Chapter 6
Confounding bias

Associations

When I started writing this chapter, I looked up synonyms of the word “association” on the computer's thesaurus, hoping to find one that will help me to explain the term. Of four listed choices only two had something to do with the statistical idea of association—connection and relationship—but neither quite fit. “Connection” sounds too physical and a “relationship”, I had been told, may develop between two people but not between two things. Variables can only have a relation—that is, an association.

Synonyms aside, what does it mean that variables A and B are associated?

It means that the two variables show a mathematical relation in a data set that enables us to guess the values of one from the values of the other, because the distribution of each variable varies according to the values of the other. Blood pressure and sex, for example, are associated in some data set if the distribution of blood pressure in men differs from that in women. Therefore, knowing that Smith is a man and Jones is a woman (both contributors of data) enables us to guess that Smith’s blood pressure is higher than Jones’s. Or the other way around: knowing that Smith’s blood pressure is higher than Jones’s enables us to guess, for instance, that the former is a man and the latter, a woman. Association does not imply a correct guess—only the possibility of a guess by using data rather than by tossing a coin.

Similarly, if two variables are not associated in some data set, knowing two values of one variable (and nothing else) add nothing to our guessing of the values of the other. We cannot make any comparative statement about the blood pressures of Smith and Jones conditional on knowledge that Smith is a man and Jones a woman, nor can we make any statement about their sex groups conditional on knowledge of their blood pressures.

It is crucial to understand that an association, or its absence, is a property of a data set rather than a universal phenomenon. With careful planning, for example, we can assemble a sample in which smoking status will be associated with eye color but not with owning a cigarette lighter, yet as far as associations are concerned there is nothing wrong with the sample or with the relations within. An association between variables A and B, or its absence, is neither false nor biased nor distorted nor spurious nor anything of the kind. It is simply a fault-free mathematical property of a data set—until we start making claims about causal reality, such as “A causes B” or “B causes A.” Researchers who call some association “spurious” are codifying a far more elaborated idea: an association that does not describe one particular cause-and-effect relation. There is nothing spurious, however, in the association itself because it never claimed to be more than what I defined two paragraphs earlier.

After understanding the idea of association, let’s find out what kind of a mathematical relation between two variables enables us to make a guess about the values of one from the values of the other. Any number of relations will do, but their form often depends on whether the variables in question are categorical or continuous. To keep things simple for the moment, we’ll discuss in this section only two possibilities: 1) one variable is binary and the other continuous; 2) both are binary. In the next section you will find an example of two continuous variables; other kinds of pairs will show up in later chapters.
When A is binary and B is continuous, their relation may take the form of different arithmetic means of B in the two categories of A. Two examples are shown in Table 6–1. In this table, smoking status (smoker or former smoker) and sex group play the role of variable A whereas FEV<sub>i</sub> (forced expiratory volume in one second) is variable B. FEV<sub>i</sub>, as you may recall, is a measure of lung function.

Table 6–1. Arithmetic mean of FEV<sub>i</sub> by smoking status and by sex in a sample of 2,476 people

<table>
<thead>
<tr>
<th>SMOKING STATUS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Smoker (N=877)</td>
</tr>
<tr>
<td>Arithmetic mean of FEV&lt;sub&gt;i&lt;/sub&gt; (liter)</td>
<td>2.77</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SEX GROUP</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female (N=1,106)</td>
</tr>
<tr>
<td>Arithmetic mean of FEV&lt;sub&gt;i&lt;/sub&gt; (liter)</td>
<td>2.51</td>
</tr>
</tbody>
</table>

FEV<sub>i</sub>: forced expiratory volume in one second

Since the mean value of FEV<sub>i</sub> is lower in smokers than in former smokers (2.77 liters versus 3.19 liters), smoking status and FEV<sub>i</sub> are associated in this sample. If we are told that one member of the sample is a smoker and another is a former smoker, we can say something about their respective FEV<sub>i</sub>—we can guess that it is smaller in the smoker. Likewise, if we are told that one person’s FEV<sub>i</sub> is smaller than another’s, we can guess that the former is a smoker and the latter no longer smokes. Neither guess will necessarily be correct, but recall that correct guesses were not required in the definition of association.

Similar reasoning applies to sex and FEV<sub>i</sub> because the arithmetic mean of FEV<sub>i</sub> is lower in women than in men (Table 6–1). If on the other hand the two means were identical, say 3.1 liter in both men and women, FEV<sub>i</sub> and sex would not have been associated: the data would not have enabled us to guess the values of one from the values of the other. Note that we could also determine whether associations exist by computing the geometric means, so it is possible to observe conflicting results. Two variables will be associated according to one kind of a mean, but not according to the other.

Consider, next, a third binary variable for which I have no data: owning a cigarette lighter. Assuming that mean FEV<sub>i</sub> is lower among members of the sample who own a cigarette lighter than among those who do not, FEV<sub>i</sub> and owning a cigarette lighter are also associated in this sample. It would be possible to guess the values of one variable from the values of the other.

From the perspective of the term association, nothing distinguishes qualitatively among the associations of FEV<sub>i</sub> with owning a cigarette lighter, with smoking status, and with sex group. All three meet the definition. But when I wrote about each association my mind (and probably yours, too) naturally wandered into the domain of causation, bringing in assumptions and expectations of what causes what. In the back of our minds
we were asking or doubting or guessing which of these associations reflect cause-and-effect and perhaps which is the cause and which is the effect. We will, of course, get to this interesting point eventually but for the time being let’s restrain our thoughts. When announcing the presence or absence of an association between two variables, we do not have to say a word about why they are associated in the sample at hand, or why they are not.

