Chapter 8
SELECTION BIAS

Suppose an investigator conducted a randomized experiment to answer the causal question "does one's looking up to the sky make other pedestrians look up too?" She found a strong association between her looking up and other pedestrians' looking up. Does this association reflect a causal effect? Well, by definition of randomized experiment, confounding bias is not expected in this study. However, there was another potential problem: The analysis included only those pedestrians that, after having been part of the experiment, gave consent for their data to be used. Shy pedestrians (those less likely to look up anyway) and pedestrians in front of whom the investigator looked up (who felt tricked) were less likely to participate. Thus participating individuals in front of whom the investigator looked up (a reason to decline participation) are less likely to be shy (an additional reason to decline participation) and therefore more likely to look up. That is, the process of selection of individuals into the analysis guarantees that one's looking up is associated with other pedestrians' looking up, regardless of whether one's looking up actually makes others look up.

An association created as a result of the process by which individuals are selected into the analysis is referred to as selection bias. Unlike confounding, this type of bias is not due to the presence of common causes of treatment and outcome, and can arise in both randomized experiments and observational studies. Like confounding, selection bias is just a form of lack of exchangeability between the treated and the untreated. This chapter provides a definition of selection bias and reviews the methods to adjust for it.

8.1 The structure of selection bias

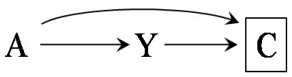


Figure 8.1

Pearl (1995) and Spirtes et al (2000) used causal diagrams to describe the structure of bias resulting from selection of individuals.

The term "selection bias" encompasses various biases that arise from the procedure by which individuals are selected into the analysis. Here we focus on bias that would arise even if the treatment had a null effect on the outcome, i.e., *selection bias under the null* (as described in Section 6.5). The structure of selection bias can be represented by using causal diagrams like the one in Figure 8.1, which depicts dichotomous treatment A, outcome Y, and their common effect C. Suppose Figure 8.1 represents a study to estimate the effect of folic acid supplements A given to pregnant women shortly after conception on the fetus's risk of developing a cardiac malformation Y (1: yes, 0: no) during the first two months of pregnancy. The variable C represents death before birth. A cardiac malformation increases mortality (arrow from Y to C), and folic acid supplementation decreases mortality by reducing the risk of malformations other than cardiac ones (arrow from A to C). The study was restricted to fetuses who survived until birth. That is, the study was conditioned on no death $C = 0$ and hence the box around the node C.

The diagram in Figure 8.1 shows two sources of association between treatment and outcome: 1) the open path $A \rightarrow Y$ that represents the causal effect of A on Y, and 2) the open path $A \rightarrow C \leftarrow Y$ that links A and Y through their (conditioned on) common effect C. An analysis conditioned on C will generally result in an association between A and Y. We refer to this induced association between the treatment A and the outcome Y as selection bias due to conditioning on C. Because of selection bias, the associational risk ratio

Figure 8.2

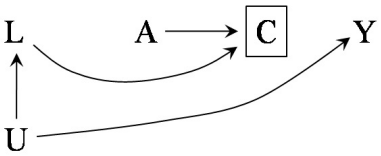


Figure 8.3

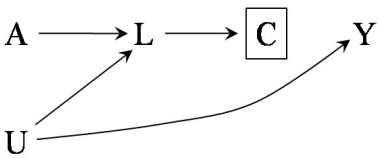


Figure 8.4

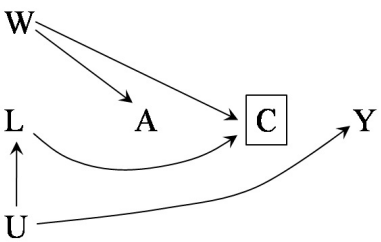


Figure 8.5

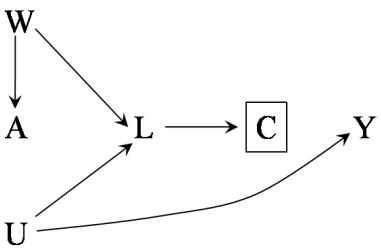


Figure 8.6

$\Pr[Y = 1|A = 1, C = 0]/\Pr[Y = 1|A = 0, C = 0]$ does not equal the causal risk ratio $\Pr\left[Y^{a=1} = 1\right]/\Pr\left[Y^{a=0} = 1\right]$; association is not causation. If the analysis were not conditioned on the common effect (collider) C, then the only open path between treatment and outcome would be $A \rightarrow Y$, and thus the entire association between A and Y would be due to the causal effect of A on Y. That is, the associational risk ratio $\Pr[Y = 1|A = 1]/\Pr[Y = 1|A = 0]$ would equal the causal risk ratio $\Pr\left[Y^{a=1} = 1\right]/\Pr\left[Y^{a=0} = 1\right]$; association would be causation.

The causal diagram in Figure 8.2 shows another example of selection bias. This diagram includes all variables in Figure 8.1 plus a node S representing parental grief (1: yes, 0: no), which is affected by vital status at birth. Suppose the study was restricted to nongrieving parents $S = 0$ because the others were unwilling to participate. As discussed in Chapter 6, conditioning on a variable S affected by the collider C also opens the path $A \rightarrow C \leftarrow Y$.

Both Figures 8.1 and 8.2 depict examples of selection bias in which the bias arises because of conditioning on a common effect of treatment and outcome: C in Figure 8.1 and S in Figure 8.2. This bias arises regardless of whether there is an arrow from A to Y, i.e., it is selection bias under the null. Remember that causal structures that result in bias under the null also cause bias when the treatment has a non-null effect. Both confounding due to common causes of treatment and outcome (see previous chapter) and selection bias due to conditioning on common effects of treatment and outcome are examples of bias under the null. However, selection bias under the null can be defined more generally as illustrated by Figures 8.3 to 8.6.

Consider the causal diagram in Figure 8.3, which represents a follow-up study of individuals with HIV infection to estimate the effect of certain antiretroviral treatment A on the 3-year risk of death Y (to reduce clutter, there is no arrow from A to Y). The unmeasured variable U represents high level of immunosuppression (1: yes, 0: no). Individuals with $U = 1$ have a greater risk of death. Individuals who drop out from the study or are otherwise lost to follow-up are censored ($C = 1$). Individuals with $U = 1$ are more likely to be censored because the severity of their disease prevents them from participating in the study. The effect of U on censoring C is mediated by the presence of symptoms (fever, weight loss, diarrhea, and so on), CD4 count, and viral load in plasma, all included in L, which could or could not be measured. (The role of L, when measured, in data analysis is discussed in Section 8.5; in this section, we take L to be unmeasured.) Individuals receiving treatment are at a greater risk of experiencing side effects, which could lead them to dropout, as represented by the arrow from A to C. The square around C indicates that the analysis is restricted to individuals who remained uncensored ($C = 0$) because those are the only ones in which Y can be assessed.

