Chapter 9
MEASUREMENT BIAS AND "NONCAUSAL" DIAGRAMS

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 weak association between her looking up and other pedestrians' looking up. Does this weak association reflect a weak causal effect? By definition of randomized experiment, confounding bias is not expected in this study. In addition, no selection bias was expected because all pedestrians' responses—whether they did or did not look up—were recorded. However, there was another problem: the investigator's collaborator who was in charge of recording the pedestrians' responses made many mistakes. Specifically, the collaborator missed half of the instances in which a pedestrian looked up and recorded these responses as "did not look up." Thus, even if the treatment (the investigator's looking up) truly had a strong effect on the outcome (other people's looking up), the misclassification of the outcome will result in a dilution of the association between treatment and the (mismeasured) outcome.

We say that there is measurement bias when the association between treatment and outcome is weakened or strengthened as a result of the process by which the study data are measured. Since measurement errors can occur under any study design—including both randomized experiments and observational studies—measurement bias need always be considered when interpreting effect estimates. This chapter provides a description of biases due to measurement error.
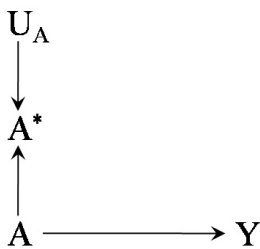
9.1 Measurement error

Figure 9.1

In previous chapters we implicitly made the unrealistic assumption that all variables were perfectly measured. Consider an observational study designed to estimate the effect of a cholesterol-lowering drug A on the risk of liver disease Y. We often expect that treatment A will be measured imperfectly. For example, if the information on drug use is obtained by medical record abstraction, the abstractor may make a mistake when transcribing the data, the physician may forget to write down that the patient was prescribed the drug, or the patient may not take the prescribed treatment. Thus, the treatment variable in our analysis data set will not be the *true* use of the drug, but rather the *measured* use of the drug. We will refer to the measured treatment as A^* (read A-star), which will not necessarily equal the true treatment A for a given individual. The psychological literature sometimes refers to A as the "construct" and to A^* as the "measure" or "indicator." The challenge in observational disciplines is making inferences about the unobserved construct (e.g., cholesterol-lowering drug use) by using data on the observed measure (e.g., information on statin use from medical records).

The causal diagram in Figure 9.1 depicts the variables A, A^*, and Y. For simplicity, we chose a setting with neither confounding nor selection bias for the causal effect of A on Y. The true treatment A affects both the outcome Y and the measured treatment A^*. The causal diagram also includes the node U_A to represent all factors other than A that determine the value of A^*. We refer to the difference between an individual's mismeasured value A^* and true value A as the *measurement error* of A for that individual. The magnitude and

Technical Point 9.1

Independence and nondifferentiality of measurement errors. For each individual, we define the measurement error of A as the difference $e_A = A^* - A$ and the measurement error of Y as the difference $e_Y = Y^* - Y$.

Let $f(\cdot)$ denote a probability density function (pdf). The measurement error e_A of treatment and the measurement error e_Y of outcome are *independent* if their joint pdf equals the product of each marginal pdf, i.e., $f(e_Y, e_A) = f(e_Y)f(e_A)$. The measurement error e_A of treatment is *nondifferential* if its pdf is independent of the outcome Y, i.e., $f(e_A|Y) = f(e_A)$. Analogously, the measurement error e_Y of the outcome is nondifferential if its pdf is independent of the treatment A, i.e., $f(e_Y|A) = f(e_Y)$.

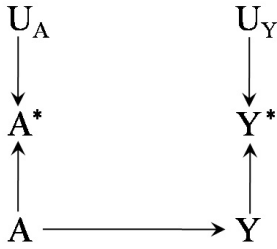Measurement error for discrete variables is known as *misclassification*.

Figure 9.2

Figure 9.3

direction of the measurement error is determined by the factors in U_A. Note that including the node U_A in the causal diagram is not strictly necessary because U_A is neither a cause shared by other variables on the diagram nor a variable that is conditioned on. We include it, however, to provide an explicit representation of the factors responsible for measurement error and for a direct comparison with the causal diagrams that we will discuss next.

Besides treatment A, the outcome Y can be measured with error too. The causal diagram in Figure 9.2 includes the measured outcome Y^*, and the factors U_Y responsible for the measurement error of Y. Figure 9.2 illustrates a common situation in practice. One wants to compute the average causal effect of the treatment A on the outcome Y, but these variables A and Y have not been, or cannot be, measured correctly. Rather, only the mismeasured versions A^* and Y^* are available to the investigator who aims at identifying the causal effect of A on Y.

Figure 9.2 also represents a setting in which there is neither confounding nor selection bias for the causal effect of treatment A on outcome Y. According to our reasoning in previous chapters, association is causation in this setting. We can compute any A-Y association measure and endow it with a causal interpretation as the effect of A on Y. For example, the associational risk ratio $\Pr[Y = 1|A = 1]/\Pr[Y = 1|A = 0]$ is equal to the causal risk ratio $\Pr[Y^{a=1} = 1]/\Pr[Y^{a=0} = 1]$. Our implicit assumption in previous chapters, which we now make explicit, was that perfectly measured data on A and Y were available.

