3.3 Controlling Confounding Bias

Whenever we undertake to evaluate the effect of one factor (X) on another (Y), the question arises as to whether we should adjust our measurements for possible variations in some other factors (Z), otherwise known as "covariates," "concomitants," or "confounders" (Cox 1958, p. 48). Adjustment amounts to partitioning the population into groups that are homogeneous relative to Z, assessing the effect of X on Y in each homogeneous group, and then averaging the results (as in (3.15)). The illusive nature of such adjustment was recognized as early as 1899, when Karl Pearson discovered what is now called *Simpson's paradox* (see Section 6.1): Any statistical relationship between two variables may be reversed by including additional factors in the analysis. For example, we may find that students who smoke obtain higher grades than those who do not smoke but, adjusting for age, smokers obtain lower grades in every age group and, further adjusting for family income, smokers again obtain higher grades than nonsmokers in every income-age group, and so on.

Despite a century of analysis, Simpson's reversal continues to "trap the unwary" [Dawid, 1979], and the practical question that it poses—whether an adjustment for a given covariate is appropriate—has resisted mathematical treatment. Epidemiologists, for example, are still debating the meaning of "confounding" (Grayson 1987; Shapiro 1997) and often adjust for wrong sets of covariates (Weinberg 1993; see also Chapter 6). The potential-outcome analyses of Rosenbaum and Rubin (1983) and Pratt and Schlaifer (1988) have led to a concept named "ignorability," which recasts the covariate selection problem in counterfactual vocabulary but falls short of providing a workable solution. Ignorability reads: "Z is an admissible set of covariates if, given Z, the value that Y would obtain had X been x is independent of X." Since counterfactuals are not observable, and since judgments about conditional independence of counterfactuals are not readily assertable from ordinary understanding of causal processes, the question has remained open: What criterion should one use to decide which variables are appropriate for adjustment?

Section 3.3.1 presents a general and formal solution of the adjustment problem using the language of causal graphs. In Section 3.3.2

we extend this result to nonstandard covariates that are affected by X and hence require several steps of adjustment. (Finally, Section 3.3.3 illustrates the use of these criteria in an example.

3.3.1 The Back-Door Criterion

Assume we are given a causal diagram G, together with nonexperimental data on a subset V of observed variables in G, and suppose we wish to estimate what effect the interventions $do(X = x)$ would have on a set of response variables Y, where X and Y are two subsets of V. In other words, we seek to estimate $P(y|\hat{x})$ from a sample estimate of $P(v)$.

We show that there exists a simple graphical test, named the "back-door criterion" in Pearl (1993b), that can be applied directly to the causal diagram in order to test if a set $Z \subseteq V$ of variables is sufficient for identifying $P(y|\hat{x})$.[5]

Definition 3.3.1 (Back-Door)
A set of variables Z satisfies the back-door *criterion relative to an ordered pair of variables (X_i, X_j) in a DAG G if:*

(i) *no node in Z is a descendant of X_i; and*

(ii) *Z blocks every path between X_i and X_j that contains an arrow into X_i.*

Similarly, if X and Y are two disjoint subsets of nodes in G, then Z is said to satisfy the back-door criterion relative to (X, Y) if it satisfies the criterion relative to any pair (X_i, X_j) such that $X_i \in X$ and $X_j \in Y$.

The name "back-door" echoes condition (ii), which requires that only paths with arrows pointing at X_i be blocked; these paths can be viewed as entering X_i through the back door. In Figure 3.4, for example, the sets $Z_1 = \{X_3, X_4\}$ and $Z_2 = \{X_4, X_5\}$ meet the back-door criterion, but $Z_3 = \{X_4\}$ does not because X_4 does not block the path $(X_i, X_3, X_1, X_4, X_2, X_5, X_j)$.

[5] This criterion may also be obtained from Theorem 7.1 of Spirtes et al. (1993). An alternative criterion, using a single d-separation test, is established in Section 3.4 (see (3.39)).