According to the rules of d-separation, conditioning on the collider C opens the path $A \rightarrow C \leftarrow L \leftarrow U \rightarrow Y$ and thus association flows from treatment A to outcome Y, i.e., the associational risk ratio is not equal to 1 even though the causal risk ratio is equal to 1. Figure 8.3 can be viewed as a simple transformation of Figure 8.1: the association between Y and C resulting from a direct effect of Y on C in Figure 8.1 is now the result of U, a common cause of Y and C. Some intuition for this bias: If a treated individual with treatment-induced side effects (and thereby at a greater risk of dropping out) did in fact not drop out ($C = 0$), then it is generally less likely that a second independent cause of dropping out (e.g., $U = 1$) was present. Therefore, an inverse association between A and U would be expected in those who did not drop out ($C = 0$). Because U is positively associated with the outcome

Y, restricting the analysis to individuals who did not drop out of this study induces an inverse association between A and Y.

The bias in Figure 8.3 is an example of selection bias that results from conditioning on censoring C, which is a common effect of treatment A and of *a cause U of* the outcome Y, rather than a common effect of treatment and outcome. We now present three additional causal diagrams that could lead to selection bias by differential loss to follow up. In Figure 8.4 prior treatment A has a direct effect on symptoms L. Restricting the study to the uncensored individuals again implies conditioning on the common effect C of A and U, thereby introducing an association between treatment and outcome. Figures 8.5 and 8.6 are variations of Figures 8.3 and 8.4, respectively, in which there is a common cause W of A and another measured variable. W indicates unmeasured lifestyle/personality/educational variables that determine both treatment (arrow from W to A) and either attitudes toward attending study visits (arrow from W to C in Figure 8.5) or threshold for reporting symptoms (arrow from W to L in Figure 8.6).

Figures 8.5 and 8.6 show examples of M-bias

We have described some different causal structures, depicted in Figures 8.1-8.6, that may lead to selection bias under the null. In all these cases, the bias is the result of selection on a common effect of two other variables in the diagram, i.e., a collider. We will use the term selection bias to refer to all biases that arise from conditioning on a common effect of two variables, one of which is either the treatment or a cause of treatment, and the other is either the outcome or a cause of the outcome. We now describe some examples of selection bias that share this structure.

More generally, selection bias can be defined as the bias resulting from conditioning on the common effect of two variables, one of which is either the treatment or associated with the treatment, and the other is either the outcome or associated with the outcome (Hernán, Hernández-Díaz, and Robins 2004).

8.2 Examples of selection bias

Consider the following examples of bias due to the mechanism by which individuals are selected into the analysis:

- *Differential loss to follow-up*: This is precisely the bias described in the previous section and summarized in Figures 8.3-8.6. It is also referred to as bias due to *informative censoring*.

- *Missing data bias,nonresponse bias*: The variable C in Figures 8.3-8.6 can represent missing data on the outcome for any reason, not just as a result of loss to follow up. For example, individuals could have missing data because they are reluctant to provide information or because they miss study visits. Regardless of the reasons why data on Y are missing, restricting the analysis to individuals with complete data ($C = 0$) may result in bias.

- *Healthy worker bias*: Figures 8.3–8.6 can also describe a bias that could arise when estimating the effect of an occupational exposure A (e.g., a chemical) on mortality Y in a cohort of factory workers. The underlying unmeasured true health status U is a determinant of both death Y and of being at work C (1: no, 0: yes). The study is restricted to individuals who are at work ($C = 0$) at the time of outcome ascertainment. (L could be the result of blood tests and a physical examination.) Being exposed to the chemical reduces the probability of being at work in the near future, either directly (e.g., exposure can cause disabling asthma), like in Figures 8.3 and 8.4, or through a common cause W (e.g., certain

The distinction between the two structures leading to lack of exchangeability is not universally made across disciplines. Conditional exchangeability is often referred as "weak ignorability" or "ignorable treatment assignment" in statistics (Rosenbaum and Rubin 1983, Rosenbaum 2002), "selection on observables" in the social sciences (Barnow, Cain, and Goldberger, 1980), and "no ommitted variable bias" or "exogeneity" in econometrics (Imbens, 2004).

Fine Point 8.1

Selection bias in case-control studies. Figure 8.1 can be used to represent selection bias in a case-control study. Suppose a certain investigator wants to estimate the effect of postmenopausal estrogen treatment A on coronary heart disease Y. The variable C indicates whether a woman in the study population (the underlying cohort, in epidemiologic terms) is selected for the case-control study (1: no, 0: yes). The arrow from disease status Y to selection C indicates that cases in the population are more likely to be selected than noncases, which is the defining feature of a case-control study. In this particular case-control study, the investigator decided to select controls ($Y = 0$) preferentially among women with a hip fracture. Because treatment A has a protective causal effect on hip fracture, the selection of controls with hip fracture implies that treatment A now has a causal effect on selection C. This effect of A on C is represented by the arrow $A \rightarrow C$. One could add an intermediate node F (representing hip fracture) between A and C, but that is unnecessary for our purposes.

In a case-control study, the association measure (the treatment-outcome odds ratio) is by definition conditional on having been selected into the study ($C = 0$). If individuals with hip fracture are oversampled as controls, then the probability of control selection depends on a consequence of treatment A (as represented by the path from A to C) and "inappropriate control selection" bias will occur. Again, this bias arises because we are conditioning on a common effect C of treatment and outcome. A heuristic explanation of this bias follows. Among individuals selected for the study ($C = 0$), controls are more likely than cases to have had a hip fracture. Therefore, because estrogens lower the incidence of hip fractures, a control is less likely to be on estrogens than a case, and hence the A-Y odds ratio conditional on $C = 0$ would be greater than the causal odds ratio in the population. Other forms of selection bias in case-control studies, including some biases described by Berkson (1946) and incidence-prevalence bias, can also be represented by Figure 8.1 or modifications of it, as discussed by Hernán, Hernández-Díaz, and Robins (2004).

exposed jobs are eliminated for economic reasons and the workers laid off) like in Figures 8.5 and 8.6.

Berkson (1955) described the structure of bias due to self-selection.

- *Self-selection bias*, *volunteer bias*: Figures 8.3-8.6 can also represent a study in which C is agreement to participate (1: no, 0: yes), A is cigarette smoking, Y is coronary heart disease, U is family history of heart disease, and W is healthy lifestyle. (L is any mediator between U and C such as heart disease awareness.) Under any of these structures, selection bias may be present if the study is restricted to those who volunteered or elected to participate ($C = 0$).

Robins, Hernán, and Rotnitzky (2007) used causal diagrams to describe the structure of bias due to the effect of pre-study treatments on selection into the study.