We now consider the more realistic setting in which only the mismeasured versions of treatment and outcome, A^* and Y^*, are available. Then there is no guarantee that the measure of association between A^* and Y^* will equal the measure of causal effect of A on Y. The associational risk ratio $\Pr[Y^* = 1|A^* = 1]/\Pr[Y^* = 1|A^* = 0]$ will generally differ from the causal risk ratio $\Pr[Y^{a=1} = 1]/\Pr[Y^{a=0} = 1]$. We say that there is *measurement bias* or *information bias*. In the presence of measurement bias, the identifiability conditions of exchangeability, positivity, and consistency are insufficient to compute the causal effect of treatment A on outcome Y.
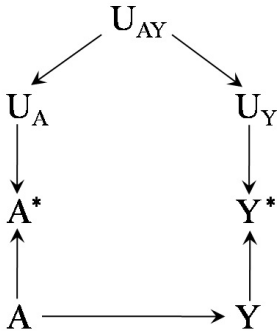
9.2 The structure of measurement error

The causal structure of confounding can be summarized as the presence of common causes of treatment and outcome, and the causal structure of selection bias can be summarized as conditioning on common effects of treatment
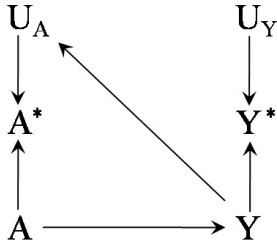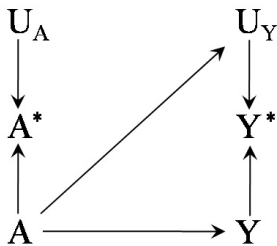
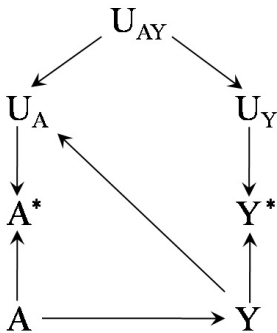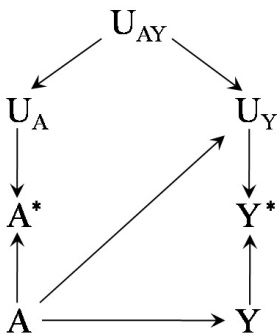and outcome (or of their causes). Measurement bias arises in the presence of measurement error, but there is no single structure to summarize measurement error. This section classifies the structure of measurement error according to two properties—*independence* and *nondifferentiality*—that we describe below (see Technical Point 9.1 for formal definitions).

The causal diagram in Figure 9.2 depicts the measurement errors U_A and U_Y for both treatment A and outcome Y, respectively. According to the rules of d-separation, the measurement errors U_A and U_Y are independent because the path between them is blocked by colliders (either A^* or Y^*). *Independent measurement errors* are expected to arise if, e.g., information on both drug use A and liver toxicity Y was obtained from electronic medical records in which data entry errors occurred haphazardly. In other settings, however, the measurement errors for exposure and outcome may be dependent, as depicted in Figure 9.3. For example, dependent measurement errors will occur if the information were obtained retrospectively by phone interview and an individual's ability to recall her medical history (U_{AY}) affected the measurement of both treatment A and outcome Y.

Figures 9.2 and 9.3 represent settings in which the factors U_A responsible for the measurement error of the treatment are independent of the true value of the outcome Y, and the factors U_Y responsible for the measurement error for the outcome are independent of the true value of treatment A. We then say that the measurement error for treatment is *nondifferential with respect to the outcome*, and that the measurement error for the outcome is *nondifferential with respect to the treatment*. The causal diagram in Figure 9.4 shows an example of independent but differential measurement error in which the true value of the outcome affects the measurement of the treatment (i.e., an arrow from Y to U_A). We now describe some examples of differential measurement error of the treatment.

Suppose that the outcome Y was dementia rather than liver toxicity, and that drug use A was ascertained by interviewing study participants. Since the presence of dementia affects the ability to recall A, one would expect an arrow from Y to U_A. Similarly, one would expect an arrow from Y to U_A in a study to compute the effect of alcohol use during pregnancy A on birth defects Y if alcohol intake is ascertained by recall after delivery—because recall may be affected by the outcome of the pregnancy. The resulting measurement bias in these two examples is often referred to as *recall bias*. A bias with the same structure might arise if blood levels of drug A^* are used in place of actual drug use A, and blood levels are measured after liver toxicity Y is present—because liver toxicity affects the measured blood levels of the drug. The resulting measurement bias is often referred to as *reverse causation bias*.

The causal diagram in Figure 9.5 shows an example of independent but differential measurement error in which the true value of the treatment affects the measurement of the outcome (i.e., an arrow from A to U_Y). A differential measurement error of the outcome will occur if physicians, suspecting that drug use A causes liver toxicity Y, monitored patients receiving drug more closely than other patients. Figures 9.6 and 9.7 depict measurement errors that are both dependent and differential, which may result from a combination of the settings described above.

