The problem of drawing causal inferences from retrospective case-control studies is considered. A model for causal inference in prospective studies is reviewed and then applied to retrospective studies. The limitations of case-control studies are formulated in terms of the level of causally relevant parameters that can be estimated in such studies. An example using data from a large retrospective study of coffee-drinking and myocardial infarctions is used to illustrate the ideas of the article.

CAUSAL INFERENCE IN RETROSPECTIVE STUDIES

PAUL W. HOLLAND
Research Statistics Group
Educational Testing Service
DONALD B. RUBIN
Harvard University

Philosophical discussions of causality often emphasize the *meaning* of causation. Scientists are usually concerned with *understanding* causal mechanisms. Purely statistical discussions of causality are substantially more limited in scope, because the unique contribution of statistics is to *measuring* causal effects and not to the understanding of causal mechanisms or to the meaning of causation. This distinction is sometimes expressed as "statistics can establish correlation, but not causation." We feel our emphasis on *measurement* is more appropriate, because it focuses on what statistical theory *can* contribute to discussions of causality. Measuring causal effects accurately without any understanding whatsoever of the causal mechanisms

---

AUTHORS' NOTE: *A version of this article titled "Causal Inference in Prospective and Retrospective Studies" was delivered at the Jerome Cornfield Memorial Session of the American Statistical Association, August 1980, in Houston. The topic of the article was especially appropriate for that session since many important contributions to the study of health effects from prospective and retrospective studies were made by Jerome Cornfield.*

EVALUATION REVIEW, Vol. 12 No. 3, June 1988 203-231

involved or of the meaning of causation is not only possible but is, of course, a commonplace experience of everyday life; that is, people are quite capable of using automobiles, ovens, calculators, and typewriters safely and effectively without any knowledge of how these devices work. Of course, careful measurements of causal effects often lead to a better understanding of the causal mechanisms involved.

In this article we first review a mathematical model for causal inferences in prospective studies that is based on the work of Rubin (1974, 1977, 1978, 1980), and developed further in Holland and Rubin (1983), Rosenbaum (1984a, 1984b, 1984c), Rosenbaum and Rubin (1983a, 1983b, 1984a, 1984b, 1985a, 1985b), Holland (1986a, 1986b) and Rubin (1986). We then apply this model to study causal inference in retrospective case-control studies.

1. CAUSAL INFERENCE IN PROSPECTIVE STUDIES

The logic of measuring causal effects is clearest in prospective studies, so we begin with that case. The essential elements of a prospective study are the following:

(1) a *population* of units, U
(2) a set, K, of well-defined causal agents (also called treatments or causes) to which each unit u can be exposed. (For notational simplicity, we consider only two causal agents, K = {t,c}.
(3) a *response* Y that can be recorded for each unit after exposure to a causal agent in K.

In a prospective study, a sample of units from U is obtained and each unit is assigned to a treatment in K. The causal agents are then applied, and later the response of each unit in the study is recorded. The intuitive notion of causal effect that we wish to describe with our model is the difference between the response measured on a unit that is exposed to cause or treatment t and the response that would have been measured *on the same unit* had it been exposed to treatment c. Thus our notion of the causal effect of a causal agent will always be relative to another causal agent, and is defined for each unit in U.

This meaning of causal effect is not foreign to statistical thinking, and is evident in the writings of R. A. Fisher (1935), Kempthorne (1952), Cochran (1965), and Cox (1958), for example (see Holland, 1986b). Although this notion of a causal effect can be *defined* for each unit in U,

in general we are not able to directly *measure* a causal effect for a single unit because having exposed a unit to t, we cannot return in time to expose the same unit to c, instead. This is the Fundamental Problem of Causal Inference, to which we shall return in the next section.

Before turning to the formal model we need to clarify the nature of the *response* Y. For our discussion we will assume that Y is dichotomous, taking on only the values 0 or 1. The extension to a general Y is straightforward. We have chosen to restrict Y to be dichotomous because it is the situation of common interest in retrospective studies.

1.1 THE FORMAL MODEL AND THE DEFINITION OF UNIT-LEVEL CAUSAL EFFECTS

In the model, instead of a single dependent variable Y we have a dependent variable, Y_k, for each of the treatments to which the unit could have been exposed. Thus if the unit is exposed to causal agent t we will record the value of Y_t for that unit. If that same unit had been exposed to causal agent c instead of t, then we would record the value of Y_c for that unit and not the value of Y_t. More formally, for two treatments, we associate the following vector with each unit in U,

$$(Y_t, Y_c), \tag{1}$$

where

$Y_k(u)$ = the response made by unit u if it is exposed to cause $k \epsilon K$.

The novel feature of this model is the introduction of several versions of the response variable Y. There is a version of Y for each of the causal agents in K, because our definition of causal effect compares Y_t (the response made if exposed to t) to Y_c (the response made if exposed to c). Rubin (1980, 1986) refers to the assumption that the vector (1) fully represents the possible values of Y under all pairings of $k \epsilon K$ with $u \epsilon U$ as the "stable unit-treatment value assumption," or the SUTVA.

The fact that each unit has a value for both Y_t and Y_c is very important because it allows us to *define* causal effects at the level of individual units. On unit u in U, the causal effect of t relative to c is a comparison of $Y_t(u)$ and $Y_c(u)$, for example the difference $Y_t(u) - Y_c(u)$.

A question that immediately arises is whether or not it is *ever possible* to expose a unit to more than one treatment and thereby directly observe

more than one component of the vector in (1). One can argue that this is never possible in principle, because once a unit has been exposed to a treatment, the unit is different from what it was before. As mentioned earlier, this is the Fundamental Problem of Causal Inference (Holland, 1986). However, the propriety of this extreme position depends on the nature of the treatments and the units under study. We will not pursue this issue further here, but will simply make the "worst-case" assumption that a unit can be exposed to at most one treatment condition. For our application to retrospective studies this assumption is adequate, since in these studies units are exposed to only one of the causal agents.

In order to relate the vector in (1) to the data that are actually observed, we introduce the variable S, where S = k if the unit is exposed to cause k; S is the "causal indicator" variable that indicates to which $k \epsilon K$ each unit is exposed.

