettinen OS, Cook EF (1981). Confounding: Essence and detection. *American Journal of Epidemiology* 1981; 114:593-603.
- Molina J, Rotnitzky A, Sued M, Robins JM (2017) Multiple robustness in factorized likelihood models. *Biometrika* 104(3):561-581.
- Neyman J (1923). On the Application of Probability Theory to Agricultural Experiments: Essay on Principles, Section 9. Translated in *Statistical Science* 1990; 5:465-480.
- Ogburn EL, VanderWeele TJ (2012). On the nondifferential misclassification of a binary confounder. *Epidemiology*; 23(3):433-439.
- Page J (2005). Doubly Robust Estimation: Structural Nested Cumulative Failure Time Models, Sc.D. dissertation, Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, MA.

- Palmer TM, Sterne JAC, Harbord RM, Lawlor DA, Sheehan NA, Meng S, Graneli R, Davey Smith G, Didelez V (2011). Instrumental variable estimation of causal risk ratios and causal odds ratios in Mendelian randomization analyses. *American Journal of Epidemiology* 173(12): 1392-1403.
- Pearl J (1988). *Probabilistic Reasoning in Intelligent Systems*. San Mateo, CA: Morgan Kaufmann.
- Pearl J (1995). Causal diagrams for empirical research. *Biometrika*; 82:669-710.
- Pearl J (2001). Direct and Indirect Effects. In: Breese J, Koller D, eds. *Proceedings of the Seventeenth Conference on Uncertainty and Artificial Intelligence*. San Francisco, CA: Morgan Kaufmann, pp. 411-420.
- Pearl J (2009). *Causality: Models, Reasoning, and Inference*, 2nd ed. New York: Cambridge University Press.
- Pearl J (2011). Understanding bias amplification. *American Journal of Epidemiology* 174(11):1223-1227.
- Pearl J (2018). Does obesity shorten life? Or is it the soda? On non-manipulable causes. *Journal of Causal Inference* 6(2); pp. 20182001.
- Pearl J (2019). On the interpretation of do(x). *Journal of Causal Inference* 7(1); pp. 20192002.
- Pearl J, Bareinboim (2014). External validity: From do-calculus to transportability Across Populations. *Statistical Science* 29(4): 579-595.
- Pearl J, Robins JM (1995). Probabilistic evaluation of sequential plans from causal models with hidden variables. In: *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*. Montreal, Canada, pp. 444-453.
- Pearson K, Lee A, Bramley-Moore L (1899). VI. Mathematical contributions to the Theory of Evolution.—VI. Genetic (Reproductive) selection: Inheritance of fertility in man, and of fecundity in thoroughbred horses. *Philosophical Transactions of the Royal Society of London, Series A* 192: 258-331.
- Peters J, Janzing D, Schölkopf B (2017). *Elements of Causal Inference: Foundations and Learning Algorithms*. Cambridge, MA: MIT Press.
- Petersen M, Schwab J, Gruber S, Blaser N, Schomaker M, van der Laan M (2014). Targeted maximum likelihood estimation for dynamic and static longitudinal marginal structural working models. *Journal of Causal Inference* 2(2):147-185.
- Picciotto S, Hernán MA, Page J, Young JG, Robins JM (2012). Structural nested cumulative failure time models for estimating the effects of interventions. *Journal of the American Statistical Association* 107(499):886-900.
- Richardson TS, Robins JM (2010). Analysis of the binary instrumental variable model. In: Dechter R, Geffner H, Halpern JY, eds. *Heuristics, Probability and Causality: A Tribute to Judea Pearl*. College Publications, UK.