- *Selection affected by treatment received before study entry*: Suppose that C in Figures 8.3-8.6 represents selection into the study (1: no, 0: yes) and that treatment A took place before the study started. If treatment affects the probability of being selected into the study, then selection bias is expected. The case of selection bias arising from the effect of treatment on selection into the study can be viewed as a generalization of self-selection bias. This bias may be present in any study that attempts to estimate the causal effect of a treatment that occurred before the study started or in which treatment includes a pre-study component. For example, selection bias may arise when treatment is measured as the lifetime exposure to certain factor (medical treatment, lifestyle behavior...) in a study that recruited 50 year-old participants. In addition to selection bias, it is also possible that there exists unmeasured confounding for the pre-study component of treatment if confounders were only measured during the study.

In addition to the biases described here, as well as in Fine Point 8.1 and Technical Point 8.1, causal diagrams have been used to characterize various

other biases that arise from conditioning on a common effect. These examples show that selection bias may occur in *retrospective studies*—those in which data on treatment A are collected *after* the outcome Y occurs—and in *prospective studies*—those in which data on treatment A are collected *before* the outcome Y occurs. Further, these examples show that selection bias may occur both in observational studies and in randomized experiments.

For example, selection bias may be induced by attempts to eliminate bias from ascertainment (Robins 2001), to estimate direct effects (Cole and Hernán 2002), and by conventional adjustment for variables affected by previous treatment (see Part III).

Take Figures 8.3 and 8.4, which could depict either an observational study or an experiment in which treatment A is randomly assigned, because there are no common causes of A and any other variable. Individuals in *both* randomized experiments and observational studies may be lost to follow-up or drop out of the study before their outcome is ascertained. When this happens, the risk $\Pr[Y = 1|A = a]$ cannot be computed because the value of the outcome Y is unknown for the censored individuals ($C = 1$). Therefore only the risk among the uncensored $\Pr[Y = 1|A = a, C = 0]$ can be computed. This restriction of the analysis to the uncensored individuals may induce selection bias because uncensored individuals who remained through the end of the study ($C = 0$) may not be exchangeable with individuals that were lost ($C = 1$).

Hence a key difference between confounding and selection bias: randomization protects against confounding, but not against selection bias when the selection occurs after the randomization. On the other hand, no bias arises in randomized experiments from selection into the study before treatment is assigned. For example, only volunteers who agree to participate are enrolled in randomized clinical trials, but such trials are not affected by volunteer bias because participants are randomly assigned to treatment only after agreeing to participate ($C = 0$). Thus none of Figures 8.3-8.6 can represent volunteer bias in a randomized trial. Figures 8.3 and 8.4 are eliminated because treatment cannot cause agreement to participate C. Figures 8.5 and 8.6 are eliminated because, as a result of the random treatment assignment, there cannot exist a common cause of treatment and any other variable.

8.3 Selection bias and confounding

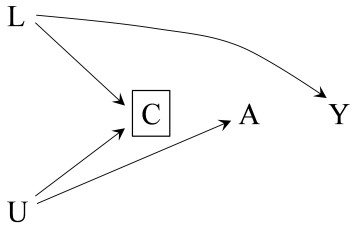


Figure 8.7

For the same reason, social scientists often refer to unmeasured confounding as *selection on unobservables*.

In this and the previous chapter, we describe two reasons why the treated and the untreated may not be exchangeable: 1) the presence of common causes of treatment and outcome, and 2) conditioning on common effects of treatment and outcome (or causes of them). We refer to biases due to the presence of common causes as "confounding" and to those due to conditioning on common effects as "selection bias." This structural definition provides a clear-cut classification of confounding and selection bias, even though it might not coincide perfectly with the traditional terminology of some disciplines. For example, statisticians and econometricians often use the term "selection bias" to refer to both types of biases. Their rationale is that in both cases the bias is due to selection: selection of individuals into the analysis (the structural "selection bias") or selection of individuals into a treatment (the structural "confounding"). Our goal, however, is not to be normative about terminology, but rather to emphasize that, regardless of the particular terms chosen, there are two distinct causal structures that lead to bias.

The end result of both structures is lack of exchangeability between the treated and the untreated—which implies that these two biases occur even under the null. For example, consider a study restricted to firefighters that aims to estimate the causal effect of being physically active A on the risk

Technical Point 8.1

The built-in selection bias of hazard ratios. The causal DAG in Figure 8.8 describes a randomized experiment of the effect of heart transplant A on death at times 1 (Y_1) and 2 (Y_2). The arrow from A to Y_1 represents that transplant decreases the risk of death at time 1. The lack of an arrow from A to Y_2 indicates that A has no direct effect on death at time 2. That is, heart transplant does not influence the survival status at time 2 of any individual who would survive past time 1 when untreated (and thus when treated). U is an unmeasured haplotype that decreases the individual's risk of death at all times. Because of the absence of confounding, the associational risk ratios $aRR_{AY_1} = \frac{\Pr[Y_1=1|A=1]}{\Pr[Y_1=1|A=0]}$ and $aRR_{AY_2} = \frac{\Pr[Y_2=1|A=1]}{\Pr[Y_2=1|A=0]}$ are unbiased measures of the effect of A on death at times 1 and 2, respectively. Even though A has no direct effect on Y_2, aRR_{AY_2} will be less than 1 because it is a measure of the effect of A on total mortality through time 2.

Consider now the time-specific hazard ratio (which, for all practical purposes, is equivalent to the rate ratio). In discrete time, the hazard of death at time 1 is the probability of dying at time 1 and thus the associational hazard ratio is the same as aRR_{AY_1}. However, the hazard at time 2 is the probability of dying at time 2 among those who survived past time 1. Thus, the associational hazard ratio at time 2 is then $aRR_{AY_2|Y_1=0} = \frac{\Pr[Y_2=1|A=1,Y_1=0]}{\Pr[Y_2=1|A=0,Y_1=0]}$. The square around Y_1 in Figure 8.8 indicates this conditioning. Treated survivors of time 1 are less likely than untreated survivors of time 1 to have the protective haplotype U (because treatment can explain their survival) and therefore are more likely to die at time 2. That is, conditional on Y_1, treatment A is associated with a higher mortality at time 2. Thus, the hazard ratio at time 1 is less than 1, whereas the hazard ratio at time 2 is greater than 1, i.e., the hazards have crossed. We conclude that the hazard ratio at time 2 is a biased estimate of the direct effect of treatment on mortality at time 2. The bias is selection bias arising from conditioning on a common effect Y_1 of treatment A and of U, which is a cause of Y_2 that opens the associational path $A \rightarrow Y_1 \leftarrow U \rightarrow Y_2$ between A and Y_2. In the survival analysis literature, an unmeasured cause of death that is marginally unassociated with treatment such as U is often referred to as a *frailty*.