The observed data from a unit u is the vector

$$(Y_S(u), S(u)). \tag{2}$$

The notation Y_S is used because it indicates that we can observe only the response of a unit to the treatment to which it is exposed, that is,

$$Y_S = Y_k \text{ if } S = k, \text{ for } k \epsilon K. \tag{3}$$

The quantity Y_S is the observed value of the response and is therefore what is usually called the "dependent variable" in statistical discussions (e.g., in an ANOVA of an experiment). We never can observe Y_k if $S \neq k$. Since we can observe only the value of Y_t or Y_c but not both, it is a consequence of the model that causal effects for individual units *are not directly measurable.*

In summary, our idealized model for a prospective causal study can be viewed as based on the following sequence of steps:

(1) determination of the population U under study
(2) determination of the set K composed of causal agents, treatments, or causes under study
(3) determination of the response variable Y to be observed
(4) consequent definition of the vector (Y_t, Y_c) for every unit in U
(5) determination of the causal indicator S for every unit in the study
(6) consequent definition of the vector of observable data (Y_S, S) for every unit in the study
(7) observation of (Y_S, S) for each unit in the study

1.2 THE THREE LEVELS OF CAUSAL INFERENCE

There are three levels or "strengths" of causal inferences that arise in practice. These are: unit-level, subpopulation-level, and population-level causal inferences. These levels are ordered by decreasing strength in the sense that knowledge of all unit-level causal inferences implies knowledge of all sub- and population-level causal inferences, and knowledge of all subpopulation causal inferences for a partition of U implies knowledge of population-level causal inferences, but not vice versa. We briefly describe each of these levels. Due to the Fundamental Problem of Causal Inference, all of these involve *indirect* estimation of causal effects.

Unit-level causal inference: The *definition* of causal effects is at the unit level and is the difference

$$Y_t(u) - Y_c(u)$$

Units are called *homogeneous* when $Y_k(u_1) = Y_k(u_2)$ for every pair of units, u_1, u_2. Homogeneous units may be encountered in laboratory research in the physical sciences. When unit homogeneity can be assumed, unit-level causal inferences are easy, since $Y_t(u) - Y_c(u) = Y_t(u_1) - Y_c(u_2)$ for every u, u_1, and $u_2 \in U$. Hence unit-level causal inference only requires the observation of Y_t on u_1 and Y_c on u_2 for one pair of units, u_1, u_2. Such situations do not require statistical methods for the estimation of causal effects.

Population-level causal inference: The population distribution of Y_t and Y_c over U are, in the dichotomous case, specified by $P(Y_t = 1)$ and $P(Y_c = 1)$. A population causal inference is a comparison of these two probabilities—or, more generally, of the distributions of Y_t and Y_c over U. A population causal inference is weaker than a unit-level causal inference because it only describes how t or c affects the distribution of Y over all of U rather than how it affects the value of Y on a given unit u in U. There are many ways to compare distributions, and two important ones for the dichotomous case are the *difference*

$$P(Y_t = 1) - P(Y_c = 1), \tag{4}$$

and the *odds-ratio,*

$$\alpha = \frac{P(Y_t = 1)}{P(Y_t = 0)} \Big/ \frac{P(Y_c = 1)}{P(Y_c = 0)} \tag{5}$$

Although (4) and (5) are equally good ways of comparing dichotomous distributions in general, they are quite different from a causal point of view. The difference (4) may be interpreted in two ways, first as a difference in probability, and second as an *average causal effect,* or ACE,

$$P(Y_t = 1) - P(Y_c = 1) = E(Y_t - Y_c) = ACE. \quad (6)$$

An ACE is the average of all the unit-level causal effects over U, and sometimes, for example when Y is continuous, an ACE can be shown to approximate *all* of the unit-level causal effects in U. This fortunate state of affairs occurs when we have a case of constant effects, i.e.

$$Y_t(u) - Y_c(u) = T \text{ for all } u \epsilon U.$$

This form of additivity does not often occur for dichotomous Y's except under very special circumstances.

The odds ratio in (5), while a useful comparison of probabilities, does not have an interpretation as an ACE, but we shall see that it arises naturally in the study of retrospective research designs. The odds ratio (5) is a comparison of the distribution of Y_t over U with that of Y_c, with no concern with how Y_t and Y_c might be related for particular units in U.

Subpopulation-level causal inference: Subpopulations of U may be defined in many ways, but we shall use only one method. It is to define subpopulations by the values of *covariates.*

A covariate is a value of a variable that is defined on each unit u in U but that is not affected by the exposure of units to causes in K. In our model, the introduction of a *variable,* X, defined on the units of U, requires the notation $X_k(u)$ to indicate that, in general, X can depend both on the unit u and on which cause, k, to which u is exposed. A *covariate* is a special type of variable for which $X_k(u) = X(u)$ for all $k \epsilon K$. Variables that are *measured prior to* the exposure of units to causal agents are always covariates, but sometimes variables measured after exposure are also covariates—for example, a drug treatment is not likely to change the value of variables such as an adult's height.

Subpopulations defined by a covariate, X, allow us to consider probabilities of the form

$$P(Y_t = 1 \mid X = x) \text{ and } P(Y_c = 1 \mid X = x).$$

Subpopulation-level causal inferences may be based on such proba-

$$P(Y_t = 1 \mid X = x) - P(Y_c = 1 \mid X = x), \tag{7}$$

or the *conditional odds ratio*

$$\alpha(x) = \frac{P(Y_t = 1 \mid X = x)}{P(Y_t = 0 \mid X = x)} \Big/ \frac{P(Y_c = 1 \mid X = x)}{P(Y_c = 0 \mid X = x)} \tag{8}$$

Just as population-level causal inferences distinguish (4) from (5), so, too, do subpopulation-level causal inferences distinguish (7) from (8) for the same reasons. Similarly, the difference (7) also can be interpreted as a conditional ACE, that is,

$$P(Y_t = 1 \mid X = x) - P(Y_c = 1 \mid X = x) = E(Y_t - Y_c \mid X = x), \tag{9}$$

whereas the odds ratio (8) has no such interpretation. Yet (8) is still useful as an intermediate type of causal inference between the unit level and population level because it can describe the way that the causal agent t changes the distribution of Y-values relative to that of the causal agent c for the subpopulation of U for which X(u) = x. If this is a relatively homogeneous population, such information can be tantamount to a unit-level causal inference, and thus a conditional ACE.

1.3 THE ROLE OF RANDOMIZATION IN PROSPECTIVE STUDIES

It is well known that randomization aids one substantially in drawing causal inferences, yet why is this so? The model developed above gives an easy answer to this question.

When randomization is used to assign units to exposure to the causal agents in K, the variable S is made statistically independent of *all* other variables defined on U. Hence in particular,

$$E(Y_k \mid S = k) = E(Y_k) \tag{10}$$

for k = t or c (or for all choices of $k \epsilon K$ when K has more than two members).

Now let us consider the data that can be observed in a prospective study, that is, $(Y_S(u), S(u))$ for u in the study.