- Richardson TS, Robins JM (2014). ACE bounds; SEMs with equilibrium conditions. *Statistical Science* 29(3):363-366.
- Richardson TS, Evans RJ, Robins JM (2010). Transparent parametrizations of models for potential outcomes. In: Bernardo JM, Bayarri MJ, Berger JO, Dawid AP, eds. *Bayesian Statistics 9*. Oxford University Press.
- Richardson TS, Robins JM (2013). Single world intervention graphs (SWIGs): A unification of counterfactual and graphical approaches to causality. Working Paper 128, Center for Statistics and the Social Sciences, Seattle, WA.
- Richardson TS, Robins JM, Wang L (2017). On modeling and estimation for the relative risk and risk difference. *Journal of the American Statistical Association* 112(519): 1121-1130.
- Richardson WS, Wilson MC, Nishikawa J, Hayward RSA (1995). The well-built clinical question: a key to evidence-based decisions. *ACP Journal Club* 123(3):A12-13.
- Robins JM (1986). A new approach to causal Inference in mortality studies with sustained exposure periods -Application to control of the healthy worker survivor effect. *Mathematical Modelling* 7:1393-1512 (errata in *Computers and Mathematics with Applications* 1987;14:917-921).
- Robins JM (1987). Addendum to "A new approach to causal inference in mortality studies with sustained exposure periods -Application to control of the healthy worker survivor effect". *Computers and Mathematics with Applications* 14 (9-12):923-945 (errata in *Computers and Mathematics with Applications* 1987;18:477).
- Robins JM (1988). Confidence intervals for causal parameters. *Statistics in Medicine* 7:773-785.
- Robins JM (1989). The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In: Sechrest L, Freeman H, Mulley A, eds. *Health Services Research Methodology: A Focus on AIDS*. U.S. Public Health Service, National Center for Health Services Research, 113-159.
- Robins JM (1993). Analytic methods for estimating HIV treatment and co-factor effects. In: *Methodological Issues of AIDS Mental Health Research*. Ostrow DG, Kessler R, eds. New York: Plenum Publishing, pp. 213-290.
- Robins JM. (1994). Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics* 23:2379-2412.
- Robins JM (1997a). Causal Inference from Complex Longitudinal Data. *Latent Variable Modeling and Applications to Causality*. Berkane M, ed. New York, NY: Springer Verlag, pp. 69-117.
- Robins JM (1997b). Structural nested failure time models. In: Survival Analysis, Andersen PK, Keiding N, Section Editors. *Encyclopedia of Biostatistics*, Armitage P, Colton T (eds). Chichester, UK: John Wiley & Sons, 4372-4389.

- Robins JM (1998a). Marginal structural models. *1997 Proceedings of the Section on Bayesian Statistical Science*. Alexandria, Virginia: American Statistical Association, 1-10.
- Robins JM (1998b). Correction for non-compliance in equivalence trials. *Statistics in Medicine* 17: 269-302.
- Robins JM (1999). Marginal structural models versus structural nested models as tools for causal inference. In: Halloran E, Berry D. *Statistical Models in Epidemiology: The Environment and Clinical Trials*. New York, Springer-Verlag: 95-134.
- Robins JM (2000). Robust estimation in sequentially ignorable missing data and causal inference models. *1999 Proceedings of the Section on Bayesian Statistical Science*. Alexandria, Virginia: American Statistical Association, pp. 6-10.
- Robins JM (2001). Data, design, and background knowledge in etiologic inference. *Epidemiology* 11(3):313-320.
- Robins JM, Greenland S (1986). The role of model selection in causal inference from nonexperimental data. *American Journal of Epidemiology* 123(3): 392-402.
- Robins JM, Greenland S (1989). Estimability and estimation of excess and etiologic fraction. *Statistics in Medicine* 8:845-859.
- Robins JM, Greenland S (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 3(2):143-155.
- Robins JM, Greenland S (2000). Comment on "Causal inference without counterfactuals." *Journal of the American Statistical Association* 95:477-82.
- Robins JM, Hernán MA, Rotnitzky A (2007). Effect modification by time-varying covariates. *American Journal of Epidemiology* 166:994-1002.
- Robins JM, Hernán MA, Siebert U (2004). Effects of multiple interventions. In: *Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors Vol II*. Ezzati M, Lopez AD, Rodgers A, Murray CJL, eds. Geneva: World Health Organization, 2004.
- Robins JM, Morgenstern H (1987). The foundations of confounding in epidemiology. *Computers & Mathematics with Applications* 14(9-12): 869-916.
- Robins JM, Orellana L, Rotnitzky A (2008). Estimation and extrapolation of optimal treatment and testing strategies. *Statistics in Medicine* 27(23):4678-721.
- Robins JM, Richardson TS (2010). Alternative graphical causal models and the identification of direct effects. In: *Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures*. P. Shrout, ed. New York, NY: Oxford University Press.