When the variables of interest are binary, their mathematical relation takes a different form. Using the same sample, Figure 6–1 displays the relation of smoking status (left panel) and sex (right panel) to a binary variable that indicates the presence or absence of chronic bronchitis, a condition characterized by chronically coughing phlegm.

Several methods may be used to describe the relation between two binary variables and we have already seen them in chapter 3. In each method we compare two numbers—two proportions or two odds (or two rates)—but without attributing causal meaning to these quantities. If the two numbers aren’t identical, then the two variables are associated in the data set.

Starting with the rows of the left table, we may compare the proportion of people with bronchitis in smokers (133/877) and former smokers (129/1,599), or the odds of bronchitis in smokers (133/744) and former smokers (129/1,470). Switching to the table’s columns, we may compare the proportion of smokers in people with bronchitis (133/262) and in people without (744/2,214), or we may compare the respective odds (133/129 and 744/1,470). Of course we can also compute and compare proportions and odds for “not having bronchitis” and for “being a former smoker.”

The good news, however, is that it doesn’t matter which pairs of numbers we choose to compare. If any pair is identical, all other pairs will be identical as well: no association between smoking status and chronic bronchitis. And if any compared pair is not identical, all other pairs will not be identical as well—the two variables will be associated in the sample.

Evidently, in this particular sample chronic bronchitis is associated with both smoking status and sex (Figure 6–1.) For example, if we are told that Smith has bronchitis and Jones does not, we may guess that Smith is a woman and Jones is a man since 68% of those with bronchitis are women (179/262) whereas only 42% of those without bronchitis are women (927/2,214). Likewise, if we are told that Smith is a smoker and Jones is not, we may guess that Smith has bronchitis and Jones does not because, for instance, the odds of
bronchitis among smokers (133/744) are higher than among former smokers (129/1,470).

**Measures of association**

Telling you now that measures of association resemble measures of effect shouldn’t be a big surprise. From a numerical standpoint, every measure of effect that was described in chapter 3, such as the mean difference, the odds ratio and the rate ratio, also carries the title “measure of association.” But there is interesting asymmetry between the two ideas. First, some measures of association never cross to the domain of effects—a correlation coefficient, for example. No reasoning enables us to claim that a correlation coefficient quantifies an effect or that a larger value of this measure implies a stronger effect. Second, most computed numbers that carry labels such as mean difference or proportion ratio or rate ratio do not qualify as measures of effect even though they are valid measures of association.

The reasons for the last statement will be fully revealed later but one simple reason can be found in the right table of Figure 6–1. After dividing the proportion of women among those with bronchitis (179/262) by the proportion of women among those without (927/2,214), we get a proportion ratio of 1.6, which is a valid measure of association. No one, however, will call this number a measure of the effect of chronic bronchitis on sex, because only in our wildest dreams might we find a mechanism by which chronic bronchitis could affect one’s sex. To sum up: a measure of association stops short of claiming anything about a causal relation behind the variables in question, only trying to inform us how well we can guess two values of one from two values of the other. To turn it into a measure of effect would always require supplemental assumptions and would often require some effort.

As you would expect, when two variables are not associated difference measures of association take the value of zero and ratio measures take the value of 1. The further we move from these “null” values in either direction, the stronger is the association and the more likely we are to guess correctly the values of one variable from the values of the other. In extreme situations of a perfect association our guess will be consistently correct. Figure 6–2 shows hypothetical examples of the two extreme situations: no association between sex and chronic bronchitis (left panel) and a perfect association between these variables (right panel).
Figure 6–2. Two hypothetical relations of sex and chronic bronchitis in a group of 2,476 people.

Since all row percentages (or odds) in the left table are identical, as are column percentages (or odds), every proportion ratio and every odds ratio is equal to 1 and every proportion difference and odds difference is equal to zero: sex and chronic bronchitis are not associated in that hypothetical data set. In contrast, ratio measures of association in the right table take the value of infinity or negative infinity and difference measures takes the value of 1 or -1—indicating an extreme (perfect) association. If we are told that Smith is a man and Jones is a woman, the right table informs us that Smith has bronchitis and Jones does not. Likewise, if we are told that Smith has bronchitis and Jones does not, we will correctly “guess” their sex.

A word about contingency tables

A 2 by 2 table, as shown in the previous section, or any similar R (rows) by C (columns) table is called a contingency table. As its name implies, a contingency table helps us to decide if the row variable and the column variable are associated—if the values of one are contingent on the values of the other. One way of quickly deciding the matter is to compare the distribution of counts in the bottom margin with the distribution of counts in each row (or alternatively, to compare counts in the right margin with counts in the columns.) Whenever the row and column variables are not associated, the distribution of one variable, found in one margin, is replicated in the strata of the other, implying no gain of knowledge about either variable from stratification on the values of the other.

Consider the left table of Figure 6–2, for instance. At the bottom of that table we see two counts: 262 people with bronchitis and 2,214 people without. Because in this example chronic bronchitis is not associated with sex, the ratio of these numbers is replicated among men (145 to 1,225 in the first row) and among women (117 to 989 in the second row.) Likewise, the ratio of 1,370 men to 1,106 women in the right margin is replicated among people with bronchitis (145 to 117 in the left column) and among those without (1,225 to 989 in the right column.) Of course, instead of comparing these ratios, we could also compare row percentages (145/1,370 = 117/1,106 = 262/2,476) or column percentages (145/262 = 1,225/2,214 = 1,370/2,476).

