Annu. Rev. Public Health 2001. 22:189–212

CONFOUNDING IN HEALTH RESEARCH

Sander Greenland[1,2] and Hal Morgenstern[1]
[1]*Department of Epidemiology, University of California, Los Angeles School of Public Health, Los Angeles, California 90095-1772; e-mail: HalM@ucla.edu*
[2]*Department of Statistics, University of California, Los Angeles College of Letters and Science, Los Angeles, California 90095*

Key Words bias, causation, collapsibility, confounder, epidemiologic methods

■ **Abstract** Consideration of confounding is fundamental to the design, analysis, and interpretation of studies intended to estimate causal effects. Unfortunately, the word confounding has been used synonymously with several other terms, and it has been used to refer to at least four distinct concepts. This paper provides an overview of confounding and related concepts based on a counterfactual model of causation. In this context, which predominates in nonexperimental research, confounding is a source of bias in the estimation of causal effects. Special attention is given to the history of definitions of confounding, the distinction between confounding and confounders, problems in the control of confounding, the relations of confounding to exchangeability and collapsibility, and confounding in randomized trials.

INTRODUCTION

Much epidemiologic and social science research is devoted to estimation of causal effects in populations and testing causal hypotheses using nonexperimental data. In such endeavors, issues of confounding invariably arise. Unfortunately, the word confounding has been used synonymously with several other terms (e.g. spurious association, fictitious association, secondary association, susceptibility bias, and Simpson's paradox), and it has been used to refer to at least four distinct concepts.

In one usage, dating to the middle of the nineteenth century, confounding is a source of bias in estimating causal effects and corresponds to lack of comparability between treatment or exposure groups (e.g. 36, 59, 103). In this usage, confounding is sometimes informally described as a mixing of effects of extraneous factors (called confounders) with the effect of interest. This usage predominates in nonexperimental research, especially in epidemiology and sociology, and is the focus of this paper.

In a second usage, originating in statistics during the past century, confounding is a synonym for noncollapsibility of an association parameter over levels of a covariate [an association is noncollapsible if its magnitude is different when adjusting

(conditioning) vs not adjusting for the covariate (e.g. 68, 96, 102)]. Sometimes this noncollapsibility definition of confounding is applied to causal parameters, i.e. causal effects instead of association measures.

In a third usage, originating in the experimental-design literature of the early twentieth century, confounding refers to inseparability of main effects and interactions under a particular study design (e.g. 10, 22). Typically, such confounding is deliberate because the interactions are not of interest to the investigator and the main effects can be estimated more efficiently. In the analysis-of-variance literature, the term aliasing is sometimes used to refer to this concept of confounding.

In a fourth usage, dating to the early nineteenth century, confounding is a type of measurement problem, resulting from inherent differences between the variables we measure and the underlying constructs of interest (e.g. 12, 20). Thus, associations observed between variables may not reflect the associations or effects of interest. This concept of confounding is sometimes described as an inferential problem in "construct validity" and is often used in psychology.

The four concepts of confounding are not always distinguished properly. In particular, the concept of confounding as a source of bias in effect estimation and the concept of it as noncollapsibility are often treated as identical. Here we provide a historical overview of these two concepts and the distinctions between them. Because these distinctions require a formal model for causal effects, we begin with a discussion of the counterfactual model of causation. We then trace the history of the concept of confounding from the writings of John Stuart Mill (58, 59) to its modern counterfactual formalization. We discuss how approaches to control for confounding fit into this formalization, and we give special attention to the relations of confounding to exchangeability and randomization. We then describe how the counterfactual model distinguishes noncollapsibility from confounding. Our penultimate section covers some issues that arise when considering confounding in studies of interventions. Given the importance of the concept in causal inference, we end with a recommendation to include more thorough discussion of confounding in all types of public-health education.

COUNTERFACTUAL MODEL OF CAUSATION

Overview

The concepts of cause and effect are central to most areas of scientific research. Thus, it may be surprising that consensus about basic definitions and methods for causal inference is limited, despite some three centuries of debate. A brief review cannot do justice to all the history and details of this debate, nor to all the schools of thought on causation. We recommend Pearl (66) for a comprehensive treatment of modern causality theory; a brief overview for the health sciences is given in Greenland (31). We focus here on one conceptualization that has proven useful in the analysis of confounding. This counterfactual or potential-outcomes approach has become common in philosophy, statistics, and epidemiology.

Since the early eighteenth century, philosophers noted serious deficiencies in common definitions of causation, and these deficiencies remain in modern usage. For example, Webster's *New Twentieth Century Dictionary* (55) offered "that which produces an effect or result" as a definition of "cause," but "to cause" is among the definitions of "produces." Informal definitions of "effect" suffer from the same circularity, because "effect" as a verb is merely a synonym for "cause," whereas "effect" as a noun is defined as a "result," which is in turn defined as an "effect" in causal contexts.

Hume (47, 48) offered another view of causation that pointed a way out of the circularity of common definitions: "We may define a cause to be an object, followed by another, . . . where, if the first object had not been, the second had never existed" (48, p. 115). Thus, by focusing on specific instances of causation, Hume asserted that an event A caused an event B if occurrence of A was necessary for occurrence of B under the observed background circumstances (e.g. see 52, 95). Essentially the same concept of causation can be found in the works of Mill (58, 59) and Fisher (21) (see also 92), as well as in later works in statistics and related fields. A typical example is from MacMahon & Pugh (54, p. 12), who state that ". . . an association may be classed as presumptively causal when it is believed that, *had the cause been* altered, the effect *would have been* changed" (italics added). The italicized phrases emphasize that the alteration of the antecedent condition ("cause") and the subsequent change in the outcome ("effect") are contrary to what was in fact observed, i.e. they are counterfactual.

The preceding definition falls short of the formalism necessary for derivation of statistical methods for causal inference. Such a formalism and derivation first appeared in Neyman (62). The basic idea is as follows: Suppose that N units (e.g. individuals, populations, or objects) are to be observed in an experiment that will assign each unit one of $K + 1$ treatments $x_0, x_1, \ldots, x_K$. The outcome of interest for unit i is the value of a response variable Y_i. Suppose that Y_i will equal y_{ik} if unit i is assigned treatment x_k. Usually, one treatment level, say x_0, is designated the reference treatment against which other treatments are to be evaluated; typically, x_0 is "no treatment," a placebo, or a standard treatment. We define the causal effect of x_k ($k \geq 1$) on Y_i relative to x_0 (the referent) to be $y_{ik} - y_{i0}$. (If the response variable is strictly positive, we may instead define the causal effect as y_{ik}/y_{i0} or $\log y_{ik} - \log y_{i0}$.) In words, a causal effect is a counterfactual contrast between the outcomes of a single unit under different treatment possibilities.

Neyman's formalism is sometimes referred to as the potential-outcomes model of causation, and it has reappeared in various guises (e.g. see 13, 17, 41, 90). By defining effects as contrasts of potential outcomes, y_{ik} gives precise meanings to words such as cause, effect, and affect. For example, "changing X from x_0 to x_k affects Y_i" is an assertion that $y_{ik} - y_{i0} \neq 0$. Note, however, that because only one of the potential outcomes y_{ik} can be observed in any one unit, an individual effect $y_{ik} - y_{i0}$ cannot be observed in isolation from the reference (baseline) outcome y_{i0}.

Counterfactual analysis can be viewed as a special type of latent-variable analysis, in which y_{ik} remains latent for any individual i who did not receive treatment

k (e.g. see 5). The potential-outcomes model can also be derived from a structural-equations approach familiar in the social sciences (65, 66).