- Robins JM, Richardson TS, Shpitser I (2022). An interventionist approach to mediation analysis. In: *Probabilistic and Causal Inference: The Works of Judea Pearl*. Geffner H, Dechter R, Halpern JY, eds. New York, NY: Association for Computing Machinery.
- Robins JM, Ritov Y (1997). Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semiparametric models. *Statistics in Medicine* 17:285-319.
- Robins JM, Rotnitzky A (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In: *AIDS Epidemiology – Methodological Issues*. Jewell N, Dietz K, Farewell V, eds. Boston, MA: Birkhäuser, pp. 297-331.
- Robins JM, Rotnitzky A, Scharfstein D (1999). Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In: Halloran ME, Berry D, eds. *Statistical Models in Epidemiology: The Environment and Clinical Trials*. New York, NY: Springer-Verlag, pp. 1-92.
- Robins JM, Rotnitzky A, Vansteelandt S (2007). Discussion on “Principal stratification designs to estimate input data missing due to death.” *Biometrics* 63(3):650-654.
- Robins JM, Rotnitzky A, Zhao LP (1994) Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* 89: 846-866
- Robins JM, Scheines R, Spirtes P, Wasserman L (2003). Uniform consistency in causal inference. *Biometrika* 90(3):491-515.
- Robins JM, Wasserman L (1997). Estimation of effects of sequential treatments by reparameterizing directed acyclic graphs. In: Geiger D, Shenoy P, eds. *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*. San Francisco, CA: Morgan Kaufmann, pp. 409-420.
- Robins JM, Weissman M (2016). Counterfactual causation and streetlamps. What is to be done? *International Journal of Epidemiology* 45(6):1830-1835.
- Rosenbaum PR (1984). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society Series A* 147:656-666.
- Rosenbaum PR (1987). Model-based direct adjustment. *Journal of the American Statistical Association* 82:387-394.
- Rosenbaum PR (2007). Interference between units in randomized experiments. *Journal of the American Statistical Association* 102:191-200.
- Rosenbaum PR (2002). *Observational Studies*, 2nd edition. New York, NY: Springer-Verlag.
- Rosenbaum PR (2005). Sensitivity analysis in observational studies. In: *Encyclopedia of Statistics in Behavioral Sciences*, Everitt BS, Howell DC T (eds). Chichester, UK: John Wiley & Sons, 4:1809-1814.

- Rosenbaum PR, Rubin DB (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* 70:41-55.
- Rothman KJ (1976). Causes. *American Journal of Epidemiology* 104:587-592.
- Rothman KJ, Greenland S, Walker AM (1980). Concepts of interaction. *American Journal of Epidemiology* 112:467-470.
- Rothman KJ, Greenland S, Poole C, Lash TL (2008). Causation and Causal Inference. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology*, 3rd ed. Philadelphia, PA: Lippincott Williams and Wilkins.
- Rubin DB (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 56:688-701.
- Rubin DB (1976). Inference and missing data (with discussion). *Biometrika* 63:581-592.
- Rubin DB (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics* 6:34-58.
- Rubin DB (1980). Discussion of “Randomized analysis of experimental data: the Fisher randomization test” by Basu D. *Journal of the American Statistical Association* 75:591-593.
- Rubin DB (2004). Direct and indirect effects via potential outcomes. *Scandinavian Journal of Statistics* 31(2):161-170.
- Rudolph KE, van der Laan MJ (2017). Robust estimation of encouragement-design intervention effects transported across sites. *Journal of the Royal Statistical Society Series B* 79(5):1509-1525.
- Samuels ML (1981). Matching and design efficiency in epidemiological studies. *Biometrika* 68:577-588.
- Santosa F, Symes WW (1986). Linear inversion of band-limited reflection seismograms. *SIAM Journal on Scientific and Statistical Computing* 7(4): 1307-1330.
- Saracci R (1980). Interaction and synergism. *American Journal of Epidemiology* 112:465-466.
- Sato T, Matsuyama Y (2003). Marginal structural models as a tool for standardization. *Epidemiology* 14:680-686.
- Scharfstein DO, Rotnitzky A, Robins JM (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association* 94(448): 1096-1120.
- Schwartz S, Gatto NM, Campbell UB (2016). Causal identification: a charge of epidemiology in danger of marginalization. *Annals of Epidemiology* 26(10):669-673.
- Shahn Z, Dukes O, Richardson D, Tchetgen Tchetgen EJ, Robins JM (2022). Structural nested mean models under parallel trends assumptions. arXiv preprint arXiv:2204.10291.

