affect Y.[20] *If both of the associational criteria (U_1) and (U_2) of Definition 6.2.2 are violated, then (X, Y) are not stably unconfounded given A_Z.*

Proof
Whenever X and Y are stably unconfounded, Theorem 6.4.3 rules out the existence of a common ancestor of X and Y in the diagram associated with the underlying model. The absence of a common ancestor, in turn, implies the satisfaction of either (U_1) or (U_2) whenever Z satisfies A_Z. This is a consequence of the d-separation rule (Section 1.2.3) for reading the conditional independence relationships entailed by a diagram.[21]
□

Theorem 6.4.4 implies that the traditional associational criteria (U_1) and (U_2) could be used in a simple operational test for stable no-confounding, a test that does not require us to know the causal structure of the variables in the domain or even to enumerate the set of relevant variables. Finding just *any* variable Z that satisfies A_Z and violates (U_1) and (U_2) permits us to disqualify (X, Y) as stably unconfounded (though (X, Y) may be incidentally unconfounded in the particular experimental conditions prevailing in the study).

Theorem 6.4.4 communicates a formal connection between statistical associations and confounding that is not based on the closed-world assumption.[22] It is remarkable that the connection can be formed under such weak set of added assumptions: the qualitative assumption that a variable may have influence on Y and is not affected by X suffices to produce a necessary statistical test for stable no-confounding.

6.5 Confounding, Collapsibility, and Exchangeability

6.5.1 Confounding and Collapsibility

Theorem 6.4.4 also establishes a formal connection between confounding and "collapsibility"—a criterion under which a measure of association remains invariant to the omission of certain variables.

Definition 6.5.1 (Collapsibility)
Let $g[P(x, y)]$ be any functional[23] *that measures the association between Y and X in the joint distribution $P(x, y)$. We say that g is* collapsible *on a variable Z if*

$$E_z g[P(x, y|z)] = g[P(x, y)].$$

[20] By "possibly affecting Y" we mean: A_Z does not contain the assumption that Z does not affect Y. In other words, the diagram associated with M must contain a directed path from Z to Y.

[21] It also follows from Theorem 7(a) in Robins (1997).

[22] I am not aware of another such connection in the literature.

[23] A *functional* is an assignment of a real number to any function from a given set of functions. For example, the mean $E(x) = \sum_x xP(x)$ is a functional, since it assigns a real number $E(X)$ to each probability function $P(x)$.

It is not hard to show that if g stands for any linear functional of $P(y|x)$—for example, the risk difference $P(y|x_1) - P(y|x_2)$—then collapsibility holds whenever Z is either unassociated with X or unassociated with Y given X. Thus, any violation of collapsibility implies violation of the two statistical criteria of Definition 6.2.2, and that is probably why many believed noncollapsibility to be intimately connected with confounding. However, the examples in this chapter demonstrate that violation of these two conditions is neither sufficient nor necessary for confounding. Thus, noncollapsibility and confounding are in general two distinct notions; neither implies the other.

Some authors tend to believe that this distinction is a peculiar property of nonlinear effect measures g, such as the odds or likelihood ratios, and that "when the effect measure is an expectation over population units, confounding and noncollapsibility are algebraically equivalent" (Greenland 1998, p. 906). This chapter shows that confounding and noncollapsibility need not correspond even in linear functionals. For example, the effect measure $P(y|x_1) - P(y|x_2)$ (the risk difference) is not collapsible over Z in Figure 6.3 (for almost every parameterization of the graph) and yet the effect measure is unconfounded (for every parameterization).

The logical connection between confounding and collapsibility is formed through the notion of *stable no-confounding*, as formulated in Definition 6.4.2 and Theorem 6.4.4. Because, any violation of collapsibility means violation of (U_1) and (U_2) in Definition 6.2.2, it also implies (by Theorem 6.4.4) violation of stable unbiasedness (or stable no-confounding). Thus we can state the following corollary.

