Statistics and Causal Inference

PAUL W. HOLLAND*

Problems involving causal inference have dogged at the heels of statistics since its earliest days. Correlation does not imply causation, and yet causal conclusions drawn from a carefully designed experiment are often valid. What can a statistical model say about causation? This question is addressed by using a particular model for causal inference (Holland and Rubin 1983; Rubin 1974) to critique the discussions of other writers on causation and causal inference. These include selected philosophers, medical researchers, statisticians, econometricians, and proponents of causal modeling.

KEY WORDS: Causal model; Philosophy; Association; Experiments; Mill's methods; Causal effect; Koch's postulates; Hill's nine factors; Granger causality; Path diagrams; Probabilistic causality.

1. INTRODUCTION

The reaction of many statisticians when confronted with the possibility that their profession might contribute to a discussion of causation is immediately to deny that there is any such possibility. "That correlation is not causation is perhaps the first thing that must be said" (Barnard 1982, p. 387). Possibly this evasive action is in response to all of those needling little headlines that pop up in the most unexpected places, for example, "If the statistics cannot relate cause and effect, they can certainly add to the rhetoric" (Smith 1980, p. 998).

One need only recall that a well-designed randomized experiment can be a powerful aid in investigating causal relations to question the need for such a defensive posture by statisticians. Randomized experiments have transformed many branches of science, and the early proponents of such studies were the same statisticians who founded the modern era of our field.

This article takes the view that statistics has a great deal to say about certain problems of causal inference and ought to play a more significant role in philosophical analyses of causation than it has heretofore. In addition, I will try to show why the statistical models used to draw causal inferences are distinctly different from those used to draw associational inferences.

The article is organized as follows. First, statistical models appropriate for associational and causal inferences will be discussed and compared. Then they will be applied to various ideas about causation that have been expressed by several writers on this subject. One difficulty that arises in talking about causation is the variety of questions that are subsumed under the heading. Some authors focus on the ultimate meaningfulness of the notion of causation. Others are concerned with deducing the causes of a given effect. Still others are interested in understanding the details of causal mechanisms. The emphasis here will be on *measuring the effects of causes* because this seems to be a place where statistics, which is concerned with measurement, has contributions to make. It is my opinion that an emphasis on the effects of causes rather than on the causes of effects is, in itself, an important consequence of bringing statistical reasoning to bear on the analysis of causation and directly opposes more traditional analyses of causation.

2. MODEL FOR ASSOCIATIONAL INFERENCE

The model appropriate for associational inference is simply the standard statistical model that relates two variables over a population. For clarity and for comparison with the model for causal inference described in the next section, however, I will briefly review association here. If I seem overly explicit in describing the model it is only because I wish to be absolutely clear on the fundamental elements of the theory presented here.

The model begins with a *population* or universe U of "units." A unit in U will be denoted by u. Units are the basic objects of study in an investigation. Examples of units are human subjects, laboratory equipment, households, and plots of land. A *variable* is simply a real-valued function that is defined on every unit in U. The value of a variable for a given unit u is the number assigned by some measurement process to u. A population of units and variables defined on these units are the basic elements of the models for both association and causation presented here. They correspond to the mathematical concepts of a set and real-valued functions defined on the elements of the set. They are the primitives of the theory and will not be further defined.

Suppose that for each unit u in U there is associated a value $Y(u)$ of a variable Y. Suppose further that Y is a variable of scientific interest in the sense that one wishes to understand why the values of Y vary over the units in U. Y is the *response variable* because of its status as a "variable to be explained." In making associational inferences one is satisfied with discovering how the values of Y are associated with the values of other variables defined on the units of U. Let A be a second variable defined on U. Distinguish A from Y by calling A an *attribute* of the units in u. Logically, however, A and Y are on an equal footing, since they are both simply variables defined on U.

All probabilities, distributions, and expected values involving variables are computed over U. A probability will mean nothing more nor less than a proportion of units in U. The expected value of a variable is merely its average value over all of U. Conditional expected values are averages over subsets of units where the subsets are defined by conditioning in the values of variables. It is in this sense that the models described here are population models.

The role of time needs to be mentioned here. Popula-

* Paul W. Holland is Director, Research Statistics Group, Educational Testing Service, Princeton, NJ 08541. A preliminary draft of this article was the basis of an invited General Methodology Lecture for the American Statistical Association, August 1985. The comments by Glymour and Granger included here were given at that session in response to that draft of this article.

Journal of the American Statistical Association
December 1986, Vol. 81, No. 396, Theory and Methods

tions of units exist within a time frame of some sort, and the measurements of characteristics of units that variables represent must also be made at particular times. For associational inference, however, the role of time is simply to affect the definition of the population of units or to specify the operational meaning of a particular variable. As we will see, in causal inference the role of time has a greater significance.

The most detailed information one can have in the model just described is the values of $Y(u)$ and $A(u)$ are all u in U. The *joint distribution* of Y and A over U is specified by $\Pr(Y = y, A = a)$ = proportion of u in U for which $Y(u) = y$ and $A(u) = a$.

The *associational parameters* are determined by this joint distribution. For example, the conditional distribution of Y given A is specified by $\Pr(Y = y \mid A = a) = \Pr(Y = y, A = a)/\Pr(A = a)$. This conditional distribution describes how the *distribution* of Y values changes over U as A varies. A typical associational parameter is the regression of Y on A, that is, the conditional expectation $E(Y \mid A = a)$.

Associational inference consists of making statistical inferences (estimates, tests, posterior distributions, etc.) about the associational parameters relating Y and A on the basis of data gathered about Y and A from units in U. In this sense, associational inference is simply descriptive statistics.

3. RUBIN'S MODEL FOR CAUSAL INFERENCE

Because experimentation is such a powerful scientific and statistical tool and one that often introduces clarity into discussions of specific cases of causation, I unabashedly draw on the language and framework of experiments for the model for causal inference. It is not that I believe an experiment is the *only* proper setting for discussing causality, but I do feel that an experiment is the *simplest* such setting. The purpose is to construct a model that is complex enough to allow us to formalize basic intuitions concerning cause and effect. The point of departure is the analysis of causal effects given in Rubin (1974, 1977, 1978, 1980). It will be sufficient for our purposes, however, to deal with a simplified, population-level version of Rubin's model. This simplified model was used in Holland and Rubin (1980) to analyze causal inference in retrospective, case-control studies used in medical research and in Holland and Rubin (1983) to analyze Lord's "analysis of covariance" paradox. I refer to this as "Rubin's model" even though Rubin would argue that the ideas behind the model have been around since Fisher. I think that Rubin (1974) was the place where these ideas were first applied to the study of causation.

This model also begins with a population of units, U. Units in the model for causal inference are the objects of study on which causes or treatments may act. The terms *cause* and *treatment* will be used interchangeably, and the notion that these terms convey is an important part of the model. It is important to realize that by using the terms cause and treatment interchangeably I do not intend to limit the discussion to the activities within a controlled randomized study. I do it to emphasize an idea that I believe receives insufficient attention in general discussions of causation. This is the fact that the effect of a cause is *always* relative to another cause. For example, the phrase "A causes B" almost always means that A causes B relative to some other cause that includes the condition "not A." The terminology becomes rather tortured if we try to stick with the usual causal language, but it is straightforward if we use the language of experiments—treatment (i.e., one cause) versus control (i.e., another cause). In Section 7 I will discuss the fundamental question of what kinds of things can be causes. The key notion, however, is the *potential* (regardless of whether it can be achieved in practice or not) for exposing or not exposing each unit to the action of a cause. For causal inference, it is critical that each unit be *potentially exposable* to any one of the causes. As an example, the schooling a student receives can be a cause, in our sense, of the student's performance on a test, whereas the student's race or gender cannot.

For simplicity it shall be assumed in this article that there are just two causes or levels of treatment, denoted by t (the treatment) and c (the control). Let S be a variable that indicates the cause to which each unit in U is exposed; that is, $S = t$ indicates that the unit is exposed to t and $S = c$ indicates exposure to c. In a controlled study, S is constructed by the experimenter. In an uncontrolled study, S is determined to some extent by factors beyond the experimenter's control. In either case, the critical feature of the notion of cause in this model is that the value of $S(u)$ for each unit *could have been different.*

The variable S is analogous to the variable A in Section 2, but with the essential difference that $S(u)$ indicates exposure of u to a specific cause, whereas $A(u)$ can indicate a property or characteristic of u. In this case the value of $A(u)$ could not have been different.

The role of time now becomes important because of the fact that when a unit is exposed to a cause this must occur at some specific time or within a specific time period. Variables now divide into two classes: pre-exposure—those whose values are determined prior to exposure to the cause; post-exposure—those whose values are determined after exposure to the cause.

The role of a response variable Y is to measure the effect of the cause, and thus response variables must fall into the post-exposure class. This gives rise to another critical element of the model. The values of post-exposure variables are potentially affected by the particular cause, t or c, to which the unit is exposed. This is nothing less than the statement that causes have effects, which is the very heart of the notion of causation. For the model to represent faithfully this state of affairs we need not a *single* variable, Y, to represent a response but *two* variables, Y_t and Y_c, to represent two potential responses. The interpretation of these two values, $Y_t(u)$ and $Y_c(u)$ for a given unit u, is that $Y_t(u)$ is the value of the response that would be observed if the unit were exposed to t and $Y_c(u)$ is the value that would be observed *on the same unit* if it were exposed to c.

The notation $Y_t(u)$ and $Y_c(u)$ is sometimes confusing

because a variable usually represents a measurement of some sort and a measurement is usually thought of as the result of a process that is applied to a unit. This is not really correct. For post-exposure variables the measurement is applied to the pairing (u, t) (i.e., u after exposure to t) or to (u, c) (i.e., u after exposure to c). A notation that more nearly expresses this joint dependence of Y on u and the exposed cause is $Y_t(u) = Y(u, t)$ and $Y_c(u) = Y(u, c)$. I shall use the Y_t, Y_c notation, however, since it leads to simpler expressions.

