Causal Inference, Path Analysis, and Recursive Structural Equations Models
Author(s): Paul W. Holland
Reviewed work(s):
Source: *Sociological Methodology*, Vol. 18 (1988), pp. 449-484
Published by: American Sociological Association
Stable URL: http://www.jstor.org/stable/271055
Accessed: 13/02/2013 18:37

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

American Sociological Association is collaborating with JSTOR to digitize, preserve and extend access to *Sociological Methodology*.

http://www.jstor.org

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

13

Causal Inference, Path Analysis, and Recursive Structural Equations Models

*Paul W. Holland**

Rubin's model for causal inference in experiments and observational studies is enlarged to analyze the problem of "causes causing causes" and is compared to path analysis and recursive structural equations models. A special quasi-experimental design, the encouragement design, is used to give concreteness to the discussion by focusing on the simplest problem that involves both direct and indirect causation. It is shown that Rubin's model extends easily to this situation and specifies conditions under which the parameters of path analysis and recursive structural equations models have causal interpretations.

1. INTRODUCTION

The perspective on causal inference developed extensively by Rubin (1974, 1977, 1978, 1980, 1986) provides a solid basis for considering issues of causal inference in complex cases, and it is the only one that is grounded where causal inferences are relatively uncontroversial—in experimental science. In Holland (1986*a*, *b*), I described this perspective and dubbed it *Rubin's model*, as I will refer to it here, too, even though a more general term, such as *the experimental model*,

I thank David Freedman, Clark Glymour, Edward Leamer, Margaret Marini, James Robins, David Rogosa, Donald Rubin, Burton Singer, and Howard Wainer for their many helpful comments on an earlier draft of this paper.

*Educational Testing Service.

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

may be more appropriate. Robins (1984, 1985, 1986) gives a closely related model in the context of epidemiological studies. One goal of this paper is to extend Rubin's model to accommodate a class of quasi-experimental procedures that are called encouragement designs by Powers and Swinton (1984). These designs involve both randomization and self-selection as well as both direct and indirect causation. Encouragement designs provide a simple yet useful "laboratory" in which the issues of direct and indirect causal relationships can be carefully examined. I hope to clarify the relationship between the systematic structure of Rubin's model and the less formal approaches of path analysis and structural equations modeling. My own experience has been that *any* discussion of causation is enriched by an analysis using Rubin's model (e.g., Holland and Rubin 1987; Rosenbaum 1987; Holland 1988).

Before proceeding, I want to make a few general comments about causation to set the stage for the subsequent discussion.

In most discussions of causation, all too little attention is given to distinguishing between the *cause* of a given effect and the *effect* of a given cause. Since the time of Aristotle, philosophers have tried to define what it means for A to be a cause of B. This activity still continues (Lewis 1986; Marini and Singer, this volume). Yet, the attribution of causation has been known to be fraught with difficulty since, at least, Hume's analysis in the mid 1700s. A statement like "A is a cause of B" is usually false because it is, at best, a tentative summary or theory of our current knowledge of the cause (or causes) of B. 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. Yet this is not quite correct either: Certain chemical reactions are the real causes, and so on, *ad infinitum*. Experiments, on the other hand, do not identify causes. Rather, experiments measure the *effects* of given causes (i.e., the effects of the experimental manipulations). The results of an experiment can be summarized by a statement of the form "An effect of A is B," but not by one of the form "A is a cause of B," unless we mean by the latter no more than the former. I would be surprised if most modern scientists were willing to equate theoretical statements like "A is a cause of B" with empirical regularities like "The effect of A is B." Theories may come and go, but old, replicable experiments never die; they are just reinterpreted.

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

The strength of Rubin's model is that it builds on the success of experimentation and focuses on the measurement of the effects of causes rather than on attempting to identify the causes of effects. Statistics has made major 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 can be done well rather than on what we might like to do, however poorly.

I do not mean to imply that the search for the causes of a phenomenon is a useless endeavor. Indeed, it is a driving force that motivates much of science. Rather, I mean that a logical analysis of the search for causes follows from an analysis of the measurement of causal effects, and it is not logically prior to this more basic activity. Defensible inferences about the *causes* of an effect are always made against a background of measured causal effects and relevant theories.

I have tried to follow several goals in writing this paper. First, I discuss population quantities rather than sample estimates of population quantities. Thus, it is best to think of the populations that occur here as large or infinite. I do not apologize for this, since such a view is implicit in most discussions of path analysis. My aim is to define causal parameters rather than to discuss ways of estimating them. Second, I may, on occasion, appear overly notational, and I apologize for that. My defense is that I wish to be very clear about what I mean, and since causation is a subtle idea, an adequate notation is essential to understanding it. Unfortunately, my notation is not identical to that usually used in path analysis or structural equations models, but it is only intended to be more explicit than these other schemes. Finally, my goal is to put the, to me, complex and intuitive models used in path analysis into a framework that I find helpful in complex problems. I hope others find it useful too.

In section 2, I define and give an example of an encouragement design that is used in the rest of this paper to focus the discussion. These designs are interesting in their own right, because they attempt to measure the effects of self-selected treatments. In section 3, I review three related topics that concern path analysis: deterministic linear systems, path analysis, and recursive structural equations models. In section 4, I extend Rubin's model to the case of encouragement designs to allow for both direct and indirect causation. I conclude the paper

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

with a short discussion. I also include an appendix, in which I apply Rubin's model to experiments and observational studies to make the paper reasonably self-contained.