Restrictions of Counterfactuals

There are several crucial restrictions that the potential-outcomes definition places on the notion of causal effects (and hence cause) (38). Most important, causal effects are defined only for comparisons of treatment levels. To state, for example, that "drinking two glasses of wine a day lengthened a person's life by 4 years" is meaningless by itself. A reference level (e.g. no wine consumption) must be at least implicit to make sense of this statement. Sometimes, in fact, the reference level requires specification of other factors that might be affected by the counterfactual condition (e.g. keeping beer and liquor consumption constant in the absence of wine consumption). Another restriction of the counterfactual model as presented here is that causes refer to factors that can be potentially manipulated, such as drug treatments, but not to fixed personal attributes such as gender and race (e.g. see 46, 49, 60, 91). Finally, implicit in most discussions of potential outcomes is that the outcome for a given unit under a specific treatment does not depend on the treatment given to any other unit, i.e. the stability assumption (17, 91, 92). This assumption is likely to be violated when the outcome is contagious or the exposure represents a set of social conditions. Fortunately, the counterfactual approach can be extended to situations in which stability is violated (40).

Objections to Counterfactuals

Counterfactual approaches are sometimes criticized because, in considering causes of past events, they invoke consideration of distributions for events that never occurred and hence cannot be observed. As a consequence, some important features of these distributions remain empirically untestable, and thus some causal inferences based on counterfactuals will depend entirely on untestable assumptions (18).

It is our view that this property of counterfactual inferences reflects a strength of the counterfactual approach, rather than a weakness. It is an unfortunate but true fact that many important causal questions are simply not answerable, at least not without employing assumptions that are untestable given ethical considerations or limitations of current knowledge and technology. Examples include assumptions of no confounding (the focus of this paper), assumptions about independence of unit-specific susceptibilities or responses, and various distributional assumptions (13, 14, 44, 65, 66, 80, 86, 87, 91). Inferences from counterfactual approaches properly reflect this harsh epistemic reality when they display sensitivity to such assumptions.

More constructively, the counterfactual approach also aids in precise formulation of assumptions needed to identify causal effects statistically, which in turn can aid in developing techniques for meeting those assumptions. The basic example

on which we focus is the assumption of exchangeability of response distributions under homogeneous treatment assignment, which is met when treatment is successfully randomized or, more generally, when treatment assignment is independent of the potential outcomes y_{ik}.

CONFOUNDING AND CONFOUNDERS

Background

Counterfactual approaches to causal inference emphasize the importance of randomization in assuring identifiability of causal effects (30, 36, 38, 62, 70, 86, 91–93). In observational studies, however, no such assurance is available, and issues of confounding become paramount.

One of the earliest systematic discussions of "confounded effects" is in Mill (59 Ch. 10) (although in Chapter 3 Mill lays out the primary issues and acknowledges Francis Bacon as a forerunner in dealing with them). There, Mill listed a requirement for an experiment intended to determine causal relations: ". . . none of the circumstances [of the experiment] that we do know shall have effects susceptible of being confounded with those of the agents whose properties we wish to study."

It should be noted that, in Mill's time, the word experiment referred to an observation in which some circumstances were under the control of the observer, as it still is used in ordinary English, rather than to the notion of a comparative trial. Nonetheless, Mill's requirement suggests that a comparison is to be made between the outcome of his experiment (which is, essentially, a trial with no control group) and what we would expect the outcome to be if the agents we wish to study had been absent. If the outcome is not what one would expect in the absence of the study agents, his requirement ensures that the unexpected outcome was not brought about by extraneous circumstances. If, however, those circumstances do bring about the unexpected outcome, and that outcome is mistakenly attributed to effects of the study agents, the mistake is one of confounding (or confusion) of the extraneous effects with the agent effects.

Much of the modern literature follows the same informal conceptualization given by Mill. Terminology is now more specific, with "treatment" used to refer to an agent administered by the investigator and "exposure" often used to denote an unmanipulated agent. The chief development beyond Mill is that the expectation for the outcome in absence of the study exposure is now almost always explicitly derived from observation of a control or reference group that is untreated or unexposed. For example, Clayton & Hills (8, p. 133) state that, in observational studies, ". . . there is always the possibility that an important influence on the outcome. . . differs systematically between the comparison [exposed and unexposed] groups. It is then possible [that] part of the apparent effect of exposure is due to these differences, [in which case] the comparison of the exposure groups is said to be *confounded*" (emphasis in the original).

As discussed below, confounding is also possible in randomized experiments because of systematic elements in treatment allocation, administration, and compliance and because of random differences between comparison groups (22, 30, 36, 38, 87).

Confounding

Attempts to quantify the above notion of confounding can be traced at least as far back as the work of Pearson et al (68) and Yule (102) on spurious correlation, but these attempts ran afoul of the absence of a formal model for causal effects [see Aldrich (1) for a review of this work]. Various mathematical formalizations of confounding have since been proposed. Perhaps the one closest to Mill's concept is based on the counterfactual model for effects. Suppose our objective is to determine the effect of applying a treatment or exposure x_1 on a parameter μ of the distribution of the outcome Y in population A, relative to applying treatment or exposure x_0. That is, we wish to contrast the marginal distributions $F_A(y_1)$ and $F_A(y_0)$ of the potential outcomes under treatments 1 and 0, using some parameter (summary) μ of the distributions. For example, population A could be a cohort of breast-cancer patients, treatment x_1 could be a new hormone therapy, x_0 could be a placebo therapy, and the parameter μ could be the expected survival or the 5-year survival probability in the cohort; μ could also be a vector or even a function, such as an entire survival curve. The population A is sometimes called the target population or index population, the treatment x_1 is sometimes called the experimental or index treatment, and the treatment x_0 is sometimes called the control or reference treatment.

Suppose that μ will equal μ_{A1} if x_1 is applied to population A, and μ will equal μ_{A0} if x_0 is applied to that population; the causal effect of x_1 relative to x_0 is defined as the change from μ_{A0} to μ_{A1}, which could be measured by the difference parameter $\mu_{A1} - \mu_{A0}$ (or by the ratio parameter μ_{A1}/μ_{A0} if μ is strictly positive). If A is observed under treatment x_1, μ will equal μ_{A1}, which is observable or estimable, but μ_{A0} will be unobservable. Suppose, however, we expect μ_{A0} to equal μ_{B0}, where μ_{B0} is the value of the outcome μ observed or estimated for a population B that was administered treatment x_0. The latter population is sometimes called a control or reference population. A comparison of population A treated with x_1 to population B treated with x_0 is an association parameter (i.e. the observable association between treatment and outcome in the combined population A and B). We say confounding is present if in fact $\mu_{A0} \neq \mu_{B0}$. When confounding is present, there would be some difference between the outcomes of populations A and B even if both populations (rather than just B) were untreated.

If confounding is present, a crude (unadjusted) association parameter obtained by substituting μ_{B0} for μ_{A0} in the effect measure will not equal the causal parameter, and the association parameter is said to be confounded. For example, if $\mu_{B0} \neq \mu_{A0}$, then $\mu_{A1} - \mu_{B0}$, (which measures the association of treatments with outcomes across the two populations) is confounded for $\mu_{A1} - \mu_{A0}$ (which measures the effect of treatment x_1 on population A). Thus, saying an association parameter

TABLE 1 Actual (observable) and counterfactual (unobservable) expected numbers[a] and average risks (R) of an outcome event ($Y = 1$) in two populations, A (in which everyone is actually exposed, $X = 1$) and B (in which everyone is actually unexposed, $X = 0$), by exposure status: examples of effect and (crude) association measures

	Population A		**Population B**	
	X = 1 Actual	**X = 0 Counterfactual**	**X = 1 Counterfactual**	**X = 0 Actual**
Outcome				
Y = 1	30	20	30	10
Y = 0	70	80	70	90
Total	100	100	100	100
Risk[b]	$R_{A1} = 0.30$	$R_{A0} = 0.20$	$R_{B1} = 0.30$	$R_{B0} = 0.10$
Effect[c]	$R_{A1} - R_{A0} = 0.10$ $R_{A1}/R_{A0} = 1.5$		$R_{B1} - R_{B0} = 0.20$ $R_{B1}/R_{B0} = 3.0$	

[a]Cell values are the expected frequencies of individuals in each population under actual and counterfactual conditions of exposure.

[b]Probability of $Y = 1$.