In contrast, the conditional hazard ratio $aRR_{AY_2|Y_1=0,U}$ is 1 within each stratum of U because the path $A \rightarrow Y_1 \leftarrow U \rightarrow Y_2$ is now blocked by conditioning on the non-collider U. Thus, the conditional hazard ratio correctly indicates the absence of a direct effect of A on Y_2. That the unconditional hazard ratio $aRR_{AY_2|Y_1=0}$ differs from the stratum-specific hazard ratios $aRR_{AY_2|Y_1=0,U}$, even though U is independent of A, shows the noncollapsibility of the hazard ratio (Greenland, 1996b). Unfortunately, the unbiased measure $aRR_{AY_2|Y_1=0,U}$ of the direct effect of A on Y_2 cannot be computed because U is unobserved. In the absence of data on U, it is impossible to know whether A has a direct effect on Y_2. That is, the data cannot determine whether the true causal DAG generating the data was that in Figure 8.8 or in Figure 8.9. All of the above applies to both observational studies and randomized experiments.

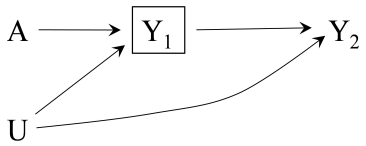


Figure 8.8

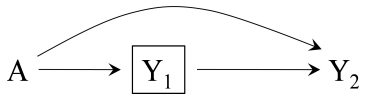


Figure 8.9

of heart disease Y as represented in Figure 8.7. For simplicity, we assume that, unknown to the investigators, A does not cause Y. Parental socioeconomic status L affects the risk of becoming a firefighter C and, through childhood diet, of heart disease Y. Attraction toward activities that involve physical activity (an unmeasured variable U) affects the risk of becoming a firefighter and of being physically active (A). U does not affect Y, and L does not affect A. According to our terminology, there is no confounding because there are no common causes of A and Y. Thus, the associational risk ratio $\Pr[Y=1|A=1]/\Pr[Y=1|A=0]$ is expected to equal the causal risk ratio $\Pr\left[Y^{a=1}=1\right]/\Pr\left[Y^{a=0}=1\right]=1$.

However, in a study restricted to firefighters ($C=0$), the associational and causal risk ratios would differ because conditioning on a common effect C of causes of treatment and outcome induces selection bias resulting in lack of exchangeability of the treated and untreated firefighters. To the study investigators, the distinction between confounding and selection bias is moot because, regardless of nomenclature, they must adjust for L to make the treated and the untreated firefighters comparable. This example demonstrates that a structural classification of bias does not always have consequences for the analysis

of a study. Indeed, for this reason, many epidemiologists use the term "confounder" for any variable L that needs to be adjusted for, regardless of whether the lack of exchangeability is the result of conditioning on a common effect or the result of a common cause of treatment and outcome.

There are, however, advantages of adopting a structural approach to the classification of sources of non-exchangeability. First, the structure of the problem frequently guides the choice of analytical methods to reduce or avoid the bias. For example, in longitudinal studies with time-varying treatments, identifying the structure allows us to detect situations in which adjustment for confounding via stratification would introduce selection bias (see Part III). In those cases, g-methods are a better alternative. Second, even when understanding the structure of bias does not have implications for data analysis (like in the firefighters' study), it could still help study design. For example, investigators running a study restricted to firefighters should make sure that they collect information on joint risk factors for the outcome Y and for the selection variable C (i.e., becoming a firefighter), as described in the first example of confounding in Section 7.1. Third, selection bias resulting from conditioning on pre-treatment variables (e.g., being a firefighter) could explain why certain variables behave as "confounders" in some studies but not others. In our example, parental socioeconomic status L would not necessarily need to be adjusted for in studies not restricted to firefighters. Finally, causal diagrams enhance communication among investigators and may decrease the occurrence of misunderstandings.

The choice of terminology usually has no practical consequences, but disregard for the causal structure may lead to apparent paradoxes. For example, the so-called Simpson's paradox (1951) was the result of ignoring the difference between common causes and common effects. Interestingly, Blyth (1972) failed to grasp the causal structure of the paradox in Simpson's example and misrepresented it as an extreme case of confounding. Because most people read Blyth's paper but not Simpson's paper, the misunderstanding was perpetuated. See Hernán, Clayton, and Keiding (2011) for details.

As an example of the last point, consider the "*healthy worker bias*", which in the previous section we described as a bias that arises from conditioning on the variable C—a common effect of (a cause of) treatment and (a cause of) the outcome. Thus the bias can be represented by the causal diagrams in Figures 8.3-8.6. However, the term "healthy worker bias" is also used to describe the bias that occurs when comparing the risk in certain group of workers with that in a group of individuals from the general population.

This second bias can be depicted by the causal diagram in Figure 7.1 in which L represents health status, A represents membership in the group of workers, and Y represents the outcome of interest. There are arrows from L to A and Y because being healthy affects job type and risk of subsequent outcome, respectively. In this case, the bias is caused by the common cause L and we would refer to it as confounding. The use of causal diagrams to represent the structure of the "healthy worker bias" prevents any confusions that may arise from employing the same term for different sources of non-exchangeability.

All the above considerations ignore the magnitude or direction of selection bias and confounding. However, it is possible that some noncausal paths opened by conditioning on a collider are weak and thus induce little bias. Because selection bias is not an "all or nothing" issue, in practice, it is important to consider the expected direction and magnitude of the bias (see Fine Point 8.2).

8.4 Selection bias and censoring

Suppose an investigator conducted a marginally randomized experiment to estimate the average causal effect of wasabi intake on the one-year risk of death ($Y = 1$). Half of the 60 study participants were randomly assigned to eating meals supplemented with wasabi ($A = 1$) until the end of follow-up or

death, whichever occurred first. The other half were assigned to meals that contained no wasabi ($A = 0$). After 1 year, 17 individuals died in each group. That is, the associational risk ratio $\Pr[Y = 1|A = 1] / \Pr[Y = 1|A = 0]$ was 1. Because of randomization, the causal risk ratio $\Pr\left[Y^{a=1} = 1\right] / \Pr\left[Y^{a=0} = 1\right]$ is also expected to be 1. (If ignoring random variability bothers you, please imagine the study had 60 million patients rather than 60.)

Unfortunately, the investigator could not observe the 17 deaths that occurred in each group because many patients were lost to follow-up, or censored, before the end of the study (i.e., death or one year after treatment assignment). The proportion of censoring ($C = 1$) was higher among patients with heart disease ($L = 1$) at the start of the study and among those assigned to wasabi supplementation ($A = 1$). In fact, only 9 individuals in the wasabi group and 22 individuals in the other group were not lost to follow-up. The investigator observed 4 deaths in the wasabi group and 11 deaths in the other group. That is, the associational risk ratio $\Pr[Y = 1|A = 1, C = 0] / \Pr[Y = 1|A = 0, C = 0]$ was $(4/9)/(11/22) = 0.89$ among the uncensored. The risk ratio of 0.89 in the uncensored differs from the causal risk ratio of 1 in the entire population: There is selection bias due to conditioning on the common effect C.