The effect of the cause t on u as measured by Y and relative to cause c is the difference between $Y_t(u)$ and $Y_c(u)$. In the model this will be represented by the algebraic difference

$$Y_t(u) - Y_c(u). \tag{1}$$

I shall call the difference (1) the causal effect of t (relative to c) on u (as measured by Y). Expression (1) is the way that the model for causal inference expresses the most basic of all causal statements. It says that treatment t causes the effect $Y_t(u) - Y_c(u)$ on unit U (relative to treatment c) or more simply that

$$t \text{ causes the effect } Y_t(u) - Y_c(u). \tag{2}$$

Causal inference is ultimately concerned with the effects of causes on specific units, that is, with ascertaining the value of the causal effect in (1). It is frustrated by an inherent fact of observational life that I call the Fundamental Problem of Causal Inference.

Fundamental Problem of Causal Inference. It is impossible to *observe* the value of $Y_t(u)$ and $Y_c(u)$ on the same unit and, therefore, it is impossible to *observe* the effect of t on u.

The emphasis is on the word *observe.* The impossibility of observing both $Y_t(u)$ and $Y_c(u)$ is self-evident in some examples and less clear in others. For example, if the unit u is a specific fourth grader, t represents a novel year-long program of study of arithmetic, c represents a standard arithmetic program, and Y is a score on a test at the end of the year, then it is evident that we could observe either $Y_t(u)$ or $Y_c(u)$ but not both. We will never observe what the effect of t was on u. On the other hand, if u is a room in a house, t means that I flick on the light switch in that room, c means that I do not, and Y indicates whether the light is on or not a short time after applying either t or c, then I might be inclined to *believe* that I can *know* the values of both $Y_t(u)$ and $Y_c(u)$ by simply flicking the switch. It is clear, however, that it is only because of the plausibility of certain assumptions about the situation that this *belief* of mine can be shared by anyone else. If, for example, the light has been flicking off and on for no apparent reason while I am contemplating beginning this experiment, I might doubt that I would know the values of $Y_t(u)$ and $Y_c(u)$ after flicking on the switch—at least until I was clever enough to figure out a new experiment!

The implicit threat of the Fundamental Problem of Causal Inference is that causal inference is impossible. But we should not jump to that conclusion too quickly. By asserting that the simultaneous *observation* of $Y_t(u)$ and $Y_c(u)$ is impossible I do not mean that knowledge relevant to these values is completely absent. It will depend on the situation considered. There are two general solutions to the Fundamental Problem, which for the sake of convenience I will label the *scientific solution* and the *statistical solution.*

The scientific solution is to exploit various homogeneity or invariance assumptions. For example, by studying the behavior of a piece of laboratory equipment carefully a scientist may come to believe that the value of $Y_c(u)$ measured at an earlier time is equal to the value of $Y_c(u)$ for the current experiment. All he needs to do now is to expose u to t and measure $Y_t(u)$ and he has overcome the Fundamental Problem of Causal Inference. Note, however, that this hypothetical scientist has made an untestable homogeneity assumption. By careful work he may convince himself and others that this assumption is right, but he can never be absolutely certain. Science has progressed very far by using this approach. The scientific solution is a commonplace aspect of our everyday life as well. We all use it to make the causal inferences that arise in our lives. These ideas are amplified in Sections 4.1 and 4.2.

The statistical solution is different and makes use of the population U in a typically statistical way. The *average causal effect*, T, of t (relative to c) over U is the expected value of the difference $Y_t(u) - Y_c(u)$ over the u's in U; that is,

$$E(Y_t - Y_c) = T. \tag{3}$$

T defined in (3) is the average causal effect. By the usual rules of probability (3) may also be expressed as

$$T = E(Y_t) - E(Y_c). \tag{4}$$

Although this does not look like much, (4) reveals that information on *different* units that *can be observed* can be used to gain knowledge about T. For example, if some units are exposed to t they may be used to give information about $E(Y_t)$ (because this is the mean value of Y_t over U), and if other units are exposed to c they may be used to give information about $E(Y_c)$. Formula (4) is then used to gain knowledge about T. The exact way that units would be selected for exposure to t or c is very important and involves all of the usual considerations of good statistical design of experiments. The important point is that the statistical solution replaces the impossible-to-observe causal effect of t on a specific unit with the possible-to-estimate *average* causal effect of t over a population of units. These ideas will be developed further in Sections 4.3 and 4.4.

The usefulness of either the scientific or the statistical solution to the Fundamental Problem of Causal Inference depends on the truth of different sets of untestable assumptions. In Section 4 I will discuss some of the typical assumptions that are often used to overcome the Fundamental Problem of Causal Inference.

It is useful to have a notation to express the fact that the causal indicator variable S determines which value, Y_t or Y_c, is observed for a given unit. If $S(u) = t$, then $Y_t(u)$ is observed, and if $S(u) = c$, then $Y_c(u)$ is observed. Thus

the observed response on unit u is $Y_{S(u)}(u)$. The *observed response variable* is, therefore, Y_S. Hence, even though the model contains three variables, S, Y_t, and Y_c, the process of observation involves only two, that is, S, Y_S. The distinction between (a) the measurement process, Y, that produces the response variable; (b) the two versions of the response variable Y_t and Y_c that corresponds to which cause the unit is exposed (and in terms of which causal effects are defined); and (c) the *observed* response variable Y_S, is very important and, often, is not made in discussions of causation. These distinctions never arise in the study of simple association, but they are crucial to the analysis of causation.

It is useful to review the model for associational inference and Rubin's model side by side to emphasize their differences. Both involve a population of units, U, and both involve two *observable* variables: (A, Y) for association and (S, Y_S) for causation. This is all, however, that they have in common. Whereas A and Y are simply variables defined on the units of U, S and Y_S presuppose a more complicated structure in order for them to apply to real situations. Two or more causes (or treatments) must be exposable to all of the units, and the response Y must be a post-exposure variable in order for the observed response Y_S to be defined. Associational inference involves the joint or conditional distributions of values of Y and A, and causal inference concerns the values $Y_t(u) - Y_c(u)$ on individual units. Causal inferences proceed from the observed values of S and Y_S and from assumptions that address the Fundamental Problem of Causal Inference but that are usually untestable. Causal inferences do not necessarily involve statistical inferences, but associational inferences almost always do.

4. SOME SPECIAL CASES OF CAUSAL INFERENCE

This section considers some simple special cases of Rubin's model for causal inference. The purpose is to show how specific assumptions added to the model allow causal inferences of particular types.

4.1 Temporal Stability and Causal Transience

One way of applying the scientific solution to the Fundamental Problem of Causal Inference is to assume that (a) the value of $Y_c(u)$ does not depend on *when* the sequence "apply c to u then measure Y on u" occurs and (b) the value of $Y_t(u)$ is not affected by the prior exposure of u to the sequence in (a). When these two assumptions are plausible it is a simple matter to measure $Y_t(u)$ and $Y_c(u)$ by sequential exposure of u to c then t, measuring Y after each exposure. The first assumption is *temporal stability*, because it asserts the constancy of response over time. The second assumption is *causal transience*, because it asserts that the effect of the cause c and the measurement process that results in $Y_c(u)$ is transient and does not change u enough to affect $Y_t(u)$ measured later. These two assumptions often apply to physical devices and are routinely made by all of us in everyday life—for example, in the "light switch" example mentioned earlier.

4.2 Unit Homogeneity

A second way of applying the scientific solution to the Fundamental Problem is to assume that $Y_t(u_1) = Y_t(u_2)$ and $Y_c(u_1) = Y_c(u_2)$ for two units u_1 and u_2. This is the assumption of *unit homogeneity*. It, too, is often applicable to work done in scientific laboratories and is also a causal workhorse of everyday life. The causal effect of t is taken to be the value of $Y_t(u_1) - Y_c(u_2)$. One way that laboratory scientists convince themselves that the units are homogeneous is to prepare them carefully so that they "look" identical in all relevant aspects. This, of course, cannot prove that the unit homogeneity assumption is valid, but it can make this assumption plausible.

4.3 Independence

In my discussion of the statistical solution to the Fundamental Problem, I did not give any specification to the way that units might be selected for observation of Y_t or Y_c. I only indicated that it was very important. Of course, the most well-known way that this occurs in experimental work is by randomization, and this section is concerned with that topic.

The supposition in using the statistical solution is that the population U does not consist of one or two units but is "large" in some sense. The observed data for each unit are values of the pair of variables (S, Y_S).

The average causal effect T is the difference between the two expected values $E(Y_t)$ and $E(Y_c)$. The observed data (S, Y_S), however, can only give us information about

$$E(Y_S \mid S = t) = E(Y_t \mid S = t) \tag{5}$$

and

$$E(Y_S \mid S = c) = E(Y_c \mid S = c). \tag{6}$$

It is important to recognize that $E(Y_t)$ and $E(Y_t \mid S = t)$ are *not* the same thing and need not have the same values in general [similarly for $E(Y_c)$ and $E(Y_c \mid S = c)$]. To state this difference in words, $E(Y_t)$ is the average value of $Y_t(u)$ over all u in U, where $E(Y_t \mid S = t)$ is the average value of $Y_t(u)$ over only those in u in U that were exposed to t. There is no reason why, in general, these two averages should be equal. For example, if $S(u) = t$ for all units for which $Y_t(u)$ is small, then $E(Y_t \mid S = t)$ will be smaller than $E(Y_t)$.

There is, however, an assumption that, if plausible, makes these two expected values equal. It is the assumption of *independence*. When units are assigned at random either to cause t or to cause c, certain physical randomization processes are carried out so that the determination of which cause (t or c) u is exposed to is regarded as statistically independent of all other variables, including Y_t and Y_c. This means that if the physical randomization is carried out correctly, then it is plausible that S is independent of Y_t and Y_c and all other variables over U. This is the *independence assumption*. If this assumption holds, then we have the basic equations

$$E(Y_t) = E(Y_t \mid S = t) \tag{7}$$

and

$$E(Y_c) = E(Y_c \mid S = c). \quad (8)$$

Hence under the independence assumption the average causal effect T satisfies the equation

$$T = E(Y_S \mid S = t) - E(Y_S \mid S = c). \quad (9)$$

The data (S, Y_S) can now be used to estimate T by taking the difference between the average value of the observed response Y_S for the units with $S = t$ and for the units with $S = c$. Hence, if randomization is possible, the average causal effect T can always be estimated. If U is large, T can be estimated with high accuracy.