In summary, we have discussed four types of measurement error: independent nondifferential (Figure 9.2), dependent nondifferential (Figure 9.3), independent differential (Figures 9.4 and 9.5), and dependent differential (Figures 9.6 and 9.7). The particular structure of the measurement error determines the methods that can be used to correct for it. For example, there is a large

Figure 9.4

Figure 9.5

Figure 9.6

Figure 9.7

The strength and direction of measurement bias. In general, measurement error will result in bias. A notable exception is the setting in which A and Y are unassociated and the measurement error is independent and nondifferential: If the arrow from A to Y did not exist in Figure 9.2, then both the A-Y association and the A^*-Y^* association would be null. In all other circumstances, measurement bias may result in an A^*-Y^* association that is either further from or closer to the null than the A-Y association. Worse, even under the independent and nondifferential measurement error structure of Figure 9.2, non-extreme measurement bias may result in A^*-Y^* and A-Y trends in opposite directions for non-dichotomous ordinal treatments and for continuous treatments. This trend reversal under independent and nondifferential measurement error occurs when the conditional mean of A^* given A is a nonmonotonic function of A. See Dosemeci, Wacholder, and Lubin (1990) and Weinberg, Umbach, and Greenland (1994) for details. VanderWeele and Hernán (2009) described a more general framework using signed causal diagrams.

The magnitude of the measurement bias depends on the magnitude of the measurement error. That is, measurement bias generally increases with the strength of the arrows from U_A to A^* and from U_Y to Y^*. Causal diagrams do not encode quantitative information, and therefore they cannot be used to describe the magnitude of the bias.

literature on methods for measurement error correction when the measurement error is independent nondifferential. In general, methods for measurement error correction rely on a combination of modeling assumptions and validation samples, i.e., subsets of the data in which key variables are measured with little or no error. The description of methods for measurement error correction is beyond the scope of this book. Rather, our goal is to highlight that the act of measuring variables (like that of selecting individuals) may introduce bias (see Fine Point 9.1 for a discussion of its strength and direction). Realistic causal diagrams need to simultaneously represent biases arising from confounding, selection, and measurement. The best method to fight bias due to mismeasurement is, obviously, to improve the measurement procedures for the variables used in our analysis.

9.3 Mismeasured confounders and colliders

Besides the treatment A and the outcome Y, the confounders L may also be measured with error. Mismeasurement of confounders may result in bias even if both treatment and outcome are perfectly measured.

To see this, consider the causal diagram in Figure 9.8, which includes the variables drug use A, liver disease Y, and history of hepatitis L. Individuals with prior hepatitis L are less likely to be prescribed drug A and more likely to develop liver disease Y. As discussed in Chapter 7, there is confounding for the effect of the treatment A on the outcome Y because there exists an open backdoor path $A \leftarrow L \rightarrow Y$, but there is no unmeasured confounding given L because the backdoor path $A \leftarrow L \rightarrow Y$ can be blocked by conditioning on L. That is, there is exchangeability of the treated and the untreated conditional on the confounder L, and one can apply IP weighting or standardization to compute the average causal effect of A on Y. The standardized, or IP weighted, risk ratio based on L, Y, and A will equal the causal risk ratio $\Pr\left[Y^{a=1} = 1\right] / \Pr\left[Y^{a=0} = 1\right]$.

Again the implicit assumption in the above reasoning is that the confounder L was perfectly measured. Suppose investigators did not have access to the
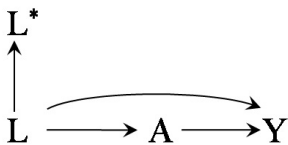

Figure 9.8

study participants' medical records. Rather, to ascertain previous diagnoses of hepatitis, investigators had to ask participants via a questionnaire. Since not all participants provided an accurate recollection of their medical history—some did not want anyone to know about it, others had memory problems or simply made a mistake when responding to the questionnaire—the confounder L was measured with error. Note that Figure 9.8 does not explicitly represent the factors U_L responsible for the measurement error of L because the particular structure of this error is not relevant to our discussion.

Investigators had data on the mismeasured variable L^* rather than on the variable L. Unfortunately, the backdoor path $A \leftarrow L \rightarrow Y$ cannot be generally blocked by conditioning on L^*. The standardized (or IP weighted) risk ratio based on L^*, Y, and A will generally differ from the causal risk ratio $\Pr\left[Y^{a=1} = 1\right] / \Pr\left[Y^{a=0} = 1\right]$. We then say that there is *measurement bias* or *information bias*. The causal diagram in Figure 9.9 shows an example of confounding of the causal effect of A on Y in which L is not the common cause shared by A and Y. Here too mismeasurement of L leads to measurement bias because the backdoor path $A \leftarrow L \leftarrow U \rightarrow Y$ cannot be generally blocked by conditioning on L^*.
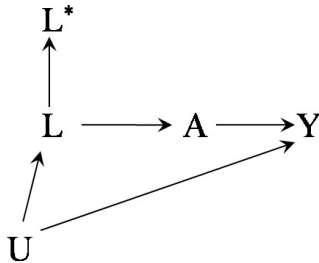

Figure 9.9

Alternatively, one could view the bias due to mismeasured confounders in Figures 9.8 and 9.9 as a form of unmeasured confounding rather than as a form of measurement bias. In fact the causal diagram in Figure 9.8 is equivalent to that in Figure 7.8. One can think of L as an unmeasured variable and of L^* as a surrogate confounder (see Fine Point 7.2). The particular choice of terminology—unmeasured confounding versus bias due to mismeasurement of the confounders—is irrelevant for practical purposes. In some settings, however, the use of mismeasured variables is sufficient to adjust for confounding. See Fine Point 9.2 for some examples.