[c]Contrasts of outcomes when everyone is exposed ($X = 1$) versus when everyone is unexposed ($X = 0$) within each population. Compare with the associations (observable contrasts between populations A and B) in the combined population: risk difference, $R_{A1} - R_{B0} = 0.20$; risk ratio, $R_{A1}/R_{B0} = 3.0$.

such as $\mu_{A1} - \mu_{B0}$ is confounded for a causal parameter such as $\mu_{A1} - \mu_{A0}$ is synonymous with saying the two parameters are not equal.

To illustrate the counterfactual definition of confounding, we take the risk (probability) of an outcome event ($Y = 1$) as the outcome parameter μ of interest. Table 1 shows the actual risks (R_{A1} and R_{B0}) and counterfactual risks (R_{A0} and R_{B1}) for two populations: A, which is entirely exposed ($X = 1$); and B, which is entirely unexposed ($X = 0$). If A is the target population, we measure the effect of the exposure on outcome risk in this population by contrasting R_{A1} with R_{A0}, e.g. by taking their difference, $0.30 - 0.20 = 0.10$, or their ratio, $0.30/0.20 = 1.5$. Because R_{A0} is unobservable, however, this counterfactual contrast is also unobservable. The association between exposure and outcome risk in the combined population ($A + B$) is a contrast between the two observable risks, R_{A1} and R_{B0}, e.g. $0.30 - 0.10 = 0.20$ (the risk difference) or $0.30/0.10 = 3.0$ (the risk ratio). Because the actual risk in the reference population B ($R_{B0} = 0.10$) differs from the counterfactual risk in the target population A ($R_{A0} = 0.20$), these two association parameters differ from their corresponding effect parameters in population A, and we say the association in the combined population is confounded for the effect in population A.

The above formalization has several interesting implications. One is that confounding depends on the outcome parameter. For example, suppose populations A and B would have a different 5-year survival probabilities μ_{A0} and μ_{B0} under placebo

treatment x_0, so that $\mu_{A1} - \mu_{B0}$ is confounded for the actual effect $\mu_{A1} - \mu_{A0}$ of treatment on 5-year survival. It is then still possible that 10-year survival ν under the placebo would be identical in both populations, i.e. ν_{A0} could still equal ν_{B0}, so that $\nu_{A1} - \nu_{B0}$ is not confounded for the actual effect of treatment on 10-year survival. (We should expect no confounding for 200-year survival because no treatment is likely to raise the 200-year survival probability of human patients above zero.)

Another important implication is that confounding depends on the target population of inference. The preceding example, with *A* as the target, had different 5-year survivals μ_{A0} and μ_{B0} for *A* and *B* under placebo therapy and, hence, $\mu_{A1} - \mu_{B0}$ was confounded for the effect $\mu_{A1} - \mu_{A0}$ of treatment on population *A*. A lawyer or ethicist may also be interested in what effect the hormone treatment x_1 would have had on population *B*. Writing μ_{B1} for the (unobserved) outcome of *B* under treatment x_1, this effect on *B* may be measured by $\mu_{B1} - \mu_{B0}$. Substituting μ_{A1} for the unobserved μ_{B1} yields $\mu_{A1} - \mu_{B0}$. This measure of association is confounded for $\mu_{B1} - \mu_{B0}$ (the effect of treatment x_1 on 5-year survival in population *B*) if and only if $\mu_{A1} \neq \mu_{B1}$. Thus, the same measure of association $\mu_{A1} - \mu_{B0}$ may be confounded for the effect of treatment on neither, one, or both of populations *A* and *B*.

Consider again the example in Table 1 in which we compared the risks of an outcome event in exposed population *A* and unexposed population *B*. If we are interested in the exposure effect in population *B*, i.e. if we now treat *B* as the target population instead of *A*, the difference and ratio effect parameters are $R_{B1} - R_{B0} = 0.30 - 0.10 = 0.20$ and $R_{B1}/R_{B0} = 0.30/0.10 = 3.0$, which are larger than the effects in population *A*. Because the actual risk in exposed population *A* ($R_{A1} = 0.30$) is equal to the counterfactual risk in target population *B* ($R_{B1} = 0.30$), these effect parameters are equal to the corresponding association parameters for the combined population (see Table 1). Thus, we say the association in the combined population is not confounded for the effect in population *B*, even though the association was confounded for the effect in population *A* (see above).

A third implication is that absence of confounding ($\mu_{A0} = \mu_{B0}$), which is a population condition, is not sufficient to identify the sharp null hypothesis of no causal effects at the unit level ($y_{i1} = y_{i0}$ for all units *i*) because causal effects of treatment may cancel out (36). For example, suppose the outcome parameter μ is the average risk of a disease during a given period, with half of persons in *A* and half in *B* having $y_{i1} = 1$ and $y_{i0} = 0$ (treatment x_1 causes disease) and half having $y_{i1} = 0$ and $y_{i0} = 1$ (treatment x_1 prevents disease). Then $\mu_{A1} = \mu_{A0} = \mu_{B0} = 1/2$, so that there is no confounding and no identifiable effect of treatment on the outcome distribution; nonetheless, every person is affected by treatment. Neyman (63) and Stone (99) make the analogous point that randomization does not identify the sharp null hypothesis.

Components of Associations

We may write the difference in the outcome parameters of populations *A* and *B* as

$$\mu_{A1} - \mu_{B0} = (\mu_{A1} - \mu_{A0}) + (\mu_{A0} - \mu_{B0}), \qquad 1.$$

which shows that $\mu_{A1} - \mu_{B0}$ is a mix of the true treatment effect $\mu_{A1} - \mu_{A0}$ and a bias term $\mu_{A0} - \mu_{B0}$ (39, 51). Nonidentifiability of the true effect $\mu_{A1} - \mu_{A0}$ follows if the bias $\mu_{A0} - \mu_{B0}$ is not identifiable, as is the case in typical epidemiologic studies (36).

By rearranging Equation 1, we may obtain $\mu_{A0} - \mu_{B0}$ as a measure of bias in $\mu_{A1} - \mu_{B0}$ due to confounding:

$$\mu_{A0} - \mu_{B0} = (\mu_{A1} - \mu_{B0}) - (\mu_{A1} - \mu_{A0}). \qquad 2.$$

When the outcome parameters μ are risks (probabilities), epidemiologists use instead the analogous ratio

$$(\mu_{A1}/\mu_{B0})/(\mu_{A1}/\mu_{A0}) = \mu_{A0}/\mu_{B0} \qquad 3.$$

as a measure of the bias due to confounding (7, 56); μ_{A0}/μ_{B0} is sometimes called the confounding risk ratio. The latter term is somewhat confusing, as it is sometimes mistakenly thought to refer to the effect of a particular confounder on risk. This is not so, although the ratio does reflect the net effect of the differences in the confounder distributions of populations *A* and *B*.

Confounders

The above formalization of confounding invokes no explicit differences (imbalances) between populations *A* and *B* with respect to circumstances or covariates that might affect μ (36). It seems intuitively clear that if μ_{A0} and μ_{B0} differ, then *A* and *B* must differ with respect to factors that affect μ. This intuition has led some authors to define confounding in terms of differences in covariate distributions among the compared populations (e.g. 99). Nonetheless, confounding, as we have defined it, is not an inevitable consequence of covariate differences; *A* and *B* may differ profoundly with respect to covariates that affect μ, yet confounding (bias in effect estimation) may be absent. In other words, a covariate difference between *A* and *B* is a necessary but not sufficient condition for confounding because the effects of the various covariate differences may balance out in such a way that no confounding is present.

Suppose now that populations *A* and *B* differ with respect to certain covariates that affect μ and that these differences have led to confounding. The responsible covariates are then termed confounders of the association measure. In the above example, with $\mu_{A1} - \mu_{B0}$ confounded for the effect $\mu_{A1} - \mu_{A0}$, the factors that led to $\mu_{A0} \neq \mu_{B0}$ are the confounders. A variable cannot be a confounder (in this sense) unless (*a*) it can causally affect the outcome parameter μ within treatment groups, and (*b*) it is distributed differently among the compared populations, i.e. there is an association between treatment (or exposure) status and the covariate in the total (combined) population [e.g. see Yule (102), who uses such terms as fictitious association rather than confounding]. Note that condition *b* does not necessarily imply that the covariate is a determinant (cause) of treatment. The two necessary conditions (*a* and *b*) are sometimes offered together as a definition of

a confounder. Nonetheless, counterexamples show that the two conditions are not sufficient for a variable with more than two levels to be a confounder as defined above (38).