It is useful to have a name for the right side of Equation (9) even when the assumption of independence does not hold. I will call it the *prima facie causal effect* of t (relative to c) and denote it by

$$T_{PF} = E(Y_t \mid S = t) - E(Y_c \mid S = c), \quad (10)$$

which is algebraically equal to the following function of the regression of Y_S on S:

$$T_{PF} = E(Y_S \mid S = t) - E(Y_S \mid S = c). \quad (11)$$

The term *prima facie causal effect* is adapted from Suppes (see Sec. 5) and used here to distinguish (11) from the *true* average causal effect, T, defined in Equation (3). The prima facie causal effect is an associational parameter for the joint distribution of the observable pair (Y_S, S). In general, the average causal effect T does not equal the prima facie causal effect T_{PF}. The assumption of independence, however, does allow the conclusion that $T = T_{PF}$, that is, Equation (9).

4.4 Constant Effect

The value of the average causal effect T is of potential interest for its own sake in certain types of studies. It would be of interest to a state education director who wanted to know what reading program would be the best to give to all of the first graders in his state. The average causal effect of the best program would be reflected in increases in statewide average reading scores.

The average causal effect T is an *average* and as such enjoys all of the advantages and disadvantages of averages. For example, if the variability in the causal effects $Y_t(u) - Y_c(u)$ is large over U, then T may not represent the causal effect of a specific unit, u_0, very well. If u_0 is the unit of interest, then T may be irrelevant, no matter how carefully we estimate it!

The assumption of *constant effect* is that the effect of t on every unit is the same, and under this assumption we have the equation

$$T = Y_t(u) - Y_c(u), \quad \text{for all } u \text{ in } U. \quad (12)$$

Hence under the assumption of constant effect T *is* the causal effect for every unit in U. This assumption is also called *additivity* in statistical models for experiments because the treatment t *adds* a constant amount T to the control response for each unit.

The assumption of constant effect makes the value of the average causal effect relevant to every unit and, therefore, allows T to be used to draw causal inferences at the unit level.

The assumption of constant effect can be partially checked in the same way that the additivity assumption is usually investigated. For example, U can be divided into subpopulations $U_1, U_2, \ldots$, and on each U_i the average causal effect can be estimated, $T_1, T_2, \ldots$. If the T_i's vary, the constant effect assumption cannot hold. If the T_i's do not vary, then the constant effect assumption may be plausible.

The constant effect assumption is implied by the unit homogeneity assumption; that is, if $Y_t(u_1) = Y_t(u_2)$ and $Y_c(u_1) = Y_c(u_2)$, then clearly $Y_t(u_1) - Y_c(u_1) = Y_t(u_2) - Y_c(u_2)$. Hence we may view the constant effect assumption as a *weakening* of the assumption of unit homogeneity.

If we make only the constant effect assumption we may not conclude that the prima facie causal effect, T_{PF}, in (10) equals the average causal effect, T, in (3). To see this observe that under constant effect we have

$$Y_t(u) = Y_c(u) + T \quad (13)$$

for all units, u. Hence

$$E(Y_t \mid S = t) = T + E(Y_c \mid S = t), \quad (14)$$

so

$$T_{PF} = T + \{E(Y_c \mid S = t) - E(Y_c \mid S = c)\}. \quad (15)$$

The term in braces in (15) is not 0 in general, that is, if the independence assumption is not true.

It is easy to show that the stronger assumption of unit homogeneity does imply equality between T and T_{PF}.

4.5 Causal Inference in Nonrandomized Observational Studies

It is beyond the scope of this article to apply the model for causal inference to nonrandomized studies. This has been done extensively, and the reader is referred to Rubin (1974, 1977, 1978), Rosenbaum (1984a,b,c), Rosenbaum and Rubin (1983a,b, 1984a,b, 1985a,b), and Holland and Rubin (1980, 1983). An important emphasis in these papers is on the ways that *pre-exposure* variables can be used to replace the independence assumption with less stringent *conditional* independence assumptions that are useful in observational studies. Rosenbaum and Rubin referred to one such assumption as "strong ignorability."

5. COMMENTS ON SELECTED PHILOSOPHERS

So much has been written about causality by philosophers that it is impossible to give an adequate coverage of the ideas that they have expressed in a short article. This section views some of these ideas in the context of Rubin's model for causal inference given in Sections 3 and 4. It makes no attempt to be exhaustive or even representative.

Aristotle distinguished four "causes" of a thing in his *Physics*: The *material* cause (that *out of which* the thing is made), the *formal* cause (that *into which* the thing is made), the *efficient* cause (that *which makes* the thing), and the

final cause (that *for which* the thing is made). It is his notion of efficient cause that is relevant to our discussion and to most discussions of causation that grow out of inquiries into the methods of science. Locke (1690) proposed these definitions: "That which produces any simple or complex idea, we denote by the general name 'cause', and that which is produced, 'effect'." Although it is evident that these definitions refer to the same kinds of things that concern the model in Section 3, they do little more than suggest that the model is not out of line with an ancient philosophical tradition. It should be noted, however, that Aristotle emphasized the *causes* of a thing rather than the effects of causes. Locke seems a little more even-handed. Bunge (1959) gave a very accessible discussion of the history of many ideas about the essential meaning of causation.

5.1 Hume

When we turn to the analysis of causation given by Hume (1740, 1748) we find a critical basis for examining Rubin's model. Hume's analysis of causality is generally regarded to be an important contribution to the literature of this subject. Hume emphasized that causation is a relation between experiences rather than one between facts. He argued that it is not empirically verifiable that the cause produces the effect, but only that the experienced event called the cause is invariably followed by the experienced event called the effect. Hume's empirical stance can be regarded as sympathetic with the classical statistical view that the role of statistics is to draw inferences about unobserved quantities on the basis of observed facts. He was also very clear about the role of untestable assumptions in drawing causal conclusions.

Hume's analysis recognized three basic criteria for causation: (a) spatial/temporal contiguity, (b) temporal succession, and (c) constant conjunction. In the analysis of the idea that A causes B this means that (a) A and B are contiguous in space and time, (b) A precedes B in time, and (c) A and B always occur (or do not occur) together.

In terms of Rubin's model the first two of Hume's criteria are easily accommodated. The criterion of spatial/temporal contiguity is expressed in the model by the action of the cause and the measurement of the effect all taking place on a common entity, the unit. Since real entities must exist in space and time the contiguity criterion is satisfied and possibly clarified by the model. Temporal contiguity is relevant to the degree that the time period involved affects the unit. Spatial contiguity is often defined by the unit itself and may not involve simple "nearness."

The issue of temporal succession is shamelessly embraced by the model as one of the defining characteristics of a response variable. The idea that an effect might *precede* a cause in time is regarded as meaningless in the model, and apparently also by Hume.

Hume's notion of constant conjunction is more difficult simply because it might not hold for many reasons. In terms of the model there are two types of reasons why it might not hold. One of these involves "measurement error," and the other is more fundamental and involves the structure of the model. Measurement error often creates violations of constant conjunction in real scientific investigations. We may think we have a case of "A and not B" but we really have a case of "A' and not B" for some A' that we mistook for A (similarly for examples of "not A and B"). In the model these "errors of measurement" can involve both the causes and the response variable that determines the effect. The other, more fundamental way that constant conjunction can fail in the model is for the constant effect assumption to fail to hold, that is, for the causal effects $Y_t(u) - Y_c(u)$ to vary with the unit u. Hence, if we disregard those cases of nonconstant conjunction that are due to measurement error, we see that Hume's third criterion requires the constant effect assumption to hold in our model. Hume would probably argue that any weakening of this assumption would allow cases that he would not call "causation" into the model. We will have to be satisfied that at least Hume's analysis fits into the model and let others judge the utility of the constant effect assumption. I should point out that the distinction between constant and variable causal effects (a) is often not easy to prove one way or the other in a particular case and (b) has been at the heart of at least one important controversy in the history of statistics (see Sec. 6).

What I see that is missing from Hume's analysis is any notion that the effect of cause is always relative to another cause. The notion that a cause could have been different from what it was and that it is this difference that defines the effect is completely missing from Hume. In Hume's analysis causes are not delineated in any way. Anything can be a cause. The importance of this point will be emphasized in Section 7. Finally, Hume does not identify the idea of an experiment as related to or important for causation.

5.2 Mill

John Stuart Mill is rather different in this regard. Mill (1843) was positively disposed toward experiments.

> Observation, in short, without experimentation (supposing no aid from deduction) can ascertain sequences and co-existences, but cannot prove causation. (p. 253)

> . . . we have not yet proved that antecedent to be the cause until we have reversed the process and produced the effect by means of that antecedent artificially, and if, when we do so, the effect follows, the induction is complete. . . . (p. 252)

Mill is credited with codifying and elaborating on several methods of experimental inquiry that had been put forth by Sir Francis Bacon 250 years earlier. Mill identified four general methods, which I now discuss.

The Method of Concomitant Variation. This method flies in the face of the distinctions that I have drawn between association and causation.

> Whatever phenomenon varies in any manner, whenever another phenomenon varies in some particular manner, is either a cause or an effect of that phenomenon, or is connected with it through some fact of causation. (p. 464)

I think that as a method of science the widespread use of this method is indisputable. Most scientists would agree

that where there is correlational smoke there is likely to be causational fire. Most would not, however, go as far as Mill's statement of the method.

Of course, even if Rubin's model does apply, the *correlation* between the observed variables S and Y_S does not say much about the causal effects or even the average causal effect, because the correlation of Y_S and S is simply another way of expressing the prima facie causal effect, T_{PF}.