Mismeasurement of confounders may also result in apparent effect modification. As an example, suppose that all study participants who reported a prior diagnosis of hepatitis ($L^* = 1$) and half of those who reported no prior diagnosis of hepatitis ($L^* = 0$) did actually have a prior diagnosis of hepatitis ($L = 1$). That is, the true and the measured value of the confounder coincide in the stratum $L^* = 1$, but not in the stratum $L^* = 0$. Suppose further that treatment A has no effect on any individual's liver disease Y, i.e., the sharp null hypothesis holds. When investigators restrict the analysis to the stratum $L^* = 1$, there will be no confounding by L because all participants included in the analysis have the same value of L (i.e., $L = 1$). Therefore they will find no association between A and Y in the stratum $L^* = 1$. However, when the investigators restrict the analysis to the stratum $L^* = 0$, there will be confounding by L because the stratum $L^* = 0$ includes a mixture of individuals with both $L = 1$ and $L = 0$. Thus the investigators will find a non-null association between A and Y as a consequence of uncontrolled confounding by L. If the investigators are unaware of the fact that there is mismeasurement of the confounder in the stratum $L^* = 0$ but not in the stratum $L^* = 1$, they could naively conclude that both the association measure in the stratum $L^* = 0$ and the association measure in the stratum $L^* = 1$ can be interpreted as effect measures. Because these two association measures are different, the investigators will say that L^* is a modifier of the effect of A on Y even though no effect modification by the true confounder L exists.
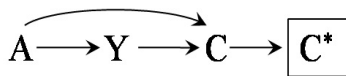
Finally, it is also possible that a collider C is measured with error. This situation is represented in the causal diagram in Figure 9.10, which is equivalent to Figure 8.2. When interested in the effect of A on Y under Figure 9.10, conditioning on the mismeasured collider C^* will generally introduce selection

Figure 9.10

Fine Point 9.2

When mismeasured confounders are not a problem. In many medical applications, measurement error in the confounders does not introduce any bias. Suppose that high blood pressure L affects both the probability of receiving antihypertensive therapy A and of having a stroke Y. Doctors and patients, however, do not make their treatment decisions based on the true blood pressure L but based on the blood pressure measurement L^* that was recorded in the doctor's office. That is, L^* is the only information about blood pressure that is accessible to decision makers.

Figure 9.11 (which is structurally equivalent to Figure 7.2) represents this situation. The possibly mismeasured L^* fully mediates the effect of L on A because any component of L that was not captured by L^* remained unknown and thus could not influence the decision to administer the treatment. It follows that the backdoor path between A and Y can be blocked by conditioning on either the true L or the measured L^*. Therefore, it is irrelevant whether investigators had access to the true L or to the measured L^*. Either variable is sufficient to adjust for confounding.

A more extreme example is shown in Figure 9.12. Under this causal diagram, having data on the true L is insufficient to adjust for confounding whereas having data on the measured L^* is sufficient to adjust for confounding. The general point is that effects can be identified whenever we have as much information in the data as the decision makers had to make their decisions, regardless of whether that information resulted from perfectly measured variables or from variables measured with error.

bias because C^* is a descendant of the collider and therefore a common effect of the treatment A and the outcome Y.

9.4 Causal diagrams without mismeasured variables?
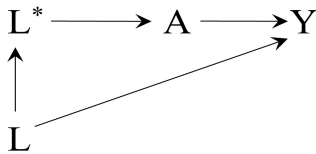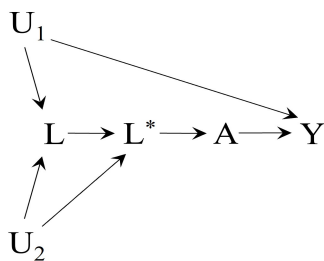

Figure 9.11

Figure 9.12

When drawing causal diagrams in previous chapters, we have been implicitly making two simplifying, and related, assumptions. In this and the next section we make those assumptions and their implications explicit.

The first assumption is that all variables on the diagram are perfectly measured. This assumption is not realistic because, in practice, measurement error is often unavoidable for treatments, outcomes, confounders, and any other variables of interest. In this chapter, we have described how causal diagrams can be used to represent mismeasured variables under different types of measurement error. We have also explored the consequences of using the mismeasured variables, which are the only ones available to investigators, for the identification of causal effects.

For example, suppose that we are interested in the effect of the treatment A on the outcome Y, but we only have data on the measured treatment A^* and the measured outcome Y^*. We have seen how measurement error of treatment or outcome may induce a noncausal association between the measured treatment A^* and the measured outcome Y^*, even if treatment A has a null effect on the outcome Y and even if there is no confounding and no selection bias. Also, in the presence of confounding, we have seen how measurement error of a confounder may prevent the measured confounder L^* from blocking a backdoor path that would be successfully blocked if we had access to the true confounder L.