Magnitude and direction of an association

Measures of an association tell us how strong it is, with no interpretation of why it is strong. For example, an association that is described by an odds ratio of 10 (or the inverse, 0.1) is stronger than one that is described by an odds ratio of 2 (or the inverse, 0.5). Sometimes, however, different measures of one association can disagree on its strength, just as the two kinds of a mean—arithmetic mean and geometric mean—can disagree about the presence of an association between a binary variable and a continuous variable. Assuming that you have learned about linear regression, Figure 6–3 shows a familiar example.
The slope of a regression line is another measure of association, extending a single mean difference between two categories to many differences between many pairs along the X-axis. As you can tell from the figure, if we regress variable B on variable A we will find a small slope—a weak association according to this measure—whereas if we reverse the order and regress variable A on B, we will find a large slope. A strong association (in fact, a perfect one) is also evident from the correlation coefficient, which in this example is 1. (It's not a ratio measure, however.) Here, the correlation coefficient obviously tells the true story about our ability to make a qualitative guess about two values of one variable (which is higher) from two values of the other. That it cannot be called a measure of effect is a different matter.

Finally, we say that variables A and B are positively associated (or directly associated) when a larger value of one implies a larger value of the other, or alternatively, a smaller value of one implies a smaller value of the other, as in Figure 6–3. If a larger value of one implies a smaller value of the other, we will say that they are negatively (or inversely) associated. Evidently, these terms should be reserved for variables whose natural values are numbers, but it’s not uncommon to hear statements such as “smoking is positively associated with chronic bronchitis.” What the speaker has in mind is an implicit coding system for two binary variables, smoking status and chronic bronchitis status, such that smoking and having chronic bronchitis take a higher value than their complementary categories. Notice that our speaker has also abbreviated the variable names in that sentence. The word “status” should have followed both “smoking” and “chronic bronchitis” because an association signifies a mathematical relation between two variables, not between one value of one variable (smoking) and one value of another (chronic bronchitis.) To realize the last point, think how meaningless the following statement is: male sex is associated with systolic blood pressure of 120-millimeter mercury.
Causes of associations

Many associations are the consequence of some causal mechanism, although unfortunately it's often not the one we are interested in. To describe mechanisms that could generate an association between two variables, A and B, or could influence its magnitude, let’s use an intuitive tool called a causal diagram—sequences of arrows and variables that follow two rules: 1) An arrow always emanates from a causal variable and points to its effect (which resides in another variable). For example, $A \rightarrow B$ is a graphical code for the statement “A causes B.” 2) An arrow abbreviates both a direct effect and indirect effects. For example,

\[
\begin{array}{c}
A \\
\end{array} \quad \begin{array}{c}
B \\
\end{array}
\]

may stand for

\[
\begin{array}{c}
A \\
\end{array} \rightarrow \begin{array}{c}
B \\
\end{array}
\]

A causal diagram is not committed to a model of causation. An arrow may be interpreted deterministically (a value of one variable is a component cause of sufficient causes of the value of another) or indeterministically (a value of one variable contributes to probabilistic realization of the value of another.)

Two other terms will prove essential: a path between two variables is any sequence of arrows that connect them, regardless of the direction of the arrowheads. For example, $A \rightarrow B \rightarrow C$, $A \leftarrow C \rightarrow B$, and $A \rightarrow C \leftarrow B$ are all paths between A and B. The second key term is crude association, or marginal association in statistical jargon. Crude association is simply the observed association between A and B in a dataset, such as a mean difference of 0.4 liter, an odds ratio of 1.4, or a rate ratio of 2. The adjectives “crude” and “marginal” serve to distinguish these numbers from any modified numbers we might introduce later in the interest of causal inquiry. Now, if the term “crude” sound reasonable to you yet you are wondering why “marginal”, just wait a little longer. I will explain it later.

Four basic mechanisms can induce an association between A and B in a data set. Three are shown in Figure 6–4 and the fourth will be presented later.

a. \[
\begin{array}{c}
A \\
\end{array} \rightarrow \begin{array}{c}
B \\
\end{array}
\]

b. \[
\begin{array}{c}
B \\
\end{array} \rightarrow \begin{array}{c}
A \\
\end{array}
\]

c. \[
\begin{array}{c}
C \\
\end{array} \rightarrow \begin{array}{c}
A \\
\end{array} \rightarrow \begin{array}{c}
B \\
\end{array}
\]

The first two diagrams (panels a and b) are probably clear, intuitively. If the values of one variable are causal assignments that affect the value of another variable, we should be able to guess the values of one from those of the other, regardless of which is the cause and which is the effect. Smoking status and having a cigarette lighter will be associated in many samples if smoking status affects the decision to buy a cigarette lighter or vice versa. Keep in mind, however, that it is possible to find a sample in which variable A is not associated with variable B even though it is a cause of B. “A causes B” is not a sufficient condition for “A is associated with B.”

The third mechanism (panel c), shaped like inverted V, illustrates a key idea that runs throughout the book: a common cause of two variables, A and B, induces an association between them even when the two variables do not make up a cause-and-effect. For example, if body weight (variable C) affects both blood pressure (variable A) and cholesterol concentration (variable B), the last two variables should be associated in many samples even if neither is a cause of the other. Another example: if sex group (C) affects both diabetes status (A) and FEV₁ (B), diabetes status and FEV₁ will be associated in many samples even if no arrow connects them.

Furthermore, if A does cause B, their shared cause C will play a role in determining the magnitude of their crude association. The crude association between A and B in a data set will reflect three effects: the effect of A on B and the two effects of C: on A and on B. And in theory, the net result might even be “null”—no crude association even when A causes B! How much effect C will have on the crude association between A and B (and what kind of effect) will depend on the strength and direction of its associations with A and B and on its distribution in the sample.