If a covariate satisfying condition *a* is time-dependent (i.e. if it can vary over time within units), condition *b* must be further restricted: (*c*) A covariate cannot be a confounder if its association with treatment status in the population is due entirely to effects of the treatment on the covariate (e.g. 38, 89). Thus, for example, the covariate may not be a confounder if it is an intermediate variable in the causal pathway between treatment and outcome, i.e. treatment affects the covariate, which affects the outcome. It is possible, however, for the same covariate at different times to be both a confounder of the treatment effect and an intermediate variable (70, 74). This situation arises when the covariate affects treatment and then treatment affects the covariate, as for example in a study of the effect of antihypertensive medication on stroke risk, with blood pressure the covariate.

Although definitions of confounder similar to that just given are common in epidemiology texts (e.g. 50, 89), they are not universal. Some authors (e.g. 57, 83) define a confounder more broadly as any variable for which adjustment is helpful in reducing bias in effect estimation. Under this broader definition, a covariate may be a confounder even if it is not a cause of the outcome, as long as the covariate is a surrogate (proxy) for such a cause. Variables that are confounders by virtue of their effects on the outcome parameter (as in the previous definition) are then called causal confounders. For example, a proxy confounder might be affected by a causal confounder and be a determinant of treatment.

It is important to recognize that the necessary conditions of a confounder discussed above apply to a source population of persons at risk of becoming study cases (i.e. of contributing an outcome event to the study data). Thus, we cannot necessarily depend on associations observed in our data to determine whether a given covariate is a confounder of a particular effect if the entire source population is not observed (as in most case-control studies). Furthermore, even if we observe the entire source population, we cannot be sure whether a covariate satisfies or fails condition *a* (i.e. whether it is a cause of the outcome) because we only observe the association of that covariate with the outcome, and that association may itself be confounded or otherwise biased as an effect estimate. For example, observing no association between a risk factor and disease status among unexposed subjects does not indicate that the factor is not a confounder, for that association may itself be confounded for the effect of that factor. Because of such problems, we must rely on prior knowledge of these associations and effects to identify confounders in a study (e.g. 36, 57, 66, 83).

Another limitation in applying conditions *a–c* to the identification of confounders in observational research is that application of these conditions to each covariate (potential confounder) must be made conditional on all other potential confounders being considered. Whether it is desirable to control for a certain covariate (to reduce bias) depends on what other covariates are being controlled by the

investigator. Thus, when we cannot identify all potential confounders and specify the causal effects among them, our definition (indeed, any definition) of confounder becomes conditional on what else has been controlled (see Sufficient Control). This complexity reflects the notion that the concept of confounding is more fundamental than is the concept of confounder (36). [For extensions of the above ideas to regression models, see chapter 20 of Rothman & Greenland (89), and Greenland et al (38).]

CONTROL OF CONFOUNDING

Control via Design

Perhaps the most obvious way to avoid confounding is to obtain a reference population B for which μ_{B0} is known to equal μ_{A0}. Among epidemiologists, such a population is sometimes said to be comparable to or exchangeable with A when considering the outcome under the reference treatment. In practice, such a population may be difficult or impossible to find. Thus, an investigator may attempt to construct such a population or to construct exchangeable index and reference populations. These constructions may be viewed as design-based methods for the control of confounding.

Restriction and Matching Perhaps no approach is more effective for preventing confounding by a known risk factor than restriction. For example, gender imbalances cannot confound a study restricted to women. Nonetheless, restriction on many factors can reduce the number of available subjects to unacceptably low levels and may greatly reduce the generalizability of results as well. Matching the treatment populations on confounders overcomes these drawbacks and, if successful, can be as effective as restriction. For example, gender imbalances cannot confound a study in which the compared exposure groups have identical proportions of women. Unfortunately, differential losses to observation may undo the initial covariate balances produced by matching. Another problem is that matches may become difficult or impossible to find if one attempts to match on more than a few factors.

Although matching on confounders can reduce bias in observational studies, the statistical advantage of matching is not to control for confounders, which can be done in the analysis without matching (see below), but to control for these confounders more efficiently (with less random error) than if matching had not been used (e.g. 89, pp. 147–61). Because the process of matching differs for cohort studies (unexposed subjects are matched to exposed subjects) and case-control studies (controls are matched to cases), the relative gain or loss in efficiency by matching differs by study design. Furthermore, in case-control studies, matching does not alter the source population, and matching on a correlate of the exposure introduces a selection bias that must be corrected in the analysis by controlling for the matching variables.

Randomization Neither restriction nor matching prevents (although they may diminish) imbalances on unrestricted, unmatched, or unmeasured covariates. In contrast, randomized treatment allocation (randomization) offers a means of dealing with confounding by covariates not explicitly accounted for by the design. It must be emphasized, however, that this solution is only probabilistic and subject to severe practical constraints. For example, protocol violations (e.g. noncompliance) and loss to follow-up may produce systematic covariate imbalances between the groups (and consequent confounding), and random imbalances may be severe, especially if the study size is small (22, 88). Blocked (stratified) randomization can help ensure that random imbalances on the blocking factors will not occur, but it does not guarantee balance of unblocked factors. Thus, even in a perfectly executed randomized trial, the no-confounding condition, $\mu_{A0} = \mu_{B0}$, is not a realistic assumption for inferences about causal effects. Successful randomization simply ensures that the difference, $\mu_{A0} - \mu_{B0}$, and hence the bias due to confounding, has expectation zero and converges to zero under the randomization distribution; it also provides a permutation distribution for causal inferences (17, 22, 86).

Exchangeability Under randomization, the parameters μ_{A0} and μ_{B0} (and μ_{A1} and μ_{B1} as well) are outcomes of a known random process and so can be treated as objective random variables (though μ_{A0} and μ_{B1} remain unobserved). Successful randomization also renders μ_{A0} and μ_{B0} unconditionally exchangeable in the subjective probabilistic sense (15), and it renders μ_{A1} and μ_{B1} exchangeable. These consequences of randomization imply that any bias due to confounding is random with a known distribution; therefore, randomization permits derivation of statistical procedures for estimating treatment effects, e.g. by substituting μ_{B0} for μ_{A0} and then allowing for random differences between μ_{A0} and μ_{B0} (73). This benefit applies regardless of what the parameter μ represents, i.e. randomization yields exchangeability for all parameters of the outcome distribution. In addition, it can be argued that randomization should lead us to use the entire (treated plus untreated) study group as the target population, rather than just the treated (exposed) group (73).

Without randomization, one can still view μ_{A0} and μ_{B0} as random variables from a Bayesian perspective, and a practical and sufficient design-based approach to confounding when estimating effects on the exposed study group (group A) is to find or construct comparison groups such that μ_{A0} and μ_{B0} are exchangeable. This perspective translates into the traditional advice to search for "natural experiments" (i.e. situations in which a compelling argument can be made that the exposure was effectively randomized by natural circumstances).

Control via Analysis

Design-based methods are often infeasible or insufficient to produce exchangeability. Thus, there has been an enormous amount of work devoted to analytic

adjustments for confounding. With a few exceptions, these methods are based on observed covariate distributions in the compared populations. Such methods will successfully control confounding only to the extent that enough confounders are accurately measured and employed in the analysis. Then, too, many methods employ parametric models at some stage, and their success thus depends on the faithfulness of the model to reality. There is a tension between the demands of adjusting for enough covariates and the dependence of the analysis on modeling assumptions. This issue cannot be covered in depth here, but a few basic points are worth noting.