Because all variables can be expected to be measured with some error, it might be argued that a causal diagram should always represent both true and measured values of all its variables. Yet, in many settings, the magnitude of the measurement error may be judged, or known, to be too small to matter. In

those settings, a causal DAG that makes no distinction between measured and true values of a variable may be preferable for simplicity. If confounding and selection bias exist under perfect measurement of all variables, then these biases will typically exist under measurement error too (though exceptions exist as shown in Fine Point 9.2). Drawing causal diagrams without measurement error allows us to focus on confounding and selection bias without being distracted by measurement issues.

Considering causal diagrams without measurement error is often a helpful first approximation. Once we have a good understanding of possible biases under perfect measurement, we can add measurement error as an additional layer of complexity. This 2-step approach to the drawing of causal diagrams helps us isolate the study of two sources of bias—confounding and selection—without being overburdened by the third one—measurement. We follow this approach in the book: when the emphasis is on confounding and selection bias, we omit the distinction between true and measured values of the variables on the causal diagram.

The second assumption we have made so far, also related to measurement, is a fundamental assumption in any causal diagram. We discuss this assumption in the next section.

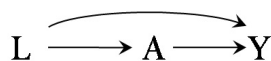## 9.5 Many proposed causal diagrams include noncausal arrows

Consider the causal diagram in Figure 9.13 (which is equal to Figure 7.1). Let A, Y, and L be three binary variables representing an antiviral treatment given to patients with COVID-19, death, and obesity (defined as body mass index greater than 30), respectively. Patients with obesity are more likely to receive treatment and to be hospitalized in the absence of treatment. Therefore, experts draw a DAG with arrows from L into both A and Y. For simplicity, we will assume that the decision to give treatment is only influenced by L and by the physician's preference, which is unrelated to any other variables on the diagram. Also for simplicity, we assume no measurement error for any variable. Specifically, the true value of L equals its measured value L^* so the latter does not need be included on the diagram.

Figure 9.13

We have previously discussed how, in causal diagrams, treatment nodes have a different status than other nodes (see Section 6.4). The reason is that meaningful quantitative causal inference about the effect of treatment A on outcome Y requires well-defined, actual or hypothetical, interventions on A. Otherwise the counterfactual outcomes Y^a remain undefined and cannot be linked to the observed outcomes Y, i.e., the consistency condition does not hold (see Chapter 3). In our example, the intervention represented by treatment A is well-defined because we have a good understanding of how antiviral treatment can be administered or withheld. Thus, the presence or absence of an arrow from A to Y is well defined too.

We now extend our discussion to nodes that are not considered a treatment, such as obesity L in Figure 9.13. As we discussed extensively in Chapter 3, interventions on obesity L on death are not well defined. Thus, the counterfactual outcomes Y^l remain undefined.

So far in this book we have ignored this problem, but many DAGs proposed in the health and social sciences have arrows emanating from variables for which well-defined interventions do not exist, like L in Figure 9.13. Those arrows, like the arrow $L \to Y$ in Figure 9.13, do not have a causal interpreta-

Fine Point 9.3

Whether interventions are well-defined depends on the outcome of interest. In the main text we said that there may exist well-defined interventions for the effect of L on A even if there are no well-defined interventions for the effect of L on Y. This statement needs a better explanation because, if we truly knew how to intervene on L to study its effect on A, then what would prevent us from using the same intervention to study its effect on Y?

This apparent contradiction results from an abuse of notation. In our example, we used the symbol L to refer to two concepts: (i) the physical quantity body weight (measured in kilos), and (ii) the recorded value of that quantity (measured in kilos). In a world without measurement error, both (i) and (ii) have the same numerical value and thus conflating both concepts has no practical impact. However, when we described in the main text a well-defined intervention for L on A, we were referring to an intervention on (ii) rather than on (i), and thus a distinction between both concepts is warranted.

Let us use the symbol L to depict the physical quantity "body weight", and the symbol L^* to depict the recording of L. If we add L^* to the causal DAG, there would be a deterministic arrow from L to L^* (assuming no measurement error) and a regular arrow from L^* to A. According to this expanded causal DAG, body weight L only affects treatment A when the doctor learns the value L^*. Therefore, if we intervene on the recorded value L^*, even if we leave the physical quantity L unchanged, the doctors' behavior would be the same as if we could somehow intervene on L. It is in this sense that we say in the main text that there are well-defined interventions on L when the outcome is A—the counterfactuals A^l are well defined—but not when the outcome is Y—the counterfactuals Y^l are not well defined.

tion. To explore the consequences of using these DAGs with noncausal arrows, we now discuss Figure 9.13 in more detail. Let us consider separately the arrows $L \to A$ and $L \to Y$.

When experts drew the arrow from $L \to A$, they were using their subject-matter knowledge. Experts knew that doctors are more likely to administer treatment to obese patients upon learning that they are obese. One could imagine well-defined interventions for the effect of L on A, such as presenting the treating physician with a patient whose body mass index differs from that of the patient for whom the treatment decision is being made. Therefore, drawing the arrow from $L \to A$ is reasonable. For additional discussion on this point, see Fine Point 9.3 after reading the next paragraph.