Why does C, a shared cause of A and B, induce or affect a crude association between A and B?

Because association has to do with being able to make guesses about one variable from another, the explanation lies in the idea of sequential guesses. If C → A, we can guess the values of C from the values of A. And if C → B, we can guess the values of B from the values of C. If we can guess C from A and B from C, we can also guess B from A or vice versa, which means that A and B are associated.

If you had taken a course in epidemiology or in research methods, you probably recognized panel c of Figure 6–4 as the phenomenon called confounding and variable C as a confounder. Confounding is one kind of bias, an idea that will be fully explained in chapter 8, but the essence is this. If reality is depicted in panel c and we estimate the effect of A on B by computing a measure of their crude association, the number we compute (say, an odds ratio of 2) is the product of a biased process—the product of confounding. It may be called a valid measure of association in the sample at hand but not a valid estimate of the effect of A on B.

We will elaborate on confounders later on but I would like to make one point now. When showing a confounder in a causal diagram, some writers use a bidirectional arrow between C and A to indicate “an association” without committing to how that association was generated. This practice is unfortunately confusing because the bidirectional arrow should never represent an arrow from A to C: if A causes C, variable C is not a confounder. In such a situation the sequence is A → C → B and we are back in panel a of
Figure 6–4 (recalling that according to our second convention $A \rightarrow C \rightarrow B$ may be abbreviated as $A \rightarrow B$.) Notice that panel a does not specify how A causes B, and when we estimate its effect on B, we are usually interested in the overall effect by all pathways, direct and indirect.

Association-to-causation: first steps

We are still living in an imaginary world where there are only three variables—A, B, and C—and only three mechanisms that induce or affect a crude association between A and B (Figure 6–4.) Suppose we are interested in estimating the effect of A on B, how do we go about getting the "right number"? How do we turn a measure of association between A and B into a measure of effect: into an unbiased estimator of a causal parameter?

Our first problem may be to rule out that B causes A (panel b) rather than A causes B (panel a)—that is, that we got the causal direction right. In some examples, we may argue that one of the two arrows is nonsensical (chronic bronchitis is not a cause of sex group) but in other examples either direction of causation may seem possible. In such situations, only the chronology of the data may help to issue the verdict. If we know, for example, that a cigarette lighter was always purchased after smoking status had been determined, any association between the two cannot be due to the effect of purchasing a cigarette lighter on smoking. Causes should precede their effects, not follow them. The same reasoning, in fact, applies to the previous example of chronic bronchitis and sex: one’s sex is determined before one gets sick.

Is it possible for causation to be circular and to produce a vicious cycle whereby A causes B and B causes A? It may well be and it is even possible to come up with theoretical examples. Nonetheless, the "second A" in the chain $A \rightarrow B \rightarrow A$ is never the same variable as the "first A". A variable that changes over time should be represented as several variables, each uniquely defined in the dimension of time, for example: $A_1 \rightarrow B \rightarrow A_2$. To the extent that time flows in one direction, self causation does not exist.

Our second key problem is variable C in panel c of Figure 6–4, or more specifically the two arrows that emanate from it. As we have already realized, these arrows allow for a sequence of guesses from variable A to variable B and thereby affect their crude association. They are the reason why measures of that crude association don’t take the "null" value when A does not cause B, nor estimate its effect on B when it does. So our task is to find a way to block the flow of guesses along the path $A \leftrightarrow C \rightarrow B$. If that path were somehow disrupted, the path $A \rightarrow B$ alone (if it exists) would have determined the magnitude of the association between A and B. And if we got the causal direction right, we would have computed a measure of association that could have been called a measure of effect—an unbiased estimator of a causal parameter.

Depending on properties of variable A and variable C, there are five main ways to break the path $A \leftrightarrow C \rightarrow B$, all of which block the left arrow ($A \leftrightarrow C$):

1) If A is a categorical variable whose values can be assigned at random, we can disconnect C from A by randomizing the causal assignments of A. If we can’t randomize or don’t want to randomize, we have the following options:

2) Select a sample in which each causal assignment of A will have the same distribution of the values of C—a procedure called matching variable A on variable C.
3) If C is a categorical variable, stratify the sample on C and compute a measure of association between A and B in each stratum of C—a procedure called *conditioning* the association on C.

4) Regardless of whether C is categorical or continuous, include it in a regression model of B on A, a method that also involves conditioning on C.

5) Use a two-step regression procedure called "inverse probability weighting."

I will explain below how randomization, matching, and conditioning through stratification serve our purpose and defer the discussion of regression until chapter 8. Inverse probability weighting will be described in chapter 21.

**Randomization: a causal diagram perspective**

In chapter 4, we saw two viewpoints of the purpose of randomization: deterministic and indeterministic. A causal diagram provides a third viewpoint and perhaps another look at the two models of causation.

If the causal assignments of A were determined at random, we may claim that no \( C \rightarrow A \) exists in the sample because the only cause of A was the flip of a coin or some other randomization procedure. The last statement is not completely true, though. As you can imagine, in a “small enough” randomized trial many C-like variables will be distributed differently within the categories of A, which means that they will be associated with A. Only if the trial is “large enough”, can we hope (but not know) that the distribution of every imaginable variable is nearly identical in every category of A, so there is little to no flow of guesses from A to B via C (Figure 6−5.)