Corollary 6.5.2 (Stable No-Confounding Implies Collapsibility)
Let Z be any variable that is not affected by X and that may possibly affect Y. Let $g[P(x, y)]$ be any linear functional that measures the association between X and Y. If g is not collapsible on Z, then X and Y are not stably unconfounded.

This corollary provides a rationale for the widespread practice of testing confoundedness by the change-in-parameter method, that is, labeling a variable Z a confounder whenever the "crude" measure of association, $g[P(x, y)]$, is not equal to the Z-specific measures of association averaged over the levels of Z (Breslow and Day 1980; Kleinbaum et al. 1982; Yanagawa 1984; Grayson 1987). Theorem 6.4.4 suggests that the intuitions responsible for this practice were shaped by a quest for a stable condition of no-confounding, not merely an incidental one. Moreover, condition A_Z in Theorem 6.4.4 justifies a requirement made by some authors that a confounder must be a causal determinant of, and not merely associated with, the outcome variable Y.

6.5.2 Confounding versus Confounders

The focus of our discussion in this chapter has been the phenomenon of confounding, which we equated with that of effect bias (Definition 6.2.1). Much of the literature on this topic has been concerned with the presence or absence of *confounders*, presuming that some variables possess the capacity to confound and some do not. This notion may be misleading if interpreted literally, and caution should be exercised before we label a variable as a confounder.

Rothman and Greenland (1998, p. 120), for example, offer this definition: "The extraneous factors responsible for difference in disease frequency between the exposed

and unexposed are called *confounders*" they go on to state that: "In general, a confounder must be associated with both the exposure under study and the disease under study to be confounding" (p. 121). Rothman and Greenland qualify their statement with "In general," and for good reason: We have seen in (in the two-coin example of Section 6.3.1) that each individual variable in a problem can be *unassociated* with both the exposure (X) and the disease (Y) under study and still the effect of X on Y remains confounded. A similar situation can also be seen in the linear model depicted in Figure 6.5. Although Z is clearly a confounder for the effect of X on Y and must

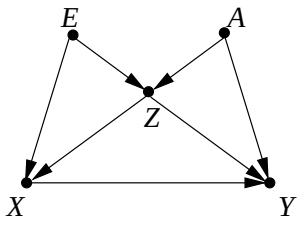

Figure 6.5: Z may be unassociated with Y and still be a confounder.

therefore be controlled, the association between Z and Y may actually vanish (at each level of X) and the association between Z and X may vanish as well. This can occur if the indirect association mediated by the path $Z \longleftarrow A \longrightarrow Y$ happens to cancel the direct association carried by the arrow $Z \longrightarrow Y$. This cancellation does not imply the absence of confounding, because the path $X \longleftarrow E \longrightarrow Z \longrightarrow Y$ is unblocked while $X \longleftarrow E \longrightarrow Z \longleftarrow A \longrightarrow Y$ is blocked. Thus, Z is a confounder that is associated neither with the exposure (X) nor with the disease (Y).

The intuition behind Rothman and Greenland's statement just quoted can be explicated formally through the notion of stability: a variable that is *stably* unassociated with either X or Y can safely be excluded from adjustment. Alternatively, Rothman and Greenland's statement can be supported (without invoking stability) by using the notion of *minimal sufficient set* (Section 3.3)—a minimal set of variables for which adjustment will remove confounding bias. It can be shown (see the end of this section) that each such sufficient set S, taken as a unit, must indeed be associated with X and be conditionally associated with Y, given X. Thus, Rothman and Greenland's condition is valid for minimal sufficient sets but not for the individual variables in a problem.

The practical ramifications of this condition are as follows. If we are given a set S of variables that is claimed to be minimally sufficient (for removing bias by adjustment), then that claim can be given a necessary statistical test: S as a compound variable must be associated both with X and with Y (given X). In Figure 6.5, for example, the minimal sufficient sets are $S_1 = \{A, Z\}$ and $S_2 = \{E, Z\}$; both must satisfy the condition stated.