The simplest methods of adjustment begin with stratification on confounders. A covariate cannot be responsible for confounding within a stratum that is internally homogeneous with respect to that covariate. This is so, regardless of whether the covariate was used to define the stratum. For example, gender imbalances cannot confound observations within a stratum composed solely of women. It would seem natural, then, to control confounding due to measured factors by simply stratifying on them all. Unfortunately, one would then confront the well-known sparse-data problem: Given enough factors, few if any strata would have subjects in both treatment groups, thereby making comparisons biased, inefficient, or impossible (38a, 79).

One solution to this sparse-data problem begins by noting that within-stratum homogeneity on a covariate is unnecessary to prevent confounding by that covariate. Within-stratum balance is sufficient, because comparisons within a stratum cannot be confounded by a covariate that is not associated with treatment within the stratum. Hence, a given stratification should be sufficient to control confounding by a set of covariates if the covariates are balanced across the strata, i.e. unassociated with treatment within the strata. Subject to any modeling restrictions used for score estimation, balance in probability for a set of covariates could be achieved by exact stratification on the estimated propensity score, where the propensity score is defined as the probability of treatment given the covariates in the combined (treated and untreated) study population (87). They further showed that this score was the coarsest score that would produce balance in probability. Stratification on the estimated propensity score thus reduces adjustment for multiple covariates to stratification on a single variable and lowers the risk of sparse-data problems if the model used for propensity scoring is correct. Unfortunately, in sparse data there may be little power to test whether the model is correct.

The most common method for avoiding sparse-data problems is to use regression models for the dependence of the outcome on the treatment and covariates; such strategies are described in many textbooks (e.g. 8, 50, 89). Hybrid methods that combine regressions on treatment and outcome have also been developed [see Robins & Greenland (82) and Rosenbaum (86) for examples]. Nonetheless, theoretical results indicate that no approach can completely solve sparse-data problems, insofar as sample size will always limit the number of degrees of freedom available for covariate adjustment (84), although flexibility in using these degrees

of freedom can be greatly improved via hierarchical regression (mixed or multilevel modeling) (32, 32a).

Sufficient Control

Without randomization, the evaluation of within-stratum or residual confounding becomes a major concern. For this purpose, we define a stratification on a set of variables as sufficient for estimation of stratum-specific causal effects if, within strata, μ_{A0} and μ_{B0} are exchangeable. Randomization ensures sufficiency of the set of measured variables not affected by treatment. In the absence of randomization, however, causal inferences become dependent on and sensitive to the assumption that the set of variables available for analysis is sufficient. It is almost always possible that this set is insufficient because some confounder essential for sufficiency has not been recorded; thus, causal inferences from observational studies almost always hinge on subject-matter priors ("judgments") about unmeasured confounders. Sensitivity of results to possible unmeasured confounders can be assessed via formal sensitivity analysis (14, 85, 86).

There are several methods for deducing the implications of background assumptions. For example, assumptions about the directions and absences of causal relations among variables (measured and unmeasured) can be conveniently encoded in a causal graph or path diagram, in which arrows (directed arcs) represent cause-effect relations. Conditional on the assumptions underlying the graph, the question of sufficiency of a set of variables (such as the set of measured variables) can be easily answered using a simple graphical algorithm called the "back-door criterion" (35, 64, 66). The same algorithm allows one to determine whether subsets of a sufficient set are themselves sufficient. Thus, by sequential deletion of variables from the original set and application of the criterion to the reduced subsets, we may identify minimally sufficient subsets (i.e. sufficient subsets with no sufficient proper subsets). The need for such identification arises, for example, in epidemiologic studies in which numerous "lifestyle" covariates (diet, physical activity, smoking and drinking habits, etc) are measured and are potential confounders of the effect under study. Here, the total set of covariates may be sufficient for control as defined above, but impractical to control in its entirety, even when using propensity score or outcome-regression methods (38). Graphical identification of sufficient subsets operates on background assumptions rather than data. An analogous statistical approach was proposed by Robins (77).

A set *S* that is sufficient for estimating stratum-specific effects will also be sufficient for estimating a summary measure of the effect of treatment on the entire target population. The converse is not true, however: Stratum-specific confounding may be in opposite directions across strata and thus "average out" within the summary measure. Consequently, a set *S* may be sufficient for estimating a summary effect even though insufficient for estimating stratum-specific effects (36). This notion is formalized in the discussion of residual confounding given by Greenland et al (38).

COLLAPSIBILITY

Consider the $I \times J \times K$ contingency table representing the joint distribution of three discrete variables X (exposure), Y (outcome), and Z (covariate), the $I \times J$ marginal table representing the joint distribution of X and Y, and the set of conditional $I \times J$ subtables (strata) representing the joint distributions of X and Y within levels of Z. Generalizing Whittemore (100) (who considered log-linear model parameters), we say a measure of association of X and Y is strictly collapsible across Z if it is constant across the strata (subtables) and this constant value equals the value obtained from the marginal table (ignoring Z).

Noncollapsibility (violation of collapsibility) is sometimes referred to as Simpson's paradox, after a celebrated article by Simpson (96). This phenomenon had been discussed by earlier authors, including Yule (102; see also 11). Some statisticians reserve the term Simpson's paradox to refer to the special case of noncollapsibility in which the conditional and marginal associations are in opposite directions, as in Yule's and Simpson's numerical examples. Simpson's algebra and discussion, however, dealt with the general case of inequality. The term collapsibility seems to have arisen in later work (see 6).

Table 2 provides some simple examples. The difference of probabilities that $Y = 1$ (the risk difference) is strictly collapsible. Nonetheless, the ratio of probabilities that $Y = 1$ (the risk ratio) is not strictly collapsible because the risk ratio varies across the Z strata, and the odds ratio is not collapsible because its marginal value does not equal the constant conditional (stratum-specific) value. Thus, collapsibility depends on the chosen measure of association.

Now suppose that a measure is not constant across the strata, but that a particular summary of the conditional measures does equal the marginal measure. This summary is then said to be collapsible across Z. As an example, in Table 2 the

TABLE 2 Examples of collapsibility and noncollapsibility in a three-way distribution[a]: X, exposure; Y, outcome; Z, covariate

	$Z = 1$		$Z = 0$		**Total**	
	$X = 1$	$X = 0$	$X = 1$	$X = 0$	$X = 1$	$X = 0$
Y = 1	0.20	0.15	0.10	0.05	0.30	0.20
Y = 0	0.05	0.10	0.15	0.20	0.20	0.30
Risk[b]	0.80	0.60	0.40	0.20	0.60	0.40
Risk difference	0.20		0.20		0.20	
Risk ratio	1.33		2.00		1.50	
Odds ratio	2.67		2.67		2.25	

[a]Cell values are proportions of the total population.

[b]Probability of $Y = 1$ given X and Z.

ratio of risks standardized to the marginal distribution of Z is

$$\{[P(Z=1)P(Y=1|X=1,Z=1)]+[P(Z=0)P(Y=1|X=1,Z=0)]\}/$$
$$\{[P(Z=1)P(Y=1|X=0,Z=1)]+[P(Z=0)P(Y=1|X=0,Z=0)]\}$$
$$=[0.50(0.80)+0.50(0.40)]/[0.50(0.60)+0.50(0.20)]$$
$$=1.50, \qquad 4.$$

which is equal to the marginal (crude) risk ratio. Thus, both the risk ratio and risk difference are collapsible in Table 2 because there is no association in the total sample between Z and X, i.e. the same proportion (50%) of persons in each stratum of Z is exposed ($X = 1$). Various tests of collapsibility and strict collapsibility have been developed for polytomous variables and multidimensional tables (3, 19, 27, 34, 100), and extensions to regression models have also been given (9, 43).

Confounding Vs Noncollapsibility

Much of the statistics literature does not distinguish between the concept of confounding as a bias in effect estimation and the concept of noncollapsibility. Nonetheless, the two concepts are distinct: For certain effect parameters, confounding may occur with or without noncollapsibility and noncollapsibility may occur with or without confounding (36, 38, 57, 101). Mathematically identical conclusions have been reached by other authors, albeit with different terminology in which noncollapsibility corresponds to "bias" and confounding corresponds to "covariate imbalance" (24, 42). As shown below, the counterfactual definition of confounding is nonparametric and specific to causal inference, whereas collapsibility depends on the choice of association parameter and requires no reference to causality, effects, or confounding.

