9 Which Comes First, Cause or Effect?

Paul W. Holland  
*Educational Testing Service, Princeton, NJ*

THE CAUSE OF AN EFFECT  
VERSUS THE EFFECT OF A CAUSE

Even though it might not seem significant to separate the question of "what is the cause of a given effect?" from that of "what is the effect of a given cause?", I believe that this simple verbal distinction reflects the wide gulf between most philosophical discussions of causation and the practice of experimental science. Aristotle began it all by identifying various notions of the meaning of a thing's cause: its *material* cause, its *efficient* cause, its *formal* cause, and its *final* cause. Hume's three criteria—constant conjunction, temporal succession, and spatial contiguity—all refer to what he believed we must observe before we conclude that A is the cause of B. Other examples are easily cited. The "principle of causality" asserts that every phenomenon has a cause.

Yet what can experimental science actually do? It does nothing more (nor less) than establish that a particular effect is the consequence of some specified cause. It does not start with effects and figure our causes, rather it starts with causes and measures their effects. John Stuart Mill wrote this of the role of 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)
I am sure that statistics have been used and misused to prove causation ever since they were first gathered. In 1861, Farr cautioned Florence Nightingale that "I must repeat my objections to intermingling Causation with Statistics" (Porter, 1986, p. 36). This admonition fell on mostly deaf ears then as it probably would today. Even Bayes' theorem is sometimes couched in causal language. Probabilities of effects given causes are turned into probabilities of causes given observed effects. What a theorem!

However, the most substantial contributions of the field of statistics to causal inference are unquestionably those of the design of comparative experiments. Randomized experiments have transformed many areas of investigation and are the original products of a few statistical giants. In this view of the contributions of statistics to causal inference it is a simple step to propose that the fundamental ideas of causal inference ought to build on the known success of experiments, that is, on the measurement of effects, rather than on the deduction of causes. Rubin's model for causal inference is an attempt to do just this (Holland, 1986, 1988a, 1988b; Holland & Rubin, 1983, 1988; Rosenbaum, 1984a, 1984b, 1984c; Rubin, 1974, 1977, 1978, 1980, 1986). My short answer to the question: Measuring effects is logically prior to discovering causes, so "effect" comes first!

**RUBIN'S MODEL**

I think it is useful to start with an abstract statement of Rubin's model and then to interpret it in terms of concrete types of studies. The logical elements of this model, in its simplest version, form a quadruple \((U, K, s, Y)\) in which \(U\) and \(K\) are sets, \(s\) is a mapping of \(U\) to \(K\) (its value denoted \(s_i\)), and \(Y\) is a real-valued function defined on the Cartesian product \(K \times U\) (its value denoted \(Y_{ki}\)). The elements of the quadruple are interpreted as follows.

\[
\begin{align*}
U & = \text{a population of units, denoted as } i \in U, \\
K & = \text{a set of causes or treatments to which the elements of } U \text{ may all be exposed, denoted as } k \in K, \\
s_i & = k, \text{ if } k \text{ is the cause to which unit } i \in U \text{ is actually exposed,} \\
Y_{ki} & = \text{the value of the response that would be observed for unit } i \text{ if } i \in U \text{ were exposed to cause } k \in K.
\end{align*}
\]

The elements of the quadruple are the primitives of this model and serve as the undefined terms. All other concepts are defined in terms of these primitives. The most basic quantity that needs definition is the observed response on each unit in \(U\). This is given by

\[
y_i = Y_{ij}.
\]
In Equation 1 $y_i$ is the value of $Y$ that is actually observed for unit $i$. Note that $y$ is a real-valued function defined on $U$ such that $y_t = Y_{it}$ if $s_t = t$ and $y_c = Y_{ct}$ if $s_t = c$. In general, $t$ and $c$ refer to two distinct elements of $K$, but in our usual experimental language they stand for the "treatment" and the "control" causes or experimental conditions.

*Causes* are taken as undefined primitives in this model, thus *effects* are defined in terms of them. The *causal effect* of cause $t$ relative to cause $c$ on unit $i$ is the value

$$T_i = Y_{it} - Y_{ct}. \tag{2}$$

Thus, $T_i$ is the increase in the value of $Y$ that would be observed for $i$ if $i$ were exposed to $t$ over the value that would be observed for $i$ if $i$ were exposed to $c$. In this notation I have adopted the convention that variables not necessarily directly observable are denoted by capital letters whereas observations are denoted by lowercase letters. I have used subscripts to denote the arguments of functions in order to depart as little as possible from standard statistical notation, that is, $y_i$ and $s_t$, and so forth.

Probably the aspect of Rubin's model that people have the most trouble with is the explicit notation for the potential response $Y_{ki}$ for each cause-unit combination rather than simply the observed response $y_i$ for each unit. However, this use of "potential responses" goes back at least as far as Neyman (1935) and is a standard tool for the analysis of randomization distributions in experimental design. One of Rubin's important contributions was to see that this idea is important for all problems of causal inference.