- Shpitser I, Pearl J (2006). Identification of joint interventional distributions in recursive semi-Markovian causal models. In: *Proceedings of the 21st National Conference on Artificial Intelligence*. Menlo Park, CA: AAAI Press, pp. 1219-1226.
- Shpitser I, Richardson TS, Robins JM (2022). Multivariate counterfactual systems and causal graphical models. In: *Probabilistic and Causal Inference: The Works of Judea Pearl*. Geffner H, Dechter R, Halpern JY, eds. New York, NY: Association for Computing Machinery.
- Shpitser I, Tchetgen Tchetgen EJ, Andrews R (2021). Modeling interference via symmetric treatment decomposition. arXiv preprint arXiv:1709.01050
- Simpson EH (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society, Series B* 13:238-241.
- Smith GCS, Pell JP (2003). Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. *British Medical Journal* 327:1459-1461.
- Sobel ME (2006). What do randomized studies of housing mobility demonstrate? *Journal of the American Statistical Association* 101: 1398-1407.
- Sofer T, Richardson DB, Colicino E, Schwartz J, Tchetgen Tchetgen EJ (2016). On negative outcome control of unobserved confounding as a generalization of difference-in-differences. *Statistical Science* 31(3): 348-361.
- Sommer A, Zeger SL (1991). On estimating efficacy from clinical trials. *Statistics in Epidemiology* 10:45-52.
- Spirtes P, Glymour C, Scheines R (2000). *Causation, Prediction and Search*, 2nd ed. Cambridge, MA: MIT Press.
- Stalnaker RC (1968). A theory of conditionals. In Rescher N, ed. *Studies in Logical Theory*. Oxford: Blackwell. Reprinted in Jackson F, ed. *Conditionals*. Oxford: Oxford University Press, 1991.
- Stensrud MJ, Young JG, Didelez V, Robins JM, Hernán MA (2020). Separable effects for causal inference in the presence of competing risks. *Journal of the American Statistical Association* 117(537):175-183.
- Stensrud MJ, Hernán MA, Tchetgen Tchetgen EJ, Robins JM, Didelez V, Young JG (2021). A generalized theory of separable effects in competing event settings. *Lifetime Data Analysis* 27(4):588-631.
- Stensrud MJ, Robins JM, Sarvet A, Tchetgen Tchetgen EJ, Young JG (2023). Conditional separable effects. *Journal of the American Statistical Association* (in press).
- Stuart EA (2010). Matching methods for causal inference. *Statistical Science* 25, 1-21.
- Swanson SA, Hernán MA (2013). How to report instrumental variables analyses (suggestions welcome). *Epidemiology* 24(3): 370-374.
- Swanson SA, Hernán MA (2014). Think globally, act globally: An epidemiologist's perspective on instrumental variable estimation. *Statistical Science* 29(3): 371-374.

- Swanson SA, Hernán MA, Miller M, Robins JM, Richardson T (2018). Partial identification of the average treatment effect using instrumental variables: Review of methods for binary instruments, treatments, and outcomes. *Journal of the American Statistical Association* 113(522):933-947.
- Swanson SA, Holme Ø, Løberg M, Kalager M, Bretthauer M, Hoff G, Aas E, Hernán MA (2015a). Bounding the per-protocol effect in randomized trials: an application to colorectal cancer screening. *Trials* 16:541.
- Swanson SA, Miller M, Robins JM, Hernán MA (2015b). Definition and evaluation of the monotonicity condition for preference-based instruments. *Epidemiology* 26:414-420.
- Swanson SA, Miller M, Robins JM, Hernán MA (2015c). Selecting on treatment: a pervasive form of bias in instrumental variable analysis. *American Journal of Epidemiology* 181(3):191-197.
- Tchetgen Tchetgen EJ (2009). A commentary on G. Molenberghs's review of missing data methods. *Drug Information Journal* 43(4):433-435.
- Tchetgen Tchetgen EJ, Rotnitzky A (2011). Double-robust estimation of an exposure-outcome odds ratio adjusting for confounding in cohort and case-control studies. *Statistics in Medicine* 30(4):335-47.
- Tchetgen Tchetgen EJ, Stefan W, Vansteelandt S, Martinussen T, Glymour M (2015). Instrumental variable estimation in a survival context. *Epidemiology* 26(3): 402-410.
- Thistlewaite D, Campbell D (1960). Regression-discontinuity analysis: An alternative to the ex-post facto experiment. *Journal of Educational Psychology* 51:309-317.
- Thompson WA (1977). On the treatment of grouped observations in life studies. *Biometrics* 33:463-470.
- Tian J, Pearl J (2002). A general identification condition for causal effects. In: *Proceedings of the 18th National Conference on Artificial Intelligence*. Menlo Park, CA: AAAI Press/MIT Press, pp. 567-573.
- Tibshirani R (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B* 58 (1): 267-288.
- van der Laan MJ, Petersen ML (2007). Causal effect models for realistic individualized treatment and intention to treat rules. *International Journal of Biostatistics* 3(1): Article 3.
- van der Laan MJ, Rubin D (2006). Targeted maximum likelihood learning. *International Journal of Biostatistics* 2(1): Article 11.
- van der Laan MJ, Gruber S (2012). Targeted minimum loss based estimation of causal effects of multiple time point interventions. *International Journal of Biostatistics* 8(1): Article 9.
- VanderWeele TJ (2009a). Concerning the consistency assumption in causal inference. *Epidemiology* 20:880-883.
- VanderWeele TJ (2009b). On the distinction between interaction and effect modification. *Epidemiology* 20:863-871.