2. ENCOURAGEMENT DESIGNS

While it is common to discuss path analysis and causal models in terms of abstract systems of variables, I find it easier to discuss issues of causal inference in the context of specific examples or classes of examples. For this reason I will use a fairly concrete quasi-experimental design, the encouragement design, as the basis of my discussion of causal theories that involve direct and indirect causation. The encouragement design is a simple and relatively clear-cut type of study in which many of the issues of direct and indirect causation arise.

I will introduce encouragement designs by giving an example that is used throughout the paper. Suppose we are interested in the effects of various amounts of study on the performance of students on a test. I will suppose that there are two experimental treatments: one that encourages a student to study for the test (t = treatment) and one that does not (c = control). After exposure to one of these treatments, a student will then study for the test for some amount of time, R. Subsequently, the student is tested and gets a test score, Y. An example of an encouragement design, similar to the one just described, is given in Powers and Swinton (1984). My first exposure to a formal analysis of encouragement designs was in Swinton (1975).

The only experimental manipulation in an encouragement design is exposure to the "encouragement conditions," which are just t or c here but could involve more than two levels, of course. Hence, using standard methods, we can measure the effect of encouragement on the amount of study, i.e. R, and on test performance, i.e. Y. However, we may also be interested in the effect of *studying* on test performance. Thus, random assignment of encouragement conditions might be possible, but the students will then self-select their own exposure levels to the amount they study, R. This self-selection is a critical feature of encouragement designs, and it is why I refer to them as a type of quasi-experimental design, after Campbell and Stanley (1963). The other critical feature of encouragement designs is the analyst's interest in measuring the causal effect of the amount of study, R, on test performance. I have chosen this example specifically because from an

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

individual student's point of view, the amount one studies is a self-imposed treatment that can be measured and over which one can exercise control. However, from the analyst's point of view, the amount a student studies is a response to the encouragement condition, as is the student's test performance. In this very special type of situation, "amount of study" plays both the role of a response and the role of a self-imposed treatment; i.e., it is both an effect and a cause.

Encouragement designs can arise in *any* study of human subjects in which the treatments or causes of interest must be voluntarily applied by the subjects to themselves. Other potential examples are medical studies that encourage voluntary healthful activities among patients or economic studies that attempt to alter people's spending behavior by various inducements. The analysis of surgical trials may involve randomization of the "intention to treat" patients, but because of clinical intervention, the actual treatment patients get may not be the one to which they were randomly assigned. This is similar to an encouragement design, but the models discussed in this paper may not be appropriate to that case, since I treat "amount of study" as a continuous variable. The general ideas are the same, however.

I suspect that encouragement designs are quite widespread but may not always be recognized. On the other hand, the special nature of these designs cannot be overemphasized. While it is plausible that "amount of study" is both an effect and a cause, this dual role is not always a plausible assumption, and ignoring this fact can lead to some rather curious causal statements. It is critical, in the analysis developed here, that those things that play the role of causes or treatments have levels that are, in principle, alterable. The statement "I could have studied but I didn't" has this flavor, but "I might have scored higher on the test but I didn't" does not. See Holland (1986*b*, sect. 7; 1988) for more emphasis on this very important point.

The basic elements of an encouragement design are, thus, (a) an experimental manipulation of "degrees" of encouragement (here, just t and c) to perform some activity, (b) measurement of the subsequent amount of the encouraged activity, (c) measurement of a final outcome or response variable, and (d) an interest in measuring the causal effect of the encouraged activity on the response variable. Encouragement designs are more often applied to human populations than to other types of experimental units because of the self-selected or voluntary nature of much of human activity.

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

3. DETERMINISTIC LINEAR SYSTEMS, PATH ANALYSIS, AND RECURSIVE STRUCTURAL EQUATIONS MODELS

In this section I review three related topics: deterministic linear systems, path analysis, and structural equations models. All three will arise in my discussion of encouragement designs in section 4. I frame this review in terms of the structure of encouragement designs.

Extended discussions of path analysis and structural equations models are numerous (e.g., Blalock 1964, 1971; Duncan 1966, 1975; Freedman 1987; Goldberger 1964; Goldberger and Duncan 1973; Heise 1975; Kenny 1979; Saris and Stronkhorst 1984; Tukey 1954; and Wright 1934). I follow Tukey (1954) in *not* standardizing the variables to have zero mean and unit standard deviation and in emphasizing regression coefficients rather than standardized regression coefficients.

3.1. *Deterministic Linear Systems*

Suppose there are two linear functions, f and g, of two variables, s and r, of the form

$$f(s) = as + d, \tag{1}$$

and

$$g(s, r) = bs + cr + d', \tag{2}$$

where a, b, and c are the important slope parameters, and d and d' are constants that play no essential role in this theory. We introduce a third variable, y, into this system via the definition

$$y = g(s, r), \tag{3}$$

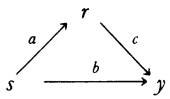
and we assume that r and s are related by the functional relationship

$$r = f(s). \tag{4}$$

Such a system captures the idea that r is functionally dependent on s and y is functionally dependent on s and r. Since f and g are linear, *changes* in y and r are determined by the slope parameters, a, b, and c. This system may be represented by the "path" diagram in Figure 1. The coefficients a, b, and c are the "path" coefficients, or the "direct effects"; i.e., a is the direct effect of s on r, c is the direct effect of r on y, and b is the direct effect of s on y.

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

FIGURE 1.



The "total effect" of s on y is found by substituting the equation for r into that for y. This yields

$$y = g(s, f(s)) = bs + c(as + d) + d',$$
$$= (b + ca)s + (cd + d'),$$

so that

$$y = (b + ac)s + d''. \tag{5}$$

Hence, the total effect of s on y is $b + ac$, which may also be calculated as the sum of the products of all the direct effects along all the paths connecting s and y in the path diagram in Figure 1; i.e., s to y yields b, and s to r to y yields ac, so the sum is $b + ac$.