From a set of observations of (Y_S, S) we can calculate the distribution of Y_S given the observed value of S. In particular we can estimate the value of

$$E(Y_k \mid S = k) \tag{11}$$

The mean value in (11) is also equal to

$$E(Y_k|S = k) \tag{12}$$

by the standard rules of conditional probability. Now suppose randomization is employed. Using (10), (11), and (12) we have this basic identity (that holds only for randomized studies in general):

$$E(Y_S|S = k) = E(Y_k) \tag{13}$$

The difference

$$E(Y_S|S = t) - E(Y_S|S = c) \tag{14}$$

is called the *prima facie* ACE (or FACE) in general. In a randomized study the FACE equals the ACE, that is,

$$E(Y_S|S = t) - E(Y_S|S = c) = E(Y_t - Y_c). \tag{15}$$

The FACE in (14) is a quantity that can be estimated from the data in *any* prospective study. In a *randomized* prospective study the FACE has *causal relevance,* since it equals the ACE.

1.4 THE ROLE OF COVARIATES IN PROSPECTIVE STUDIES

When a covariate is available in a study, the observed data for each unit in the study is expanded from (Y_S, S) to

$$(Y_S, S, X) \tag{16}$$

where X is the covariate (possibly vector-valued). From observed values of these data for the units in the study we can, in principle, estimate the regression of Y_S on S and X, that is,

$$E(Y_S|S =k, X = x). \tag{17}$$

The quantity in (17), however, equals

$$E(Y_k|S = k, X = x) \tag{18}$$

from the standard rules of conditional probability. But (18), as it stands, is much like (12) in not having any causal relevance. Rosenbaum and Rubin (1983) define a special condition that generalizes randomization and that gives (18) causal relevance when it holds. It is the condition of *strong ignorability*. Treatment assignment (i.e., the distribution of S given X and Y_k) is *strongly ignorable* if (a) given X, S is independent of Y_k for $k \epsilon K$, and (b) $P(S = k|X) > 0$ for all $k \epsilon K$. This is a stronger condition than *ignorability* defined by Rubin (1978) for Bayesian inference.

If strong ignorability holds, then (18) becomes

$$E(Y_k|X = x), \tag{19}$$

and the difference

$$E(Y_t|S = t, X = x) - E(Y_c|S = c, X = x), \tag{20}$$

which is the conditional FACE, equals the conditional ACE; that is, (20) equals

$$E(Y_t|X = x) - E(Y_c|X = x). \tag{21}$$

Note that by averaging over the *conditional* ACE in (21) we obtain the ACE in (15), that is,

$$E[E(Y_t - Y_c|X)] = E(Y_t - Y_c). \tag{22}$$

Hence the condition of strong ignorability is less restrictive than that of randomization, but it still allows us to measure the average causal effect in a prospective study using the data that are available. Of course, the plausibility of the assumption of strong ignorability needs to be considered carefully in any real application. For a more detailed discussion of these issues see Rubin (1977), Holland and Rubin (1983), and Rosenbaum (1984a).

2. CAUSAL INFERENCE IN RETROSPECTIVE CASE-CONTROL STUDIES

The structure of a retrospective case-control study is considerably different from the general prospective study discussed in the preceding

section. In a case-control study, a population of units is divided into those who have a particular symptom or disease of interest (i.e., the "cases") and those who do not have the symptom or disease (i.e., the "controls"). Samples (random samples, in principle) of cases and controls are selected from this population, and information about each selected person is obtained to ascertain (a) the level of exposure to the particular causal agents of interest and (b) other medically relevant information that may be used to define subpopulations of units.

The response variable for a case-control study is the dichotomous variable that indicates whether or not the unit is a "case" or a "control," that is,

$$Y_S = \begin{cases} 1 \text{ if unit is a case} \\ 0 \text{ if unit is a control.} \end{cases}$$

Case-control studies are *retrospective* because they begin at the endpoint of a prospective study (i.e., observations of the response variable for each unit in the study) and then look back in time to discover the causal agent to which each unit has been exposed (i.e., the value of the causal indicator S). In retrospective studies, the basic groups are cases ($Y_S = 1$) and controls ($Y_S = 0$) with S measured on each sampled unit, whereas in prospective studies the basic groups are exposed ($S = t$) and not exposed ($S = c$), with Y_S measured on each sampled unit. In addition to this fundamental difference between case-control and prospective studies, two other differences should be mentioned. First, since the investigator can only *collect* data on prior exposure to the causal agents of interest, it is impossible to use randomization to *assign* units to the causal agents. Thus case-control studies are never randomized. Prospective studies, on the other hand, may or may not employ randomization depending on the amount of control that is possible. Second, the populations studied in case-control studies usually consist of survivors only, because it is often impossible to obtain comparable data on individuals who are deceased. This limitation may have consequences for the interpretability of the results of a case-control study. An excellent reference for case-control studies is Breslow and Day (1980).

Although, in principle, it is almost always possible to design a prospective version of a case-control study, it is often much more expensive than the case-control study. There are several reasons for this:

(a) prospective studies require large sample sizes in which the "cases" are rare (e.g., when Y_S = represents a rare disease) and (b) prospective studies may involve long time spans before relevant data become available. Hence it is likely that case-control studies will always be an attractive possibility for many types of scientific investigations, especially in the early stages of the research. It is therefore important to know their limitations, to design them as well as possible and to analyze the data collected in such studies correctly. Our goal in the present article is to illuminate all of these points by applying the model for causal inference outlined in Section 1 to case-control studies.

2.1 THE STANDARD TWO-WAY TABLE

In analyzing data from a case-control study it is customary to form and draw conclusions from the two-way table of counts illustrated in Table 1. We assume that this table is formed by randomly sampling m_{1+} "cases" (units with $Y_S = 1$) from the subpopulation of cases, and randomly sampling m_{0+} "controls" (units with $Y_S = 0$) from the subpopulation of controls.

In Table 1, m_{yk} is the number of units in the study for which $Y_S = y$ and $S = k$. For example, m_{1c} is the number of "cases" in the study that were observed exposed to causal agent c. Before examining this table of sample data, let us consider the population table that underlies it. Table 2 gives the population proportion of people with exposure to t or c among all those who are cases or controls. These population values are denoted by

$$r_{yk} = P(S = k | Y_S = y). \tag{23}$$

The corresponding sample ratio

$$\hat{r}_{yk} = m_{myk}/m_{y+} \tag{24}$$

estimates r_{yk}. We shall call the $\{r_{yk}\}$ the *retrospective probabilities* of the study. They are "retrospective" because the conditioning is on an event that occurs *after* the event whose probability is being assessed.