- VanderWeele TJ (2010a). Empirical tests for compositional epistasis. *Nature Reviews Genetics* 11:166.
- VanderWeele TJ (2010b). Sufficient cause interactions for categorical and ordinal exposures with three levels. *Biometrika* 97:647-659.
- VanderWeele TJ (2015). *Explanation in Causal Inference. Methods for Mediation and Interaction*. New York, NY: Oxford University Press.
- VanderWeele TJ, Hernán MA (2006). From counterfactuals to sufficient component causes and vice versa. *European Journal of Epidemiology* 21:855-858.
- VanderWeele TJ, Hernán MA (2009). Results on differential and dependent measurement error of the exposure and the outcome using signed directed acyclic graphs. *American Journal of Epidemiology* 175(12):1303-1310.
- VanderWeele TJ, Hernán MA. Causal inference under multiple versions of treatment. *Journal of Causal Inference* 2013; 1(1):1-20.
- VanderWeele TJ, Hernán MA, Robins JM (2008). Causal directed acyclic graphs and the direction of unmeasured confounding bias. *Epidemiology* 19:720-728.
- VanderWeele TJ, Arah OA (2011). Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology* 22:42-52.
- VanderWeele TJ, Robins JM (2007a). The identification of synergism in the sufficient-component-cause framework. *Epidemiology* 18:329-339.
- VanderWeele TJ, Robins JM (2007b). Four types of effect modification. A classification based on directed acyclic graphs. *Epidemiology* 18; 561-568.
- VanderWeele TJ, Robins JM (2007c). Directed acyclic graphs, sufficient causes, and the properties of conditioning on a common effect. *American Journal of Epidemiology* 166; 1096-1104.
- VanderWeele TJ, Robins JM (2008). Empirical and counterfactual conditions for sufficient cause interactions. *Biometrika* 95:49-61.
- VanderWeele TJ, Robins JM (2012). Stochastic counterfactuals and stochastic sufficient causes. *Statistica Sinica* 22:379-392.
- VanderWeele TJ, Shpitser I (2013). On the definition of a confounder. *Annals of Statistics* 41(1): 196-220.
- VanderWeele TJ, Tchetgen Tchetgen EJ, Cornelis M, Kraft P (2014). Methodological challenges in Mendelian randomization. *Epidemiology* 25(3):427-435.
- Vansteelandt S, Goetghebeur E (2003). Causal inference with generalized structural mean models. *Journal of the Royal Statistical Society - Series B* 65:817-835.
- Wald A (1940). The fitting of straight lines if both variables are subject to error. *Annals of Mathematical Statistics* 11:284-300.

- Walker AM (1991). *Observation and Inference: An Introduction to the Methods of Epidemiology*. Newton Lower Falls, MA: Epidemiology Resources, Inc.
- Wang L, Meng X, Richardson TS, Robins JM (2022). Coherent modeling of longitudinal causal effects on binary outcomes. *Biometrics* (in press).
- Wang L, Tchetgen Tchetgen E (2018). Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables. *Journal of the Royal Statistical Society Series B* 80 (Part 3):531-550.
- Wang Y, Zubizarreta JR (2020). Minimal dispersion approximately balancing weights: Asymptotic properties and practical considerations. *Biometrika* 107, 93-105.
- Wasserman L (2004). *All of Statistics: A Concise Course in Statistical Inference*. New York, NY: Springer.
- Weinberg CR, Umbach DM, Greenland S (1994). When will nondifferential misclassification of an exposure preserve the direction of a trend? *American Journal of Epidemiology* 140:565-571.
- Wen L, Young JG, Robins JM, Hernán MA (2021). Parametric g-formula implementations for causal survival analyses. *Biometrics* 77(2):740-753.
- Wen L, Hernán MA, Robins JM (2022). Multiply robust estimators of causal effects for survival outcomes. *Scandinavian Journal of Statistics* 49:1304-1328.
- Wen L, Sarvet AL, Stensrud MJ (2023). Causal effects of intervening variables in settings with unmeasured confounding. arXiv preprint arXiv:2305.00349.
- Westreich D, Edwards JK, Lesko CR, Stuart E, Cole SR (2017). Transportability of trial results using inverse odds of sampling weights. *American Journal of Epidemiology* 186(8):1010-1014.
- Wold H (1954). Causality and econometrics *Econometrica* 22(2):162-177.
- Wooldridge JM (2010). *Econometric Analysis of Cross section and Panel Data*. Cambridge, MA: MIT Press, 2nd edition.
- Young JG, Cain LE, Robins JM, O'Reilly E, Hernán MA (2011). Comparative effectiveness of dynamic treatment regimes: an application of the parametric g-formula. *Statistics in Biosciences* 3:119-143.
- Young JG, Hernán MA, Robins JM (2014). Identification, estimation and approximation of risk under interventions that depend on the natural value of treatment using observational data. *Epidemiologic Methods* 3(1):1-19.
- Young JG, Stensrud MJ, Tchetgen Tchetgen EJ, Hernán MA (2019). A causal framework for classical statistical estimands in failure time settings with competing events. *Statistics in Medicine* 39(8):1199-1236.
- Yule GU (1903). Notes on the theory of association of attributes in statistics. *Biometrika* 2:121-134.