I use the phrase *deterministic linear system* to refer to a path diagram that arises from a set of nonstochastic linear equations like the ones just described.

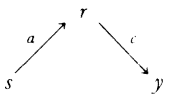
Viewed simply as a visual representation of a deterministic linear system, path diagrams are easy to understand and can help keep track of the bookkeeping that is associated with the total effects of one variable on another. The real appeal of path diagrams arises in systems that involve more than three variables, but the basic ideas are already present in systems with three.

The path diagram in Figure 1 is not the only one we could draw using three variables, but it is relevant to encouragement design in the following way. If $s = 1$ or 0 as there is encouragement or not and if r denotes the resulting amount of study and y denotes the subsequent test score, then the parameters a, b, and c have the following interpretation. The change in amount of study due to encouragement to study is a, and $b + ac$ is the change in test scores due to encouragement to study. The change in test scores due to a unit change in the amount of study within each level of the encouragement condition, s, is c.

When one of the coefficients a, b, or c is zero, it is customary to delete the corresponding arrow from the path diagram. For example, if $b = 0$, we have the diagram in Figure 2.

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

FIGURE 2.



In the encouragement-design example, the path diagram in Figure 2 would be interpreted as displaying no effect of encouragement on test scores except through its effect on studying. We will return to this idea later in section 4.

Deterministic linear systems not only motivate the nondeterministic linear models of path analysis and structural equations models but also play a role in what I call the ALICE "causal model" in section 4.2.

3.2. *Path Analysis*

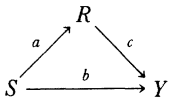
Deterministic linear systems do not really describe data, except in certain special circumstances, usually in the physical sciences. Suppose instead that there is a population U of "units" and that for each unit u in U we can obtain measurements on three numerical variables, $S(u)$, $R(u)$, and $Y(u)$. In our application, the units are students; $S(u) = 1$ if u is encouraged to study, and $S(u) = 0$ if otherwise; $R(u)$ is the amount that u studies; and $Y(u)$ is u's test score.

As u varies over U, $(S(u), R(u), Y(u))$ forms a trivariate distribution. This distribution can be used to define quantities such as the conditional expectation of R given S, $E(R|S = s)$. This conditional expectation is the average value of R for those units in U for which $S(u) = s$. The conditional expectation, $E(Y|R = r, S = s)$, has a similar definition in terms of averages over U. The expected value $E(Y|R = r, S = s)$ is the "true" regression function of Y on R and S in the sense that it is what one is trying to estimate by a least squares regression fit of Y regressed on R and S. However, in general, $E(Y|R = r, S = s)$ need not be linear in r and s.

In the example of an encouragement design, there is a natural "causal order" to the variables S, R, and Y: S comes first, then R, and then Y. A path analysis uses a causal ordering to focus on certain regression functions; in the encouragement design, they are the two described above: $E(R|S = s)$ and $E(Y|S = s, R = r)$. Suppose, for sim-

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

FIGURE 3.



plicity, that they are both linear, i.e., that

$$E(R|S=s)=f(s)=as+d \tag{6}$$

and

$$E(Y|S=s,R=r)=g(s,r)=bs+cr+d'. \tag{7}$$

This defines a deterministic linear system, as described above, when we identify y with $g(s,r)$ and equate r and $f(s)$. We may associate the path diagram in Figure 1 with this system, but because we are dealing with the measurements S, R, and Y rather than the abstract variable s, r, and y, we relabel the nodes of the graph with S, R, and Y, as in Figure 3. The path coefficients in Figure 3 are just the (population) linear regression coefficients that may be estimated by a (linear) regression of R on S, and of Y on S and R. The same terminology is used as before for the direct effects: The regression coefficients are the direct effects. The "total effect" of S on Y, i.e. $b+ac$, can be interpreted as the coefficient of S in the regression of Y on S alone:

$$\begin{aligned} E(Y|S) &= E(E(Y|S,R)|S) \\ &= E(bS+cR+d'|S) \\ &= bS+cE(R|S)+d' \\ &= bS+c(aS+d)+d' \\ &= (b+ac)S+dc+d'. \end{aligned} \tag{8}$$

I will use the phrase *empirical path diagram* to refer to any path diagram constructed from a causal ordering and the implied set of linear regression functions. An empirical path diagram is, therefore, simply the result of computing certain regression coefficients and arranging them in the appropriate places in the diagram.

In path analysis, the causal order is given, often rather vaguely, by some sort of theory. One nice feature of encouragement designs is that the causal order of "S before R before Y" is consistent with the way the data might be collected and with our intuition about study

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

and test performance. Once given, the causal order tells us which (linear) regression functions to estimate from the data. In the estimated regression functions, the coefficient of each independent variable is interpreted as the "effect" of that independent variable on the dependent variable. Thus, in (7), c is the "effect" of studying on test performance. This usage is typical of the casual causal talk (lamented by Rogosa [1987]) that often accompanies regression analyses. In section 4, I will show how causal effects can be precisely defined within Rubin's model and how specific assumptions are needed to conclude that regression coefficients are equal to causal effects.

A causal ordering is not sufficient to justify interpreting a regression coefficient as a causal effect. A causal ordering only tells the analyst which regression functions to estimate.