Noncollapsibility Without Confounding Table 3 gives the response distributions under treatments x_1 and x_0 for a hypothetical target population A, and the response distribution under treatment x_0 for a hypothetical reference population B. Suppose A receives treatment x_1, B receives x_0, we wish to estimate the effect that receiving x_1 rather than x_0 had on A, and Z is unaffected by treatment. If we take the odds of response as the outcome parameter μ, ignoring the covariate Z, we get $\mu_{A1} = 0.6/(1-0.6) = 1.50$, and $\mu_{A0} = \mu_{B0} = 0.4/(1-0.4) = 0.67$. Hence, there is no confounding of the odds ratio by Z: $\mu_{A1}/\mu_{A0} = \mu_{A1}/\mu_{B0} = 1.50/0.67 = 2.25$ (just as there is no confounding of the risk ratio and the risk difference by Z). Nonetheless, the covariate Z is associated with response in A and B. Furthermore, the odds ratio is not collapsible over Z: Within levels of Z, the odds ratios, comparing A under treatment x_1 to either A or B under x_0,

TABLE 3 Distribution of responses (Y) for hypothetical index population A under treatments x_1 and x_0, and for reference population B under treatment x_0: example of noncollapsibility without confounding of the odds ratio

	Response probability ($Y = 1$) if		
Stratum	$\mathbf{X = x_1}$	$\mathbf{X = x_0}$	**Stratum size**
Population A			
$Z = 1$	0.8	0.6	1000
$Z = 0$	0.4	0.2	1000
Total	0.6	0.4	
Population B			
$Z = 1$	NU[a]	0.6	1000
Z = 0	NU	0.2	1000
Total	NU	0.4	

[a]NU, Not used in example.

are $(0.8/0.2)/(0.6/0.4) = (0.4/0.6)/(0.2/0.8) = 2.67$, which is higher than the unconditional (crude) odds ratio of 2.25 obtained when Z is ignored.

The preceding example illustrates a peculiar property of the odds ratio as an effect measure: Treatment x_1 (relative to x_0) elevates the odds of response by 125% in population A, yet within each stratum of Z it raises the odds by 167%. If Z is associated with response conditional on treatment but unconditionally unassociated with treatment, the stratum-specific odds ratios must be farther from 1 than the unconditional odds ratio if the latter is not 1 (25, 42). This phenomenon is often interpreted as a "bias" in the unconditional odds ratio, but in fact there is no bias if one takes care to not misinterpret the unconditional effect as an estimate of the stratum-specific or individual effects (29, 57).

Confounding Without Noncollapsibility To create a numerical example in which the odds ratio is collapsible and yet is confounded for the overall effect, we need only modify Table 3 slightly, e.g. by changing the stratum size for $Z = 0$ in population B to 1500. With this change, the proportion with $Z = 1$ in population B drops from $1000/2000 = 0.5$ to $1000/2500 = 0.4$, the unconditional response probability in population B under treatment x_0 drops from 0.4 to $0.4(0.6) + 0.6(0.2) = 0.36$, and the unconditional response odds μ_{B0} in population B under x_0 becomes $0.36/(1 - 0.36) = 0.5625$. Thus, $\mu_{B0} = 0.5625 < 0.67 = \mu_{A0}$, with consequent confounding of the odds ratio by Z: μ_{A1}/μ_{A0}, the true effect, equals 2.25 (as before), which is less than the unconditional odds ratio $\mu_{A1}/\mu_{B0} = 1.50/0.5625 = 2.67$. (Similarly, the risk difference and risk ratio are also confounded.) Nonetheless, this unconditional odds ratio equals the stratum-specific odds ratios in population A, which are unchanged from the previous example.

Conditions for Equivalence

The example in Table 3 shows that when μ is the odds of the outcome, μ_{A0} may equal μ_{B0} (no confounding) even when the odds ratio is not collapsible over the confounders. Conversely, the modified example shows that we may have $\mu_{A0} \neq \mu_{B0}$ even when the odds ratio is collapsible. A probabilistic explanation of the discrepancy between nonconfounding and collapsibility is that μ_{A0} will equal μ_{B0} whenever Z is sufficient for control and is unconditionally unassociated with treatment, as in Table 3, whereas collapsibility of the odds ratio will occur whenever Z is unassociated with treatment conditional on response, as in the modified example (6). Thus, the discrepancy is just a consequence of the nonequivalence of unconditional and conditional associations.

If the effect measure is the difference or ratio of response proportions, results from Gail (24) imply that this measure will be collapsible over Z if Z has the same distribution in A and B (i.e. if Z and treatment are unconditionally unassociated). It follows that, when examining such measures, the above phenomena (noncollapsibility without confounding and confounding without noncollapsibility) cannot occur if Z is sufficient for control. More generally, when the effect measure can be expressed as the average effect on population members, the conditions for noncollapsibility and confounding will be identical, provided the covariates in question form a sufficient set for control. In such cases, noncollapsibility and confounding become equivalent, which may explain why the two concepts are often not distinguished. The nonequivalence of the two concepts for odds ratios simply reflects the fact that the unconditional effect of a treatment on the odds is not the average treatment effect on population members (29).

CONFOUNDING IN INTERVENTION STUDIES: Further Issues

In this section we briefly discuss some special issues of confounding that arise in studies of interventions, such as clinical trials and natural experiments.

Adjustment in Randomized Trials

Some controversy has existed about adjustment for random covariate imbalances in randomized trials. Although Fisher asserted that randomized comparisons were "unbiased," he also pointed out that they could be confounded in the sense used here (e.g. 22, p. 49). Fisher's use of the word unbiased was based on what would be expected before the randomization was carried out; therefore, it is of little guidance for analysis of a given trial. Some arguments for accounting for the actual result of the randomization process are given in Greenland & Robins (36) and Robins & Morgenstern (83). Other arguments for adjustment in randomized trials have been given by Rothman (88), Miettinen & Cook (57), and Senn (94).

Intent-to-Treat Analysis

In a randomized trial, noncompliance can easily lead to confounding in comparisons of the groups actually receiving treatments x_1 and x_0. One somewhat controversial solution to noncompliance problems is intent-to-treat analysis, which defines the comparison groups *A* and *B* by treatment assigned rather than treatment received. Detractors of intent-to-treat analysis consider it an attempt to define away a serious problem, especially when treatment received is the treatment of scientific interest. Supporters of intent-to-treat analysis emphasize that intent-to-treat tests (tests of assigned-treatment effects) remain valid tests of received-treatment effects under broader conditions than conventional tests of received-treatment effects [for a discussion of these and related issues, see Goetghebeur & van Houwelingen (28)].

A crucial point is that confounding can affect even intent-to-treat analyses. For example, apparently random assignments may not be random, as when blinding is insufficient to prevent the treatment providers from protocol violations or when there is differential loss to follow-up. Even when these problems do not occur, random imbalances remain possible. A more subtle problem is that noncompliance can produce bias away from the null in an intent-to-treat analysis of a trial that examines whether two treatments are equivalent (i.e. an equivalence trial) (78). To illustrate, suppose treatments *A* and *B* are both 100% effective and thus completely equivalent with respect to their effect on the outcome, so that the equivalence null is satisfied. Suppose, however, that treatment *A* causes a harmless but unpleasant flushing sensation, whereas treatment *B* does not; consequently, compliance is 70% for *A* but 100% for treatment *B*. Then the intent-to-treat test will reject the null hypothesis of equivalence solely because of the lower compliance with treatment *A*. Thus, in this example, noncompliance confounds the intent-to-treat analysis away from the correct null hypothesis of equivalence. Many authors have proposed instrumental-variable methods to adjust for possible bias due to noncompliance (e.g. 2, 4, 77) [see Greenland (33) for a nontechnical overview of these methods].