![Randomization procedure](image)

Figure 6−5. A causal diagram showing the lack of association between C and A, given randomization and a large sample

Why does the samples size matter? How exactly does the addition of randomized units eventually eliminate any association between C and A? These are interesting questions for which I have no simple answer. It is like asking how repeated throwing of a fair dice eventually produces a sequence in which one-sixth of the throws shows “3-up.” Ten throws may not, but ten thousand will. The best I can offer is the idea that randomization is a force too, like causal propensities: a force that operates in the dimension of counts and depends on counts for its full expression. As the number of randomized people increases, the left arrow gradually takes over until at some unknown number—a large enough trial—any association of A with any C is practically negligible.

In the language of indeterminism, a trial is large enough when the background propensity to cause B due to any possible variable is similar across the randomized assignments of A. For the indeterminist, such a trial set the stage for assuming that a
measure of the crude association is also a measure of effect. For the determinist the story is different, however. A large enough randomized trial is one whose intrinsic causal parameter (say, \( P_{\text{causative}} - P_{\text{preventive}} \)) is estimated well enough because the four deterministic classes are distributed similarly in the two treatment groups (chapter 4). And this condition depends on a large enough sample. The deterministic difficulty with a small, randomized trial has to do with imprecise estimation of hypothetical, deterministic outcomes—not with confounders.

**Matching on C**

If an association or its absence is a property of a sample, why not exploit samples in the interest of causal inquiry? In fact, we already did. As we have just seen, randomizing enough times to the causal assignments of A will yield a sample in which no C-like variable will be associated with A, thereby breaking the flow of guesses along the path \( A \leftarrow C \rightarrow B \). Matching on C serves exactly the same purpose: to construct a sample in which C and A will not be associated even though C is a cause of A (Figure 6–6.)

![Figure 6–6. A causal diagram showing the lack of association between C and A, given matching on C](image)

It is easiest to explain the idea of matching when both C and A are binary variables. Suppose A is smoking status (smoker or former smoker) and C is sex group, a variable that may affect the decision to quit smoking. Matching smoking status on sex means selecting a sample in which the distribution of sex is identical in smokers and former smokers. Since sex is a binary variable, the idea of identical distributions may be expressed in any of several ways: identical proportion of men (or women) in the two groups, identical ratio of men-to-women, or identical ratio of women-to-men. Evidently, matching on sex will generate a sample in which sex and smoking status will not be associated even if the former causes the latter.

It is simple to extend the idea of matching to any C-variable that takes a finite number of values, for example, race group. But when C is continuous or when A is continuous, exact matching becomes technically difficult to achieve, and sometimes impossible. It is also difficult to assemble a sample in which A is matched on several C-like variables. In short, matching would have been a great idea if causal reality involved only categorical A variables and only a handful of C variables, all of which were categorical, too.

Matching shows some resemblance to randomization but the differences are far more important than any similarity. First, the left arrow in the randomization diagram (Figure 6–5) represents a true cause of A (a researcher assigning the values of A) whereas the left
arrow in the matching diagram (Figure 6–6) represents the "force" of sample selection. To emphasize the difference, I drew a dashed line for the matching arrow instead of a solid line. Second, in the case of randomization, the X mark on the arrow C→A indicates that the effect has been eliminated, whereas in the case of matching the X mark only indicates that an association was abolished by selecting a special kind of a sample. Indeed, C remains a cause of A (in a deterministic sense or in an indeterministic sense.) Third, randomization thrives on a large sample whereas matching does not. Nowhere in the rationale of matching did we need to argue that matching more people was essential to blocking the association of C with A.

Before turning to the next method for disrupting the path A←C→B, I should mention a common confusion between the rationale of matching as described above and the rationale of matching in a case-control study. Matching controls to cases on a confounder like C serves a very different purpose and by itself induces bias, not eliminates confounding. More on this matter in chapter X.

Conditioning on C

One extreme method of matching variable A on variable C is to construct a sample in which C takes the same value for every value of A. For example, if C is sex group and A is smoking status (smoker or former smoker), we can match smoking status on sex group by selecting a sample in which both smokers and former smokers are men. The definition of matching is obviously met in this sample: sex is “distributed” identically in smokers and former smokers—men only. Alternatively, we can match smoking status on sex by selecting a sample of women.

But why work hard on selecting a matched sample of either kind when both samples are waiting for us within any study that included both men and women? To match smoking status on sex, all that we have to do is to stratify on sex—to split the sample into two, separating the men from the women. In each sex group, smoking status and sex will be perfectly matched and no guesses could flow along the path A←C→B (Figure 6–7.)
Figure 6–7. Causal diagrams showing de-confounding from C (sex group) by stratification.

It would have been enough to block the arrow from C to A, as in planned matching or randomization, but notice that stratification has also abolished the association of C with B. Again, similar to matching, an X on an arrow indicates the blocking of an association (by selection of sub-samples) and not interference with causal reality.

Assuming a three-variable world, stratification on sex has generated two samples in which some measures of the crude association between A and B qualify as measures of the effect of A on B. But a new problem has arisen. We now have two estimates of that effect—one in men and another in women—and unless we are extremely lucky, the two estimates will differ: maybe a little, but maybe a lot. If they turn out “substantially different” we will find ourselves back in chapter 5, inadvertently looking at effect-modification by sex—even though we had no intention to look for it! Recall that we were only trying to eliminate confounding by sex.

At this junction we have to decide whether the difference between the sex-specific estimates is both substantial and credible. If both adjectives hold, we should discard our original goal of de-confounding and present heterogeneity of effect by sex: the causal parameter takes one value for men and another for women. Otherwise, we’ll assume that both numbers estimate a single value of the causal parameter, and if so—our best estimate of that value would be some average of the two. What difference is substantial and how we decide on credibility will be discussed in other chapters, but as I have said before, don’t expect a recipe for deciding on matters of credibility in science. In chapter 8, for instance, you will see that a widely used recipe is loaded with flaws.