Index

- accelerated failure time model, 226
- administrative end of follow-up, 221
- ancillarity, 140
- antagonism, 66
- as-treated analysis, 302
- association, 10
 - measures, 11
- attributable fraction, 40, 68
- backdoor criterion, 87
 - generalized, 255
- backdoor path, 85
- balancing score, 196
- Bayesian interval, 133
- bias, 81, 134, 135
 - amplification, 240
 - collider, 117
 - confounding, *see* confounding
 - healthy worker, 86, 105
 - information, *see* measurement
 - bias
 - M-bias, *see* M-bias
 - measurement, *see*
 - measurement bias
 - missing data, *see* missing data
 - nonresponse, 105
 - of traditional methods, 263
 - selection, *see* selection bias
 - self-selection, 106
 - systematic, *see* systematic
 - bias
 - under the alternative, 82
 - under the null, 82
 - volunteer, 106
- blip function, 287
- bootstrapping, 174
- bounds for causal effect, 206
 - natural, 206
 - sharp, 206
- causal DAG, 72, 125
- feedback cycles, 262
- for time-varying treatments, 251
- causal diagrams, 71
 - augmented, 116
 - DAG, *see* causal DAG
 - signed, 90
 - twin, 95
- causal discovery, 81, 144
- causal effect, *see* effect
- censoring, 109, 166, 226
 - administrative, 221
 - artificial, 234
 - as a time-varying treatment, 291
 - informative, 105
 - loss to follow-up, 105, 222
- channeling, *see* confounding
- cloning, 310
- collapsibility, 51, 56
- collider, 75, 91, 237
 - mismeasured, 123
- compatibility interval, 136, *see*
 - confidence interval
- competing event, 114, 222
- compliance types, *see* principal strata
- compositional epistasis, 63
- conditionality principle, 138, 139
- confidence interval, 132
 - anticonservative, 133
 - asymptotic, 133
 - calibrated, 133
 - conservative, 133
 - exact, 133
 - frequentist, 133
 - honest, 134
 - large-sample, 133
 - of doubly robust machine learning estimators, 243
 - small-sample, 133

- valid, 133
- Wald, 132
- confounder, 90
 - mismeasured, 122
 - on causal pathway, 266
 - surrogate, 94, 129
 - time-varying, 259
 - traditional definition, 93, 95
- confounding, 12, 82, 85
 - by indication, 86
 - strength and direction, 90
 - structure, 85
 - time-varying, 260
 - unmeasured, 88
- confounding adjustment, 96
 - sufficient set, 88, 96
- consistency, 4, 33
 - for censoring, 114
 - in causal diagrams, 80
 - sequential, 254, 255
- continuity principle, 140
- conventional methods, 97
- counterfactual outcomes, 4
 - deterministic, 9
 - nondeterministic, 10
- counterfactual response type, 60
- Cox proportional hazards model, 226
- cross-fitting, 242
- cross-validation, 236, 238
- crossover experiment, 16, *see*
 - randomized experiment
- cumulative incidence, *see* risk
- curse of dimensionality, 142, 159
- d-separation, 78
- difference-in-differences, 100, 297
- direct effect
 - natural, *see* pure direct effect
 - controlled, 305
 - principal stratum, 306
 - pure, 306
 - separable, 321
- directed acyclic graph, 71, *see*
 - causal DAG
- dose-response curve, 149, 164
- doubly robust estimator, 175
 - augmented inverse probability
 - weighted, 178, 179
 - based on g-estimation, 192
 - for time-varying treatments, 280, 284
 - plug-in, 176
- doubly robust machine learning
 - estimator, 242
- effect
 - average causal effect, 5
 - conditional, 51
 - direct, *see* direct effect
 - in the compliers, 213, 218
 - in the treated, 46, 48, 161, 170
 - in the untreated, 52
 - individual causal effect, 4, 31
 - measures, 7
 - on additive scale, 8
 - on multiplicative scale, 8
 - population causal effect, 7
- effect modification, 18, 44, 50
 - additive, 44, 165
 - in causal diagrams, 83
 - in marginal structural models, 165
 - in structural nested models, 288
 - multiplicative, 44
 - qualitative, 44
 - with propensity scores, 200
- effect modifier, 44
 - causal, 46, 84
 - surrogate, 46, 84
- effect-measure modification, 44
- estimand, 132
- estimate, 9, 132
- estimator, 9, 132
 - closed form, 190, 290
 - consistent, 9, 132, 135
 - doubly robust, *see* doubly
 - robust estimator
 - exactly unbiased, 135
 - Fisher consistent, 151
 - nonparametric, 150
 - parametric, 149
 - systematically biased, 135
- etiologic fraction, 40
- excess fraction, 40, 68
- exchangeability, 14, 29
 - and confounding, 87
 - conditional, 18, 30
 - expressed parametrically, 182
 - for censoring, 113
 - full, 15
 - in causal diagrams, 72
 - marginal, 18
 - mean, 15
 - partial, 48