In this development we must emphasize the importance of representing the observed value of the response as Y_S. For example, in (23) it

TABLE 1
The Standard Two-Way Table in Retrospective Studies Showing the Sample Distribution of Cases and Controls Observed for Each Causal Agent

		Causal Agent		
		$S = t$	$S = c$	*Total*
Cases	$Y_S = 1$	m_{1t}	m_{1c}	m_{1+}
Controls	$Y_S = 0$	m_{0t}	m_{0c}	m_{0+}
	Total	m_{+t}	m_{+c}	m_{++}

TABLE 2
The Population Table of Retrospective Probabilities r_{yk} Related to the Sample Table in Table 1

		Causal Agent		
		$S = t$	$S = c$	*Total*
Cases	$Y_S = 1$	$r_{1t} = P(S = t \mid Y_S = 1)$	$r_{1c} = P(S = c \mid Y_S = 1)$	1
Controls	$Y_S = 0$	$r_{0t} = P(S = t \mid Y_S = 0)$	$r_{0c} = P(S = c \mid Y_S = 0)$	1

would be incorrect to condition on $Y_k = y$ since Y_k is the response made if exposed to cause k, whereas Y_S is the observed response. Because Y_S is being conditioned on in Table 2, it is sometimes said that in a case-control study *exposure* is the dependent variable and *diagnosis* (i.e., case or control) is the independent variable. This description is neither helpful nor of scientific interest, and we will not describe the situation in these terms.

If we consider the weakest level of causal inference, that is, a population-level causal inference, then the causal parameters are the marginal probabilities $P(Y_t = 1)$ and $P(Y_c = 1)$. Thus the retrospective probabilities in (23) are not in themselves of any causal interest, because, at the very least, they do not address the correct events. However, by applying the usual rules of probability, we may reverse the roles of S and Y_S in (23) and obtain more interesting probabilities. The result of this

reversal is the accepted justification for looking at Table 1 (see Cornfield, 1956).

2.2 RELATING RETROSPECTIVE AND PROSPECTIVE PROBABILITIES

To reverse the roles of S and Y_S we make use of Bayes's theorem to obtain

$$P(Y_S = y \mid S = k) = P(S = k \mid Y_S = y) \frac{P(Y_S = y)}{P(S = k)} \tag{25}$$

However,

$$P(Y_S = y \mid S = k) = P(Y_k = y \mid S = k) \tag{26}$$

so it follows that

$$P(Y_k = y \mid S = k) = P(S = k \mid Y_S = y) \frac{P(Y_S = y)}{P(S = k)}$$

The probability $P(Y_k = y|S = k)$ is "prospective" because the conditioning event occurs prior in time to the event whose probability is being assessed. We denote these prospective probabilities by

$$p_{yk} = P(Y_k = y|S = k) \tag{27}$$

Hence the retrospective and prospective probabilities are related by

$$p_{yk} = r_{yk} \frac{a_y}{b_k} \tag{28}$$

where

$$a_y = P(Y_S = y),$$

and

$$b_k = P(S = k)$$

$$= \sum_y P(S = k \mid Y_S = y)\ P(Y_S = y),$$

or

$$b_k = \sum_y r_{yk}\ a_y \tag{29}$$

Hence the prospective probabilities p_{yk} can be determined from the retrospective probabilities r_{yk} and the overall proportion of cases and controls in the population, a_y, via

$$p_{yk} = \frac{r_{yk}\ a_y}{\sum_z r_{zk}\ a_z} \tag{30}$$

We illustrate the array of "prospective" probabilities of (26) and (27) in Table 3.

The cross-product ratio for Table 2 may be expressed as:

$$\alpha_{ret} = \frac{r_{1t}}{r_{1c}} \Big/ \frac{r_{0t}}{r_{0c}} = \frac{P(S = t \mid Y_S = 1)}{P(S = c \mid Y_S = 1)} \Big/ \frac{P(S = t \mid Y_S = 0)}{P(S = c \mid Y_S = 0)} \tag{31}$$

The cross-product ratio for Table 3 may be expressed as:

$$\alpha_{pro} = \frac{p_{1t}}{p_{0t}} \Big/ \frac{p_{1c}}{p_{0c}} = \frac{P(Y_t = 1 \mid S = t)}{P(Y_t = 0 \mid S = t)} \Big/ \frac{P(Y_c = 1 \mid S = c)}{P(Y_c = 0 \mid S = c)} \tag{32}$$

Because Tables 2 and 3 are related via row and column multiplication, see equation (28), it is well known (e.g., Bishop et al., 1975) and easily shown that the two cross-product ratios α_{ret} and α_{pro} are equal.

2.3 WHY THE STANDARD TWO-WAY TABLE IS MISLEADING

Even though the equality of α_{ret} and α_{pro} is the usual reason one ever looks at the data in Table 1, the crucial question of how it relates to the causal parameters of interest remains, that is, $P(Y_t = 1)$, $P(Y_c = 1)$ or the

TABLE 3
The Population Table of Prospective Probabilities

		Causal Agents	
		$S = t$	$S = c$
Cases	$Y = 1$	$p_{1t} = p(Y_t = 1 \mid S = t)$	$p_{1c} = p(Y_c = 1 \mid S = c)$
Controls	$Y = 0$	$P_{0t} = P(Y_t = 0 \mid S = t)$	$P_{0c} = P(Y_c = 0 \mid S = c)$
	Total	1	1

average causal effects in (6) or (9). The answer is that, without an additional assumption, the cross-product ratio, α_{pro}, has *no causal relevance.*

The crucial assumption is randomization, because then S is independent of Y_k and

$$P(Y_k = 1 | S = k) = P(Y_k = 1)$$

and hence we have

$$\alpha_{pro} = \frac{P(Y_t = 1)}{P(Y_t = 0)} \Big/ \frac{P(Y_c = 1)}{P(Y_c = 0)} \quad \text{(i.e., } = \alpha). \tag{33}$$

Hence randomization implies that α_{pro} equals α defined in (5), which is a population-level causal parameter. However, as we stated earlier, retrospective studies are never randomized, so that the assumption that S is independent of Y_k is dubious in most cases. Thus there is generally no value to examining the data in Table 1 from the point of view of using it to estimate causal parameters.

2.4 THE ROLE OF COVARIATES IN RETROSPECTIVE STUDIES

If there is a covariate X (possibly a vector) that is measured on each unit in the study, then we may form a table like Table 1 for each value of X. Let m_{ykx} be the number of units in the study for which $Y_S = y$, $S = k$,

and X = x. These are arrayed in Table 4 for X = x. We suppose that at each value of X, the data arise from a random sample of cases and a random sample of controls, not necessarily with the same sampling rates. This sampling scheme includes matched case-control pairs, where the cases are randomly sampled from the population of cases, and for each sampled case with X = x, a matching control with X = x is found.