To illustrate the steps just described, we’ll de-confound a crude association between smoking status (A) and FEV₁ (B) from confounding by sex (C). Although I prefer to estimate the effect on FEV₁ by a geometric mean ratio, let’s use the arithmetic mean difference, which is simpler.
The crude association can be calculated from the data in Table 6–1: a mean difference of 0.42 liter (3.19–2.77). Figure 6–8 shows the process of de-confounding by stratification on sex.

<table>
<thead>
<tr>
<th></th>
<th>Smoker (N=877)</th>
<th>Former smoker (N=1,599)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.77 liter</td>
<td>3.19 liter</td>
<td></td>
</tr>
</tbody>
</table>

**Mean difference**

<table>
<thead>
<tr>
<th></th>
<th>Smoker (N=400)</th>
<th>Former smoker (N=970)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.28 liter</td>
<td>3.55 liter</td>
<td></td>
</tr>
</tbody>
</table>

**Mean difference**

<table>
<thead>
<tr>
<th></th>
<th>Smoker (N=477)</th>
<th>Former smoker (N=629)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.34 liter</td>
<td>2.63 liter</td>
<td></td>
</tr>
</tbody>
</table>

**Mean difference**

**De-confounded mean difference**

0.28 liter

Figure 6–8. An example of de-confounding by stratification on sex followed by computing an average mean difference over the two sex groups.

Luckily, the estimated effect of smoking status on FEV$_1$ turned out fairly similar in men and in women: 0.27 liter in men and 0.29 liter in women. Assuming we declare no effect modification by sex (on an additive scale), we can go on to compute a de-confounded estimate by taking their average (Figure 6–8.) A simple average of 0.28 liter will do at the moment, but in later chapters you will find a better-rationalized method called a weighted average. To sum up: if reality is fully captured in panel c of Figure 6–4, 0.28 liter is the estimated effect on FEV$_1$ of the causal contrast between smoking and former smoking. Keep in mind, however, earlier discussions about the distinction between a theoretical causal variable and the variable at hand (chapter 2), about the use of an additive scale to measure effects (chapter 3), and about the possibility of effect-modification by other variables (chapters 4 and 5.) No matter how hard we try to simplify a hypothetical world, hidden complexity always seems to remain.
Computing a measure of association between A and B in each stratum of C is called conditioning the association on C. We allow the measure of interest to depend on C and possibly vary by the "conditions" (categories) that C defines, such as "male sex" and "female sex". Conditioning, therefore, plays a dual role in causal inquiry: On one hand it is a method to search for effect-modification by C, and on the other hand, it is the first step in de-confounding from C. Again, if we try to de-confound and incidentally observe effect-modification, we forget about confounding, label C an effect-modifier, and present stratum-specific estimates. Otherwise, we conclude by collapsing the stratum-specific estimates into one number—their average. This reasoning explains why a variable C either modifies the effect of A on B or confounds it—but not both. That we may have hard time deciding what it does is another matter.

Another example of conditioning followed by collapsing is shown in Figure 6–9. Here, I conditioned the association between smoking status and chronic bronchitis status on sex group, and used the odds ratio for chronic bronchitis as a measure of association. Fortunately, again, the sex-specific estimates turned out similar (1.84 and 1.75) so we shouldn't worry too much about effect modification. After taking a simple average, we get an odds ratio of 1.79, an estimate of a causal parameter in our three-variable world.

Figure 6–9. An example of de-confounding by stratification on sex followed by computing an average over the two sex groups

A few sections earlier, I mentioned that crude association is also called marginal association, without explaining why. Equipped with Figure 6–9 and the term conditional
association, I can now clarify the term. First, marginal is the opposite of conditional; it is that one association between A and B that is not conditioned on any other variable. Second, the adjective originated in contingency tables which relate one margin (the distribution of one variable) to another margin (the distribution of the other) through the Table’s cells. In that sense, all three tables in Figure 6–9 display a "marginal" association but the bottom two already claimed a more informative adjective: conditional. You may also view the crude association as a contingency table that is formed by summing stratified tables. No matter on which variable we stratify, we will always get back to the margins of the crude association by summing the stratum-specific marginal counts—hence the synonym "marginal association".

**Conditioning on two confounders or more**

No one knows, of course, all the details of any causal web, but many times we are able to draw more than a single confounding path (A⇐C⇒B) and to add names of intermediary variables on the paths C⇒A and C⇒B. Continuing with the example of smoking status and FEV₁, Figure 6—10 shows a causal diagram that contains two confounders, age and sex; two intermediary variables on the path from smoking status to FEV₁ via age; and one intermediary variable on the path via sex. (To simplify, I displayed age, smoking drive and physical activity as binary variables.) This diagram is undoubtedly a naive representation of reality, but it does contain some grains of truth such as the effect of asthma on smoking status or the effect of age on physical activity. (It also contains some falsehood such as no arrow between asthma and FEV₁, which will be added later.) At any rate, let’s assume that our diagram depicts not only truth but the whole truth.

![Causal Diagram](image-url)

Figure 6–10. A causal diagram showing two confounders, sex and age, along with three intermediary variables
To de-confound a measure of association between smoking status and \( \text{FEV}_1 \), we now have to block two confounding paths. But we are not obligated to condition on age and sex. Conditioning on either smoking drive or physical activity will also block the flow of guesses from smoking status to \( \text{FEV}_1 \) via age, whereas conditioning on asthma will also block sequential guesses via sex. In general, a crude association between A and B that is induced or affected by a path like

\[ A \leftarrow U_n \leftarrow \ldots \leftarrow U_2 \leftarrow U_1 \leftarrow C \rightarrow V_1 \rightarrow V_2 \rightarrow \ldots \rightarrow V_n \rightarrow B \]

can be blocked by conditioning on any \( U \) or any \( V \). We don’t have to condition on the confounder \( C \).