- sequential, *see* sequential
 - exchangeability
- exclusion restriction, 204, 215, 300
- exogeneity, 14, 105
- experimental treatment
 - assumption, *see*
 - positivity
- exposure, *see* treatment
- faithfulness, 79
- finest causally interpreted
 - structural tree graph, 73
 - FFRCISTG, 74, 88
- frailty, 108
- front door criterion, 101
- front door formula, 101, 295, 296
- functional form, 149
- g-computation, *see* g-formula
- g-computation algorithm formula, *see* g-formula
- g-estimation, 187, 192
 - for survival analysis, 231
 - for time-varying treatments, 283
- g-formula, 101
 - as a simulation, 272
 - big, 294
 - for a density, 275
 - for survival analysis, 230
 - for time-varying treatments, 271
 - front door formula, 295
 - general expression, 275
 - ICE, 285
 - mediation formula, 320
 - parametric, 175, 274
 - plug-in, 175, 274
 - representations, 285
- g-methods, 97, 271
- g-null paradox, 280
- g-null test, 264
- grace period, 311
- Hajek estimator, 160
- hazard, 223
 - discrete time, 223
- hazard ratio, 223
 - built-in selection bias, 225
 - via a logistic model, 227
 - via a proportional hazards model, 225
- heterogeneity of treatment effects, *see* effect modification
- hidden variable, 126
- homogeneity in IV estimation, 208
- homoscedasticity, 149
- Horvitz-Thompson estimator, 160
- identifiability, 27, 29
 - conditions, 28
 - nonparametric, 29
- identification, 131
 - partial, 206
- ignorability, 28, 105
- independence, 11
 - conditional, 76
 - cross-world, 88, 319
 - mean independence, 11
- instrument, 203
 - bias amplification, 240
 - candidate, 205
 - causal, 204
 - surrogate, 204
 - weak, 205, 215
- instrumental conditions, 203, 204, 214
 - falsification tests for, 205
- instrumental variable, *see* instrument
- instrumental variable estimation, 206, 314
 - additive structural mean models, 209
 - multiplicative structural mean models, 210
 - usual estimand, 207
 - Wald estimator, 207
- intention-to-treat analysis, 300
 - modified, 301
 - pseudo-, 301
- intention-to-treat effect, 300, 304
 - observational analog, 308
- interaction, 57
 - additive, 57
 - biologic, *see* sufficient cause
 - interaction
 - in causal diagrams, 84
 - multiplicative, 59
 - subadditive, 59
 - submultiplicative, 59
 - sufficient cause, 65
 - superadditive, 59
 - supermultiplicative, 59
- interference, 5, 50
- intervention, *see* treatment
- joint, 57

- sufficiently well-defined, 35, 80, 114, 125, 312
- inverse probability weighting
 - augmented, 179
- inverse probability weighting, 22
 - augmented, 178
 - for censoring, 167, 292
 - for survival analysis, 228
 - for time-varying treatments, 276
 - nonstabilized, 158, 276, 277
 - stabilized, 161, 167, 168, 276, 277
 - vs. standardization, 173
 - with models, 158
- iterated conditional expectation, *see* g-formula ICE
- Kaplan-Meier curve, 224, 226
- karma, *see* treatment
- kernel function, 155
- lasso regression, 236
- least squares
 - ordinary, 149, 279
 - two-stage, 207
 - weighted, 159, 279
- link function, 155
- linkage disequilibrium, 86
- local average treatment effect, *see* effect in the compliers
- M-bias, 91, 105, 239, 274
- machine learning algorithms, 241
- Mantel-Haenszel method, 53
- marginal structural model, 163
 - faux, 166, 193
 - for the mean, 163
 - for time-varying treatments, 278
 - logistic, 165
 - semiparametric, 184
 - vs. structural nested model, 184, 288
- matching, 51, 79
 - with propensity scores, 198
- measurement bias, 82
 - strength and direction, 122
- measurement error, 119
 - independent, 120, 121
 - nondifferential, 120, 121
- mediation formula, 318
 - and the g-formula, 320
- mediator, 76, 239, 305
- Mendelian randomization, 205
- misclassification, 120
- missing data, 18, 105, 166
- model, 149
 - for the mean outcome, 171
 - generalized additive, 155
 - linear, 149
 - marginal structural, *see* marginal structural model
 - misspecification, 150
 - multiplicative survival, 116
 - parametric, 149
 - parsimonious, 151
 - predictive, 200
 - rank-preserving, 186
 - saturated, 150
 - semiparametric, 155, 183, 226
 - structural, 199
 - structural nested, *see* structural nested model
- monotonicity, 62, 303
 - in IV estimation, 211, 218
- multiply robust estimator, 283, 284
- negative outcome controls, 100
- nonparametric structural equation model, 73
 - NPSEM-IE, 74, 88
- nuisance parameter, 194
- null hypothesis
 - of no average causal effect, 5, 44
 - sharp causal null, 6, 44, 186
- null preservation, 164, 190, 301
- number needed to treat, 8
- observational study, 27, 72
- omitted variable bias, 105
- on-treatment analysis, 302
- outcome regression, 193
- overadjustment for mediators, 239
- overfitting, 238
- per protocol analysis, 302
- per-protocol effect, 300, 304
 - observational analog, 308
- placebo, 300
- placebo controls, *see* negative outcome controls
- plan, *see* treatment strategy
- point estimate, *see* estimate