Note that, although this test can be used for screening sets claimed to be minimally sufficient, it does not constitute a test for detecting confounding. Even if we find a set S in a problem that is associated with both X and Y, we are still unable to conclude that X and Y are confounded. Our finding merely qualifies S as a candidate for minimally sufficient status in case confounding exists, but we cannot rule out the possibility that

the problem is unconfounded to start with. (The sets $S = \{E, A\}$ or $S = \{Z\}$ in Figure 6.1 illustrate this point.) Observing a discrepancy between adjusted and unadjusted associations (between X and Y) does not help us either, because (recalling our discussion of collapsibility) we do not know which—the preadjustment or postadjustment association—is unbiased (see Figure 6.4).

Proof of Necessity

To prove that (U_1) and (U_2) must be violated whenever Z stands for a minimally sufficient set S, consider the case where X has no effect on Y. In this case, confounding amounts to a nonvanishing association between X and Y. A well-known property of conditional independence, called *contraction* (Section 1.1.5), states that violation of (U_1), $X \not\perp\!\!\!\perp S$, together with sufficiency, $X \perp\!\!\!\perp Y|S$, implies violation of minimality, $X \not\perp\!\!\!\perp Y$:

$$X \not\perp\!\!\!\perp S \;\&\; X \perp\!\!\!\perp Y|S \Rightarrow X \not\perp\!\!\!\perp Y.$$

Likewise, another property of conditional independence, called *intersection*, states that violation of (U_2), $S \not\perp\!\!\!\perp Y|X$, together with sufficiency, $X \perp\!\!\!\perp Y|S$, also implies violation of minimality, $X \not\perp\!\!\!\perp Y$.

$$S \not\perp\!\!\!\perp Y|X \;\&\; X \perp\!\!\!\perp Y|S \Rightarrow X \not\perp\!\!\!\perp Y.$$

Thus, both (U_1) and (U_2) must be violated by any minimally sufficient set S (thus replacing Z in Definition 6.2.2).

Note, however, that intersection holds only for strictly positive probability distributions, which means that the Rothman-Greenland condition may be violated if deterministic relationships hold among some variables in a problem. This can be seen from a simple example in which both X and Y stand in a one-to-one functional relationship to a third variable, Z. Clearly, Z is a minimally sufficient set yet is not associated with Y given X; once we know the value of X, the probability of Y is determined, and would no longer change with learning the value of Z.

6.5.3 Exchangeability versus Structural Analysis of Confounding

Students of epidemiology complain bitterly about the confusing way in which the fundamental concept of confounding has been treated in the literature. A few authors have acknowledged the confusion (e.g. Greenland and Robins 1986; Wickramaratne and Holford 1987; Weinberg 1993)) and have suggested new ways of looking at the problem that might lead to more systematic analysis. Greenland and Robins (GR), in particular, have recognized the same basic principles and results that we have expounded here in Sections 6.2 and 6.3. Their analysis represents one of the few bright spots in the vast literature on confounding in that it treats confounding as an unknown causal quantity that is not directly measurable from observed data. They further acknowledge (as do Miettinen and Cook 1981) that the presence or absence of confounding should not be equated with absence or presence of collapsibility and that confounding should not be regarded as a parameter-dependent phenomenon.

However, the structural analysis presented in this chapter differs in a fundamental way from that of GR, who have pursued an approach based on judgment of "exchangeability." In Section 6.1, we have encountered a related notion of exchangeability, one with which Lindley and Novick (1981) attempted to view Simpson's paradox; GR's idea of exchangeability is more concrete and more clearly applicable. Conceptually, the connection between confounding and exchangeability is as follows. If we undertake to assess the effect of some treatment, we ought to make sure that any response differences between the treated and the untreated group is due to the treatment itself and not to some intrinsic differences between the groups that are unrelated to the treatment. In other words, the two groups must resemble each other in all characteristics that have bearing on the response variable. In principle, we could have ended the definition of confounding at this point, declaring simply that the effect of treatment is unconfounded if the treated and untreated groups resemble each other in all relevant features. This definition, however, is too verbal in the sense that it is highly sensitive to interpretation of the terms "resemblance" and "relevance." To make it less informal, GR used De Finetti's twist of hypothetical permutation; instead of judging whether two groups are similar, the investigator is instructed to imagine a hypothetical *exchange* of the two groups (the treated group becomes untreated, and vice versa) and then to judge whether the observed data under the swap would be distinguishable from the actual data.