The sample ratios

$$\hat{r}_{ykx} = m_{mkx}/m_{y+x} \tag{34}$$

estimate the population *retrospective probabilities*

$$r_{ykx} = P(S = k \mid Y_S = y, X = x). \tag{35}$$

We may again apply Bayes's theorem to reverse the roles of S and Y_S in (35) as we did in (25). This yields

$$P(Y_S = y \mid S = k, X = x) = P(S = k \mid Y_S = y, X = x) \frac{P(Y_S = y \mid X = x)}{P(S = k \mid X = x)}. \tag{36}$$

However,

$$P(Y_S = y \mid S = k, X = x) = P(Y_k = y \mid S = k, X = x) \tag{37}$$

so that

$$P(Y_k = y \mid S = k, X = x) = P(S = k \mid Y_S = y, X = x) \frac{P(Y_S = y \mid X = x)}{P(S = k \mid X = x)} \tag{38}$$

Again the probability $P(Y_k = y \mid S = k, X = x)$ is "prospective," and we denote it by

$$p_{ykx} = P(Y_k = y \mid S = k, X = x), \tag{39}$$

and, as before, the retrospective and prospective probabilities are related by

$$p_{ykx} = r_{ykx} \frac{a_{yx}}{b_{kx}} \tag{40}$$

TABLE 4
The Distribution of Cases and Controls in the Sample Observed for Each Causal Agent, at X = x

		Value of X = x		
		Causal Agent		
		S = t	*S = c*	*Total*
Cases	$Y_S = 1$	m_{1tx}	m_{1cx}	m_{1+x}
Controls	$Y_S = 0$	m_{0tx}	m_{0cx}	m_{0+x}
	Total	m_{+tx}	m_{+cx}	m_{++x}

where

$$a_{yx} = P(Y_S = y | X = x)$$

and

$$b_{kx} = P(S = k | X = x) = \sum_y P(S = k | Y_S = y, X = x)\, P(Y_S = y | X = x)$$

or

$$b_y = \sum_y r_{ykx}\, a_{yx}$$

Again, p_{ykx} can be determined from r_{ykx} and a_{yx}, the proportion of cases and controls among those units with X = x, via

$$p_{ykx} = \frac{r_{ykx}\, a_{yx}}{\sum_z r_{zkx}\, a_{zx}} \tag{41}$$

The odds ratio for the tables of retrospective probabilities underlying the sample in Table 4 is

$$\alpha_{ret}(x) = \frac{r_{1tx}}{r_{1cx}} \Big/ \frac{r_{0tx}}{r_{0cx}} \tag{42}$$

The corresponding odds ratio for the prospective probabilities is

$$a_{pro}(x) = \frac{p_{1tx}}{p_{0tx}} \Big/ \frac{p_{1cx}}{p_{0cx}} \tag{43}$$

As before, these two odds ratios are equal, that is,

$$\alpha_{pro}(x) = \alpha_{ret}(x). \tag{44}$$

Under strong ignorability we have

$$P(Y_k = y | S = k, X = x) = P(Y_k = y | X = x), \tag{45}$$

so that

$$\alpha_{pro}(x) = \frac{P(Y_t = 1 | X = x)}{P(Y_t = 0 | X = x)} \Big/ \frac{P(Y_c = 1 | X = x)}{P(Y_c = 0 | X = x)} \quad (\text{i.e.} = \alpha(x)) \tag{46}$$

Hence, when strong ignorability holds, $\alpha_{pro}(x)$ equals $\alpha(x)$ as defined in (8), which is a subpopulation causal parameter.

It may happen that $\alpha(x)$ as defined in (8) and (46) does not depend on x—this is the case of no interaction of X with the effect of the causal agents on the distribution of Y-values. In such a situation we will denote this common value by α_o. Note that $\alpha_{pro}(x)$ can be constant in x regardless of the plausibility of strong ignorability—these are two, quite different, assumptions. The constancy of $\alpha_{pro}(x)$ is testable with the data (Y_S, S, X) whereas strong ignorability is not.

Unfortunately, there is no simple relation between the population causal parameter

$$\alpha = \frac{P(Y_t = 1)}{P(Y_t = 0)} \Big/ \frac{P(Y_c = 1)}{P(Y_c = 0)} \tag{47}$$

and α_o even when strong ignorability and the assumption of no interaction between X and the effect of t holds. Nevertheless, α_o is a causally interesting parameter itself: α_o is the amount by which the odds for $Y_t = 1$ are increased over the odds that $Y_c = 1$ in each X-stratum of U, and, thus, α_o is a causal parameter. Since α_o is specific to each X-stratum

of U, it provides causal information about the effects of t relative to c in U that is at a more detailed level than the overall population level. However, in general it is not as strong a causal inference as a unit-level causal inference.

Our conclusion is that in a case-control study the simple two-way table (Table 1) generally holds no causal interest even for a matched case-control study (i.e., $M_{1+x} = m_{0+x}$). It is crucial to stratify on covariates and to estimate $\alpha_{ret}(x) = \alpha_{pro}(x)$—which is a causally relevant parameter, $\alpha(x)$, under strong ignorability. If the stratified table exhibits constant odds ratios, then, assuming strong ignorability, this parameter equals α_0 and gives the amount that t increases the proportion of units in each X-stratum that are "cases" relative to c. This "amount of increase" is in terms of the odds corresponding to the proportions. Thus, for example, for a given value of the proportion $P(Y_c = 1|X = x)$, we may calculate $P(Y_t = 1|X = x)$ via the formula

$$P(Y_t = 1|X = x) = \frac{\alpha_0 P(Y_c = 1|X = x)}{P(Y_c = 0|X = x) + \alpha_0 P(Y_c = 1|X = x)}. \quad (48)$$

Comparing this to the given value of $P(Y_c = 1|X = x)$ leads to a causal inference about the effect of the causal agent when $X = x$. In general, however, the stratified table will not exhibit a constant odds ratio and then the values of $\alpha(x)$ are the causally relevant parameters (under strong ignorability), and can be used in place of α_0 in (48) to calculate $P(Y_t = 1|X = x)$ from $P(Y_c = 1|X = x)$.