More generally, not everything can be a "cause" in the sense used in the model, but Mill's method of concomitant variation can be applied to cases for which only association is appropriate. That this can result in nonsense discussions of causation is well known.

Method of Difference. This method is almost an exact statement of what we mean by a causal effect, even though it is couched in a more general language and its proposed use is to identify causes and effects.

> If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common save one, that one occurring in the former; the circumstances in which alone the two instances differ, is the effect, or the cause, or an indispensable part of the cause of the phenomenon. (p. 452)

If we restrict our attention to the following interpretation of the elements of this quotation we see a fairly straightforward definition of causal effect: "phenomenon under investigation" occurs—$Y = 1$; "phenomenon under investigation" does not occur—$Y = 0$; "the circumstance in which the instances differ"—when present = t, when absent = c. Then $Y_t(u) = 1$ denotes the fact that when the circumstance was present the phenomenon occurs, and $Y_c(u) = 0$ denotes the fact that when the circumstance was absent the phenomenon did not occur. The equality of all other circumstances is modeled by considering the same unit. Thus $Y_t(u) - Y_c(u) = 1$, so the causal effect of the circumstance on the unit is 1 and corresponds to Mill's statement that the circumstance is "the cause or an indispensable part of the cause of the phenomenon."

Mill also considered reversing the process to look for causes of given effects. This is a well-known scientific technique—for example, it occurs often in epidemiological studies of public health problems. It is beyond the scope of this article to apply the model to such a case, but some work along this line can be found in Hamilton (1979) and Holland and Rubin (1980).

The Method of Residues. This method also applies fairly simply to the model. Its statement is

> Subduct from any phenomenon such part as is known by previous inductions to be the effect of certain antecedents, and the residue of the phenomenon is the effect of the remaining antecedents. (p. 460)

To place this into the context of the model let the antecedents (i.e., causes) be denoted by a = those whose effect is known and b = the remaining antecedents.

The causal effect of ab relative to a is simply $Y_{ab}(u) - Y_a(u)$, which is the residue Mill tells us to compute. I regard Mill's method of residues to be a nearly explicit, early statement of the definition of causal effect.

The Method of Agreement. Usually this method is discussed first because it is so clearly a part of scientific investigations. I have left it to the end because it requires the introduction of the notion of a "null effect." The method is stated as follows:

> If two or more instances of a phenomenon under investigation have only one circumstance in common, the circumstance in which alone all the instances agree, is the cause (or effect) of the given phenomenon. (p. 451)

Although it looks like a method for identifying the cause of a phenomenon, it is clear to anyone who has ever used the method of agreement that all that the method really does is to *rule out* possible causes. It is this aspect of the method of agreement that fits into the model.

If, as in the discussion of the method of difference, we let $Y = 1$ (or 0) denote the occurrence (or not) of "the phenomenon under investigation," and then if the phenomenon occurs when the cause t occurs and also when the cause t does not occur, that is, c, we have

$$Y_t(u) = 1 \quad \text{and} \quad Y_c(u) = 1,$$

so

$$Y_t(u) - Y_c(u) = 0.$$

Hence the causal effect of t is 0; that is, t is a cause with a *null effect.* The principle of causality states that every phenomenon has a cause; that is, every effect has a cause. Every practicing experimentalist can attest to the fact that the reverse is not true—experiments fail. Causes do not necessarily have effects. Null effects are the stuff from which null hypotheses are made!

My conclusion is that Mill's thinking, being driven by an experimental model, is in reasonably close agreement with the model of Section 3. He is close to the idea that the effect of a cause is always relative to another cause, unlike Hume. Like Hume, however, he does not restrict the notion of cause in any way. For Hume and Mill any phenomenon can be a cause. Finally, like Hume, Mill does not consider variation (i.e., either unit inhomogeneity or variable causal effects) in any serious way.

5.3 Suppes

Variation is an explicit consideration in Patrick Suppes's (1970) *probabilistic theory of causality.* Suppes's goal was to improve upon Hume's analysis, specifically the constant conjunction criterion, because

> . . . in restricting himself to the concept of constant conjunction, Hume was not fair to the use of causal notions in ordinary language and experience. (p. 10)

Like Hume, Suppes puts no restriction on what causes and effects *are* save only that they be expressible as events that occur in time. Thus Suppes uses the language of stochastic processes to formalize his theory. He explained the intuitive idea of his theory as follows:

> Roughly speaking, the modification of Hume's analysis I propose is to say that one event is the cause of another if the appearance of the first event is followed with a high probability by the appearance of the second, and there is no third event that we can use to factor out the probability relationship between the first and second events. (p. 10)

Suppes expressly adopted the temporal succession cri-

terion that all causes precede their effects in time. He first defined a *prima facie cause* of an event as an event that temporally precedes it and that is positively associated with it. He then defined a *spurious cause* of an effect (i.e., an event) as a prima facie cause of the effect that is, in fact, conditionally independent of the effect given a second event that is temporally prior to the prima facie cause and that is conditionally positively associated with the effect given the prima facie cause. This is what he meant by "factoring out" a probability relationship. A *genuine cause* is a prima facie cause that is not spurious.

More precisely Suppes's definitions are as follows:

(S1) If $r < s$ denote two time values, the event C_r is a *prima facie* cause of the event E_s if

$$\Pr(E_s \mid C_r) > \Pr(E_s). \tag{16}$$

(S2) C_r is a *spurious cause* of E_s if C_r is a prima facie cause of E_s and for some $q < r < s$ there is an event D_q such that

$$\Pr(E_s \mid C_r, D_q) = \Pr(E_s \mid D_q) \tag{17}$$

and

$$\Pr(E_s \mid C_r, D_q) \geq \Pr(E_s \mid C_r). \tag{18}$$

(S3) C_r is a *genuine cause* of E_s if C_r is a prima facie cause of E_s but C_r is not a spurious cause of E_s.

In all of these definitions the probabilities of the events used in the conditioning statements are assumed to be positive. Suppes also considered other issues, such as direct and indirect causes, but (S1)–(S3) are the main elements of his theory.

It is clear that Suppes's analysis is quite different from that given in Section 3. He defined the cause of an effect rather than the effect of a cause. Like Hume and Mill he placed no general restriction on the nature of a cause other than that it be expressible as an event that occurs prior in time to the effect. There is no explicit place for units in Suppes's stochastic process model—they are buried in the probability space on which the events he considered are defined. Hence Suppes does not have the machinery to express the effect of a cause in a particular case. His model describes average behavior, not individual behavior.

At bottom, Suppes's notion of a genuine cause is simply a correlation between a cause and effect that will not go away by "partialling out" legitimate competing causes. In a sense then for Suppes all genuine causes are only temporarily so as they await the cleverness of the analyst to identify the proper conditioning event that will render null their association with the effect. Although this may, indeed, describe much informal scientific practice, it does not appear to me to get to the heart of the notion of causation, which, I believe, Rubin's model does.

Suppes's theory, however, does capture some useful ideas, and because it is stated with precision it is a fairly easy task to relate these ideas to Rubin's model.

In what follows, all probabilities and expectations are computed over the population U of units.

Earlier, his notion of a *prima facie* cause was translated into the *prima facie causal effect* as follows. The association between the observed response Y_S and the causal indicator S can be measured by the difference in the average value of the response between the units exposed to t and those exposed to c. We have called this the *prima facie causal effect* of t (relative to c), that is,

$$T_{PF} = E(Y_S \mid S = t) - E(Y_S \mid S = c). \tag{19}$$

We have seen that the association between cause and effect that defines a prima facie cause *is* a causal effect under certain conditions that have wide use in science, but T_{PF} is not always a causal effect. This is why Suppes defined *prima facie* causes.

I will finish this section by showing what happens when we apply Suppes's notion of a spurious cause to the context of a randomized experiment. This will shed some light on the relation of his theory to Rubin's model.

If the response variable Y is a 0/1 indicator, then we may keep the discussion in terms of the event terminology that Suppes used. Thus $\{Y_S = 1\}$ corresponds to E_s and $\{S = t\}$ corresponds to C_r, and I will discuss the meaning of the event D_q subsequently.

Consider Equation (17) from (S2). For a randomized experiment it is

$$\Pr(Y_S = 1 \mid S = t, D_q) = \Pr(Y_S = 1 \mid D_q). \tag{20}$$

By using the usual rules for handling conditional probabilities we may express (20) as follows:

$$\{\Pr(Y_t = 1 \mid S = t, D_q) - \Pr(Y_c = 1 \mid S = c, D_q)\} \times \Pr(S = c \mid D_q) = 0. \tag{21}$$

Hence the only way that Equation (20) can hold is for either

$$\Pr(S = t \mid D_q) = 1 \tag{22}$$

or

$$\Pr(Y_t = 1 \mid S = t, D_q) = \Pr(Y_c = 1 \mid S = c, D_q). \tag{23}$$

If D_q is an event that occurs prior in time to the exposure of the units to t or c, then I will assume that D_q is determined by the values of *pre-exposure* variables defined on the units in U. Now suppose that the assumption of independence holds so that S is statistically independent of Y_t, Y_c and of the pre-exposure variables that define D_q. Furthermore, suppose that

$$0 < \Pr(S = t) < 1, \tag{24}$$

so each unit has positive probability of being exposed to either cause. The independence assumption and (24) then imply that (22) cannot hold and that Equation (17), therefore, reduces to

$$\Pr(Y_t = 1 \mid D_q) = \Pr(Y_c = 1 \mid D_q). \tag{25}$$

Because Y is an indicator variable we can rewrite (25) in terms of an average causal effect; that is,

$$T(D_q) = E(Y_t - Y_c \mid D_q) = 0. \tag{26}$$

The average causal effect $T(D_q)$ in (26) is the average

causal effect over all units in U for which the event D_q occurs. Hence we see that Suppes's condition (17) for a spurious cause reduces to the condition

$$T(D_q) = 0 \tag{27}$$

in a randomized experiment. The other condition that Suppes required in (S2) is inequality (18), which is, in the present context, equivalent to

$$\Pr(Y_S = 1 \mid S = t, D_q) \geq \Pr(Y_S = 1 \mid S = t). \tag{28}$$

Under randomization this becomes

$$\Pr(Y_t = 1 \mid D_q) \geq \Pr(Y_t = 1). \tag{29}$$

If we put (29) and (27) together with the condition that t be a prima facie cause we find that the treatment in a randomized experiment is a spurious cause of the effect if and only if it has a positive average causal effect, but a subpopulation of units can be identified on the basis of pre-exposure variables (a) on which the average causal effect is 0 and (b) for which the response under t is more likely to occur than it is for all of U. I think that part (a) is more accurately described as a null effect in the subpopulation and part (b) is unrelated to the notion of cause. The existence of a subpopulation on which the effect is null while the overall effect is positive is an example of nonconstant conjunction in Hume's sense. It would be called an *interaction* by most statisticians.