One can justifiably ask what has been gained by this mental exercise, relative to judging directly if the two groups are effectively identical. The gain is twofold. First, people are quite good in envisioning dynamic processes and can simulate the outcome of this swapping scenario from basic understanding of the processes that govern the response to treatment and the factors that affect the choice of treatment. Second, moving from judgment about resemblance to judgment about probabilities permits us to cast those judgments in probabilistic notation and hence to invite the power and respectability of probability calculus.

Greenland and Robins made an important first step toward this formalization by bringing notation closer to where judgment originates—the human understanding of causal processes. The structural approach pursued in this book takes the next, natural step: formalizing the causal processes themselves.

Let A and B stand (respectively) for the treated and untreated groups, and let $P_{A1}(y)$ and $P_{A0}(y)$ stand (respectively) for the response distribution of group A under two hypothetical conditions, treatment and no treatment.[24] If our interest lies in some parameter μ of the response distribution, we denote by μ_{A1} and μ_{A0} the values of that parameter in the corresponding distribution $P_{A1}(y)$ and $P_{A0}(y)$, with μ_{B1} and μ_{B0} defined similarly for group B. In actuality, we measure the pair (μ_{A1}, μ_{B0}); after the hypothetical swap, we would measure (μ_{B1}, μ_{A0}). We define the groups to be *exchangeable* relative to parameter μ if the two pairs are indistinguishable, that is, if

$$(\mu_{A1}, \mu_{B0}) = (\mu_{B1}, \mu_{A0}).$$

In particular, if we define the causal effect by the difference $\mathrm{CE} = \mu_{A1} - \mu_{A0}$, then exchangeability permits us to replace μ_{A0} with μ_{B0} and so obtain $\mathrm{CE} = \mu_{A1} - \mu_{B0}$,

[24] In $do(\cdot)$ notation, we would write $P_{A1}(y) = P_A(y|do(X = 1))$.

which is measurable because both quantities are observed. Greenland and Robins thus declare the causal effect CE to be *unconfounded* if $\mu_{A0} = \mu_{B0}$.

If we compare this definition to that of (6.10), $P(y|do(x)) = P(y|x)$, we find that the two coincide if we rewrite the latter as $\mu[P(y|do(x))] = \mu[P(y|x)]$, where μ is the parameter of interest in the response distribution. However, the major difference between the structural and the GR approaches lies in the level of analysis. Structural modeling extends the formalization of confounding in two important directions. First, (6.10) is not submitted to direct human judgment but is derived mathematically from more elementary judgments concerning causal processes.[25] Second, the input judgments needed for the structural model are both qualitative and stable.

A simple example will illustrate the benefits of these features. Consider the following statement (Greenland 1998):

(Q^*) "if the effect measure is the difference or ratio of response proportions, then the above phenomenon—noncollapsibility without confounding—cannot occur, nor can confounding occur without noncollapsibility."

We have seen in this chapter that statement (Q^*) should be qualified in several ways and that, in general, noncollapsibility and confounding are two distinct notions—neither implying the other, regardless of the effect measure (Section 6.5.1). However, the question we wish to discuss here is methodological: What formalism would be appropriate for validating, refuting, or qualifying statements of this sort? Clearly, since (Q^*) makes a general claim about all instances, one counterexample would suffice to refute its general validity. But how do we construct such a counterexample? More generally, how do we construct examples that embody properties of confounding, effect bias, causal effects, experimental versus nonexperimental data, counterfactuals, and other causality-based concepts?