There is a *fundamental problem with causal inference* because it is impossible in principle to simultaneously observe both of the values, $Y_{it}$ and $Y_{ct}$, on a single unit, $i$. Thus, the causal effect $T_i$ is in principle not directly observable. Many techniques of experimental science are aimed at overcoming this fundamental problem. In Holland (1986) I discuss some of these techniques more extensively than I can here, however, I briefly mention two familiar ones—unit homogeneity and randomization—that, in my opinion, are the twin pillars that support causal inference in controlled experimentation.

**Unit Homogeneity**

In a scientific laboratory care is exercised to prepare homogeneous samples of material (i.e., units) for study. This allows the experimenter to assume (at some level of accuracy) that $Y_{ki} = Y_{kj}$ for all experimental units $i$ and $j$ and all relevant causes or treatments, $k$ in $K$. This implies that the value of $T_i$ defined in Equation 2 equals the difference $Y_{ii} - Y_{cj}$ for a pair of distinct units $i$ and $j$. We can observe $Y_{ii}$ for $i$ and $Y_{cj}$ for a second unit, $j$, so the causal effect becomes directly measurable—all because of the assumed homogeneity of the experimental units.
However, the assumption of unit homogeneity is an assumption that cannot be verified directly. Of course, experience may lead a scientist to believe that it is valid for some purposes but this never proves the validity of an assumption, only its utility.

**Randomization**

When we examine the difference in average \( Y \)-values between groups of units, one group exposed to \( t \) and the other to \( c \), we are estimating the quantity

\[
\text{FACE} = E(\mathbf{Y}|s = t) - E(\mathbf{Y}|s = c).
\]

(3)

The FACE is the *prima facia average causal effect* and can always be estimated by the differences between the treatment and control group average \( y \)-values. The FACE must be distinguished from the *average causal effect (ACE)* defined by

\[
\text{ACE} = E(Y_t - Y_c|s = t) = E(T|s = t).
\]

(4)

The ACE is the average of the causal effects \( T \) over those units in the population \( U \) who were exposed to \( t \). The ACE, being an average of causal effects, is a causal parameter, whereas the FACE describes the association between the observed variables \( y_i \) and \( s_i \). The ACE describes causation whereas the FACE describes “mere association.”

We may express the FACE in terms that more closely relate it to the ACE as follows.

\[
\text{FACE} = E(\mathbf{Y}|s = t) - E(\mathbf{Y}|s = c) = E(Y_t|s = t) - E(Y_c|s = c)
\]

\[
= E(Y_t|s = t) - E(Y_c|s = c)
\]

\[
= E(Y_t|s = t) - E(Y_c|s = t) + E(Y_t|s = t) - E(Y_c|s = c)
\]

so

\[
\text{FACE} = \text{ACE} + \text{BIAS},
\]

(5)

where

\[
\text{BIAS} = E(Y_c|s = t) - E(Y_c|s = c).
\]

(6)

The term BIAS involves the “counterfactual expectation” \( E(Y_c|s = t) \), which is the average value of \( Y \) that would have been observed in the treatment group had those units all been exposed to the control treatment, instead. It is *counterfactual* because the event in the conditioning, \( s = t \), precludes observing the variable being averaged, \( Y_c \)—see Glymour (1986) for more discussion of Rubin’s model and counterfactuals.

When a large number of units are randomly assigned to treatments, \( s \), becomes a random variable that is statistically independent of each \( Y_k \) considered as functions on \( U \), for each \( k \) in \( K \). In this case the counterfactual expectation \( E(Y_c|s = t) \) equals the value \( E(Y_c|s = c) \) and, in Equation 6, BIAS = 0 so that the
ACE and the FACE are equal. The equality of the FACE and the ACE under randomization is one reason why randomization is such a powerful tool for causal inference. Even when units are not homogeneous, randomization creates equality between a causal parameter (ACE) and an associational parameter (FACE). Note, however, that knowledge of the causal effect $T_i$ for unit $i$ is sacrificed for knowledge of the average of $T_i$ over all $i \in U$ exposed to $t$.

BEYOND EXPERIMENTS

The desire to make causal inferences in controlled and/or randomized experiments is only the first step. Many branches of science cannot do such studies and yet the measurement of causal effects is still a goal of researchers in these fields. One purpose of Rubin’s model is to show how the experimental paradigm can shed light on causal inference in such fields. This model carries the general message that the counterfactual expectation, $E(Y_{i|S = s})$, is the key quantity about which we must make critical assumptions that, in the absence of randomization, are often untested or even untestable. This section briefly covers three topics that arise once we leave the relative intellectual safety of controlled experiments—prospective observational studies, retrospective case-control studies, and posttreatment concomitant variables.