3.3. *Structural Equations Models*

In some areas of applied statistics—notably econometrics, but also parts of sociology, psychometrics, and even political science—it has become standard practice to use a framework that is, in a sense, more general than the conditional expectations and regression functions of path analysis, just described. These are called both structural equations models and simultaneous equations models. Instead of formulating a causal ordering or causal model for the encouragement design in terms of the regression functions, $E(R|S = s)$ and $E(Y|S = s, R = r)$, a structural equations model for such a design would be expressed as the following system of two equations:

$$R(u) = d + aS(u) + \varepsilon_1(u), \tag{9}$$

and

$$Y(u) = d' + bS(u) + cR(u) + \varepsilon_2(u). \tag{10}$$

In (9) and (10), S, R, and Y are as before, but $\varepsilon_1(u)$ and $\varepsilon_2(u)$ are new variables defined for all u's in U, so that equations (9) and (10) hold exactly for each u. The ε_1 and ε_2 are called error or disturbance terms; they take up the slack in the empirical relationship between R and S and between Y and R and S. The system (9) and (10) is "recursive," in the language of structural equations, because $Y(u)$ does not occur on the right-hand side of equation (9) (Goldberger 1964). The disturbance terms differ from the variables Y, R, and S in that they are unobserva-

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

ble. The three-variable system $(S(u), R(u), Y(u))$ is thus enlarged to a five-variable system $(S(u), R(u), Y(u), \varepsilon_1(u), \varepsilon_2(u))$, which defines a five-dimensional, multivariate distribution as u varies over U. This is what is meant by saying that S, R, Y, ε_1, and ε_2 are random variables. The causal interpretation of structural equations models, such as (9) and (10), is based on the following extension of the notion of "effect" in regression discussed earlier. For example, in equation (9), a is the "effect" of S on R, and ε_1 is that part of R that is determined by all other relevant causes that are not measured (see Goldberger [1964] for an explicit statement along these lines). Thus, the equation $R = d + aS + \varepsilon_1$ is a tidy totaling of the effects of all causes, both measured (i.e. S) and unmeasured (i.e. ε_1), on R. The point of view that underlies such an interpretation of equation (9) is that the value of $R(u)$ is "caused," in some sense, by numerous factors including $S(u)$. This sense of causation is quite unclear because it makes vague references to the "causes" of the value of a variable, i.e. of $R(u)$, rather than to measuring the effect of the experimental manipulation described by $S(u)$. In section 4.4, I show how Rubin's model can be used to give causal interpretation of the parameters of models like (9) and (10) in some situations.

It is easy to show that without further assumptions on the joint distribution of the disturbance terms, ε_1 and ε_2, with R and S, equations (9) and (10) cannot necessarily be interpreted as conditional expectations. This is often discussed in econometrics as the conditions under which ordinary least squares estimates give unbiased estimates of structural parameters (e.g., Goldberger 1964). For example, if we assume equation (9) and compute $E(R|S = s)$, we get

$$\begin{aligned} E(R|S = s) &= E(d + aS + \varepsilon_1|S = s) \\ &= d + as + E(\varepsilon_1|S = s). \end{aligned}$$

Thus, for $E(R|S = s) = as + d$, we need the joint distribution of ε_1 and S over U to satisfy

$$E(\varepsilon_1|S = s) = 0, \quad \text{for } s = 0, 1. \tag{11}$$

A sufficient condition for this is the independence of ε_1 and S and the usual zero-expected-value condition, $E(\varepsilon_1) = 0$, for ε_1. Similarly, for

$$E(Y|S = s, R = r) = d' + bs + cr, \tag{12}$$

we need to satisfy the following condition:

$$E(\varepsilon_1|S = s, R = r) = 0, \text{ for all } s \text{ and } r.$$

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

Structural equations models like (9) and (10) may be regarded as more general than the regression functions (6) and (7) precisely because we may impose assumptions on the distribution of the disturbance terms, ε_1 and ε_2, that *do not* necessarily result in a correspondence between the equations (9) and (10) and the regression functions (6) and (7). Unfortunately, since ε_1 and ε_2 are unobservable, it is not always evident how to verify assumptions made about them. For example, why should ε_1 be independent of S over U when by definition $\varepsilon_1 = R - aS - d$, i.e., when the very definition of ε_1 involves S? Such assumptions must be justified by considerations that go beyond the empirical data.

Structural equations models do little more to justify the causal interpretation of their coefficients than the causal orderings of path analysis. In both approaches, such causal interpretations are established by fiat rather than by deduction from more basic assumptions. Rubin's model, as I will show in the next section, allows one to formally state assumptions about unit-level causal effects that imply causal interpretations of regression coefficients and structural parameters.

4. A CAUSAL MODEL FOR ENCOURAGEMENT DESIGNS

The appendix gives an overview of Rubin's model as applied to randomized experiments and observational studies. In this section, I will extend that model to accommodate the added complexity of encouragement designs, with two levels of encouragement, t and c. I have tried to write this extension of Rubin's model so that the reader does not need to refer to the appendix to understand it, except for amplification of a few points.

4.1. *The General Model*

The key property of encouragement designs is that there is one cause—i.e. encouragement (indicated, as before, by $S(u) = 1$ or 0 as u is either exposed to t or c)—that affects another cause—i.e. amount of study (indicated by R)—and that these two causes, in turn, can affect the response of interest—i.e. test performance (indicated by Y). However, the mathematical structure of R and Y is really quite different

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

from that used in section 3, where R and Y were both simply regarded as functions of u alone, $R(u)$ and $Y(u)$.