In probability theory, if we wish to refute a general statement about parameters and their relationship we need only present one density function f for which that relationship fails to hold. In propositional logic, in order to show that a sentence is false, we need only present one truth table T that satisfies the premises and violates the conclusions. What, then, is the mathematical object that should replace f or T when we wish to refute causal claims like statement (Q^*)? The corresponding object used in the exchangeability framework of Greenland and Robins is a counterfactual contingency table (see e.g. Greenland 1999b, p. 905, or Figure 1.7 in Section 1.4.4). For instance, to illustrate confounding, we need two such tables: one describing the hypothetical response of the treated group A to both treatment and nontreatment, and one describing the hypothetical response of the untreated group B to both treatment and non-treatment. If the tables show that the parameter μ_{A0}, computed from the hypothetical response of the treated group to no treatment, differs from μ_{B0}, computed from the actual response of the untreated group, then we have confounding on our hands.

Tables of this type can easily be constructed for simple problems involving one treatment and one response variable, but they become a nightmare when several covariates are involved or when we wish to impose certain constraints on those covariates.

[25] Recall that the $do(\cdot)$ operator is defined mathematically in terms of equation deletion in structural equation models; consequently, the verification of the nonconfounding condition $P(y|do(x)) = P(y|x)$ in a given model is not a matter of judgment but a subject of mathematical analysis.

For example, we may wish to incorporate the standard assumption that a covariate Z does not lie on the causal pathway between treatment and response, or that Z has causal influence on Y, but such assumptions cannot conveniently be expressed in counterfactual contingency tables. As a result, the author of the claim to be refuted could always argue that the tables used in the counterexample may be inconsistent with the agreed assumptions.[26]

Such difficulties do not plague the structural representation of confounding. In this formalism, the appropriate object for exemplifying or refuting causal statements is a "causal model," as defined in Chapter 3 and used throughout this book. Here, hypothetical responses (μ_{A0} and μ_{B0}) and contingency tables are not the primitive quantities but rather are derivable from a set of equations that already embody the assumptions we wish to respect. Every parameterization of a structural model implies (using the $do(\cdot)$ operator) a specific set of counterfactual contingency tables that satisfies the input assumptions and exhibits the statistical properties displayed in the graph. For example, any parameterization of the graph in Figure 6.3 generates a set of counterfactual contingency tables that already embodies the assumptions that Z is not on the causal pathway between X and Y and that Z has no causal effect on Y, and almost every such parameterization will generate a counterexample to claim (Q^*). Moreover, we can also disprove (Q^*) by a casual inspection of the diagram and without generating numerical counterexamples. In Figure 6.3, for example, shows vividly that the risk difference $P(y|x_1) - P(y|x_2)$ is not collapsible on Z and, simultaneously, that X and Y are (stably) unconfounded.

The difference between the two formulations is even more pronounced when we come to substantiate, not refute, generic claims about confounding. Here it is not enough to present a single contingency table; instead, we must demonstrate the validity of the claim for all tables that can possibly be constructed in compliance with the input assumptions. This task, as the reader surely realizes, is a hopeless exercise within the framework of contingency tables; it calls for a formalism in which assumptions can be stated succinctly and in which conclusions can be deduced by mathematical derivations. The structural semantics offers such formalism, as demonstrated by the many generic claims proven in this book (examples include Theorem 6.4.4 and Corollary 6.5.2).

As much as I admire the rigor introduced by Greenland and Robins's analysis through the framework of exchangeability, I am thoroughly convinced that the opacity and inflexibility of counterfactual contingency tables are largely responsible for the slow acceptance of GR framework among epidemiologists and, as a byproduct, for the lingering confusion that surrounds confounding in the statistical literature at large. I am likewise convinced that formulating claims and assumptions in the language of structural models will make the mathematical analysis of causation accessible to rank-and-file researchers and thus lead eventually to a total and natural disconfounding of confounding.

[26] Readers who attempt to construct a counterexample to statement (Q^*) using counterfactual contingency tables will certainly appreciate this difficulty.