Prospective Observational Studies

This category includes many types of studies in which the active experimenter, who controls the assignment of causes to units, is replaced by a passive observer who can only record which unit turned out to be exposed to each relevant cause or treatment.

Rubin’s model can express this by the lack of the plausibility of the assumption that $s_i$ is independent of the variables $Y_{ik}$ for $k \in K$ and hence, in Equation 6, BIAS $\neq 0$. A technique that may sometimes replace the independence assumption, and that has been studied the most so far, concerns “generalized covariate adjustments.” When randomization is absent, the FACE and the ACE are no longer equal, so that the mean difference in responses between the treatment and control groups is no longer an unbiased estimate of the ACE. Suppose that, in addition to the variables $y_i$ and $s_i$, we also observe the value of a “covariate” $x_i$ on each unit $i$ in $U$. In order for the notation $x_i$ to be appropriate, the value $x_i$ must not be influenced by the particular cause or treatment to which $i$ is exposed. This assumption might be plausible in some circumstances (i.e., a drug treatment will not change your sex or age). In other cases we can only insure that it is true by measuring $x_i$ prior to the exposure of units to causes. When $x_i$ does not depend on the value of $s_i$, $x_i$ is a covariate. The assumption that $s_i$ is independent of all relevant variables (i.e., randomization) is stronger than the weaker assumption
that $Y_i$ and $s_i$ are *conditionally independent* given the value of $x_i$. This weaker, conditional independence assumption is often made in observational studies and leads to the estimation of the *covariate-adjusted, prima facia average causal effect* defined as

$$C\text{-FACE} = E(E(y|s = t, x) - E(y|s = c, x)|s = t). \quad (7)$$

We have

$$E(y|s = t, x) = E(Y_i|s = t, x), \quad (8)$$

and under the conditional independence assumption

$$E(y|s = c, x) = E(Y_i|s = c, x) = E(Y_i|s = t, x). \quad (9)$$

From these equalities it follows that $C\text{-FACE} = ACE$. Thus, under the conditional independence assumption estimates of the $C\text{-FACE}$ are estimates of the $ACE$. When $E(y|s, x)$ is estimated using a linear regression model, the coefficient of $s$ has a causal interpretation, that is, as an estimate of an $ACE$, whereas the coefficients of the other independent variables in $x$ do not necessarily have a causal interpretation.

**Retrospective Case-Control Studies**

Medical studies have made increasing use of the case-control design. In such studies, the elements of Rubin's model have the following interpretation. $y_i = 1$ or 0 as patient $i$ is a case (has the disease under study) or a control (does not have the disease). $U$ is the population of patients under study and $K$ is the level of exposure of the causative agent being investigated (e.g., amount of smoking, use of oral contraceptives, etc.). In case-control studies the data often consist of $y_i$, $s_i$, and a vector of covariates, $x_i$. Care must be taken to insure that the values of the covariates are not affected by exposure to the levels of the causative agent.

Case-control data are gathered in a retrospective way, so all that can be guaranteed in such studies is the observation of data from the joint distribution of $s_i$ and $x_i$ given either $y_i = 1$ or $y_i = 0$. Thus, we can observe data on the conditional distribution $p(k, x|y) = P(x = k, x|y)$. When $K$ has two elements, $t$ and $c$, the population probabilities underlying these data may be arrayed into a series of $2 \times 2$ tables of the form

$$
\begin{array}{c|cc}
  & y = 1 & y = 0 \\
  s = t & p(t, x|1) & p(t, x|0) \\
  s = c & p(c, x|1) & p(c, x|0)
\end{array}
$$

(10)

one for each value of $x$. The crossproduct ratio of the table in Equation 10 is
9. WHICH COMES FIRST, CAUSE OR EFFECT?

\[ \alpha(x) = \frac{p(t, x|1) p(c, x|0)}{p(c, x|1) p(t, x|0)}. \]  

(11)

By Bayes' theorem, the $2 \times 2$ table in Equation 10 has the same crossproduct ratio as this one

<table>
<thead>
<tr>
<th>$s$</th>
<th>$y = 1$</th>
<th>$y = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t$</td>
<td>$\pi(1</td>
<td>t, x)$</td>
</tr>
<tr>
<td>$c$</td>
<td>$\pi(1</td>
<td>c, x)$</td>
</tr>
</tbody>
</table>

(12)

where $\pi(y|k, x) = P(Y_s = y|s = k, x)$.