AN EXAMPLE

THE DATA

The following data are taken from a case-control study of the relationship between coffee drinking and the occurrence of myocardial infarctions (MI) by Jick et al. (1973). We use these data for illustrative purposes only. A total of 24,741 patients were classified as "cases" (had an MI) or "controls" (did not have an MI). Table 5 shows the standard two-way table that presents the cases and controls cross-classified by the potential causal agents under study—self-reported daily coffee con-

sumption. Although our previous notation has considered only two causal agents, Table 5 presents four, a control (0 cups per day) and three levels of the amount of coffee drinking; the extensions needed to handle this extra complexity are simple. The odds ratios estimated in Table 5 are defined by

$$\alpha^{k}_{pro} = \frac{P(Y_k = 1 | S = k)}{P(Y_k = 0 | S = k)} \Big/ \frac{P(Y_1 = 1 | S = 1)}{P(Y_1 = 0 | S = 1)}, \tag{49}$$

for k = 2, 3, 4, that is, α^{k}_{pro} is the odds ratio for level k of coffee drinking relative to the control of no coffee drinking.

Table 5 suggests a modest increase in the risk of MI among persons who drink coffee. The odds ratios range from 1.5 to 1.8. The odds ratios exhibited in Table 5 are not monotone in the amount of self-reported coffee drinking, and the effect seems to be almost as strong for persons who drink 1-2 cups per day as for those drink 6 or more cups per day.

Table 5, however, does not take various background variables into account and, as we have discussed earlier, is therefore likely to be misleading because it is not reasonable to believe that the drinking of coffee is randomly assigned and therefore independent of Y_k, k = 1, 2, 3, 4.

In addition to the variables

S = level of self-reported coffee intake

and

Y = case or control,

the following set of variables were also available on all patients in the study.

A = Age: 6 levels: 20-29, 30-39, . . . , 70-79
G = Gender: 2 levels: male, female
C = Smoking: 3 levels: other, ex-smoker, current smoker
O = Other heart disease: 2 levels: yes, no

In addition, because the data were collected from 24 suburban Boston hospitals, a fifth variable, H = hospital, was included in the

TABLE 5
Cross-Tabulation of Self-Reported Daily Coffee Intake (S) by Cases and Controls (Y) for 24,741 Patients

		S = 1 *0 cup/day*	*S = 2* *1-2 cups*	*S = 3* *3-5 cups*	*S = 4* *6+ cups*	*Total*
$Y_S = 1$	MI cases	128	269	147	86	630
$Y_S = 0$	non-MI controls	6918	9371	5290	2532	24111
	Total	7046	9640	5437	2618	24741

NOTE: Estimated raw odds-ratios, α^k_{pro}, relative to k = 1. α^2_{pro} 1.551; α^3_{pro} 1.502; α^4_{pro} 1.836

analysis (with 24 levels). This results in a covariate X, which takes on $6 \times 2 \times 3 \times 2 \times 24 = 1728$ values, so that when Table 5 is stratified on X we obtain a seven-way contingency table with $2 \times 4 \times 1728 = 13{,}824$ cells. With a total of 24,741 observations, this gives us about 1.9 observations per cell—a very sparse table indeed! Many approaches to simplifying this sort of situation are possible, for example, see Breslow and Day (1980). We shall use log-linear contingency table models: (a) because of their direct relationship to odds ratios, (b) because they allow us to adjust for the effect of all of the covariates simultaneously, and (c) because they allow us to smooth the sparse, seven-dimensional table.

3.2 LOG-LINEAR MODELS FOR THIS PROBLEM

Let X = (A, G, C, O, H) denote our complete vector of covariates. The logs of the retrospective probabilities r_{ykx} from (35) may be expressed as

$$\log(r_{ykx}) = u + u_{1(y)} + u_{2(k)} + u_{3(x)} + u_{12(y,k)} + u_{13(y,x)} + u_{23(k,x)} + u_{123(y,k,x)} \tag{50}$$

where the u-terms in (50) are assumed to satisfy the usual ANOVA-like identifying constraints, $u_{1(+)} = u_{2(+)} = 0$, etc. (Bishop et al., 1975). We need to express the odds ratios

$$\alpha^k_{ret}(x) = \frac{r_{1kx}}{r_{11x}} \Big/ \frac{r_{0kx}}{r_{01x}}, \tag{51}$$

in terms of the u-terms in (50). It is easy to show that the following equation holds:

$$\alpha^k_{ret}(x) = \alpha^k_{ret} \exp\{u_{123(l,k,x)} - u_{123(l,l,x)} - u_{123(0,k,x)} + u_{123(0,l,x)}\} \quad (52)$$

where

$$\alpha^k_{ret} = \exp\{u_{12(l,k)} - u_{12(l,l)} - u_{12(0,k)} + u_{12(0,1)}\} \quad (53)$$

From (52) it follows that the hierarchical log-linear model specified by setting all $u_{123} = 0$ corresponds to the assumption that

$$\alpha^k_{ret}(x) = \alpha^k_{ret}, \quad (54)$$

for all x. Thus we may investigate the question of whether or not the odds ratio, $\alpha^k_{ret}(x)$, depends on x by testing for three-way interaction of the various covariates in X and with Y_S and S. Furthermore, if a model where all $u_{123} = 0$ is acceptable, the estimated u_{12}-terms may be used to obtain estimates of α^k_{ret}. If we are willing to make the assumptions necessary to ensure that α^k_{ret} is the causally relevant parameter discussed in Section 3.4, that is, α^k_0, then we may test $\alpha^k_{ret} = 1$ (i.e., no effect of different levels of the causal agent) by testing that $u_{12} = 0$. This test will adjust for the distribution of the covariates in the several exposure groups. In the remainder we assume strong ignorability and refer to α^k_0 rather than α^k_{ret}.

3.3 SIMPLIFYING THE ANALYSIS

As described above it may seem as though we are considering the whole $2 \times 4 \times 1728$ table, but one important feature of the use of log-linear models is that they do not force this when there are insufficient data to do so. Instead we break up the joint distribution of X = (A, G, C, O, H) into various marginal distributions and expand the model in (50) to make use of them. In the present example we expand the table to the full seven dimensions, fit all u-terms involving Y_S and/or S but not X, and only fit effects for the following pairs and triples of variables involving X:

(u_{23}) HS/AS/GCS/GOS/COS/

$$(u_{13})\ \mathrm{HY/AGY/ACY/AOY/GCY/GOY/COY/}$$
$$(u_3)\ \mathrm{HA/HG/HC/HO/AGC/AGO/ACO/GCO} \qquad (55)$$

The u-terms in parentheses indicate which terms in (50) have been expanded in the seven-way table.