The causal diagram in Figure 8.3 depicts the relation between the variables L, A, C, and Y in the randomized trial of wasabi. U represents atherosclerosis, an unmeasured variable, that affects both heart disease L and death Y. Figure 8.3 shows that there are no common causes of A and Y, as expected in a marginally randomized experiment, and thus there is no need to adjust for confounding to compute the causal effect of A on Y. On the other hand, Figure 8.3 shows that there is a common cause U of C and Y. The presence of this backdoor path $C \leftarrow L \leftarrow U \rightarrow Y$ implies that, were the investigator interested in estimating the causal effect of censoring C on Y (which is null in Figure 8.3), she would have to adjust for confounding due to the common cause U. The backdoor criterion says that such adjustment is possible because the measured variable L can be used to block the backdoor path $C \leftarrow L \leftarrow U \rightarrow Y$.

The causal contrast we have considered so far is "the risk if everybody had been treated", $\Pr\left[Y^{a=1} = 1\right]$, versus "the risk if everybody had remained untreated", $\Pr\left[Y^{a=0} = 1\right]$, and this causal contrast does not involve C at all. Why then are we talking about confounding for the causal effect of C? It turns out that the causal contrast of interest needs to be modified in the presence of censoring or, in general, of selection. Because selection bias would not exist if everybody had been uncensored $C = 0$, we would like to consider a causal contrast that reflects what would have happened in the absence of censoring.

Let $Y^{a=1,c=0}$ be an individual's counterfactual outcome if he had received treatment $A = 1$ *and* he had remained uncensored $C = 0$. Similarly, let $Y^{a=0,c=0}$ be an individual's counterfactual outcome if he had not received treatment $A = 0$ *and* he had remained uncensored $C = 0$. Our causal contrast of interest is now "the risk if everybody had been treated and had remained uncensored", $\Pr\left[Y^{a=1,c=0} = 1\right]$, versus "the risk if everybody had remained untreated and uncensored", $\Pr\left[Y^{a=0,c=0} = 1\right]$.

For example, we may want to compute the causal risk ratio $\mathrm{E}\left[Y^{a=1,c=0}\right] / \mathrm{E}\left[Y^{a=0,c=0}\right]$ or the causal risk difference $\mathrm{E}\left[Y^{a=1,c=0}\right] - \mathrm{E}\left[Y^{a=0,c=0}\right]$.

Often it is reasonable to assume that censoring does not have a causal effect on the outcome (an exception would be a setting in which being lost to follow-up prevents people from getting additional treatment). Because of the lack of effect of censoring C on the outcome Y, one might imagine that the definition of causal effect could ignore censoring, i.e., that we could omit the superscript $c = 0$. However, omitting the superscript would obscure the fact that considerations about confounding for C become central when computing the causal effect of A on Y in the presence of selection bias. In fact, when

In causal diagrams with no arrow from censoring C to the observed outcome Y, we could replace Y by the counterfactual outcome $Y^{c=0}$ and add arrows $Y^{c=0} \longrightarrow Y$ and $C \longrightarrow Y$.

conceptualizing the causal contrast of interest in terms of $Y^{a,c=0}$, we can think of censoring C as just another treatment. That is, the goal of the analysis is to compute the causal effect of a joint intervention on A and C. To eliminate selection bias for the effect of treatment A, we need to adjust for confounding for the effect of treatment C.

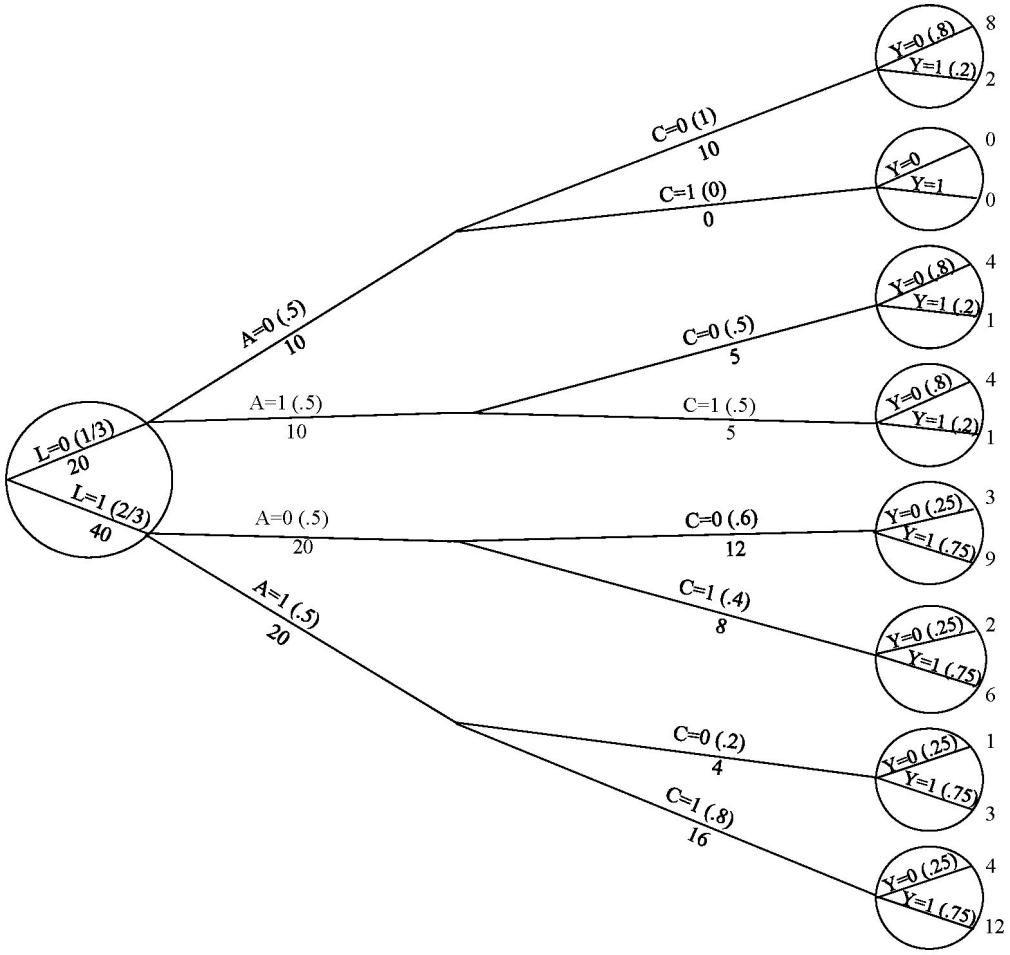
Since censoring C is now viewed as a treatment, it follows that we will need to (i) ensure that the identifiability conditions of exchangeability, positivity, and consistency hold for C as well as for A, and (ii) use analytical methods that are identical to those we would have to use if we wanted to estimate the effect of censoring C. Under these identifiability conditions and using these methods, selection bias can be eliminated via analytic adjustment and, in the absence of measurement error and confounding, the causal effect of treatment A on outcome Y can be identified. The next section explains how to do so.