The primary tools used in the analysis of retrospective studies are various types of covariate adjustments (Breslow & Day, 1980) that all make the conditional independence assumption discussed earlier. That assumption implies the equality

\[ \pi^*(y|x) = \pi(y|k, x) \]

where $\pi^*(y|x) = P(Y_s = y|s = t, x)$. The corresponding $2 \times 2$ table is obtained by replacing $\pi$ by $\pi^*$ in Equation 12. This substitution results in the crossproduct ratio

\[ \alpha^*(x) = \frac{\pi^*(1|x) \pi^*(0|x)}{\pi^*(0|x) \pi^*(1|x)}. \]  

(13)

Hence, standard log-linear model techniques (Bishop, Fienberg, & Holland, 1975) may be used to study $\alpha^*(x)$ by analysis of the data in Equation 10 in case-control studies when the conditional independence assumption is plausible.

Note that $\alpha^*(x)$ compares the distribution (conditional on $x$) of $Y_t$ to that of $Y_c$ among those exposed to $t$; but neither $\alpha^*(x)$ nor its average over $x$ is an ACE describing the average of $T$, over some part of $U$. However, $\alpha^*(x)$ is still a useful "causal" parameter because it describes how the distribution of $Y$-values changes in response to exposure of units to $t$ or $c$. The ACE is a stronger causal parameter because it describes how the average unit is changed by exposure to $t$ or $c$. In this sense, causal inferences in case-control studies are inherently weaker than those available in prospective observational studies satisfying exactly the same conditional independence assumption. Case-control studies are discussed more extensively from this point of view in Holland and Rubin (1988).

When Is a Covariate Not a Covariate?

Rubin's model can help one make important distinctions that are sometimes unknowingly swept under the rug. A good example is the use of posttreatment concomitant variables. When a second variable, $X$ (besides $Y$), enters the system,
Rubin's model must be expanded to \((U, K, s, y, x)\) where the new element, \(x\), is also a function (possibly vector-valued) on \(K \times U\). \(X_{ik}\) is defined on \(K \times U\) to reflect the fact that the value of \(X_{ik}\) can depend both on \(k\) and \(i\). However, as mentioned earlier, a covariate is a function of \(i\) alone, that is, \(X_{ik} = x_{i}\). Rosenbaum (1984b) discussed the role of posttreatment concomitant variables in experimental and observational studies. A posttreatment concomitant is a variable whose value is measured after the unit \(i\) has been exposed to the cause \(s_i\), which is not a covariate in the sense used here if it is possible that \(X_{ik}\) is affected by \(k\) as well as by \(i\). Rosenbaum showed that adjustments for posttreatment concomitant are usually more of a problem than a solution when one is estimating an average causal effect and should not be done without a great deal of care. When \(X_{ik}\) depends strongly on \(k\), the bias introduced by adjusting for \(X\) may overwhelm any other advantage such an adjustment might have had.

WHAT ABOUT PATH ANALYSIS?

Much of what Rubin's model tells us is simply common sense to statisticians steeped in a tradition of experimentation. Although its goal is to help us understand how to analyze nonexperimental data, there is no magic. We are left with the problem of contemplating a series of untestable but plausible assumptions about our data and checking the sensitivity of our conclusions to these assumptions. That is the message of the model, pure and simple. But the model goes beyond this platitude because it focuses our attention on crucial quantities such as the counterfactual expectation, \(E(Y_i|S = i)\), or on basic assumptions such as the stability assumption or the stable unit-treatment value assumption SUTVA (Rubin, 1980, 1986). How much less attractive is such an approach than the promise of path analysis! In path analysis, the cold bones of correlation are turned into the warm flesh of causation with direct, indirect, total, and partial causal pathways. Diagrams emerge; papers can now be written about unseen causal mechanisms winding their ways through latent variables and finally emerging as the potent explanation of observed correlations. However, it is all the same data so it may be hard to understand how there is such a gulf between the dreary analyses of Rubin's model and the artistic appeal of path diagrams. It is, I fear, another manifestation of that unbridgeable gulf between the measurement of the effects of causes and the identification of the causes of given effects. What can pass for a cause in path analysis might never get a moment's notice in an experiment. Students' scores on a test can cause their future action (Saris & Stronkhorst, 1984). Gender or race are just as good as anything else as causes in path analysis, although no experimenter would know how to manipulate them. Holland (1988a) gave a detailed description of the indirect causes in path analysis using an expansion of Rubin's model for research designs in which we cannot directly manipulate the causes of interest. Holland (1988b) also used Rubin's
model to address the causal status of gender in the application of statistics to employment discrimination.

Statistical science has made strong contributions to issues of causal inference when it has addressed the problem of measuring the effects of causes. It does less useful things when its methodology claims to identify the causes of effects. Rubin’s model focuses our attention on what we can do well rather than on what we might like to do, however poorly.

ACKNOWLEDGMENT

This chapter is based on an earlier version published in 1986 in *The New York Statistician*, 38, 1–6, a publication of the New York Area Chapter of the American Statistical Association.

REFERENCES


Rosenbaum, P. R. (1984b). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society (A)*, 147, 656–666.