6. COMMENTS FROM A FEW STATISTICIANS

This section is devoted to a brief examination of the writings of a few statisticians to see in what way the idea of multiple versions of the response, that is, Y_t and Y_c, has appeared before. I find that many people have difficulty with the idea of distinguishing Y_t and Y_c from Y or Y_S and perhaps this look at earlier work may help clarify this assumption. Unfortunately, the exact idea is never stated explicitly, so there is a need for a certain amount of detective work to find it. I hope I will not be held guilty of wrongly reinterpreting the work of others.

A fairly clear statement of this idea was given by Kempthorne (1952) in a discussion of the analysis of randomized block designs. (A randomized block design is a typical agricultural experimental plan in which larger tracts of land, called blocks, are each subdivided into p plots and then one of the experimental treatments is applied at random to each of the p plots within each block.) For example, Kempthorne (1952, p. 136) first defined *yields* as follows: "We shall denote the yield with treatment k . . . on plot j . . . of block i . . . by y_{ijk}." He then wrote:

> In fact we do not observe the yield of treatment k on plot j but merely the yield of treatment k on a randomly chosen plot in the block. . . . we denote the observed yield of treatment k in block i by y_{ik}. (p. 137)

It seems evident from the two quotations that the y_{ijk} in the first refers to different versions of the response—one for each k—on each combination (i, j) of plot within block. The y_{ik} in the second quotation is the value of y_{ijk} for that plot to which treatment k is actually applied in block i.

It is not difficult to make the following translation of Kempthorne's notation. The units are the "plots," so the units need two subscripts for identification; that is, u_{ij} is the jth plot within block i. The yield of treatment k on the unit u_{ij} is $y_{ijk} = Y_k(u_{ij})$, where $Y_k(u)$ is the value of the response that is observed if u is exposed to treatment k. The randomization process picks one of the treatments to apply to unit u_{ij}, and this can be indicated by $S(u_{ij})$; that is, if treatment k is applied to unit u_{ij} then $S(u_{ij}) = k$. The observed yield on u_{ij} is

$$y_{ijS(u_{ij})} = Y_{S(u_{ij})}(u_{ij}).$$

The plot in block i to which treatment k is applied can be denoted by j_k so that the observed yield of treatment k on block i is

$$y_{ik} = Y_k(u_{ij_k}).$$

In D. R. Cox's (1958) book on the planning of experiments he defined *true treatment effects* in an experiment in almost exactly the same way that we have defined causal effects. In an experiment with treatments T_1, T_2, he defined the true treatment effects as the difference between "the observation obtained on any unit when, say, T_1 is applied" and "the observation that would have been observed had, say, T_2 been applied" (p. 15). Hence Cox appears to have accepted the idea that the response of a unit could be one value, $Y_t(u)$, if the unit were exposed to t and another, possibly different value, $Y_c(u)$, if the unit were exposed to c. Cox also made the assumption of constant effect in defining true treatment effects. His reasons for this are not clear but appear to be primarily technical rather than conceptual. He did not reject the idea of variable causal effects, however, and discussed ways in which causal effects might depend "on the value of some supplementary measurement that can be made on each unit" (p. 18).

Curiously, R. A. Fisher, who founded the modern theory of experimental design, never dealt directly with the idea of multiple versions of the response. Instead, he gave examples that are so laced with specific details that it is not always clear what level of generality he meant to convey. For example, in the first article in which Fisher (1926) attempted to set out the principles of the design of field experiments in agriculture we find this question in a discussion of a hypothetical experiment to evaluate the apparent productive value of treating a given acre of ground with a manurial treatment:

> What reason is there to think that, even if no manure had been applied, the acre which actually received it would not still have given the higher yield? (p. 504)

It is fairly clear in this quotation that he could consider the possibility that had a different treatment (i.e., no manure) been applied to the field the resulting yield might have been the same. This clearly concerns the null hypothesis of no treatment effect and, more generally, Fisher came closest to the idea of multiple versions of the response in his discussions of the relationship between the null hypothesis and randomization.

The earliest explicit reference that I have found to multiple versions of the response is Neyman (1935). In his paper (read before the Industrial and Agricultural Re-

search Section of the Royal Statistical Society in March of 1935) Neyman gave an explicit statement of the idea of multiple versions of the response (which is for Neyman the yield from an experimental plot of land in an agricultural experiment). Unfortunately, Neyman's discussion also introduced the notion of a stochastic element that is added to Y to allow for "technical errors" that are due to inaccuracies of experimental technique. If we ignore this problem of measurement error and assume zero "technical errors," then Neyman's definition of a "true yield" explicitly refers to multiple versions of the response. "Thus $X_{ij}(k)$ will mean the 'true' yield of the kth object obtainable from the plot (i, j)" (p. 110; by "object" Neyman means treatment). His notation is very similar to that used by Kempthorne. To put it into the notation of Section 3, the units are the plots, u_{ij}, and $X_{ij}(k) = Y_k(u_{ij})$, where $Y_k(u)$ is defined as in the previous discussion of Kempthorne.

Neyman also had an explicit expression for the average value of $X_{ij}(k)$ over all of the units, u_{ij}. It is $X_{..}(k)$. In the notation of Section 3 this is $X_{..}(k) = E(Y_k)$. Hence it is clear that by the time Neyman was writing the idea of multiple versions of the response, one for each treatment, was established. It seems to have been used by writers concerned about the details of the effects of randomization in specific experimental plans (e.g., Cox 1958; Kempthorne 1952) but is generally not a part of the standard statistical notation of many other writers [an exception is Hamilton (1979)].

The Neyman (1935) reference is also relevant to the model in Section 3 because of the controversy between Fisher and Neyman that it engendered. The controversy revolves around the choice of null hypothesis in experiments such as randomized block designs. Fisher was quite clear that the null hypothesis that he proposed is that the causal effect (as we have defined it) is 0 *for each unit.* For example, in the famous discussion at the end of Neyman (1935) Fisher first quoted Neyman, as follows:

> . . . this bias vanishes when . . . the objects compared are reacting to differences in soil fertility in exactly the same manner. . . . This is not always true. (p. 153)

Then Fisher wrote:

> However, it was always true in the case for which it was required, namely, when the hypothesis to be tested was true, that differences of treatment made no difference to the yields. (p. 157)

Then Neyman, in responding to Fisher's remarks, emphasized his interest in what I would call the average causal effect.

> 'Our purpose in the field experiment consists in comparing numbers such as $X_{..}(k)$, or the average true yields which our objects are able to give when applied to the whole field.' It is seen that this problem is essentially different from what Professor Fisher suggested. So long as the average yields of any treatments are identical, the question as to whether these treatments affect *separate* yields on *single* plots seems to be uninteresting and academic. (p. 173)

Fisher's sardonic reply indicates that, at least, he agreed that Neyman stated their differences clearly. "It may be foolish, but that is what the z test was designed for, and the only purpose for which it has been used" (p. 173).

Evidently, I would conclude that Neyman's null hypothesis is one of zero average causal effect, that is, $E(Y_t - Y_c) = 0$, whereas Fisher's is one of zero causal effect for all units, that is, $Y_t(u) - Y_c(u) = 0$ for all $u \in U$.

7. WHAT CAN BE A CAUSE?

It may seem very extreme to some to limit the notion of *cause* to the sense used in Section 3. Aristotle set the stage for this, however, by distinguishing more than one meaning to the word *cause.* It might be better to ask, what can be an "efficient cause" in his sense? Evidently even this restriction did not limit the notion of cause for such thinkers as Hume and Mill. Anything can be a cause for them—or, at least, a potential cause.

Put as bluntly and as contentiously as possible, in this article I take the position that causes are only those things that could, in principle, be treatments in experiments. The qualification "in principle" is important because practical, ethical, and other considerations might make some experiments infeasible, that is, limit us to contemplating *hypothetical experiments.* For example, in the medical and social world we might be able to conceive of an experiment, but no one would ever try to carry it out. Instead, we might have to wait for a "natural experiment" to occur. "Observational study" is the term used by statisticians (e.g., Cochran 1983) to refer to studies for which "The objective is to study the causal effects of certain agents" but "For one reason or another the investigator can not . . . impose on . . . or withhold from the subject, a treatment whose effects he desires to discover" (p. 1).

I believe that the notion of cause that operates in an experiment *and* in an observational study *is the same.* The difference is in the degree of control an *experimenter* has over the phenomena under investigation compared with that which an *observer* has. In Rubin's model this is expressed by the joint distribution of S with Y_t and Y_c. Total control can make S independent of Y_t and Y_c.

It may bother some readers that I have been using the term "experiment" in a very restricted sense—though one that is common in the study of the design of experiments. For example, experiments in chemistry in which a substance is analyzed into its component ingredients or in which ingredients are combined with each other to synthesize a new substance often may not have clearly identifiable units, treatments, and response variables. My view is that in such experiments the Aristotelian notion of *material cause* is often more relevant than that of *efficient cause*, and hence such experiments are not concerned with the notion of cause that is discussed in this article.

To return to the question of what can be a cause let me consider three examples of statements that involve the word *cause* but that vary in its exact usage.