3.4 RESULTS

If we fit the log-linear model indicated by the pairs and triples of variables in (55) and then delete the YS terms and refit the model, we obtain a likelihood-ratio test of $\alpha^k_0 = 1$ for k = 2, 3, 4. The value of the likelihood-ratio statistic is 12.3, which, under the null hypothesis, has three degrees of freedom. Thus this analysis supports the conclusion that at least one α^k_0 is not 1, and thus that there is a relationship between coffee consumption and myocardial infarctions. The estimated values of α^k_0 are

$\hat{\alpha}^2_0$	$\hat{\alpha}^3_0$	$\hat{\alpha}^4_0$	
1.188	1.235	1.719	(56)

as opposed to the raw odds ratios given in Table 5. These adjusted odds ratios are monotonic in the amount of coffee consumed, with the major effect for higher levels of coffee consumption. If we are willing to assume strong ignorability of the distribution of coffee consumption and diagnostic status (case or control) given the set of covariates in X, then these estimated odds ratios are the subpopulation-level causal effects described in Section 2.4.

To study the question of whether $\alpha_{ret}(x) = \alpha_{pro}(x)$ varies with x, we fit five additional models, each of which supplements (55) with one of these triples of variables: HSY, ASY, GSY, CSY, or OSY. The likelihood-ratio statistics for these models, the degree of freedom, and attained significance levels are given in Table 6.

None of these interactions are strong enough to be statistically significant at conventional levels. This result contradicts previous analysis of these data that found an interaction of the effect of coffee drinking on diagnostic status with these variables, (Miettinen, 1976).

TABLE 6
Summary of Study of Dependence of a^k_{ret} (x) on x

Interaction of YS with	*df*	*LR Statistic*	*Level Attained*
H	69	79.39	.20
A	15	10.31	.80
G	3	2.5	.47
C	6	8.83	.18
O	3	3.97	.25

4. DISCUSSION AND SUMMARY

4.1 RISK FACTORS: CAUSAL AGENTS VERSUS ATTRIBUTES

In medical studies the term "risk factor" is used to lump "true" causal agents such as smoking or coffee drinking, which can be altered, with individual attributes such as age and sex, which cannot be altered. We think it is wise to distinguish carefully between these two types of risk factors and to reserve discussions of causation to include only the former. The model used in this article presupposes a response value Y_t if the unit is exposed to t and a value Y_c if it is exposed to c. When t or c is construed to be an attribute of a unit (e.g., a person's sex) it is entirely unclear how to define both Y_t *and* Y_c on each unit. In this sense it is meaningless to speak of estimating the causal effects of attributes of units. This does not mean that attributes have no predictive value, since prediction is simply a consequence of association between variables, which does not necessarily involve notions of cause. Our definition of a causal agent is much stricter than some definitions used by economists, for example, Granger Causality (Granger, 1969). Granger labels as a *cause* any predictor of Y that adds independent information to the prediction. We believe this is too generous a definition of causality, not only misusing the language but possibly leading researchers away from the study of the effects of manipulations that are possible—see Holland (1986) and Rubin (1986).

4.2 RANDOMIZATION AND STRONG IGNORABILITY

One useful feature of the model developed for prospective studies in Section 1 is that it clarifies the importance of randomization in causal

studies. The statistical independence of Y_k and S that randomization ensures is very important but not always appreciated by writers on the subject. For example, it is often asserted that there is some sort of difficulty in resolving randomization of treatments to units with the Bayesian/likelihood/modeling framework (Basu, 1980, Kempthorne, 1976, and Kruskal, 1980, but argued otherwise by Rubin, 1978, 1980). One possible source of confusion is that the independence of S and (Y_t, Y_c) does not imply that S is independent of the observed response Y_S except in very special circumstances, for example, when $Y_t = Y_c$ for all units. However, the equation

$$E(Y_S|S = t) - E(Y_S|S = c) = E(Y_t - Y_c) \tag{57}$$

which is a consequence of randomization, has an impact on both Bayesians and frequentists alike. This is simply because it states that a population parameter that can be estimated with observed data, that is, the FACE, $E(Y_S|S = t) - E(Y_S|S = c)$, equals a population parameter that has causal relevance, that is, the ACE, $E(Y_t - Y_c)$.

The assumption of strong ignorability is a crucial one for causal inferences in retrospective studies. Because such studies are never randomized, strong ignorability appears to be one of the few constructs available to us for using data from retrospective studies to draw the type of conclusions we might try to make in a prospective study. There are two reasons why we might be willing to assume strong ignorability even when the stronger assumption of randomization is absurd. First of all, if each X-stratum contains a very homogeneous set of units who tend to respond very similarly to t or c then it can be shown that strong ignorability will hold approximately. Second, we may be willing to make the assumption of ignorability because there is nothing in the observed data to contradict it. This is a subtle point and needs elaboration. Suppose that we assumed that S is randomized and, therefore, independent of Y_k and all other variables, including X. Then we could check this assumption by examining the observed distribution of X given S. Under randomization X and S are independent, so

$$P(X = x|S = k) = P(X = x) \tag{58}$$

Equation (58) can be checked with a simple chi-square test for homogeneous proportions, and rejection indicates that S is *not* independent of X and therefore not randomized. However, if we assume

that S is *conditionally* independent of Y_k given $X = x$ then we cannot use the observed distribution of X given $S = k$ to disprove this assumption. Hence, strong ignorability is the *strongest* independence assumption that we can make that is not contradictable by the data if we restrict ourselves to (Y_S, S, X).

One point that should be emphasized is that the causal parameters that can be estimated in retrospective studies are more limited than those that can be estimated in prospective studies, even when we are willing to make the strong ignorability assumption in both cases. For retrospective studies we can estimate

$$\alpha(x) = \frac{P(Y_t = 1 | X = x)}{P(Y_t = 0 | X = x)} \Big/ \frac{P(Y_c = 1 | X = x)}{P(Y_c = 0 | X = x)}$$

However in prospective studies we may estimate $\alpha(x)$ but also

$$E(Y_t - Y_c),$$

or

$$E(Y_t - Y_c | X = x),$$

and

$$\alpha = \frac{P(Y_t = 1)}{P(Y_t = 0)} \Big/ \frac{P(Y_c = 1)}{P(Y_c = 0)}$$

Thus in comparing the results of prospective and retrospective studies of the same causal agents it is important to be sure that estimates of comparable parameters are being considered.

4.3 THE ROLE OF MATCHING IN PROSPECTIVE AND RETROSPECTIVE STUDIES

We close with a comment on an alternative way to describe the difference between prospective and retrospective studies from the point of view of matching.