To begin, the amount that u studies depends, potentially, on u and on the encouragement condition to which u is exposed, so that R is really a function of u and s, where $s = t$ or c, i.e. $R(u, s)$. Thus,

$$\begin{aligned} R(u, t) &= \text{amount } u \text{ studies if encouraged to study,} \\ R(u, c) &= \text{amount } u \text{ studies if not encouraged to study.} \end{aligned} \tag{13}$$

Let $K = \{t, c\}$; then K is the set of encouragement conditions and R is a real-valued function on $U \times K$.

What about Y? The test performance of u depends, potentially, on u, on whether u is encouraged to study or not (s), and on the amount of time u studies (r). Hence, Y is a function of u, s, and r ($Y(u, s, r)$). Thus,

$$\begin{aligned} Y(u, t, r) &= \text{test score for } u \text{ if } u \text{ is encouraged to study} \\ &\qquad \text{and } u \text{ studies for } r \text{ hours,} \\ Y(u, c, r) &= \text{test score for } u \text{ if } u \text{ is } \textit{not} \text{ encouraged to study} \\ &\qquad \text{and } u \text{ studies for } r \text{ hours.} \end{aligned} \tag{14}$$

The variable $S(u)$ depends only on u, as it did in section 3, since $S(u)$ indicates whether u is exposed to t or to c. I will engage in a slight abuse of notation and use $S(u) = t$ or c to index the encouragement condition to which u is exposed and $S(u) = 1$ or 0 to indicate the same thing when I need $S(u)$ to be a treatment indicator or dummy variable in a regression function, as in section 3.

In summary, the model for an encouragement design is a quintuple (U, K, S, R, Y), where U and K are sets, S maps U to K, R is a real-valued function of (u, s), and Y is a real-valued function of (u, s, r).

A subscript notation is useful, and we let

$$R_s(u) = R(u, s), \tag{15}$$

and

$$Y_{sr}(u) = Y(u, s, r). \tag{16}$$

Some people find such an explicit notation—i.e. $R(u, s)$ and $Y(u, s, r)$—loathsome, but I do not see how one can precisely define the elusive concepts that underlie causal inference without them. $R(u, s)$ and $Y(u, s, r)$ are not directly observable for all combinations

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

of u, s, and r. This is the main reason why causal inference is difficult and involves something more than merely the study of associations. In section 3, I used $R(u)$ and $Y(u)$ to denote the values of R and Y that are observed for unit u. This standard notation is actually misleading because it does not reveal the causal structure of the problem. In terms of $S(u)$, $R(u, s)$, and $Y(u, s, r)$, the *observed values* of R and Y are properly defined as follows:

$$R_S(u) = R(u, S(u)) = \text{the observed } R\text{-response}, \tag{17}$$

and

$$Y_{SR_S}(u) = Y(u, S(u), R(u, S(u))) = \text{the observed } Y\text{-response}. \tag{18}$$

The use of "multiple versions" of the dependent variable—e.g. R_t and R_c, Y_{tr} and Y_{cr}—goes back to Neyman (1935) in the experimental design literature and is often implicit in the early work of Fisher (1926). See Holland (1986*b*, sect. 6) for more on the history of this notation.

The dependence of $R(u, s)$ and $Y(u, s, r)$ on the unit, u, is the way that Rubin's model accommodates individual variation in response to causes. This individual variation is just another way of conceptualizing the idea that the value of a response, say Y, depends both on causes that are measured, like s and r, and on other factors that affect u's responses in various ways.

The data obtained from any unit u in an encouragement design is the triple

$$\big(S(u), R_S(u), Y_{SR_S}(u)\big). \tag{19}$$

In an encouragement design, the values of $S(u)$ are under experimental control, so that the value of $S(u)$ for each u can be determined by randomization. When U is infinite, randomization implies that $S(u)$ is statistically independent of $R_s(u)$ and $Y_{sr}(u)$ over U for any choices of s and r. When U is finite and large, randomization implies that the independence of S and R_s and of S and Y_{sr} over U holds approximately. This is discussed in more detail in the appendix. An important difference between the variables R_s and R_S and between Y_{sr} and Y_{SR_S} is that, except in very special circumstances, randomization does not imply that the observed variables R_S or Y_{SR_S} are statistically independent of S over U even though R_s and Y_{sr} are. For example,

$$P(R_S = r | S = t) = P(R_t = r | S = t) = P(R_t = r),$$

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

and unless $P(R_t = r) = P(R_c = r)$, it follows that $P(R_S = r|S = t) \neq P(R_S = r)$. (The "probabilities," $P(R_S = r|S = t)$, $P(R_t = r)$, etc., are to be interpreted simply as proportions of unit in U [see the appendix].) Thus, randomization may be used to justify the assumption that S and $\{R_s, Y_{sr}$, for all $s, r\}$ are independent but not that S and the *observed values*, R_S and Y_{SR_S}, are independent.

There are four types of unit-level causal effects in this system: three different effects of encouragement (t) and one effect of studying (R). Thus, t can affect both R and Y, and two of the t effects are defined as follows:

$$R_t(u) - R_c(u) = \text{the causal effect of } t \text{ on } R, \tag{20}$$

$$Y_{tR_t(u)}(u) - Y_{cR_c(u)}(u) = \text{the causal effect of } t \text{ on } Y. \tag{21}$$

The definition in (20) is interpreted as the increment in the amount that unit u would study if encouraged to study over how much u would study if not encouraged. The definition in (21) is similar in that it is the increment in the test score u would receive if u were encouraged to study (and studied for $R_t(u)$ hours) over the test score u would receive if u were not encouraged to study (and studied for $R_c(u)$ hours).