We seem to have found a general method to de-confound a crude association, regardless of the number of confounders. When seeking to estimate the effect of A on B, draw a causal diagram that depicts all conjectural confounding paths that connect A to B, with as much detail as possible, and then simultaneously stratify the association of A with B on all confounders or their substitutes on a confounding path. Under the assumptions displayed in the diagram, the method of stratification and collapsing will produce a de-confounded estimator of the causal parameter.

As you might imagine, life isn’t so simple. Besides the possibility of encountering effect modification, other difficulties may be waiting for us along the way. To stratify on sex and age alone (even if binary), would mean to divide the sample into four—young men, young women, old men, old women—and to have one fourth of the sample in some strata, or even less. With several confounders at hand, we may quickly find ourselves facing the problem of sparse data. For example, smoking status and \( \text{FEV}_1 \) may have no association in the stratum of young, wealthy, well educated men with genotype \( Z \)—other than due to the effect of smoking status—but only ten people may belong to that stratum. What credibility will the mean difference in \( \text{FEV}_1 \) between smokers and former smokers carry in that stratum? What contribution should it have on the average mean difference over all strata?

Sometimes we may be able to solve the problem of sparse data by a different method of conditioning (regression) but there is another fundamental problem. Confounding paths might not be nicely separated as in Figure 6-10. The variables on one path may be causes of variables on another and in that case, we should take extra care when choosing variables for conditioning. As you will see in the next sections, some choices of conditioning could make things worse: create confounding paths that did not exist, produce bias rather than eliminate it.

**Colliders**

When the arrows that emanate from two variables, A and B, point to variable C (that is, \( A \rightarrow C \leftarrow B \)), we call C a collider on the path from A to B because the two arrowheads collide at C. Unlike a confounding path, however, a path that contains a collider does not induce an association between A and B: there is no flow of guesses across the collider. To understand intuitively why, let’s divert to a simple example of two causes of car accidents: brake condition and street condition (Figure 6-11). Evidently, in this figure the variable "accident" is a collider.

Assume that we have a data set for 1,000 cars that contains three variables: the condition of the car’s brakes, the condition of the streets in the owner’s town, and
whether the car was involved in an accident in the owner’s town. Is brake condition associated with street condition in this data set? Can we guess one from the other?

![Diagram of Brake condition (bad, good) → Accident (yes, no) → Street condition (bad, good)]

Figure 6–11. A causal diagram showing a path that contains a collider.

If we are told that the brakes of car A were bad and the brakes of car B were good, we can guess that the former was involved in an accident and the latter was not. But if we now guess (based on the brake condition) that car A was involved in an accident and car B was not, can we also guess the condition of the streets in each owner’s town? Of course not. Knowledge of the car brake condition does not deliver knowledge of the street condition and vice versa. This logical result resonates with our intuition that an effect does not induce an association between its causes.

Before developing the idea of a collider further, it may be helpful to contrast Figure 6–11 with Figure 6–12 which shows a common cause of brake condition and street condition: the wealth of the town.

![Diagram of Brake condition (bad, good) → Town status (poor, wealthy) → Street condition (bad, good)]

Figure 6–12. A causal diagram showing a common cause of brake condition and street condition.

This figure is not new. It is, again, the third mechanism for inducing an association between two variables (Figure 6–4, panel c) according to the following logic: If we are told that the brakes of car A were bad and the brakes of car B were good, we can guess
that the former came from a poor town and the latter, from a wealthy town. Since the wealth of a town affects street condition, we can also guess that A was driven on poorly maintained streets and B was driven on well-maintained streets. In other words, we can guess the street condition from the brake condition and vice versa. To sum up: a common cause of two variables induces an association between them but a common effect of them does not—unless we condition on that common effect…

**Conditioning on a collider**

Although a collider does not induce an association between the colliding variables, conditioning on it will create such an association. To understand intuitively why, consider again the example of brake condition, street condition, and accidents (Figure 6–13).

![Diagram of Brake condition, Street condition, and Accident](image)

**Figure 6–13. An example of conditioning on a common effect (a collider).**
By conditioning on the collider “accident”, we form two strata: one contains data on cars that were involved in an accident and the other contains data on cars that were not. The question at hand is whether brake condition and street condition are associated within each stratum: can we guess the values of one variable from the values of the other? The answer is, yes—at least partially.

Let’s start with the stratum on the left of the figure (cars involved in accidents.) If we are told that the brakes of car A were bad and those of car B were good, we cannot make any guess about the condition of the street in which A was driven. With bad brakes followed by an unfortunate accident, the street condition could have been either good or bad. But the situation is different for car B. Being told that its brakes were good and knowing that it was involved in an accident (from stratification) allow us to guess that the street condition was bad. A flow of guesses from the condition of the car brakes to the condition of the street has become partially possible—the two variables are associated within this stratum. One similar guess is also possible within the stratum of cars that were not involved in an accident: if we are told that the brakes of a car were good, we can guess that the street condition was also good.

This example is, of course, qualitative but its principles equally apply to measures of association. A path from A to B that contains at least one collider does not affect their crude association. If we condition, however, on all colliders in that path, compute stratum-specific measures of association, and collapse these estimates into one average, we will alter the crude association between A and B, just like de-confounding. With one important difference, however: instead of de-confounding a measure of association, we will generate collider-dependent confounding. Conditioning on a collider is therefore a fourth mechanism for inducing or affecting an association.