Were the experts justified in drawing the arrow $L \to Y$? Suppose the experts know that obese patients are more likely to die, but cannot propose well-defined interventions for the effect of L and Y. Then the arrow $L \to Y$ has no causal interpretation. In this chapter and generally throughout the book, we restrict the term causal DAG to DAGs for which all arrows have a well-defined causal interpretation. Therefore, under our restriction and the current state of knowledge, Figure 9.13 is a "noncausal" diagram. We have placed the word "noncausal" in quotes as a reminder that many papers in the causal literature continue to define DAGs that include both causal and noncausal arrows as causal diagrams. See Fine Point 9.4 for more discussion on "noncausal" diagrams.
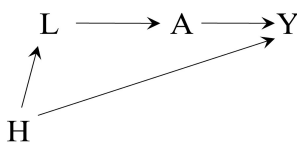
Now suppose that a second group of experts proposes an alternative interpretation for the known fact that "obese patients are more likely to die": obesity is a surrogate or proxy for a hidden factor H that has a causal effect on Y. The corresponding causal diagram is Figure 9.14 (structurally identical to Figure 7.2), which includes unmeasured and possibly unknown variables H such as genetic factors, metabolic factors related to body fat stores, microbiota, and probably others not yet discovered by science. According to Figure 9.14, the effect of the factors H on the treatment A is fully mediated through L.

Figure 9.14

Fine Point 9.4

"Noncausal"diagrams with well-defined statistical interpretations. Consider a DAG representation for an FFR-CISTG model that assumed (i) the treatment A was the only variable with well-defined interventions with counterfactuals (M^a, Y^a) and (ii) the distribution of the variables on the DAG factored according to the DAG, i.e., each variable was conditionally independent of its non-descendants given its parents. In these DAGs, the arrows do not, in general, have a causal interpretation. Rather, the arrows simply encode, via d-separation, the conditional independencies satisfied by the variables on the diagram and on the associated SWIG (see Technical Point 21.12). Under this noncausal interpretation of DAGs, the $L \to Y$ arrow need not be removed in Figure 9.13 just because interventions on L are not well-defined.

Richardson and Robins (2013) pointed out serious difficulties with this approach. Specifically, if arrows have no causal interpretation, then there is no reason to expect the distribution of the study variables to be representable by (i.e., factor according to) any (incomplete) DAG. This difficulty is magnified in the presence of unmeasured variables U because then it is not even possible to empirically check some conditional independencies from the observed data. For example, the front door graph of Figure 7.14 implies that Y is independent of A given M and U. Any investigator who chooses Figure 7.14 as the appropriate causal DAG must have had substantive reasons for postulating this conditional independence. Indeed many researchers might find it hard to imagine what those reasons could be other than the belief that A and Y share a common cause U and that the causal effect of A on Y is completely mediated by M—a belief that endows every arrow on the diagram with a causal interpretation. See also Fine Point 9.5.

One can alternatively view the use of noncausal arrows as a response by researchers who are skeptical of the strong causal claim that M-counterfactuals Y^m exist. This "noncausal" approach interprets Figure 7.14 as representing the hypothesized statistical independence $Y \perp\!\!\!\perp A|M$ among the observed variables that would hold in a future trial in which A is randomly assigned. Thus, if this independence fails to hold in the future trial, the justification of the strong causal claim that M-counterfactuals exist has been falsified along with the structure in Figure 7.14.

That is, there is no direct arrow from H to A. This assumption is reasonable if, for example, the only information used to make real world treatment decisions is obesity L.

Like Figure 9.13, Figure 9.14 encodes the (reasonable) assumptions that there are well-defined interventions for the effect of obesity L on treatment A—this assumption is represented by the arrow $L \to A$—and that there is no direct effect of L on Y not through A. Therefore, the associated counterfactuals are A^l and Y^a, which imply the existence of a well-defined intervention of L on Y with $Y^l = Y^{A^l}$. That is, Y^l is equal to Y^a for a equal to the counterfactual A^l. In the language of Technical Point 6.2, we say that Y^l is obtained from Y^a and A^l by recursive substitution. There is no contradiction in Y^l being well-defined for the causal diagram of Figure 9.14 but not for Figure 9.13 because, assuming Figure 9.14 is a causal diagram, Figure 9.13 is not, as Figure 9.13 fails to include the common cause H of L and Y. See also Fine Point 9.6.

Figure 9.14 also encodes the assumption that well-defined interventions exist for the effect of the unknown factors H on both L and Y. That is, the arrows $H \to L$ and $H \to Y$ on Figure 9.14 imply that there exist counterfactuals L^h and Y^h, respectively, associated with a well-defined intervention on H. It may seem counterintuitive for us to stipulate simultaneously that (i) the current state of knowledge is insufficient to even characterize many of the factors in H, and (ii) well-defined interventions exist for the effect of H on both L and Y. To explain why we might do so and hence regard Figure 9.14 as a causal diagram under the current state of knowledge, let us define H more precisely.