Theorem 3.3.2 (Back-Door Adjustment)
If a set of variables Z satisfies the back-door criterion relative to (X, Y), then the causal effect of X on Y is identifiable and is given by the formula

$$P(y|\hat{x}) = \sum_z P(y|x, z)P(z). \tag{3.21}$$

The summation in (3.21) represents the standard formula obtained under adjustment for Z; variables X for which the equality in (3.21) is valid were named "conditionally ignorable given Z" in Rosenbaum and Rubin (1983). Reducing ignorability conditions to the graphical criterion of Definition 3.3.1 replaces judgments about counterfactual dependencies with judgments about the structure of causal processes, as represented in the diagram. The graphical criterion can be tested by systematic procedures that are applicable to diagrams of any size and shape. The criterion also enables the analyst to search for an optimal set of covariate—namely, a set Z that minimizes measurement cost or sampling variability [Tian 1998]. The use of a similar graphical criterion for identifying path coefficients in linear structural equations is demonstrated in Chapter 5. Applications to epidemiological research are given in Greenland et al. (1999a), where the set Z is called "sufficient set" for control of confounding.

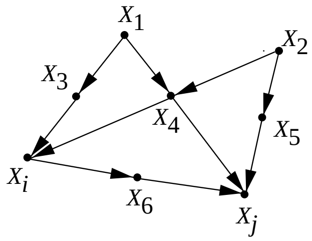

Figure 3.4: A diagram representing the back-door criterion; adjusting for variables $\{X_3, X_4\}$ (or $\{X_4, X_5\}$) yields a consistent estimate of $P(x_j|\hat{x}_i)$.

Proof of Theorem 3.3.2

The proof originally offered in Pearl (1993b) was based on the observation that, when Z blocks all back-door paths from X to Y, setting $(X = x)$ or conditioning on $X = x$ has the same effect on Y. This can best be seen from the augmented diagram G' of Figure 3.2, to which the intervention arcs $F_X \rightarrow X$ were added. If all back-door paths from X to Y are blocked, then all paths from F_X to Y must go through the children of X, and those would be blocked if we condition on X. The implication is that Y is independent of F_X given X,

$$P(y|x, F_X = do(x)) = P(y|x, F_X = \text{idle}) = P(y|x), \qquad (3.22)$$

which means that the observation $X = x$ cannot be distinguished from the intervention $F_X = do(x)$.

Formally, we can prove this observation by writing $P(y|\hat{x})$ in terms of the augmented probability function P' in accordance with (3.11) and conditioning on Z to obtain:

$$\begin{aligned} P(y|\hat{x}) = P'(y|F_x) &= \sum_z P'(y|z, F_x)P'(z|F_x) \\ &= \sum_z P'(y|z, x, F_x)P'(z|F_x). \qquad (3.23) \end{aligned}$$

The addition of x to the last expression is licensed by the implication $F_x \Rightarrow X = x$. To eliminate F_x from the two terms on the right-hand side of (3.23), we invoke the two conditions of Definition 3.3.1. Since F_x consists of root nodes with children restricted to X, it must be independent of all nondescendants of X, including Z. Thus, condition (i) yields

$$P'(z|F_x) = P'(z) = P(z)$$

Invoking now the back-door condition (ii), together with (3.22), permits us to eliminate F_x from (3.23), thus proving (3.21). □

3.3.2 The front-door criterion

Condition (i) of Definition 3.3.1 reflects the prevailing practice that "the concomitant observations should be quite unaffected by the treatment"

(Cox 1958, p. 48). This section demonstrates how concomitants that *are* affected by the treatment can be used to facilitate causal inference. The emerging criterion, named the front-door criterion in Pearl (1995a), will constitute the second building block of the general test for identifying causal effects (Section 3.4).

Consider the diagram in Figure 3.5, which represents the model of Figure 3.4 when the variables $X_1, \ldots, X_5$ are unobserved and $\{X_i, X_6, X_j\}$ are relabeled $\{X, Z, Y\}$, respectively. Although Z does not satisfy any of the back-door conditions, measurements of Z can nevertheless enable consistent estimation of $P(y|\hat{x})$. This will be shown by reducing the expression for $P(y|\hat{x})$ to formulas that are computable from the observed distribution function $P(x, y, z)$.

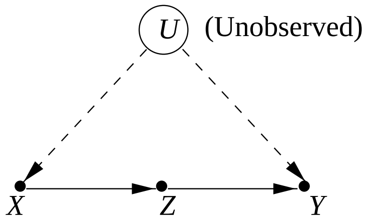

Figure 3.5: A diagram representing the front-door criterion. A two-step adjustment for Z yields a consistent estimate of $P(y|\hat{x})$.