In addition to (20) and (21), to specify the ALICE model in the next section, we need to define the effect of t on Y for fixed r, i.e., the effect of t on $Y(\cdot, \cdot, r)$. This is

$$Y_{tr}(u) - Y_{cr}(u) = \text{the causal effect of } t \text{ on } Y(\cdot, \cdot, r). \tag{22}$$

Definition (22) is the "pure" effect of encouragement on test scores because it is the increment in u's test score when u studies r hours and is encouraged to study, compared with u's test score when u studies r hours but is not encouraged to study. Definition (22) is an explicit statement of the idea that the amount u studies is a self-selected treatment that can differ from what actually occurs, i.e., from the particular values $R_t(u)$ and $R_c(u)$ that appear in definition (21). The idea behind (22) is quite subtle and central to the notion of indirect causation. In the studying example used throughout this paper, it may be plausible to suppose that causal effects defined in (22) are all zero, but I will not make that assumption at this stage of the development so that I can apply the model to other cases in which these causal effects might not be zero.

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

The amount of study, R, can affect only Y, and the effect of R is defined as follows:

$$Y_{sr}(u) - Y_{sr'}(u) = \text{the effect of } R = r \text{ relative to } R = r' \text{ on } Y(\cdot, s, \cdot). \tag{23}$$

Definition (23) is also an explicit statement of the idea that amount of study is a self-selected treatment and can differ from the amount the student did study; i.e., r could have taken on values other than $R_t(u)$ and $R_c(u)$. In (23), the encouragement condition is fixed, s, and the causal effect of R is the change in test score that results when u studies r versus r' hours.

These four types of causal effects, i.e.,

$$R_t(u) - R_c(u), Y_{tR_t(u)}(u) - Y_{cR_c(u)}(u), Y_{tr}(u) - Y_{cr}(u)$$

$$\text{and } Y_{sr}(u) - Y_{sr'}(u),$$

are all defined on each unit and express the effect of encouragement and of studying on the behavior of individual students.

The key feature of Rubin's model is its use of unit-level causal effects as the basic building blocks for defining all other causal parameters. (Rogosa [1987] also emphasizes the importance of models that start at the level of individual units and build up.) Unit-level causal effects are never directly observable because of what I call the fundamental problem of causal inference (see the appendix), but they may be used to define causal parameters that can be estimated or measured with data.

Averaging each of the four types of unit-level causal effects over U results in the important causal parameters called *average causal effects*, or ACEs. The four ACEs are

$$\text{ACE}_{tc}(R) = E(R_t - R_c), \tag{24}$$

$$\text{ACE}_{tc}(Y) = E(Y_{tR_t} - Y_{cR_c}), \tag{25}$$

$$\text{ACE}_{tc}(Y(\cdot, \cdot, r)) = E(Y_{tr} - Y_{cr}), \tag{26}$$

and

$$\text{ACE}_{rr'}(Y(\cdot, s, \cdot)) = E(Y_{sr} - Y_{sr'}). \tag{27}$$

In (24)–(27) and below, we use $E(\ )$ to denote expectation or average over U. The ACEs are typically the only causal parameters that can be estimated with data. Under some conditions, such as those defined by

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

the ALICE model discussed in section 4.2, an ACE may be interpreted as a unit-level causal effect, but in general it is not.

The ACEs must be distinguished from the *prima facie average causal effects*, or FACEs, which are defined in terms of the observables $S(u)$, $R_S(u)$, and $Y_{SR_S}(u)$. The four FACEs are the following differences in regression functions:

$$\text{FACE}_{tc}(R) = E(R_S|S=t) - E(R_S|S=c), \tag{28}$$

$$\text{FACE}_{tc}(Y) = E(Y_{SR_S}|S=t) - E(Y_{SR_S}|S=c), \tag{29}$$

$$\begin{aligned}\text{FACE}_{tc}(Y(\cdot,\cdot,r)) &= E(Y_{SR_S}|S=t, R_S=r) \\ &\quad - E(Y_{SR_S}|S=c, R_S=r),\end{aligned} \tag{30}$$

$$\begin{aligned}\text{FACE}_{rr'}(Y(\cdot,s,\cdot)) &= E(Y_{SR_S}|S=s, R_S=r) \\ &\quad - E(Y_{SR_S}|S=s, R_S=r').\end{aligned} \tag{31}$$

Because they are based on the observables, the FACEs are associational parameters rather than causal parameters. They are *prima facie* ACEs rather than ACEs because they may or may not equal their corresponding ACEs, depending on whether certain assumptions are met. Causal inference in Rubin's model means inference about causal parameters, such as the ACEs. Such inferences must be made from observable data and hence the FACEs play an important role. For example, consider $\text{FACE}_{tc}(R)$. Since S is independent of R_t and R_c by assumption (a consequence of the random assignment of the encouragement conditions), we have

$$\begin{aligned}\text{FACE}_{tc}(R) &= E(R_S|S=t) - E(R_S|S=c) \\ &= E(R_t|S=t) - E(R_c|S=c) \\ &= E(R_t) - E(R_c) \\ &= \text{ACE}_{tc}(R).\end{aligned} \tag{32}$$

Thus, because of random assignment, the causal parameter, $\text{ACE}_{tc}(R)$, and the associational parameters, $\text{FACE}_{tc}(R)$, are equal. Similarly, one may show that

$$\text{FACE}_{tc}(Y) = \text{ACE}_{tc}(Y), \tag{33}$$

also because of the random assignment of encouragement.

