6.2 Why There Is No Statistical Test For Confounding, Why Many Think There Is, and Why They Are Almost Right

6.2.1 Introduction

Confounding is a simple concept. If we undertake to estimate the effect[8] of one variable (X) on another (Y) by examining the statistical association between the two, we ought to ensure that the association is not produced by factors other than the effect under study. The presence of spurious association, due for example to the influence of extraneous variables, is called *confounding* because it tends to confound our reading and to bias our estimate of the effect studied. Conceptually, therefore, we can say that X and Y are confounded when there is a third variable Z that influences both X and Y; such a variable is then called a *confounder* of X and Y.

As simple as this concept is, it has resisted formal treatment for decades, and for good reason: The very notions of "effect" and "influence"—relative to which "spurious association" must be defined—have resisted mathematical formulation. The empirical definition of effect as an association that *would* prevail in a controlled randomized experiment cannot easily be expressed in the standard language of probability theory, because that theory deals with static conditions and does not permit us to predict, even from a full specification of a population density function, what relationships would prevail if conditions were to change—say, from observational to controlled studies. Such predictions require extra information in the form of causal or counterfactual assumptions which are not discernible from density functions (see Sections 1.3 and 1.4). The $do(\cdot)$ operator used in this book was devised specifically for distinguishing and managing this extra information.

These difficulties notwithstanding, epidemiologists, biostatisticians, social scientists and economists[9] have made numerous attempts to define confounding in statistical terms, partly because statistical definitions—free of theoretical terms of "effect" or "influence"—can be expressed in conventional mathematical form and partly because such definitions may lead to practical tests of confounding and thereby alert investigators to possible bias and need for adjustment. These attempts have converged in the following basic criterion.

Associational Criterion

Two variables X and Y are not confounded if and only if every variable Z that is not affected by X is either:

(U_1) *unassociated with X or*

(U_2) *unassociated with Y, conditional on X.*

This criterion, with some variations and derivatives (often avoiding the "only if" part), can be found in almost every epidemiology textbook (Schlesselman 1982; Roth-

[8] We will confine the use of the terms "effect," "influence," and "affect" to their causal interpretations; the term "association" will be set aside for statistical dependencies.

[9] In econometrics, the difficulties have focused on the notion of "exogeneity" (Engle et al. 1983; Leamer 1985; Aldrich 1993) which stands essentially for "no confounding" (see Section 5.4.3).

man 1986; Rothman and Greenland 1998) and in almost every article dealing with confounding. In fact, the criterion has become so deeply entrenched in the literature that authors (e.g. Gail 1986; Hauck et al. 1991; Becher 1992; Steyer et al. 1996) often take it to be the *definition* of no-confounding, forgetting that ultimately confounding is useful only so far as it tells us about effect bias.[10]

The purpose of this and the next section is to highlight several basic limitations of the associational criterion and its derivatives. We will show that the associational criterion neither ensures unbiased effect estimates nor follows from the requirement of unbiasedness. After demonstrating, by examples, the absence of logical connections between the statistical and the causal notions of confounding, we will define a stronger notion of unbiasedness, called "stable" unbiasedness, relative to which a modified statistical criterion will be shown necessary and sufficient. The necessary part will then yield a practical test for stable unbiasedness that, remarkably, does not require knowledge of all potential confounders in a problem. Finally, we will argue that the prevailing practice of substituting statistical criteria for the effect-based definition of confounding is not entirely misguided, because stable unbiasedness is in fact (i) what investigators have been (and perhaps should be) aiming to achieve and (ii) what statistical criteria can test.

6.2.2 Causal and Associational Definitions

In order to facilitate the discussion, we shall first cast the causal and statistical definitions of no-confounding in mathematical forms.[11]

Definition 6.2.1 (No-Confounding; Causal Definition)
Let M be a causal model of the data-generating process—that is, a formal description of how the value of each observed variable is determined. Denote by $P(y|do(x))$ the probability of the response event $Y = y$ under the hypothetical intervention $X = x$, calculated according to M. We say that X and Y are not confounded *in M if and only if*

$$P(y|do(x)) = P(y|x) \tag{6.10}$$

for all x and y in their respective domains, where $P(y|x)$ is the conditional probability generated by M.

For the purpose of our discussion here, we take this causal definition as the meaning of the expression "no confounding." The probability $P(y|do(x))$ was defined in Chapter 3 (Definition 3.2.1, also abbreviated $P(y|\hat{x})$); it may also be interpreted as the conditional probability $P^*(Y = y|X = x)$ corresponding to a controlled experiment in which X

[10] Hauck et al. (1991) dismiss the effect-based definition of confounding as "philosophic" and consider a difference between two measures of association to be a "bias." Grayson (1987) even goes so far as to skate that the change-in-parameter method, a derivative of the associational criterion, is the only fundamental definition of confounding (see Greenland et al. 1989 for critiques of Grayson's position).

[11] For simplicity, we will limit our discussion to unadjusted confounding; extensions involving measurement of auxiliary variables are straightforward and can be obtained from Section 3.3. We also use the abbreviated expression "X and Y are not confounded," though "the effect of X on Y is not confounded" is more exact.

is randomized. We recall that this probability can be calculated from a causal model M either directly, by simulating the intervention $do(X = x)$, or (if $P(x, s) > 0$) via the adjustment formula (equation (3.19))

$$P(y|do(x)) = \sum_s P(y|x, s)P(s),$$

where S stands for any set of variables, observed as well as unobserved, that satisfy the back-door criterion (Definition 3.3.1). Equivalently, $P(y|do(x))$ can be written $P(Y(x) = y)$, where $Y(x)$ is the potential-outcome variable as defined in (3.51) or in Rubin (1974). We bear in mind that the operator $do(\cdot)$, and hence also effect estimates and confounding, must be defined relative to a specific causal or data-generating model M because these notions are not statistical in character and cannot be defined in terms of joint distributions.

Definition 6.2.2 (No-Confounding; Associational Criterion)
Let T be the set of variables in a problem that are not affected by X. We say that X and Y are not confounded *in the presence of T if each member Z of T satisfies at least one of the following conditions:*

(U_1) *Z is not associated with X (i.e., $P(x|z) = P(x)$);*

(U_2) *Z is not associated with Y, conditional on X (i.e., $P(y|z, x) = P(y|x)$).*

Conversely, X and Y are said to be confounded *if any member Z of T violates both (U_1) and (U_2).*

Note that the associational criterion in Definition 6.2.2 is not purely statistical in that it invokes the predicate "affected by" which is not discernible from probabilities but rests instead on causal information. This exclusion of variables that are affected by treatments (or exposures) is unavoidable and has long been recognized as a necessary judgmental input to every analysis of treatment effect in observational and experimental studies alike (Cox 1958, p. 48; Greenland and Neutra 1980). We shall assume throughout that investigators possess the knowledge required for distinguishing variables that are affected by the treatment X from those that are not. We shall then explore what additional causal knowledge is needed, if any, for establishing a test of confounding.

6.3 How the Associational Criterion Fails

We will say that a criterion for no-confounding is *sufficient* if it never errs when it classifies a case as no-confounding and *necessary* if it never errs when it classifies a case as confounding. There are several ways that the associational criterion of Definition 6.2.2 fails to match the causal criterion of Definition 6.2.1. Failures with respect to sufficiency and necessity will be addressed in turn.