The joint distribution associated with Figure 3.5 can be decomposed (equation (3.6)) into

$$P(x, y, z, u) = P(u)P(x|u)P(z|x)P(y|z, u). \tag{3.24}$$

From (3.12), the intervention $do(x)$ removes the factor $P(x|u)$ and induces the postintervention distribution

$$P(y, z, u|\hat{x}) = P(y|z, u)P(z|x)P(u). \tag{3.25}$$

Summing over z and u then gives

$$P(y|\hat{x}) = \sum_z P(z|x) \sum_u P(y|z, u)P(u). \tag{3.26}$$

In order to eliminate u from the r.h.s. of (3.26), we use the two conditional independence assumptions encoded in the graph of Figure 3.5:

$$\begin{aligned} P(u|z, x) &= P(u|x), & (3.27) \\ P(y|x, z, u) &= P(y|z, u). & (3.28) \end{aligned}$$

This yields the equalities

$$\begin{aligned}\sum_u P(y|z,u)P(u) &= \sum_x \sum_u P(y|z,u)P(u|x)P(x) \\ &= \sum_x \sum_u P(y|x,z,u)P(u|x,z)P(x) \\ &= \sum_x P(y|x,z)P(x) \qquad (3.29)\end{aligned}$$

and allows the reduction of (3.26) to a form involving only observed quantities:

$$P(y|\hat{x}) = \sum_z P(z|x) \sum_{x'} P(y|x',z)P(x'). \qquad (3.30)$$

All factors on the r.h.s. of (3.30) are consistently estimable from nonexperimental data, so it follows that $P(y|\hat{x})$ is estimable as well. Thus, we are in possession of an identifiable nonparametric estimand for the causal effect of X on Y whenever we can find a mediating variable Z that meets the conditions of (3.27) and (3.28).

Equation (3.30) can be interpreted as a two-step application of the back-door formula. In the first step, we find the causal effect of X on Z; since there is no back-door path from X to Z, we simply have

$$P(z|\hat{x}) = P(z|x).$$

Next, we compute the causal effect of Z on Y, which we can no longer equate with the conditional probability $P(y|z)$ because there is a back-door path $Z \leftarrow X \leftarrow U \rightarrow Y$ from Z to Y. However, since X blocks (d-separates) this path, X can play the role of a concomitant in the back-door criterion, which allows us to compute the causal effect of Z on Y in accordance with (3.21), giving $P(y|\hat{z}) = \sum_{x'} P(y|x',z)P(x')$. Finally, we combine the two causal effects via

$$P(y|\hat{x}) = \sum_z P(y|\hat{z})P(z|\hat{x}),$$

which reduces to (3.30).

We summarize this result by a theorem after formally defining the assumptions.

Definition 3.3.3 (Front-Door)
A set of variables Z is said to satisfy the front-door *criterion relative to an ordered pair of variables (X, Y) if:*

(i) *Z intercepts all directed paths from X to Y;*

(ii) *there is no back-door path from X to Z; and*

(iii) *all back-door paths from Z to Y are blocked by X.*

Theorem 3.3.4 (Front-Door Adjustment)
If Z satisfies the front-door criterion relative to (X, Y) and if $P(x, z) > 0$, then the causal effect of X on Y is identifiable and is given by the formula

$$P(y|\hat{x}) = \sum_{z} P(z|x) \sum_{x'} P(y|x', z)P(x'). \tag{3.31}$$

The conditions stated in Definition 3.3.3 are overly restrictive; some of the back-door paths excluded by conditions (ii) and (iii) can actually be allowed provided they are blocked by some concomitants. For example, the variable Z_2 in Figure 3.1 satisfies a front-door-like criterion relative to (X, Z_3) by virtue of Z_1 blocking all back-door paths from X to Z_2 as well as those from Z_2 to Z_3. To allow the analysis of such intricate structures, including nested combinations of back-door and front-door conditions, a more powerful symbolic machinery will be introduced in Section 3.4, one that will sidestep algebraic manipulations such as those used in the derivation of (3.30). But first let us look at an example illustrating possible applications of the front-door condition.

3.3.3 Example: Smoking and the Genotype Theory

Consider the century-old debate on the relation between smoking (X) and lung cancer (Y) (Spirtes et al. 1993, pp. 291-302). According to many, the tobacco industry has managed to forestall antismoking legislation by arguing that the observed correlation between smoking and lung cancer could be explained by some sort of carcinogenic genotype (U) that involves inborn craving for nicotine.