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

The other two FACEs involve R_S, whose distribution is not under experimental control. First consider $\mathrm{FACE}_{tc}(Y(\cdot,\cdot,r))$

$$\begin{aligned} &= E(Y_{SR_S}|S=t, R_S=r) - E(Y_{SR_S}|S=c, R_S=r) \\ &= E(Y_{tr}|S=t, R_t=r) - E(Y_{cr}|S=c, R_c=r) \\ &= E(Y_{tr}|R_t=r) - E(Y_{cr}|R_c=r). \end{aligned} \tag{34}$$

In general, this does not equal $E(Y_{tr}) - E(Y_{cr})$. Therefore, we cannot use $\mathrm{FACE}_{tc}(Y(\cdot,\cdot,r))$ for $\mathrm{ACE}_{tc}(Y(\cdot,\cdot,r))$ without additional assumptions.

Next consider $\mathrm{FACE}_{rr'}(Y(\cdot,s,\cdot))$

$$\begin{aligned} &= E(Y_{SR_S}|S=s, R_S=r) - E(Y_{SR_S}|S=s, R_S=r') \\ &= E(Y_{sr}|S=s, R_s=r) - E(Y_{sr'}|S=s, R_s=r') \\ &= E(Y_{sr}|R_s=r) - E(Y_{sr'}|R_s=r'). \end{aligned} \tag{35}$$

Again, in general this does not equal the corresponding ACE, i.e. $E(Y_{sr}) - E(Y_{sr'})$, which is the average causal effect of studying on test performance that interests us in an encouragement design.

What can we conclude so far? First, assuming random assignment of the encouragement conditions to units, the FACEs based on the conditional expectations of R_S and of Y_{SR_S} given S are equal to their corresponding ACEs and thus have causal interpretations as ACEs. These FACEs would be estimated, in practice, by treatment-control mean differences for R_S and Y_{SR_S}, respectively. This result is not surprising, and related material is discussed in the appendix. The other two FACEs, those based on the conditional expectation of Y_{SR_S} given both S and R_S, do not equal their corresponding ACEs, in general; and in particular, without further assumptions it is not true that the "effect of studying" on test performance that one would obtain from a regression analysis of Y_{SR_S} on S and R_S can be interpreted as an average causal effect over U.

4.2. *The ALICE Model*

In Rubin's model, a causal theory specifies, or partially specifies, values for $R(u,s)$ and $Y(u,s,r)$. An important causal theory that I find helpful in understanding the relationship between this extension of Rubin's model and path analysis and structural equations models is

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

what I call the *a*dditive, *li*near, *c*onstant *e*ffect, or ALICE, model. It is given by three equations that involve unit-level causal effects:

$$R_t(u) - R_c(u) = \rho, \tag{36}$$

$$Y_{tr}(u) - Y_{cr}(u) = \tau, \tag{37}$$

$$Y_{sr}(u) - Y_{sr'}(u) = \beta(r - r'). \tag{38}$$

In this model, the effects of t and r on Y for a given unit, u, are *additive*, the effect of values of r on Y enters *linearly*, and the causal effects of t on R and Y and of r on Y are constant, not depending on the unit; i.e., this is a causal theory with *constant effects* (see the appendix for more on constant effects). Equations (36)–(38) involve three of the four unit-level causal effects in (20)–(23). The fourth one, i.e. (21), can be expressed in terms of the other three:

$$Y_{t_{R_t}}(u) - Y_{c_{R_c}}(u) = \tau + \rho\beta. \tag{39}$$

In (36), ρ is the (constant) number of hours that encouragement increases each student's amount of study. In (39), $\tau + \rho\beta$ is the (constant) improvement in test scores due to encouragement to study. In (37), τ is the (constant) amount that encouragement increases the test scores of a student who always studies r. In (38), β is the (constant) amount that studying one hour more increases a student's test scores.

The ALICE model in (36)–(38) is equivalent to these two functional relations of the variables s and r for each fixed unit, u:

$$R_s(u) = R_c(u) + \rho s, \tag{40}$$

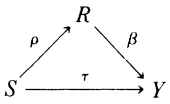
$$Y_{sr}(u) = Y_{c0}(u) + \tau s + \beta r. \tag{41}$$

On the right-hand sides of (40) and (41), s is a 0/1 variable, $Y_{c0}(u)$ is the test performance of u if u is not encouraged to study and doesn't study, and $R_c(u)$ is the amount u studies if not encouraged to study. The values of $Y_{c0}(u)$ and $R_c(u)$ will vary from student to student and are the vehicle for introducing unit heterogeneity into this model (see the appendix on unit homogeneity).

For a fixed unit, (40) and (41) are a deterministic linear system involving the functions $Y_{sr}(u)$ and $R_s(u)$ of the variables s and r. If we equate r and $R_s(u)$, then we have a deterministic linear system, and as in section 3, we may associate the path diagram of Figure 4 with it. I have left off the subscripts for R and Y in Figure 4 to emphasize that it does not describe empirical relationships in data; i.e., it is not an

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

FIGURE 4.



empirical path diagram. Rather, it is a theory about the values of $Y(u, s, r)$ and $R(u, s)$; i.e., it is a causal model or a causal theory.

The ALICE model may appear to be an extremely strong model, yet we shall see presently that it is not strong enough to ensure that the regression coefficients of path analysis have the desired causal interpretations.

The parameters of the ALICE model (ρ, β, and τ) may be used to express the four ACEs of the model. These are

$$\mathrm{ACE}_{tc}(R) = \rho, \tag{42}$$

$$\mathrm{ACE}_{tc}(Y) = \tau + \beta\rho, \tag{43}$$

$$\mathrm{ACE}_{tc}(Y(\cdot, \cdot, r)) = \tau, \tag{44}$$

$$\mathrm{ACE}_{rr'}(Y(\cdot, s, \cdot)) = \beta(r - r'). \tag{45}$$

The ALICE causal model is an example of a "constant effect" model (see the appendix). Consequently, in the ALICE model, the ACEs are interpretable as unit-level causal effects. This is seen by comparing (42)–(45) with (36)–(39).