When, as in Figure 9.13, there exists a variable L on a proposed causal diagram for which Y^l is ill-defined, we introduce a high-dimensional parent H

Fine Point 9.5

A connection to the front door formula. Figure 9.14 is precisely the front door diagram in Figure 7.14 with $L \to A$ substituted for $A \to M$. Hence $\mathrm{E}\left[Y^l\right]$ is identified by the front door formula in Technical Point 7.4 with L, l, l' substituted for A, a, a' and A, a substituted for M, m.

Suppose a researcher modifies Figure 9.14 by adding a direct $L \to Y$ arrow. The justification is that, when dietitians and physical therapists are referred an individual who has a high value of L, they provide additional ancillary care (such as dietary advice, exercises to prevent deep vein thrombosis, etc.), which is not recorded in the medical record available to the researchers. The counterfactual variable Y^l in the unmodified diagram in Figure 9.14 differs from the variable Y^l in the modified diagram with a direct $L \to Y$ arrow. In the unmodified graph, the contrast $Y^l - Y^{l'}$ is the causal effect of l versus l' via the causal pathway $L \to A \to Y$. In contrast, in the modified graph, $Y^l - Y^{l'}$ is the total effect along the two pathways $L \to Y$ and $L \to A \to Y$. In fact, in contrast to the unmodified graph, $\mathrm{E}\left[Y^l\right]$ is not identified under the modified graph.

One might conjecture that the effect of l versus l' along the pathway $L \to A \to Y$ should be the same in the two graphs and thus remain identified by the front door formula. This conjecture is indeed true, although we must defer a proof until we study the identification of path-specific effects in Chapter 23.

of L that encodes all unmeasured—known and unknown—causal determinants of L (other than any known, but unmeasured, variables U already present on the diagram). H may be a direct cause and thus a parent of other variables such as Y as well. We regard L as the lower-dimensional effect of, or surrogate for, H that has been recorded for data analysis. In our obesity example, the continuous variable body mass index L is an effect of the largely unknown factors H that regulate body weight. In some instances, L may be thought of as a deterministic (many to one) function of H.

Even though we are ignorant of the precise intervention on the components (many unknown) of H that affect Y, we nonetheless assume that there exist factors H that have a causal effect on Y. Because current knowledge does not rule out the existence of well-defined interventions for the effect of H on Y, the arrow $H \to Y$ is tentatively justified and therefore we consider Figure 9.14 to be a causal diagram.

In this section, we distinguished between H, unmeasured common causes between two variables when the effect of one of the variables on the other is ill-defined, and U, unmeasured common causes between two variables when the effect of one of the variables on the other is well-defined. In most of the book, we do not make this distinction and simply use U to represent all unmeasured variables.

9.6 Does it matter that many proposed diagrams include noncausal arrows?

We have seen that the original DAG (Figure 9.13) proposed by the experts is not a causal DAG because one of its arrows (the one from L to Y) cannot be causally interpreted. Yet, despite being causally wrong, this DAG is adequate for causal inference about the effect of treatment A on the outcome Y: regardless of whether we use the noncausal DAG in Figure 9.13 or the causal DAG in Figure 9.14, we conclude that all backdoor paths between A and Y are blocked by conditioning on L.

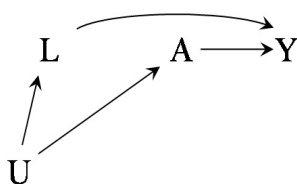Thus the DAG in Figure 9.13, which lacks the node U and includes the non-

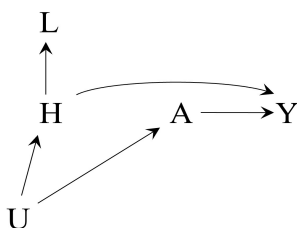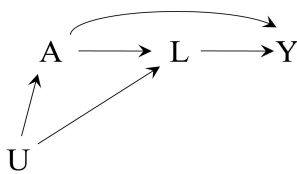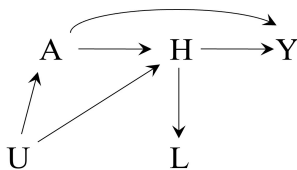causal arrow $L \rightarrow Y$, correctly guides data analysis because adjusting for L is all that is needed in the identifying formula—standardization or IP weighting in our example—for the average causal effect of A on Y. This conclusion is expected because, in both DAGs, there are no arrows from unmeasured variables into treatment A (see also Fine Point 9.2). This oversimplified scenario illustrates a general issue: in realistic complex settings, expert knowledge is rarely good enough to draw a causal diagram in which all of its components are known.

In fact, many DAGs proposed in the health and social sciences are actually noncausal DAGs because they lack the hidden variables H that make the DAG causal. By including a variable without well-defined interventions for its effect on its descendants (like obesity and its descendant death), we are effectively declaring that our DAG is noncausal. It is, however, possible that the identifying formula for the causal effect of interest is the same when derived from the noncausal DAG and from the causal DAG with hidden nodes. This is exactly what happened in our example: if Figure 9.14 is the correct DAG, using the DAG in Figure 9.13 will result in the same identifying formula.