The amount of tar (Z) deposited in a person's lungs is a variable that promises to meet the conditions listed in Definition 3.3.3, thus fitting the structure of Figure 3.5. To meet condition (i), we must assume that smoking cigarettes has no effect on the production of lung cancer except as mediated through tar deposits. To meet conditions (ii) and (iii), we must assume that, even if a genotype is aggravating the production of lung cancer, it nevertheless has no effect on the amount of tar in the lungs except indirectly (through cigarette smoking). Likewise, we must assume that no other factor that affects tar deposit has any influence on smoking. Finally, condition $P(x, z) > 0$ of Theorem 3.3.4 requires that high levels of tar in the lungs be the result not only of cigarette smoking but also of other factors (e.g., exposure to environmental pollutants) and that tar may be absent in some smokers (owing perhaps to an extremely efficient tar-rejecting mechanism). Satisfaction of this last condition can be tested in the data.

To demonstrate how we can assess the degree to which cigarette smoking increases (or decreases) lung-cancer risk, we will assume a hypothetical study in which the three variables X, Y, Z, were measured simultaneously on a large, randomly selected sample of the population. To simplify the exposition, we will further assume that all three variables are binary, taking on true (1) or false (0) values. A hypothetical data set from a study on the relations among tar, cancer, and cigarette smoking is presented in Table 3.1.

It shows that 95% of smokers and 5% of nonsmokers have developed high levels of tar in their lungs. Moreover, 81% of subjects with tar deposits have developed lung cancer, compared to only 14% among those with no tar deposits. Finally, within each of these two groups (tar and no-tar), smokers show a much higher percentage of cancer than nonsmokers.

These results seem to prove that smoking is a major contributor to lung cancer. However, the tobacco industry might argue that the table tells a different story—that smoking actually decreases one's risk of lung cancer. Their argument goes as follows. If you decide to smoke, then your chances of building up tar deposits are 95%, compared to 5% if you decide not to smoke. In order to evaluate the effect of tar deposits, we look separately at two groups, smokers and nonsmokers. The

		$P(x, z)$	$P(Y = 1 \mid x, z)$
	Group Type	Group Size	% of Cancer Cases
		(% of Population)	in Group
$X = 0,\ Z = 0$	Non-smokers, No tar	47.5	10
$X = 1,\ Z = 0$	Smokers, No tar	2.5	90
$X = 0,\ Z = 1$	Non-smokers, Tar	2.5	5
$X = 1,\ Z = 1$	Smokers, Tar	47.5	85

Table 3.1:

table shows that tar deposits have a protective effect in both groups: in smokers, tar deposits lower cancer rates from 90% to 85%; in non-smokers, they lower cancer rates from 10% to 5%. Thus, regardless of whether I have a natural craving for nicotine, I should be seeking the protective effect of tar deposits in my lungs, and smoking offers a very effective means of acquiring those deposits.

To settle the dispute between the two interpretations, we now apply the front-door formula (equation (3.31)) to the data in Table 3.1. We wish to calculate the probability that a randomly selected person will develop cancer under each of the following two actions: smoking (setting $X = 1$) or not smoking (setting $X = 0$).

Substituting the appropriate values of $P(z|x)$, $P(y|x, z)$, and $P(x)$, we have

$$\begin{aligned} P(Y = 1|do(X = 1)) &= .05(.10 \times .50 + .90 \times .50) \\ &\quad +.95(.05 \times .50 + .85 \times .50) \\ &= .05 \times .50 + .95 \times .45 = .4525, \\ P(Y = 1|do(X = 0)) &= .95(.10 \times .50 + .90 \times .50) \\ &\quad +.05(.05 \times .50 + .85 \times .50) \\ &= .95 \times .50 + .05 \times .45 = .4975. \quad (3.32) \end{aligned}$$

Thus, contrary to expectation, the data prove smoking to be somewhat beneficial to one's health.

The data in Table 3.1 are obviously unrealistic and were deliberately crafted so as to support the genotype theory. However, the purpose of this exercise was to demonstrate how reasonable qualitative assumptions about the workings of mechanisms, coupled with nonexperimental data, can produce precise quantitative assessments of causal effects. In reality, we would expect observational studies involving mediating variables to refute the genotype theory by showing, for example, that the mediating consequences of smoking (such as tar deposits) tend to increase, not decrease, the risk of cancer in smokers and nonsmokers alike. The estimand of (3.31) could then be used for quantifying the causal effect of smoking on cancer.