CONCLUSION

Concepts of confounding have been discussed by philosophers and scientists for centuries. It is only in more recent decades, however, that precise formal definitions of these concepts have emerged. These developments underscore the importance of subject-matter (prior) knowledge in making causal inferences from observational data, and they make explicit the distinction between counterfactual and collapsibility-based concepts of confounding: The counterfactual definition of confounding is nonparametric and specific to causal inference, whereas collapsibility depends on the choice of association parameter and requires no reference to causality or effects. Given its importance to causal inference, we recommend a more thorough discussion of confounding in all types of public-health education.

Most of our discussion has assumed that both the treatment variable and the confounders can be fully characterized by fixed covariates. Further subtleties can arise when these variables are time-dependent (see 67, 70–72, 77). We also have not considered issues of confounding in separating indirect and direct effects of treatments or exposures on outcome, i.e. effects mediated vs effects not mediated by measured covariates (for discussions of these issues, see 65, 67, 70, 77, 81, 82).

We wish to end on the cautionary note that confounding is but one of many problems that plague studies of cause and effect. Biases of comparable or even greater magnitude can arise from measurement errors, selection (sampling) biases, and systematically missing data, as well as from model-specification errors. Even when confounding and other systematic errors are absent, individual causal effects will remain unidentified by statistical observations (37, 38, 80). It remains a serious challenge to create a theory that can encompass all these problems coherently and also yield practical methods for data analysis.

ACKNOWLEDGMENT

This article was adapted from "Confounding and collapsibility in causal inference" (38) by S. Greenland, J. Robins, and J. Pearl, which appeared in volume 14 of *Statistical Science*, pp. 29–46. Copyright 1999, Institute of Mathematical Statistics, reproduced by permission.

LITERATURE CITED

1. Aldrich J. 1995. Correlations genuine and spurious in Pearson and Yule. *Stat. Sci.* 10:364–76
2. Angrist JD, Imbens GW, Rubin DB. 1996. Identification of causal effects using instrumental variables. *J. Am. Stat. Assoc.* 91:444–72
3. Asmussen S, Edwards D. 1983. Collapsibility and response variables in contingency tables. *Biometrika* 70:567–78
4. Balke A, Pearl J. 1997. Bounds on treatment effects from studies with imperfect compliance. *J. Am. Stat. Assoc.* 92:1171–76
5. Berkane M, ed. 1997. *Latent Variable Modeling and Applications to Causality*. New York: Springer-Verlag
6. Bishop YMM, Fienberg SE, Holland PW. 1975. *Discrete Multivariate Anal.: Theory and Practice*. Cambridge, MA: MIT Press
7. Bross IDJ. 1967. Pertinency of an extraneous variable. *J. Chronic Dis.* 20:487–95
8. Clayton D, Hills M. 1993. *Statistical Models in Epidemiology*. New York: Oxford Univ. Press
9. Clogg CC, Petkova E, Haritou A. 1995. Statistical methods for comparing regression coefficients between models. *Am. J. Sociol.* 100:1261–305
10. Cochran WG, Cox GM. 1950. *Experimental Designs*. New York: Wiley
11. Cohen MR, Nagel E. 1934. *An Introduction to Logic and the Scientific Method*. New York: Harcourt, Brace
12. Cook TD, Campbell DT. 1979. *Quasi-Experimentation: Design and Anal. Issues for Field Settings*. Boston: Houghton Mifflin
13. Copas JB. 1973. Randomization models

for matched and unmatched 2 × 2 tables. *Biometrika* 60:467–76
14. Copas JB, Li HG. 1997. Inference for non-random samples. *J. R. Stat. Soc. Ser. B* 59:55–95
15. Cornfield J. 1976. Recent methodological contributions to clinical trials. *Am. J. Epidemiol.* 104:408–21
16. Cornfield J, Haenszel W, Hammond WC, Lilienfeld AM, Shimkin MB, Wynder EL. 1959. Smoking and lung cancer: recent evidence and a discussion of some questions. *J. Natl. Cancer Inst.* 22:173–203
17. Cox DR. 1958. *The Planning of Experiments*. New York: Wiley
18. Dawid AP. 2000. Causal inference without counterfactuals. *J. Am. Stat. Assoc.* 95:407–48
19. Ducharme GR, LePage Y. 1986. Testing collapsibility in contingency tables. *J. R. Stat. Soc. Ser. B* 48:197–205
20. Farr W. 1974 (1837). Vital statistics or statistics of health, sickness, diseases, and death. In *Mortality in Mid 19th Century Britain*, pp. 589–601. London: Gregg Int.
21. Fisher RA. 1918. The causes of human variability. *Eugenics Rev.* 10:213–20
22. Fisher RA. 1935. *The Design of Experiments*. Edinburgh: Oliver & Boyd
23. Frydenberg M. 1990. Marginalization and collapsibility in graphical statistical models. *Ann. Stat.* 18:790–805
24. Gail MH. 1986. Adjusting for covariates that have the same distribution in exposed and unexposed cohorts. In *Modern Statistical Methods in Chronic Disease Epidemiology*, ed. SH Moolgavkar, RL Prentice, pp. 3–18. New York: Wiley
25. Gail MH, Wieand S, Piantadosi S. 1984. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika* 71:431–44
26. Galles D, Pearl J. 1998. An axiomatic characterization of causal counterfactuals. *Found. Sci.* 4:151–82
27. Geng Z. 1989. Algorithm AS 299. Decomposability and collapsibility for log-linear models. *Appl. Stat.* 38:189–97
28. Goetghebeur E, van Houwelingen H., eds. 1998. Analyzing noncompliance in clinical trials. *Stat. Med.* 17:247–389
29. Greenland S. 1987. Interpretation and choice of effect measures in epidemiologic analyses. *Am. J. Epidemiol.* 125:761–68
30. Greenland S. 1990. Randomization, statistics, and causal inference. *Epidemiology* 1:421–29
31. Greenland S. 2000. Causal analysis in the health sciences. *J. Am. Stat. Assoc.* 95:286–89
32. Greenland S. 2000. When should epidemiologic regression use random coefficients? *Biometrics* 56:915–21
32a. Greenland S. 2000. Principles of multilevel modelling. *Int. J. Epidemiol.* 29:158–67
33. Greenland S. 2000. An introduction to instrumental variables for epidemiologists. *Int. J. Epidemiol.* 29:722–29
34. Greenland S, Mickey RM. 1988. Closed-form and dually consistent methods for inference on collapsibility in 2 × 2 × K and 2 × J × K tables. *Appl. Stat.* 37:335–43
35. Greenland S, Pearl J, Robins JM. 1999. Causal diagrams for epidemiologic research. *Epidemiology* 10:37–48
36. Greenland S, Robins JM. 1986. Identifiability, exchangeability, and epidemiological confounding. *Int. J. Epidemiol.* 15:413–19
37. Greenland S, Robins JM. 1988. Conceptual problems in the definition and interpretation of attributable fractions. *Am. J. Epidemiol.* 128:1185–97
38. Greenland S, Robins JM, Pearl J. 1999. Confounding and collapsibility in causal inference. *Stat. Sci.* 14:29–46
38a. Greenland S, Schwartzbaum JA, Finkle WD. 2000. Small-sample and sparse-data problems in conditional logistic regression. *Am. J. Epidemiol.* 151:531–39