But noncausal DAGs may be misleading. Consider the DAG in Figure 9.15. Declaring that this DAG is a causal DAG implies that we believe there are well-defined interventions for the effect of L on Y (because there is a direct arrow from L to Y). Conversely, in the absence of such well-defined interventions, the DAG is noncausal and we can only hope that the identifying formula based on L happens to be correct.

Following our reasoning above, Figure 9.16 depicts our measured variable L is actually a surrogate of an unknown variable H for which well-defined interventions exist. In Figure 9.15 the backdoor path between A and Y is blocked by L, but in Figure 9.16 it is not.

Therefore, if investigators unaware of the status of their measured variable as a surrogate confounder L proposed Figure 9.15 instead of Figure 9.16, they would reach the incorrect conclusion that L is sufficient to block all backdoor paths, as discussed in Section 9.3. When drawing causal DAGs, we need to think carefully whether the variables that happen to be measured are also the variables for which well-defined interventions exist. Otherwise, lack of attention to the noncausal arrows of a DAG may give us false confidence in the validity of our effect estimates. See Fine Point 9.6 for another example.

We have come a long way since we introduced causal diagrams in Chapter 6. Causal DAGs and SWIGs are formidable tools for investigators to organize and communicate their causal assumptions but, as we have seen in this chapter, these diagrams are subject to the same practical constraints that are inherent to causal inference in general. Specifically, a causal arrow $X \rightarrow Y$ cannot be meaningfully interpreted in the absence of well-defined interventions for the effect of X on Y.

When proposing a causal DAG, we need to think carefully about the interpretation of each of its arrows. A scientifically blind acceptance of DAGs with noncausal arrows may lead to incorrect conclusions for causal inference. This level of scrutiny is unnecessary for causal diagrams representing an electrical circuit in which all interventions are well-defined, but indispensable for the causal diagrams proposed in the health and social sciences.

Figure 9.15

Figure 9.16

Figure 9.17

Figure 9.18

The above discussion simplifies the concept of well-defined intervention for pedagogic purposes. As discussed in Chapter 3, no intervention is perfectly well-defined but, for some interventions, the scientific consensus is that they

Fine Point 9.6

From noncausal diagrams to causal diagrams. Suppose some investigators interested in the effect of A on Y proposed the DAG in Figure 9.17. To draw this DAG, they relied on prior knowledge about the temporal order of the variables and the following two facts: (i) the measured variable L is associated with Y, and (ii) there is one set of unmeasured, but known, factors U that affect A and are associated with L. As discussed in Fine Point 7.4, the effect of A on Y is not identifiable if Figure 9.17 is the true causal diagram because L is a descendant of A.

Upon further reflection, the investigators realize that there are no well-defined interventions for the effect of L on Y. Therefore, the arrow $L \rightarrow Y$ is not a causal arrow and their DAG is not causal. To transform Figure 9.17 into a causal diagram, they add the hidden variable H with a causal arrow into Y and represent L as a surrogate of H. Figure 9.18 depicts their revised DAG. The effect of A on Y is not identifiable if Figure 9.18 is the true causal diagram because the backdoor path $A \leftarrow U \rightarrow H \rightarrow Y$ cannot be blocked by any measured variable.

Note that the investigators kept the arrow $U \rightarrow A$, which implies that they believe that there are well-defined interventions for the effect of U and A. They also redirected the arrows from U and A to L in Figure 9.17 towards H in Figure 9.18. This implies that the investigators believe that there must exist well-defined interventions for the effect of both U and A on H and that the effects of both U and A on L are fully mediated by their effect on H. (Even if L were a deterministic function of H, which is compatible with Figure 9.18, conditioning on L would not block paths through H because H is not a deterministic function of L, as it is of higher dimensionality and complexity than L.)

The inheritance by H in Figure 9.18 of all the arrows into L in Figure 9.17 is not always warranted. For example, U may have direct effect on L as in the causal DAG in Figure 9.19 rather than on H as in Figure 9.18. If, in fact, Figure 9.19 is the true causal DAG then the effect of A on Y would be identifiable because there are no open backdoor paths between A and Y. Note that, in Figure 19.9, the surrogate L cannot be a deterministic function of H as it is also affected by U. An example where, as in Figure 9.19, U has a direct effect on the surrogate L but no direct effect on H is the following: U denotes a physician's decision to order a particular diagnostic test, L the result of the test, and H the underlying biological determinants of the test result.
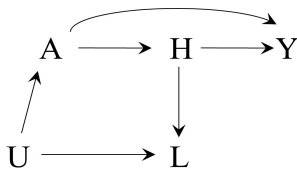

Figure 9.19

are sufficiently well-defined. We have argued that quantitative causal inference fundamentally relies on the (admittedly fuzzy) concept of sufficiently well-defined interventions. Therefore, in the remainder of this book, except for several well sign-posted exceptions, we will assume all DAGs are strictly causal in the sense that every arrow on the DAG represents a sufficiently well-defined intervention. That is, each arrow is associated with an intervention that can be specified in such a way that no meaningful vagueness remains based on current scientific knowledge. In a few years or decades, we may find out that our beliefs, and thus our causal diagrams, were incorrect.