- policy, *see* treatment, *see* treatment strategy
- pooling, 53
- population stratification, 86
- possible worlds, 37
- positivity, 25, 32
 - for censoring, 113
 - for g-formula, 272
 - for propensity score, 196, 198
 - in causal diagrams, 78
 - sequential, 254, 255
 - structural, 170
 - violations, 163
- potential outcomes, *see* counterfactual outcomes
- pragmatic trial, 303
- predicted value, 149
- predictive algorithm, 236
 - black-box, 236
 - interpretable, 236
- principal strata, 211
- prognostic score, 196
- propensity score, 97, 159, 195
- proximal causal inference, 100
- pseudo-population, 22, 158, 161
- pure direct effect, 317

- random error, 10
- random variability, 8
- randomization, 13
 - conditional, 17
- randomization-based inference, 135
- randomized experiment, 13
 - conditionally randomized, 17, 30, 72
 - crossover, 31
 - double-blind, 203, 300
 - ideal, 14
 - marginally randomized, 17, 29, 72
 - placebo-controlled, 203, 300
 - sequentially, 253
- rank preservation, 185, 286
 - additive, 186
 - local, 286
- regime, *see* treatment strategy
- regression discontinuity design, 219
- relevance condition in IV
 - estimation, 204, 214
- residual, 149
- restriction, 51

- reverse causation, 86
- ridge regression, 236
- risk, 223
- robust variance, 160, 279

- sample splitting, 242
- sampling variability, 9, 134
- selection bias, 103
 - in case-control studies, 106
 - in hazard ratios, 108
 - strength and direction, 117
 - structure, 103
 - under the alternative, 82, 238
 - under the null, 82, 103, 237
- selection on observables, 105
- selection on unobservables, 107
- selection without bias, 115
- sensitivity analysis, 177, 245
 - for model misspecification, 154
 - for selection bias, 158
 - for unmeasured confounding, 90
 - in instrumental variable estimation, 217
 - using g-estimation, 189
- sequential emulation, 310
- sequential exchangeability, 254, 256
 - full, 256
 - static, 257
 - unconditional, 254
- shrinkage methods, 237
- Simpson's paradox, 109
- single-world intervention graphs, *see* SWIG
- smoothing, 151, 225
- stable-unit-treatment-value
 - assumption, *see* SUTVA
- standard error, 132
- standardization, 20
 - vs. inverse probability weighting, 173
 - of the mean outcome, 172
 - with models, 169
 - with propensity scores, 197
- stratification, 18
 - as a form of adjustment, 49
 - bias for time-varying treatments, 265
 - to identify effect modification, 45
 - with propensity scores, 197

- structural nested model, 183
 - accelerated failure time model, 231
 - cumulative failure time model, 231, 232
 - cumulative survival time model, 231, 232
 - logistic, 185
 - mean model, 183, 284
 - multiplicative, 185
 - saturated, 287
 - vs. marginal structural model, 184, 288
 - with two or more parameters, 190
- sufficient cause, 62
- sufficient-component cause, 63
- super-population, 9, 132, 137
- survival analysis, 221
 - models, 226
 - with time-varying treatments, 293
- SUTVA, 5, 6
- SWIG, 95
 - arrows from intervention nodes, 258
 - for time-varying treatments, 255
- synergism, 66
- systematic bias, 81, 135, 136
- target experiment, *see* target trial
- target trial, 39
 - emulation, 307
 - protocol components, 307
 - with sustained strategies, 303
- targeted maximum likelihood estimation, *see* TMLE
- targeted minimum loss-based estimation, *see* TMLE
- time zero, 309
- TMLE, 179, 281, 283
- total indirect effect, 318
- transportability, 48, 50
- treatment, 3
 - history, 249, 274
 - mismeasured, 269
 - multiple versions, 6, 34, 50
 - natural value of, 96
 - strategy, *see* treatment strategy
 - time-varying, 249
- treatment strategy, 250
 - deterministic, 251
 - dynamic, 250, 252, 278
 - optimal, 251
 - random, 251
 - static, 251
- treatment variation irrelevance, 6
- treatment-confounder feedback, 261
- variable selection, 235
 - in regression models, 237