In prospective matching, units exposed to t and to c are matched on X, whereas in retrospective matching a unit that is a case is matched on X with a unit that is a control. Suppose for simplicity that S is independent of Y_k given X so that at each level of X we have a randomized experiment, that is, the experiment is a randomized block design with the blocks defined by X. Prospective matching reconstructs the randomized block experiment by creating matched pairs of units exposed to t and to c. The average matched-pair difference is an unbiased estimate of the treatment effect for the population defined by the values of X in the matched pairs. Thus prospective matching on X perfectly controls for X in this population whenever both members of each matched pair have the same values of X.

In contrast, retrospective matching on X in general cannot perfectly control for X because it does not reconstruct the randomized block experiment. In each matched pair, one member is a case and one member is a control; to reconstruct the randomized block experiment, one member must be exposed and one unexposed, which generally does not occur when one member is a case and the other a control. Thus summaries from the case-control matched sample such as the odds ratio do not represent an estimate of a causal effect for which X has been controlled, even when all matched pairs are exactly matched with respect to X. With retrospective matches, we really need to estimate the odds ratio in each matched pair, and this requires building a model relating Y_1, Y_2 to X and S such as we have illustrated in Section 3.

REFERENCES

BASU, D. (1980) "Randomization analysis of experimental data: the Fisher Randomization Test." J. of the Amer. Stat. Assn. 75: 575-582.

BISHOP, Y.M.M., S. E. FIENBERG, and P. W. HOLLAND (1975) Discrete Multivariate Analysis: Theory and Practice. Cambridge: MIT Press.

BRESLOW, N. E. and N. E. DAY (1980) Statistical Methods in Cancer Research, Vol. 1: The Analysis of Case-Control Studies. Lyon, France: International Agency for Research on Cancer.

COCHRAN, W. G. (1965) "Planning of observational studies of human populations." J. of the Royal Stat. Society series A, 128, part 2: 234-255.

CORNFIELD, J. (1951) "A method of estimating comparative rates from clinical data, application to cancer of the lung, breast and cervix." J. of the National Cancer Institute 11: 1269-1275.

CORNFIELD, J. (1956) "A statistical problem arising from retrospective studies." Proceedings of the Third Berkeley Symposium 4: 135-148.

COX, D. R. (1958) Planning of Experiments. New York: John Wiley.

FISHER, R. A. (1935) The Design of Experiments. Edinburgh: Oliver & Boyd.

GRANGER, C.W.J. (1969) "Investigating causal relations by econometric models and cross-spectral methods." Econometrica 37: 424-438.

HOLLAND, P. W. (1986a) "Which comes first, cause or effect?" New York Statistician 38: 1-6.

HOLLAND, P. W. (1986b) "Statistics and causal inference." J. of the Amer. Stat. Assn. 81: 945-970.

HOLLAND, P. W. and D. B. RUBIN (1983) "On Lord's Paradox," in Wainer and Messick (eds.) Principles of Modern Psychological Measurement. Hillsdale, NJ: Lawrence Erlbaum.

JICK, H. et al. (1973) "Coffee and mycardial infarction." New England J. of Medicine 289: 63-67.

KEMPTHORNE, O. (1952) The Design and Analysis of Experiments. New York: John Wiley.

KEMPTHORNE, O. (1976) Discussion of "On rereading R. A. Fisher" by Leonard J. Savage. Annals of Statistics 4: 495-497.

KRUSKAL, W. (1980) "The significance of Fisher." J. of the Amer. Stat. Assn. 75: 1019-1030.

MIETTINEN, O. S. (1976) "Stratification by a multivariate confounder score." Amer. J. of Epidemiology 104: 609-620.

ROSENBAUM, P. R. (1984a) "From association to causation in observational studies: the role of tests of strongly ignorable treatment assignment." J. of the Amer. Stat. Assn. 79: 41-48.

ROSEMBAUM, P. R. (1984b) "The consequences of adjustment for a concomitant variable that has been affected by the treatment." J. of the Royal Stat. Society series A, 147: 656-666.

ROSENBAUM, P. R. (1984c) "Conditional permutation tests and the propensity score in observational studies." J. of the Amer. Stat. Assn. 79: 565-574.

ROSENBAUM, P. R. and D. B. RUBIN (1983a) "The central role of the propensity score in observational studies for causal effects." Biometrika 70: 41-55.

ROSEMBAUM, P. R. and D. B. RUBIN (1983b) "Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome." J. of the Royal Stat. Society series B, 45: 212-218.

ROSENBAUM, P. R. and D. B. RUBIN (1984a) Discussion of Pratt and Schlaifer "On the nature and discovery of structure." J. of the Amer. Stat. Assn. 79: 26-28.

ROSENBAUM, P. R. and D. B. RUBIN (1984b) "Reducing bias in observational studies using subclassification on the propensity score." J. of the Amer. Stat. Assn. 79: 516-524.

ROSENBAUM, P. R. and D. B. RUBIN (1985a) "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score." Amer. Statistician 39: 33-38.

ROSENBAUM, P. R. and D. B. RUBIN (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." J. of Educ. Psychology 66: 688-701.

RUBIN, D. B. (1977) "Assignment of treatment group on the basis of a covariate." J. of Educ. Statistics 2: 1-26.
RUBIN, D. B. (1978) "Bayesian inference for causal effects: the role of randomization." Annals of Statistics 6: 34-58.
RUBIN, D. B. (1980) Discussion of "Randomization analysis of experimental data: the Fisher Randomization Test," by D. Basu. J. of the Amer. Stat. Assn. 75: 591-593.
RUBIN, D. B. (1986) "Which ifs have causal answers." J. of the Amer. Stat. Assn. 81, 961-962.

Paul W. Holland is Distinguished Research Scholar and Chairman of the Research Statistics Group at Educational Testing Service, Princeton, New Jersey 08541. He has contributed widely to the applications of statistics to the social sciences, including categorical data analyses, social networks, robust regression, test equating, item bias, item response theory, and causal inference in nonexperimental research. He is a fellow of the American Statistical Association, the Institute of Mathematical Statistics, and the American Association for the Advancement of Science.

Donald B. Rubin is Professor and Chairman of the Department of Statistics, Harvard University, Cambridge, MA 02138. He has contributed to the theory and application of statistics to many areas, including causal inference in randomized and nonrandomized studies, methods such as matched sampling for controlling bias in observational studies, techniques such as the EM algorithm and multiple imputation for dealing with missing data, and the application of Bayesian and empirical Bayesian techniques to social science data. He is a fellow of the American Statistical Association, the Institute of Mathematical Statistics, the International Statistical Institute, the American Association for the Advancement of science, the John Simon Guggenheim Foundation, and the Woodrow Wilson Society.