It is helpful, perhaps, to think of a collider as the opposite of a confounder. A confounder induces a path of association between A and B and that association will be abolished by conditioning on the confounder. In contrast, a collider blocks a path of potential association between A and B and that association will be established by conditioning on the collider.

Back to our example of smoking status and FEV, (Figure 6−14). If you compare this figure to Figure 6−10, you will see that I added two arrows: 1) an arrow that emanates from asthma and points to FEV, which turned asthma into a confounder. 2) an arrow that emanates from sex and points to physical activity. This arrow has turned physical activity into a collider on any path that includes the segment sex → physical activity ← age, though this variable continues to contribute to confounding paths. (A collider on one path can contribute to confounding on another.) In the new diagram, age, sex, and asthma are confounders, generating an association along two sides of the triangles they form with smoking status and FEV, Sex also generates confounding via asthma and physical activity.

With the new diagram at hand, let’s see what might go wrong when we try to de-confound a measure of the crude association between smoking status and FEV,
As we know, a confounding path will be blocked by conditioning on any intermediary variable, so it may be tempting to condition on physical activity because it is located at the junction of two paths: age-induced and sex-induced. For complete de-confounding, however, we still have to block two other paths: smoking status $\leftarrow$ asthma $\rightarrow$ FEV$_1$ and smoking status $\leftarrow$ asthma $\leftarrow$ sex $\rightarrow$ FEV$_1$. We could condition on asthma and sex, the sources of confounding, but conditioning on asthma alone seems to be just as good because this variable resides on both paths. At the end, we might decide to de-confound by conditioning on physical activity and asthma.

Well, that will be a serious mistake. By conditioning on physical activity—a collider—we will induce an association between the colliding variables age and sex, and thereby open a confounding path along the dotted line in Figure 6–15. And conditioning on asthma alone, rather than on asthma and sex, will leave it open! As a result, the so-called de-confounded measure of association between smoking status and FEV$_1$ will be confounded.

The lesson from this example is this. If you decide to condition on a collider, you should draw a line that connects the colliding variables (as well as their causes, which also collide, indirectly). Then, check what new confounding paths may have opened and properly block them as well. Of course, in this example we could have avoided the trouble by simply conditioning on age, asthma, and sex.
Collider-related thoughts

While confounding and confounders have been recognized for a long time, the idea of a collider is fairly new. It was developed in the 1980s for applications in computer science and artificial intelligence, found its way to epidemiology in the late 1990s, and goes deeper than what I presented above. For example, after setting the diagram and committing to the variables and arrows, you can follow a graphical algorithm for deciding whether a set of variables would suffice for de-confounding and even identify the smallest possible set. If interested, you will find the details in a book and several articles (Suggested Readings.)

Once you understand the collider idea, however, the method of choosing variables and checking for errors should work, as it worked in our example. In addition, you may be able to avoid common pitfalls by not conditioning on colliders, if possible; by conditioning on causes of B (the effect of interest), rather than on causes of A alone (the exposure); by preferring confounders to intermediary variables on their paths; by conditioning liberally (under sample-size constraints); and by not conditioning on effects of A or B. In fact, much of what I have just suggested agrees with common practice which predates colliders' theory and causal diagrams.

One classical pitfall is shown in Figure 6–16. After conditioning on the center variable C, the variables $U_1$ and $U_2$ become associated and we open the path $A \leftarrow U_1 \rightarrow U_2 \rightarrow B$. To block this new path and fully de-confound a measure of the crude association between A and B, we also have to condition on $U_1$ or $U_2$.

But there is a deeper message in this Figure. We know the pitfall of conditioning on C alone, and its remedy, only because we knew the names of $U_1$ and $U_2$ and drew arrows to
connect them to C, A, and B. If we didn’t know about these variables, we would have conditioned on C alone and then compute a so-called de-confounded estimate of the causal parameter that would have been biased—just like the measure of the crude association.

![Figure 6-16. A causal diagram in which conditioning on a collider (C) generates an association between U₁ and U₂ and, therefore, opens a new confounding path.](image)

Which brings us back to the following truism that is nonetheless worth repeating: Nothing can assure us that a measure of association is a measure of effect—neither a non-randomized study nor the largest randomized trial. It is always possible that we conditioned on a variable like C above and unknowingly created a new confounding path, or that we did not condition on all confounders, or that unknown confounders were associated with the exposure despite randomization and a huge sample. So—you may be asking—what distinguishes scientific causal inquiry from witchcraft, or how do we single one of several methods as a better science? The answer is this: the reasoning for what we do; not what we do. A subtle but key distinction that is easy to miss.

Consider the following example. Suppose that to estimate the effect of A on B, we can select variables for conditioning by one of three methods: 1) by expressing all of our causal assumptions in a diagram, including colliders; 2) by ignoring colliders and just choose every C that forms a triangle with A and B, or its surrogate; 3) by a regression method called forward stepwise selection (ignoring the details). No matter which method we choose, we cannot justify our choice by claiming that it will get us closer to the truth than other methods—it might not. But we can rationalize our choice by explaining why other choices are worse. For example: If we accept the theory of causal diagrams, it would be silly (irrational) to select variables for conditioning by stepwise regression. And if we accept the theory of colliders, it would be silly to pretend that they don’t exist. Of course, someone could choose stepwise regression and come up with counter-arguments for why it would be silly to draw a causal diagram. The critical debate is science, too—at its best.