8.5 How to adjust for selection bias

Though selection bias can sometimes be avoided by an adequate design (see Fine Point 8.1), it is often unavoidable. For example, loss to follow up, self-selection, and, in general, missing data leading to bias can occur no matter how careful the investigator. In those cases, the selection bias needs to be explicitly corrected in the analysis. This correction can sometimes be accomplished by IP weighting (or by standardization), which is based on assigning a weight W^C to each selected individual ($C = 0$) so that she accounts in the analysis not only for herself, but also for those like her, i.e., with the same values of L and A, who were not selected ($C = 1$). The IP weight W^C is the inverse of the probability of her selection $\Pr[C = 0|L, A]$.

We have described IP weights to adjust for confounding, $W^A = 1/f(A|L)$, and selection bias. $W^C = 1/\Pr[C = 0|A, L]$. When both confounding and selection bias exist, the product weight $W^A W^C$ can be used to adjust simultaneously for both biases under assumptions described in Chapter 12 and Part III.

To describe the application of IP weighting for selection bias adjustment consider again the wasabi randomized trial described in the previous section. The tree graph in Figure 8.10 presents the trial data. Of the 60 individuals in the trial, 40 had ($L = 1$) and 20 did not have ($L = 0$) heart disease at the time of randomization. Regardless of their L status, all individuals had a 50/50 chance of being assigned to wasabi supplementation ($A = 1$). Thus 10 individuals in the $L = 0$ group and 20 in the $L = 1$ group received treatment $A = 1$. This lack of effect of L on A is represented by the lack of an arrow from L to A in the causal diagram of Figure 8.3. The probability of remaining uncensored varies across branches in the tree. For example, 50% of the individuals without heart disease that were assigned to wasabi ($L = 0$, $A = 1$), whereas 60% of the individuals with heart disease that were assigned to no wasabi ($L = 1$, $A = 0$), remained uncensored. This effect of A and L on C is represented by arrows from A and L into C in the causal diagram of Figure 8.3. Finally, the tree shows how many people would have died ($Y = 1$) both among the uncensored and the censored individuals. Of course, in real life, investigators would never know how many deaths occurred among the censored individuals. It is precisely the lack of this knowledge which forces investigators to restrict the analysis to the uncensored, opening the door for selection bias. Here we show the deaths in the censored to document that, as depicted in Figure 8.3, treatment A is marginally independent of Y, and censoring C is independent of Y within levels of L. It can also be checked that the risk ratio in the entire population (inaccessible to the investigator) is 1 whereas the risk ratio in the uncensored (accessible to the investigator) is 0.89.



Let us now describe the intuition behind the use of IP weighting to adjust for selection bias. Look at the bottom of the tree in Figure 8.10. There are 20 individuals with heart disease ($L = 1$) who were assigned to wasabi supplementation ($A = 1$). Of these, 4 remained uncensored and 16 were lost to follow-up. That is, the conditional probability of remaining uncensored in this group is 1/5, i.e., $\Pr[C = 0|L = 1, A = 1] = 4/20 = 0.2$. In an IP weighted analysis the 16 censored individuals receive a zero weight (i.e., they do not contribute to the analysis), whereas the 4 uncensored individuals receive a weight of 5, which is the inverse of their probability of being uncensored (1/5). IP weighting replaces the 20 original individuals by 5 copies of each of the 4 uncensored individuals. The same procedure can be repeated for the other branches of the tree, as shown in Figure 8.11, to construct a pseudo-population of the same size as the original study population but in which nobody is lost to follow-up. (We let the reader derive the IP weights for each branch of the tree.) The associational risk ratio in the pseudo-population is 1, the same as the risk ratio $\Pr\left[Y^{a=1,c=0} = 1\right] / \Pr\left[Y^{a=0,c=0} = 1\right]$ that would have been computed in the original population if nobody had been censored.

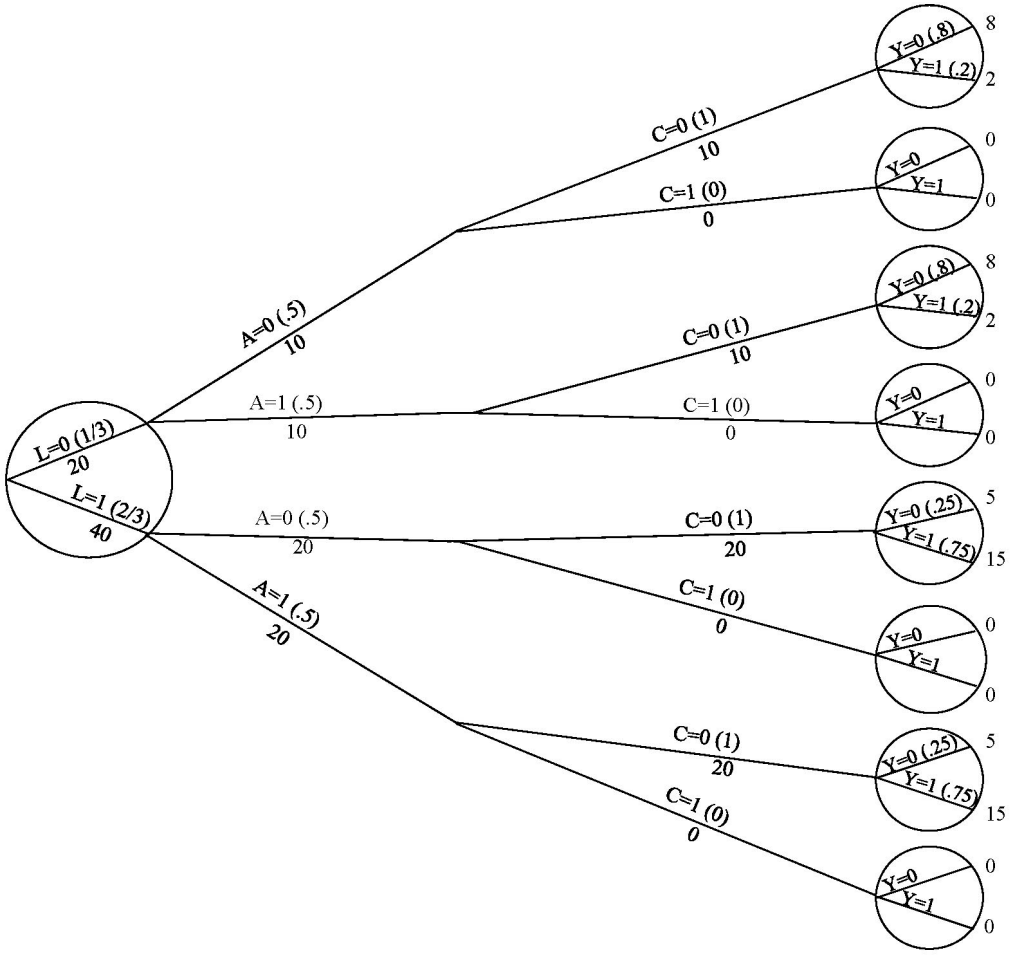


Figure 8.11

The association measure in the pseudo-population equals the effect measure in the original population if the following three identifiability conditions are met.