(A) She did well on the exam because she is a woman.

(B) She did well on the exam because she studied for it.

(C) She did well on the exam because she was coached by her teacher.

I think that these statements, even though they are perfectly understandable English sentences, vary in the mean-

ing of the "because" in each. In each, the effect, using the term loosely, is the same—doing well on an exam. The causes, again using the term loosely, are different. In (A) the "cause" is ascribed to an attribute she possesses. In (B) the "cause" is ascribed to some voluntary activity she performed, and in (C) it is ascribed to an activity that was imposed on her.

An attribute cannot be a cause in an experiment, because the notion of *potential exposability* does not apply to it. The only way for an attribute to change its value is for the unit to change in some way and no longer be the same unit. Statements of "causation" that involve attributes as "causes" are always statements of association between the values of an attribute and a response variable across the units in a population. In (A) all that is meant is that the performance of women on the exam exceeds, in some sense, that of men.

Examples of the confusion between attributes and causes fill the social science literature. Saris and Stronkhorst (1984) gave the following example of a causal hypothesis: "Scholastic achievement affects the choice of secondary school" (p. 13). These authors clearly intended for this hypothesis to state that an *attribute* of a student (i.e., scores on tests, performance in primary school) can *cause* (i.e., affect) the student's choice of a particular type of secondary school. It is difficult to conceive of how scholastic achievement could be a treatment in an experiment and, therefore, be a "cause" in the sense used in this article. A somewhat stronger statement of my point was given by Kempthorne (1978, p. 15): "It is epistemological nonsense to talk about one trait of an individual *causing* or determining another trait of the individual."

At the other extreme is Example (C). This is easily interpreted in terms of the model. The interpretation is that had she not been coached by her teacher she would not have done as well as she did. It implies a comparison between the responses to two causes, even though this comparison is not explicitly stated.

Example (B) is just one of many types of examples in which the applicability of the model is not absolutely clear, and it shows one reason why arguments over what constitutes a proper causal inference can rage without any definitive resolution.

In (B) the problem arises because of the voluntary aspect of the supposed cause—studying for the exam. It is not clear that we could expose a person to studying or not in any verifiable sense. We might be able to *prevent* her from studying, but that would change the sense of (B) to something much more like (C). We could operationally define studying as so many hours of "nose in book," but that just defines an attribute we could measure on a subject. In my opinion the application of the model to statement (B) is problematical and not easily resolved. The voluntary nature of much of human activity makes causal statements about these activities difficult in many cases.

The voluntary aspect of the "cause" in (B) is not the only source of difficulty in deciding on the applicability of Rubin's model to specific problems. It is, however, a common source of difficulty.

The general problem, I think, is in deciding when something is an *attribute* of units and when it is a *cause* that can act on units. In the former case all that can be discussed is association, whereas in the latter case it is possible, at least, *to contemplate* measuring causal effects.

One may view Fisher's (1957) attack on those who used the association between smoking and lung cancer as evidence of a "causal link" between them as an example of the difficulty in deciding whether or not smoking is an attribute or a cause. Certainly the data that began this debate are purely associational. Doll and Hill's studies (1950, 1952, 1956) ascertained only smoking status and lung cancer status on sets of subjects. Fisher argued that smoking might only be indicative of certain genetic differences between smokers and nonsmokers and that these genetic differences could be related to the development or not of lung cancer. Fisher (1957) did feel that "a good prima facie case had been made for further investigation."

The response to Fisher's criticism can also be viewed as attempting to show that smoking should be thought of in causal terms rather than as indicative of a genetic attribute of subjects. For example, among his responses to Fisher, McCurdy (1957) pointed out that lung cancer rates increase with the *amount* of smoking and that subjects who stopped smoking had lower lung cancer rates than those who did not. Both of these arguments can be viewed as emphasizing the causal aspects of smoking—one can do more or less of it and one might stop doing it. A discussion of the entire debate was given by Cook (1980).

8. COMMENTS ON CAUSAL INFERENCES IN VARIOUS DISCIPLINES

This section will briefly consider discussions of causation in three disciplines—medicine, economics, and "causal modeling." In each case an attempt will be made to relate the discussion to Rubin's model for causal inference, but no attempt is made to be exhaustive or even representative in the selection of topics considered.

8.1 Causation and Medicine

We begin with a simple, yet basic, example from medicine—the establishment of specific bacteria as the cause of specific infectious diseases. Yerushalmy and Palmer (1959) described the situation in the following terms:

> Almost from the very beginning, when bacteria were first found to cause disease, bacteriologists felt the need for a set of rules to act as guideposts in investigation of bacteria as possible causal agents in disease. (p. 28)

These two authors described three postulates formulated by the great bacteriologist, Robert Koch, who discovered, among other things, the tuberculosis bacillus in 1882. Koch's postulates [also called the Koch–Henle postulates, Evans (1978)] are simple, no-nonsense criteria for deciding when a microscopic organism is implicated in a disease. According to Yerushalmy and Palmer (1959), "while there is no single formulation of Koch's postulates—they can be stated as consisting essentially of the following:

I. The organism must be found in all cases of the disease in question.

II. It must be isolated from patients and grown in pure culture.
III. When the pure culture is inoculated into susceptible animals or man, it must reproduce the disease." (p. 30)

Rubin's model applies rather clearly to Postulates I and III. Postulate I is simply Mill's method of agreement applied to this problem. It ensures that there are no data to support a null causal effect in this case—that is, if there were bona fide cases of the disease in which the organism was not present, along with other cases of the disease in which it was, then assuming unit homogeneity we would have an estimate of zero causal effect for the presence of the organism relative to its absence. Postulate III is like the light switch example—put in the organism and the disease occurs. The validity of this postulate stems from the unstated assumption that had the animal or human not been inoculated with the culture the disease would not have been expected to occur. Note that the word "susceptible" has crept in, presumably to deal with the inevitable "nonconstant conjunction" of real laboratory work—in this case, the immune system.

Koch's second postulate relates more to good experimental techniques than to causal inference. If the organism is isolated from patients and grown in pure culture, then when it comes time to inoculate animals or people with it the experimenter knows what the inoculant is in fairly exact terms. In a sense, Postulate II is a way of minimizing measurement error in the treatment (t) that is exposed to the units.

Medicine is more difficult when the biological theory is less well developed. As an example I now consider several suggestions made by Sir Austin Bradford Hill to those who might wish to separate association from causation in the study of the environment and disease. He had spent a lifetime in public health and was among the first to argue, quantitatively, for the causal link between smoking and lung cancer (Doll and Hill 1950, 1952, 1956). Hill (1965) named nine factors that he felt were useful in such work for deciding that the most likely interpretation of an observed association is causation. I will consider these in an order that differs from Hill's.

Temporality. "Which is the cart and which the horse?" (Hill 1965, p. 297). Hill felt that while the time sequence of events, cause preceding effect, might not be difficult to establish in many cases, "it certainly needs to be remembered, particularly with selective factors at work in industry" (p. 298). Clearly, temporal succession is a given for Hill.

Experiment. In this category Hill placed the occasional "natural experiment" that gives strong evidence for causation. He had in mind the effect of preventative actions taken to reduce the incidence of the disease. Do they work? If a person stops smoking does he lower his risk of lung cancer? Hill clearly views such "experiments" in the same way Mill viewed the production of an effect by artificially introducing the presumed causal agent—strong causal evidence when you can find it.

Biological Gradient. By this Hill referred to evidence that showed an increasing disease rate as exposure to the agent in question intensified. Both experiment and biological gradient may be viewed as emphasizing the causal nature of the proposed causal agent, as discussed in the previous section.

Plausibility, Coherence, Analogy. I have grouped these three together because they all refer to the prior knowledge that the epidemiologist would need to consider. Is the suspected causation biologically *plausible*? Is it *coherent* in the sense of not being seriously in conflict with known facts? Is it *analogous* to known causal relations for similar agents and diseases? These factors, although important in some cases, all reflect the state of relevant scientific knowledge and do not directly translate into aspects of the model of Section 3. In particular Hill felt that it was unwise to place undue emphasis on these because of the relatively poor state of relevant biological knowledge in many cases of interest.

Although Hill felt that the six factors listed above were important from time to time, they were the six least significant factors on his list. He felt that the three most important factors are the *strength*, *consistency*, and *specificity* of the association in question.

Strength. This is Hill's first factor—"First upon my list I would put the strength of the association" (p. 295). This may be viewed as simple acceptance of Mill's method of concomitant variation in practical terms or of the scientific utility of the prima facie causal effect. Although there is no guarantee for this, it is often more likely that a larger prima facie causal effect will hold up when a controlled study is performed than will a smaller prima facie causal effect. A relevant result in this regard is the inequality given in Cornfield et al. (1959) that bounds the influence of unmeasured factors on the relative risk (a form of prima facie causal effect).

Consistency. Hill's second significant factor concerns the generality of the association across populations of units. This might be viewed as a weakened form of constant conjunction. At the very least, an association that is present in one population and absent in another suggests variable causal effects. I think that there is a clear bias against calling variable causal effects "causal" by scientists, even though those who must deal with heterogeneous units, such as humans, will generally agree that it is usually too much to expect constant effects in the real world.

Specificity. Hill's third factor refers to specific causes having specific effects.

> If . . . the association is limited to specific workers and to particular sites and types of disease and there is no association between the work and other modes of dying, then clearly that is a strong argument in favor of causation. (p. 297)

I think that specificity is related to the believability of the independence assumption. The lack of an association between the exposure of a person to a particular work place and the causes of that person's death supports the independence assumption in a relevant way (but does not prove

the assumption is valid). Since the independence assumption implies that the prima facie causal effect equals the average causal effect, *specificity*, in conjunction with the strong association, may well be convincing evidence of a strong causal connection. Lack of specificity, however, does not disprove the independence assumption in many cases, and this explains why lack of specificity is not regarded as a serious problem by Hill.