39. Groves ER, Ogburn WF. 1928. *American Marriage and Family Relationships*. New York: Holt
40. Halloran ME, Struchiner CJ. 1995. Causal inference for infectious diseases. *Epidemiology* 6:142–51
41. Hamilton MA. 1979. Choosing a parameter for 2 × 2 table or 2 × 2 × 2 table analysis. *Am. J. Epidemiol.* 109:362–75
42. Hauck WW, Neuhas JM, Kalbfleisch JD, Anderson S. 1991. A consequence of omitted covariates when estimating odds ratios. *J. Clin. Epidemiol.* 44:77–81
43. Hausman J. 1978. Specification tests in econometrics. *Econometrica* 46:1251–71
44. Heckman JJ, Hotz VJ. 1989. Choosing among alternative nonexperimental methods for estimating the impact of social programs: the case of manpower training. *J. Am. Stat. Assoc.* 84:862–74
45. Hill AB. 1965. The environment and disease: association or causation? *Proc. R. Soc. Med.* 58:295–300
46. Holland PW. 1986. Statistics and causal inference. *J. Am. Stat. Assoc.* 81:945–70
47. Hume D. 1888 (1739). *A Treatise of Human Nature*. Oxford, UK: Oxford Univ. Press
48. Hume D. 1988 (1748). *An Enquiry Concerning Human Understanding*. LaSalle: Open Court
49. Kaufman JS, Cooper RS. 1999. Seeking causal explanations in social epidemiology. *Am. J. Epidemiol.* 150:113–20
50. Kelsey JL, Whittemore AS, Evans AS, Thompson WD. 1996. *Methods in Observational Epidemiology*. New York: Oxford Univ. Press. 2nd ed.
51. Kitagawa EM. 1955. Components of a difference between two rates. *J. Am. Stat. Assoc.* 50:1168–94
52. Lewis D. 1973. Causation. *J. Philos.* 70:556–67
53. Lewis D. 1973. *Counterfactuals*. Oxford, UK: Blackwell
54. MacMahon B, Pugh TF. 1967. Causes and entities of disease. In *Preventive Medicine*, ed. DW Clark, B MacMahon, pp. 11–18. Boston: Little, Brown
55. McKechnie JL, ed. 1979. *Webster's New Twentieth Century Dictionary*. New York: Simon & Schuster
56. Miettinen OS. 1972. Components of the crude risk ratio. *Am. J. Epidemiol.* 96:168–72
57. Miettinen OS, Cook EF. 1981. Confounding: essence and detection. *Am. J. Epidemiol.* 114:593–603
58. Mill JS. 1862. *A System of Logic, Ratiocinative and Inductive*. London: Parker, Son & Bowin. 5th ed.
59. Mill JS. 1956 (1843). *A System of Logic, Ratiocinative and Inductive*. London: Longmans, Green
60. Morgenstern H. 1997. Defining and explaining race effects. *Epidemiology* 8:609–11
61. Neuhaus JM, Kalbfleisch JD, Hauck WW. 1991. A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *Int. Stat. Rev.* 59:25–35
62. Neyman J. 1923. Sur les applications de la thar des probabilities aux experiences Agaricales: essay des principle. Transl. D Dabrowska, T Speed, 1990, *in Stat. Sci.* 5:463–72 (From French)
63. Neyman J. 1935. Statistical problems in agricultural experimentation. *J. R. Stat. Soc.* 2(Suppl.):107–80
64. Pearl J. 1995. Causal diagrams for empirical research. *Biometrika* 82:669–710
65. Pearl J. 1997. On the identification of nonparametric structural models. See Ref. 5, pp. 29–68
66. Pearl J. 2000. *Causality*. New York: Cambridge Univ. Press
67. Pearl J, Robins JM. 1995. Probabilistic evaluation of sequential plans from causal models with hidden variables. In *Uncertainty in Artificial Intelligence*, ed. P Besnard, S Hanks, 11:444–53. San Francisco: Morgan-Kaufman
68. Pearson K, Lee A, Bramley-Moore L.

1899. Mathematical contributions to the theory of evolution. VI. Genetic (reproductive) selection: inheritance of fertility in man, and of fecundity in thorough-bred racehorses. *Philos. Transact. R. Soc. London Ser. A* 192:257–330

69. Prentice RL, Kalbfleisch JD. 1988. Author's reply. *Biometrics* 44:1205
70. Robins JM. 1986. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math. Model.* 7:1393–512
71. Robins JM. 1987. Addendum to "A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect." *Comput. Math. Appl.* 14:923–45
72. Robins JM. 1987. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. *J. Chronic Dis.* 40(Suppl. 2):139–61S
73. Robins JM. 1988. Confidence intervals for causal parameters. *Stat. Med.* 7:773–85
74. Robins JM. 1989. The control of confounding by intermediate variables. *Stat. Med.* 8:679–701
75. Robins JM. 1995. Discussion of "Causal diagrams for empirical research" by J. Pearl. *Biometrika* 82:695–98
76. Robins JM. 1995. An analytic method for randomized trials with informative censoring. *Lifetime Data Anal.* 1:241–54
77. Robins JM. 1997. Causal inference from complex longitudinal data. See Ref. 5, pp. 69–117
78. Robins JM. 1998. Correction for noncompliance in equivalence trials. *Stat. Med.* 17:269–302
79. Robins JM, Greenland S. 1986. The role of model selection in causal inference from nonexperimental data. *Am. J. Epidemiol.* 123:393–402
80. Robins JM, Greenland S. 1989. The probability of causation under a stochastic model for individual risks. *Biometrics* 46:1125–38
81. Robins JM, Greenland S. 1992. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 3:143–55
82. Robins JM, Greenland S. 1994. Adjusting for differential rates of prophylaxis therapy for PCP in high versus low dose AZT treatment arms in an AIDS randomized trial. *J. Am. Stat. Assoc.* 89:737–49
83. Robins JM, Morgenstern H. 1987. The mathematical foundations of confounding in epidemiology. *Comput. Math. Appl.* 14:869–916
84. Robins JM, Ritov Y. 1997. Toward a curse-of-dimensionality appropriate (CODA) asymptotic theory for semi-parametric models. *Stat. Med.* 16:285–319
85. Robins JM, Rotnitzky A, Scharfstein DO. 1999. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In *Statistical Models in Epidemiology*, ed. E Halloran, pp. 1–94. New York: Springer-Verlag
86. Rosenbaum PR. 1995. *Observational Studies*. New York: Springer-Verlag
87. Rosenbaum PR, Rubin DB. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70:41–55
88. Rothman KJ. 1977. Epidemiologic methods in clinical trials. *Cancer* 39:1771–75
89. Rothman KJ, Greenland S. 1998. *Modern Epidemiology*. Philadelphia: Lippincott-Raven. 2nd ed.
90. Rubin DB. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* 66:688–701
91. Rubin DB. 1978. Bayesian inference for causal effects: the role of randomization. *Ann. Stat.* 7:34–58
92. Rubin DB. 1990. Comment: Neyman (1923) and causal inference in experiments and observational studies. *Stat. Sci.* 5:472–80

93. Rubin DB. 1991. Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism. *Biometrics* 47:1213–34
94. Senn S. 1989. Covariate imbalance and random allocation in clinical trials. *Stat. Med.* 8:467–75
95. Simon HA, Rescher N. 1966. Cause and counterfactual. *Philos. Sci.* 33:323–40
96. Simpson EH. 1987 (1951). The interpretation of interaction in contingency tables. In *The Evolution of Epidemiologic Ideas*, ed. S Greenland, pp. 103–7. Chestnut Hill, MA: ERI
97. Slud EV, Byar DP, Schatzkin DP. 1988. Dependent competing risks and the latent-failure model. *Biometrics* 44:1203–4
98. Stalnaker RC. 1968. A theory of conditionals. In *Studies in Logical Theory*, ed. N Rescher. Oxford, UK: Blackwell
99. Stone R. 1993. The assumptions on which causal inference rest. *J. R. Stat. Soc. Ser. B* 55:455–66
100. Whittemore AS. 1978. Collapsing multidimensional contingency tables. *J. R. Stat. Soc. Ser. B* 40:328–40
101. Wickramaratne PJ, Holford TR. 1987. Confounding in epidemiologic studies: the adequacy of the control group as a measure of confounding. *Biometrics* 43:751–65
102. Yule GU. 1903. Notes on the theory of association of attributes in statistics. *Biometrika* 2:121–34
103. Zizek F. 1913. *Statistical Averages: A Methodological Study*. Transl. WM Persons. New York: Holt (From German)

Annual Review of Public Health
Volume 22, 2001

CONTENTS