First, the average outcome in the uncensored individuals must equal the unobserved average outcome in the censored individuals with the same values of A and L. This provision will be satisfied if the probability of selection $\Pr[C = 0|L = 1, A = 1]$ is calculated conditional on treatment A and on all additional factors that independently predict both selection and the outcome, that is, if the variables in A and L are sufficient to block all backdoor paths between C and Y. Unfortunately, one can never be sure that these additional factors were identified and recorded in L, and thus the causal interpretation of the resulting adjustment for selection bias depends on this untestable *exchangeability* assumption.

Second, IP weighting requires that all conditional probabilities of being uncensored given A and the variables in L must be greater than zero. Note this *positivity* condition is required for the probability of being uncensored ($C = 0$) but not for the probability of being censored ($C = 1$) because we are

not interested in inferring what would have happened if study individuals had been censored, and thus there is no point in constructing a pseudo-population in which everybody is censored. For example, the tree in Figure 8.10 shows that $\Pr[C = 1|L = 0, A = 0] = 0$, but this zero does not affect our ability to construct a pseudo-population in which nobody is censored.

The third condition is consistency, including *sufficiently well-defined interventions*. IP weighting is used to create a pseudo-population in which censoring C has been abolished, and in which the effect of the treatment A is the same as in the original population. Thus, the pseudo-population effect measure is equal to the effect measure had nobody been censored. This effect measure may be relatively well defined when censoring is the result of loss to follow up or nonresponse, but not when censoring is defined as the occurrence of a *competing event*. For example, in a study aimed at estimating the effect of certain treatment on the risk of Alzheimer's disease, death from other causes (cancer, heart disease, and so on) is a competing event. Defining death as a form of censoring is problematic: we might not wish to base our effect estimates on a pseudo-population in which all other causes of death have been removed, because it is unclear even conceptually what sort of intervention would produce such a population. Also, no feasible intervention could possibly remove just one cause of death without affecting the others as well.

A *competing event* is an event that prevents the outcome of interest from happening. A typical example of competing event is death because, once an individual dies, no other outcomes can occur.

Finally, one could argue that IP weighting is not necessary to adjust for selection bias in a setting like that described in Figure 8.3. Rather, one might attempt to remove selection bias by stratification (i.e., by estimating the effect measure conditional on the L variables) rather than by IP weighting. Stratification could yield unbiased conditional effect measures within levels of L because conditioning on L is sufficient to block the backdoor path from C to Y. That is, the conditional risk ratio

$$\Pr[Y = 1|A = 1, C = 0, L = l] \,/\, \Pr[Y = 1|A = 0, C = 0, L = l]$$

can be interpreted as the effect of treatment among the uncensored with $L = l$. For the same reason, under the null, stratification would work (i.e., it would provide an unbiased conditional effect measure) if the data can be represented by the causal structure in Figure 8.5. Stratification, however, would not work under the structure depicted in Figures 8.4 and 8.6.

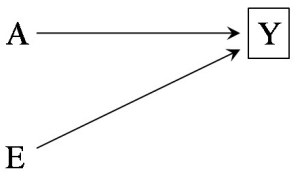


Figure 8.12

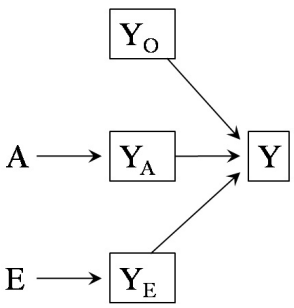


Figure 8.13

Take Figure 8.4. Conditioning on L blocks the backdoor path from C to Y but also opens the path $A \rightarrow L \leftarrow U \rightarrow Y$ from A to Y because L is a collider on that path. Thus, even if the causal effect of A on Y is null, the conditional (on L) risk ratio would be generally different from 1. And similarly for Figure 8.6. In contrast, IP weighting appropriately adjusts for selection bias under Figures 8.3-8.6 because this approach is not based on estimating effect measures conditional on the covariates L, but rather on estimating unconditional effect measures after reweighting the individuals according to their treatment and their values of L.

This is the first time we discuss a situation in which stratification cannot be used to validly compute the causal effect of treatment, even if the three conditions of exchangeability, positivity, and consistency hold. We will discuss other situations with a similar structure in Part III when considering the effect of time-varying treatments.

8.6 Selection without bias

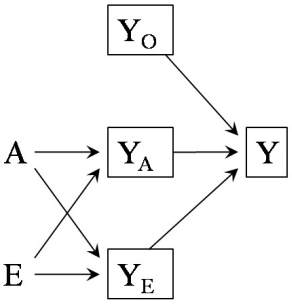


Figure 8.14

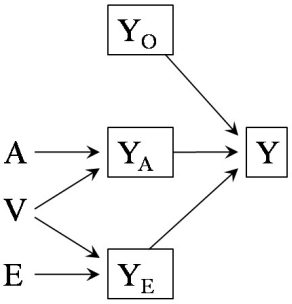


Figure 8.15

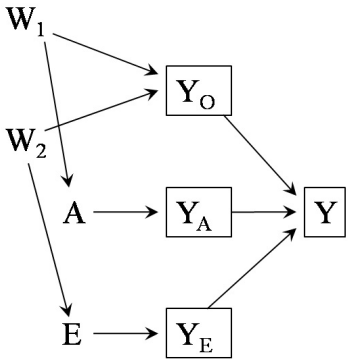


Figure 8.16

The causal diagram in Figure 8.12 represents a hypothetical study with dichotomous variables surgery A, certain genetic haplotype E, and death Y. According to the rules of d-separation, surgery A and haplotype E are (i) marginally independent, i.e., the probability of receiving surgery is the same for people with and without the genetic haplotype, and (ii) associated conditionally on Y, i.e., the probability of receiving surgery varies by haplotype when the study is restricted to, say, the survivors ($Y = 0$).

Indeed conditioning on the common effect Y of two independent causes A and E always induces a conditional association between A and E in at least one of the strata of Y (say, $Y = 1$). However, there is a special situation under which A and E remain conditionally independent within the other stratum (say, $Y = 0$).

Suppose A and E affect survival through totally independent mechanisms in such a way that E cannot possibly modify the effect of A on Y, and vice versa. For example, suppose that the surgery A affects survival through the removal of a tumor, whereas the haplotype E affects survival through increasing levels of low-density lipoprotein-cholesterol levels resulting in an increased risk of heart attack (whether or not a tumor is present). In this scenario, we can consider 3 cause-specific mortality variables: death from tumor Y_A, death from heart attack Y_E, and death from any other causes Y_O. The observed mortality variable Y is equal to 1 (death) when Y_A or Y_E or Y_O is equal to 1, and Y is equal to 0 (survival) when Y_A and Y_E and Y_O equal 0. The causal diagram in Figure 8.13, an expansion of that in Figure 8.12, represents a causal structure linking all these variables. We assume data on underlying cause of death (Y_A, Y_E, Y_O) are not recorded and thus the only measured variables are those in Figure 8.12 (A, E, Y).