> In short, if specificity exists we may be able to draw conclusions without hesitation; if it is not apparent, we are not thereby necessarily left sitting on the fence. (p. 297)

Of course, specificity does not *guarantee* that the independence assumption is valid, but it does not directly contradict this assumption in the way that a lack of specificity does.

8.2 Granger Causation in Economics

The primary source of data that is available to economists is so-called "time series" data in which measurements of a variable or set of variables are made repeatedly on an economic entity over time. For such data, Granger (1969) developed a particular notion of causality that some economists have found useful in their analyses.

In my opinion, however, Granger's essential ideas involving causation do not require the time-series setting he adopted. I will try to restate his theory in terms of the types of models used in Sections 2 and 3—that is, variables defined on a population of units. Granger formulated his theory around the idea of prediction—a "cause" ought to improve our ability to predict an effect in a probabilistic system. In Granger's theory a variable *causes* another variable; that is, the values of one variable improve one's ability to predict the future values of another variable. The only important way that his theory used the time-series setting was to separate variables into those whose values are determined prior to, at, or after a given point in time. I will simply adopt these temporal distinctions in the definitions of the variables that arise. Granger (1969, p. 430) clearly accepted the idea of temporal succession in his analysis: "In the author's opinion there is little use in the practice of attempting to discuss causality without introducing time." It is the past values of a variable that cause, in Granger's sense, the future values of another variable.

Although Granger originally formulated his theory in terms of one variable causing another, later writers (e.g., Florens and Mouchart 1985) restated it in terms of noncausality and I will follow that approach. In reformulating his theory I will also shift from his emphasis on a particular type of predictor, that is, "the optimum, unbiased, least-squares predictor" (p. 428), to the more generally applicable notion of conditional statistical independence. This means that instead of limiting attention to the inability of a specific predictor to predict the values of a variable, I will use the stronger condition that *no* predictor can predict the desired values. Although this is a stronger type of noncausality than Granger defined I do not believe that this unduly distorts Granger's theory and it certainly generalizes its applicability—indeed, see Granger (1980).

If X, Y, and Z denote three (possibly vector-valued) variables defined on a population, then X and Y are *conditionally independent* given Z if

$$\Pr(Y = y \mid X = x, Z = z) = \Pr(Y = y \mid Z = z). \quad (30)$$

Conditional independence is a strong form of the idea that the values of X are unable to predict the values of Y, given the values of Z.

In Granger's time-series setting, the value of Y is determined at some time point s, and the values of X and Z are determined at or prior to some other time point $r < s$. I will say that *X is not a Granger cause of Y* (relative to the information in Z) if X and Y are conditionally independent given Z. Thus X is a Granger cause of Y if different values of X lead to different predictive distributions of Y given both X *and* the information in Z, that is, if X helps predict Y even when Z is taken into consideration.

Viewed in this way, Granger noncausality is very much like Suppes's notion of a spurious cause. Both involve the inability of the spurious cause to predict a future event or value given certain other information.

How might Granger's ideas be applied to the setting in Section 3? It is natural to make the following identification of Granger's setting with elements of Rubin's model.

Granger	Rubin's Model
Y	Y_S
X	S
Z	A set of pre-exposure variables also called Z.

The conditional independence condition is

$$\Pr(Y_S = y \mid S = t, Z) = \Pr(Y_S = y \mid Z),$$

and this reduces to

$$0 = \{\Pr(Y_t = y \mid S = t, Z) - \Pr(Y_c = y \mid S = c, Z)\} \times \Pr(S = c \mid Z). \quad (31)$$

In a randomized experiment

$$\Pr(S = c \mid Z) = \Pr(S = c),$$

which we assume lies strictly in (0, 1). Hence Equation (31) reduces to

$$\Pr(Y_t = y \mid S = t, Z) = \Pr(Y_c = y \mid S = c, Z). \quad (32)$$

But under randomization S is independent of Y_t, Y_c, and Z, so Equation (31) becomes

$$\Pr(Y_t = y \mid Z) = \Pr(Y_c = y \mid Z), \quad (33)$$

which, in turn, implies that

$$E(Y_t \mid Z) = E(Y_c \mid Z) \quad (34)$$

for all values of Z. If we define the average causal effect on the subpopulation specified by $Z = z$ as

$$T(z) = E(Y_t - Y_c \mid Z = z), \quad (35)$$

then Equation (34) says that if S is not a Granger cause of Y_S relative to Z, then $T(z) = 0$ for all values of z. Hence in a randomized experiment Granger noncausality implies

zero average causal effect on all subpopulations defined by the values of Z. Conversely, it is easy to see that if t has a null effect on all units, then in a randomized experiment S will not be a Granger cause of Y_S relative to *any* Z that is a pre-exposure variable.

Although Granger causality has some intuitively satisfying properties with respect to Rubin's model, it fails, in my opinion, to get to the heart of the notion of causality in the same way that Suppes's theory of causality fails. Granger's "causes" are always only temporarily in that category. If an analyst simply gathers more information, that is, changes Z, an X that was once a Granger cause of Y might be shown to be only a spurious cause in exactly the same spirit as in Suppes's theory.

8.3 Causal Models in Social Science

No discussion of causal inference would be complete without some reference to the expanding literature on causal modeling, that is, Blalock (1971), Goldberger and Duncan (1973), Duncan (1975), and Saris and Stronkhorst (1984). Little work has been done to relate Rubin's model to those used in the causal modeling literature—an exception is Rosenbaum (1984b), in which the average causal effect in a population is related to coefficients that arise in certain linear path models. The relationship between these two types of models is a natural research topic, since both causal models and Rubin's model were developed to deal with the same problem—causal inference in nonexperimental research.

In this section I will hint at some possible points of contact between the path diagrams that are used in causal modeling and the model used in this article. I think that this is a large topic, and I can only scratch its surface here.

Path diagrams are used to represent visually causal relationships among a set of variables. For example, if X causes Y this is expressed by the diagram

$$X \rightarrow Y. \tag{36}$$

From the point of view adopted in this article some diagrams like (36) are meaningful and some are not. For example, if A is an attribute of units and Y is a response variable, then

$$A \rightarrow Y \tag{37}$$

is meaningless. On the other hand, if S indicates exposure to causes and Y_S is an observed response variable, then

$$S \rightarrow Y_S \tag{38}$$

is a meaningful diagram.

What happens when we add a third variable to this system? There are several possibilities. If A is an attribute, then it is either a pre- or post-exposure variable. In the first case we might denote this as

$$A \qquad S \rightarrow Y_S \tag{39}$$

to indicate the time flow but without any arrow from A to S or Y_S. In the second case the value of A might be affected by exposure to the cause and we would need to indicate that by subscripting A, A_t, and A_c. This suggests the diagram

$$S \rightarrow (A_S, Y_S). \tag{40}$$

It indicates that S changes the values of both A and Y. This is the situation analyzed by Rosenbaum (1984b).

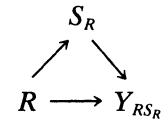
The other possibility is that the third variable is an indicator, R, of a second set of causes, say t' and c'. If the R causes act on the units at the same time that the S causes do, then we can combine R and S into a single causal indicator (R, S). Y must then be doubly subscripted to indicate the responses to the various (R, S) combinations, that is, Y_{RS}. This can be denoted by the diagram

$$(R, S) \rightarrow Y_{RS}. \tag{41}$$

The fact that the R causes and the S causes act at the same time is not really important for Diagram (41). It really says that the R causes do not affect exposure to the S causes, and vice versa. We get an essentially new case, however, when, for example, the R causes act temporally prior to the S causes and they affect the exposure of units to the S causes. This requires that S be subscripted by t' or c', that is,

$$S_{t'}(u) \qquad \text{and} \qquad S_{c'}(u). \tag{42}$$

Although it is a mouthful, here is what $S_{t'}(u)$ denotes: $S_{t'}(u)$ is the S cause that u is exposed to if u was earlier exposed to the R cause t'. The following path diagram expresses this situation:



(43)

Diagram (43) indicates that R changes the values of S and Y and that S changes the value of Y. R has, potentially, both a direct and an indirect (i.e., through S) effect on Y.

An example may help clarify the meaning of (43). Suppose that we wish to measure the effect of studying certain material on the performance on a particular test. We might be able to *encourage* or *not encourage* students to study the material—these are the R causes, t' and c'. We might then be able to ascertain whether the students *did* or *did not* study the material—these are the S causes, t and c. The response variable is the score Y on the test given subsequent to these events. Diagram (43) indicates that encouragement can affect studying and possibly the test scores and that studying can affect the scores. For example, one might hypothesize that encouragement really does not affect test scores directly. This would be expressed in the model by

$$Y_{t's}(u) - Y_{c's}(u) = 0 \tag{44}$$

for all u in U and $s = t$ or c. For more on "encouragement designs" see Powers and Swinton (1984).

The essential point I wish to make about these diagrams is that they are easily interpreted in terms of Rubin's model when they are not causally meaningless. The causal model literature has not been careful in separating meaningful and meaningless causal statements and path diagrams, in

my opinion. For a similar view see Kempthorne (1978). One expects that the application of Rubin's model will help clarify the meaning of complex causal models and their path diagrams.

9. SUMMARY

This article has covered a variety of topics that involve causation, but there are a few general points that, I think, are important enough to emphasize in summary.

First of all, I believe it is very helpful to try to see what experiments (as the term is used by statisticians) tell us about causation. I have emphasized three ideas about causation on which statistical experiments focus our attention.

1. The analysis of causation should begin with studying the effects of causes rather than the traditional approach of trying to define what the cause of a given effect is.
2. Effects of causes are always relative to other causes (i.e., it takes two causes to define an effect).
3. Not everything can be a cause; in particular, attributes of units are never causes.

Let me make a few brief comments on each of these important ideas.

Traditional analyses of causation start by looking for the cause of an effect. I think that looking for causes of effects is a worthwhile scientific endeavor, but it is not the proper perspective in a theoretical analysis of causation. Moreover, I would hold that the "cause" of a given effect is always subject to revision as our knowledge about the phenomenon increases. For example, do bacteria cause disease? Well, yes . . . until we dig deeper and find that it is the toxins the bacteria produce that really cause the disease; and this is really not it either. Certain chemical reactions are the real causes . . . and so on, ad infinitum. The effect of a cause may be difficult to measure in some circumstances, but it is, at least, precisely definable—as done in Section 3. It is for this reason that I believe that formal theories of causation must begin with the effects of given causes rather than vice versa.

That an effect requires two causes for its definition is obvious in the context of an experiment but never seems to get much recognition by those who discuss causation in general terms. This is probably an important contribution of statistical thinking to discussions of causation. Experiments without control comparisons are simply not experiments. Those who think in terms of physical science experiments may have some difficulty with this idea, but I believe that it is true of any experiment.

That everything has a cause is sometimes called the law of causality, but it does not imply that everything can be a cause. The experimental model eliminates many things from being causes, and this is probably very good, since it gives more specificity to the meaning of the word *cause*. Donald Rubin and I once made up the motto

NO CAUSATION WITHOUT MANIPULATION

to emphasize the importance of this restriction. Although many people balk at the idea that causes might be limited in some way, this idea is a simple consequence of the structure of the model in Section 3. Unless both $Y_t(u)$ *and* $Y_c(u)$ can be defined, in principle, it is impossible to define the causal effect $Y_t(u) - Y_c(u)$. For an attribute $A(u)$ we can define $Y_a(u)$ for all u for which $A(u) = a$, and we can define $Y_b(u)$ for all u for which $A(u) = b$. Attributes are functions, however, and $A(u)$ is either a or b (or neither) but not both a and b for any unit, u. Hence $Y_a(u) - Y_b(u)$ cannot be defined for any unit, u, and attributes are not causes in the sense that causal effects cannot be defined for them.

The second set of important general points I wish to summarize concern the immediate consequences of Rubin's model. There are two consequences I wish to emphasize.

1. The difference between the *model* (S, Y_t, Y_c) and the process of observation (S, Y_S).
2. The Fundamental Problem of Causal Inference—only Y_t or Y_c but not both can be observed on any unit u.

These two consequences are really the same thing said in different ways. It is a great mistake to confuse Y_t or Y_c with Y_S, and yet this is done all the time. It is also a mistake to conclude from the Fundamental Problem of Causal Inference that causal inference is impossible. What *is* impossible is causal inference without making untested assumptions. This does not render causal inference impossible, but it does give it an air of uncertainty. It is the same uncertainty discussed by Hume. The strength of a model like Rubin's is that it allows us to make these assumptions more explicit than they usually are. When they are explicitly stated the analyst can then begin to look for ways to evaluate or to partially test them.

ACKNOWLEDGMENTS

I first learned about the causal model in Section 3 from the person I consider its originator, Donald Rubin. Don's work in this area is always a source of inspiration for me. Lindsey Churchill read an early draft of this article and made numerous suggestions that have improved and focused both my thinking and the article in substantial ways. Paul Rosenbaum has, very generously, given me the benefit of his insight into causal inference on many occasions. Ben King encouraged me to put the ideas in this article together as a General Methodology Lecture for the 1985 meetings of the ASA. My other colleagues at ETS—Henry Braun, Donald Rock, Dorothy Thayer, and Howard Wainer—are always a source of intelligence and keen criticism. Lynne Steinberg, as an ETS postdoctoral fellow during 1984–1985, spent many hours explaining to me how causation works in experimental psychology. Finally, Kathy Fairall's good nature and many skills insured the timely production of the manuscript for the 1985 meeting of the ASA.

[*Received October 1985. Revised January 1986.*]

REFERENCES

Barnard, G. A. (1982), "Causation," in *Encyclopedia of Statistical Sciences* (Vol. 1), eds. S. Kotz, N. Johnson, and C. Read, New York: John Wiley, pp. 387–389.

Blalock, H. M., Jr. (ed.) (1971), *Causal Models in the Social Sciences*, Chicago: Aldine-Atherton.
Bunge, M. (1959), *Causality and Modern Science* (3rd ed.), New York: Dover Publications.
Cochran, W. G. (1983), *Planning and Analysis of Observational Studies*, New York: John Wiley.
Cook, R. D. (1980), "Smoking and Lung Cancer," in *R. A. Fisher: An Appreciation*, eds. S. Fienberg and D. Hinkley, New York: Springer-Verlag.
Cornfield, J., Haenszel, W., Hammond, E. C., Lilienfeld, A. M., Shimkin, M. B., and Wynder, E. L. (1959), "Smoking and Lung Cancer: Recent Evidence and a Discussion of Some Questions," *Journal of the National Cancer Institute*, 22, 173–203.
Cox, D. R. (1958), *The Planning of Experiments*, New York: John Wiley.
Doll, R., and Hill, B. (1950), "Smoking and Carcinoma of the Lung," *British Medical Journal*, 2, September 30, 739–748.
——— (1952), "A Study of the Aetiology of Carcinoma of the Lung," *British Medical Journal*, 2, December 13, 1272–1286.
——— (1956), "Lung Cancer and Other Causes of Death in Relation to Smoking," *British Medical Journal*, 2, November 10, 1071–1081.
Duncan, O. D. (1975), *Introduction to Structural Equation Models*, New York: Academic Press.
Evans, A. S. (1978), "Causation and Disease: A Chronological Journey," *American Journal of Epidemiology*, 108, 249–258.
Fisher, R. A. (1926), "The Arrangement of Field Experiments," *Journal of Ministry of Agriculture*, 33, 503–513.
——— (1957), "Letter to the Editor," *British Medical Journal*, 2, July 6, 43.
Florens, J. P., and Mouchart, M. (1985), "A Linear Theory for Noncausality," *Econometrica*, 53, 157–175.
Goldberger, A. S., and Duncan, O. D. (1973), *Structural Equation Models in the Social Sciences*, New York: Seminar Press.
Granger, C. W. J. (1969), "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica*, 37, 424–438.
——— (1980), "Testing for Causality: A Personal Viewpoint," *Journal of Economic Dynamics and Control*, 2, 329–352.
Hamilton, M. A. (1979), "Choosing a Parameter for 2 × 2 Table or 2 × 2 × 2 Table Analysis," *American Journal of Epidemiology*, 109, 362–375.
Hill, A. B. (1965), "The Environment and Disease: Association or Causation," *Proceedings of the Royal Society of Medicine*, 58, 295–300.
Holland, P. W., and Rubin, D. B. (1980, "Causal Inference in Prospective and Retrospective Studies," address given at the Jerome Cornfield Memorial Session of the American Statistical Association Annual Meeting, August.
——— (1983), "On Lord's Paradox," in *Principals of Modern Psychological Measurement*, eds. H. Wainer and S. Messick, Hillsdale, NJ: Lawrence Erlbaum.
Hume, D. (1740), *A Treatise on Human Nature.*
——— (1748), *An Inquiry Concerning Human Understanding.*
Kempthorne, O. (1952), *The Design and Analysis of Experiments*, New York: John Wiley.
——— (1978), "Logical, Epistemological and Statistical Aspects of Nature–Nurture Data Interpretation," *Biometrics*, 34, 1–24.
Locke, J. (1690), *An Essay Concerning Human Understanding*, Book II, Chapter XXVI.
McCurdy, R. (1957), "Letter to the Editor," *British Medical Journal*, 2, July 20.
Mill, J. S. (1843), *A System of Logic.*
Neyman, J. (with Iwaszkiewicz, K., and Kolodziejczyk, S.) (1935), "Statistical Problems in Agricultural Experimentation" (with discussion), *Supplement of Journal of the Royal Statistical Society*, 2, 107–180.
Powers, D. E., and Swinton, S. S. (1984), "Effects of Self-Study for Coachable Test Item Types," *Journal of Educational Measurement*, 76, 266–278.
Rosenbaum, P. R. (1984a), "From Association to Causation in Observational Studies: The Role of Tests of Strongly Ignorable Treatment Assignment," *Journal of the American Statistical Association*, 79, 41–48.
——— (1984b), "The Consequences of Adjustment for a Concomitant Variable That Has Been Affected by the Treatment," *Journal of the Royal Statistical Society*, Ser. A, 147, 656–666.
——— (1984c), "Conditional Permutation Tests and the Propensity Score in Observational Studies," *Journal of the American Statistical Association*, 79, 565–574.
Rosenbaum, P. R., and Rubin, D. B. (1983a), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41–55.
——— (1983b), "Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study With Binary Outcome," *Journal of the Royal Statistical Society*, Ser. B, 45, 212–218.
——— (1984a), Discussion of "On the Nature and Discovery of Structure," by J. W. Pratt and R. Schlaifer, *Journal of the American Statistical Association*, 79, 26–28.
——— (1984b), "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score," *Journal of the American Statistical Association*, 79, 516–524.
——— (1985a), "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score," *The American Statistician*, 39, 33–38.
——— (1985b), "The Bias Due to Incomplete Matching," *Biometrics*, 41, 103–116.
Rubin, D. B. (1974), "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66, 688–701.
——— (1977), "Assignment of Treatment Group on the Basis of a Covariate," *Journal of Educational Statistics*, 2, 1–26.
——— (1978), "Bayesian Inference for Causal Effects: The Role of Randomization," *The Annals of Statistics*, 6, 34–58.
——— (1980), Discussion of "Randomization Analysis of Experimental Data: The Fisher Randomization Test," by D. Basu, *Journal of the American Statistical Association*, 75, 591–593.
Saris, W., and Stronkhorst, H. (1984), *Causal Modelling in Nonexperimental Research*, Amsterdam: Sociometric Research Foundation.
Smith, R. Jeffrey (1980), "Government Says Cancer Rate Is Increasing," *Science*, 227, 998–1002.
Suppes, P. C. (1970), *A Probabilistic Theory of Causality*, Amsterdam: North-Holland.
Yerushalmy, J., and Palmer, C. E. (1959), "On the Methodology of Investigations of Etiologic Factors in Chronic Diseases," *Journal of Chronic Diseases*, 10, 27–40.