Because the arrows from Y_A, Y_E and Y_O to Y are deterministic, conditioning on observed survival ($Y = 0$) is equivalent to simultaneously conditioning on $Y_A = 0$, $Y_E = 0$, and $Y_O = 0$ as well, i.e., conditioning on $Y = 0$ implies $Y_A = Y_E = Y_O = 0$. As a consequence, we find by applying d-separation to Figure 8.13 that A and E are conditionally independent given $Y = 0$, i.e., when conditioning on collider $Y = 0$, the path between A and E through Y is blocked by conditioning on the non-colliders Y_A, Y_E and Y_O. On the other hand, conditioning on death $Y = 1$ does not imply conditioning on any specific values of Y_A, Y_E and Y_O as the event $Y = 1$ is compatible with 7 possible unmeasured events: $(Y_A = 1, Y_E = 0, Y_O = 0)$, $(Y_A = 0, Y_E = 1, Y_O = 0)$, $(Y_A = 0, Y_E = 0, Y_O = 1)$, $(Y_A = 1, Y_E = 1, Y_O = 0)$, $(Y_A = 0, Y_E = 1, Y_O = 1)$, $(Y_A = 1, Y_E = 0, Y_O = 1)$, and $(Y_A = 1, Y_E = 1, Y_O = 1)$. Thus, A and E are associated given $Y = 1$, i.e., when conditioning on collider $Y = 1$, the path between A and E through Y is not blocked.

In contrast with the situation represented in Figure 8.13, the variables A and E will not be independent conditionally on $Y = 0$ when one of the situations represented in Figures 8.14-8.16 occur. If A and E affect survival through a common mechanism, then there will exist an arrow either from A to Y_E or from E to Y_A, as shown in Figure 8.14. In that case, A and E will be dependent within both strata of Y. Similarly, if Y_A and Y_E are not independent because of a common cause V as shown in Figure 8.15, A and E will be dependent within both strata of Y. Finally, if the causes Y_A and Y_O, and Y_E and Y_O, are not independent because of common causes W_1 and W_2 as shown in Figure 8.16, then A and E will also be dependent within both strata of Y. When the data can be summarized by Figure 8.13, we say that the data

Technical Point 8.2

Multiplicative survival model. When the conditional probability of survival $\Pr[Y = 0|E = e, A = a]$ given A and E is equal to a product $g(e)h(a)$ of functions of e and a, we say that a multiplicative survival model holds. A multiplicative survival model

$$\Pr[Y = 0|E = e, A = a] = g(e)h(a)$$

is equivalent to a model that assumes the survival ratio $\Pr[Y = 0|E = e, A = a] / \Pr[Y = 0|E = e, A = 0]$ does not depend on e and is equal to $h(a)$. The data follow a multiplicative survival model when there is no interaction between A and E for $Y = 0$ on the multiplicative scale. A proof that Figure 8.13 represents a multiplicative survival model proceeds as follows:

$\Pr[Y = 0|E = e, A = a] =$

$\Pr[Y_O = 0, Y_A = 0, Y_E = 0|E = e, A = a] = \Pr[Y_O = 0] \Pr[Y_A = 0|A = a] \Pr[Y_E = 0|E = e]$,

where the first equality is by determinism and the second by the DAG factorization.

Now set $g(e) = \Pr[Y_E = 0|E = e]$ and $h(a) = \Pr[Y_O = 0] \Pr[Y_A = 0|A = a]$. Note if $\Pr[Y = 0|E = e, A = a] = g(e)h(a)$, then $\Pr[Y = 1|E = e, A = a] = 1 - g(e)h(a)$ does not follow a multiplicative mortality model. Hence, when A and E are conditionally independent given $Y = 0$, they will be conditionally dependent given $Y = 1$.

Augmented causal DAGs, introduced by Hernán, Hernández-Díaz, and Robins (2004), can be extended to represent the sufficient causes described in Chapter 5 (VanderWeele and Robins, 2007c).

follow a *multiplicative survival model* (see Technical Point 8.2).

What is interesting about Figure 8.13 is that by adding the unmeasured variables Y_A, Y_E and Y_O, which functionally determine the observed variable Y, we have created an *augmented causal diagram* that succeeds in representing both the conditional independence between A and E given $Y = 0$ and the their conditional dependence given $Y = 1$.

In summary, conditioning on a collider always induces an association between its causes, but this association could be restricted to certain levels of the common effect. In other words, it is theoretically possible that selection on a common effect does not result in selection bias when the analysis is restricted to a single level of the common effect. Collider stratification is not always a source of selection bias.

Fine Point 8.2

The strength and direction of selection bias. We have referred to selection bias as an "all or nothing" issue: either bias exists or it doesn't. In practice, however, it is important to consider the expected direction and magnitude of the bias.

The direction of the conditional association between 2 marginally independent causes A and E within strata of their common effect Y depends on how the two causes A and E interact to cause Y. For example, suppose that, in the presence of an undiscovered background factor U that is unassociated with A or E, having either $A = 1$ or $E = 1$ is sufficient and necessary to cause death (an "or" mechanism), but that neither A nor E causes death in the absence of U. Then among those who died ($Y = 1$), A and E will be negatively associated, because it is more likely that an individual with $A = 0$ had $E = 1$ because the absence of A increases the chance that E was the cause of death. (Indeed, the logarithm of the conditional odds ratio $OR_{AE|Y=1}$ will approach minus infinity as the population prevalence of U approaches 1.0.) This "or" mechanism was the only explanation given in the main text for the conditional association of independent causes within strata of a common effect; nonetheless, other possibilities exist.

For example, suppose that in the presence of the undiscovered background factor U, having both $A = 1$ and $E = 1$ is sufficient and necessary to cause death (an "and" mechanism) and that neither A nor E causes death in the absence of U. Then, among those who die, those with $A = 1$ are more likely to have $E = 1$, i.e., A and E are positively correlated. A standard DAG such as that in Figure 8.12 fails to distinguish between the case of A and E interacting through an "or" mechanism from the case of an "and" mechanism. Causal DAGs with sufficient causation structures (VanderWeele and Robins, 2007c) overcome this shortcoming.

Regardless of the direction of selection bias, another key issue is its magnitude. Biases that are not large enough to affect the conclusions of the study may be safely ignored in practice, whether the bias is upwards or downwards. Generally speaking, a large selection bias requires strong associations between the collider and both treatment and outcome. Greenland (2003) studied the magnitude of selection bias under the null, which he referred to as *collider-stratification bias*, in several scenarios.