We see that the "total effect" of S on Y in Figure 4 (i.e. $\tau + \beta\rho$) is an ACE. In addition, for the ALICE model, ρ, τ, and β can all be interpreted as ACEs. What about the FACEs? From the results of the previous subsection, we know that because of randomization,

$$\mathrm{FACE}_{tc}(R) = \mathrm{ACE}_{tc}(R) = \rho \tag{46}$$

and

$$\mathrm{FACE}_{tc}(Y) = \mathrm{ACE}_{tc}(Y) = \tau + \beta\rho. \tag{47}$$

The other two FACEs are more complicated. They may be shown to be given by the following formulas:

$$\mathrm{FACE}_{tc}(Y(\cdot, \cdot, r)) = \tau + \mu_c(r - \rho) - \mu_c(r), \tag{48}$$

and

$$\mathrm{FACE}_{tc}(Y(\cdot, s, \cdot)) = \beta(r - r') + \mu_s(r) - \mu_s(r'), \tag{49}$$

where

$$\mu_s(r) = E(Y_{c0} | R_s = r) \text{ for } s = t, c. \tag{50}$$

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

Thus, the two remaining FACEs both equal their corresponding ACEs plus biases that involve the regression of Y_{c0} on R_s, i.e. $\mu_s(r)$.

The value of $\mu_c(r)$ is the average value of test scores for students when they are not encouraged to study *and* they do not study, for all those students who would study an amount r when they are not encouraged to study. Thus, $\mu_c(r)$ is a "counterfactual" regression because Y_{c0} and R_s can *never* be simultaneously observed except when $R_s = 0$. Hence, $\mu_c(r)$ is inherently unobservable, and assumptions made about it have no empirical consequences that can be directly tested. The function $\mu_c(r)$ is a complicated quantity and one that is not easily thought about. Suppose, for simplicity, that it is linear, i.e., that

$$\mu_c(r) = \gamma + \delta r. \tag{51}$$

A positive δ means that the more a student would study when not encouraged, the *higher* he or she would score on the test without studying and without encouragement. A negative δ means that the more a student would study when not encouraged, the *lower* he or she would score without studying and without encouragement.

The quantities computed in path analysis are the conditional expectations

$$E(R_S|S) = E(R_c) + \rho S \tag{52}$$

and

$$E(Y_{SR_S}|S, R_S) = \mu_c(R_S - \rho S) + \tau S + \beta R_S, \tag{53}$$

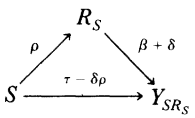
in which S is a 1/0 indicator variable. If we make the untestable assumption that $\mu_c(r)$ is linear, e.g. (51), then (53) becomes

$$E(Y_{SR_S}|S, R_S) = \gamma + (\tau - \delta\rho)S + (\beta + \delta)R_S. \tag{54}$$

Equations (52) and (54) are both linear and may be combined into the empirical path diagram in Figure 5.

Comparing Figures 5 and 4, we see that even if the ALICE model holds and $\mu_c(r)$ is linear, the estimated path coefficients are

FIGURE 5.



This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

biased estimates of the causal effects τ and β unless $\mu_c(r)$ does not depend on r (i.e. $\delta = 0$). Furthermore, these problems stem from the inhomogeneity of the units with respect to the values of $R_c(u)$ and $Y_{c0}(u)$. This inhomogeneity is, I believe, the proper way to view the "disturbance terms" of the structural equations model (9) and (10) in section 3 (see section 4.4). One nice thing is that while the *direct effects* are not the same in Figures 5 and 4, the total effects are: Both equal $\tau + \rho\beta$.

4.3. *Two Different Ways to Estimate the Causal Effect of the Encouraged Activity*

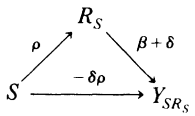
The message of the previous subsection is that the effect of study on test performance cannot be estimated by the usual regression methods of path analysis without making untestable assumptions about the counterfactual regression function, $\mu_c(r)$. If we assume that $\mu_c(r)$ is constant, then the biases shown in Figure 5 vanish and the usual path coefficients may be interpreted as causal effects, i.e. as ACEs. However, because $\mu_c(r)$ is so difficult to think about, there is little reason to believe that it is constant. Nor can δ be easily assessed as either positive or negative, since, in this example, there are reasons why it might be either: If students who study a lot tend to be those who do well even when they don't study, then δ is positive; but if those who study a lot are those who need to study, then δ is negative.

An alternative approach is to suppose that encouragement, of and by itself, has no effect on Y. In the studying example, this might be a plausible assumption. This corresponds to the restriction that

$$\tau = 0. \tag{55}$$

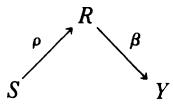
Now the empirical path diagram becomes that in Figure 6, and the path diagram for the causal theory becomes that in Figure 7. The total effect of S on Y_{SR_S} is now $\rho\beta$, whereas the total effect of S on R_S is ρ.

FIGURE 6.



This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

FIGURE 7.



Hence,

$$\beta = \frac{\rho\beta}{\rho} = \frac{\text{total effect of } S \text{ on } Y_{SR_S}}{\text{total effect of } S \text{ on } R_S}. \tag{56}$$

This is also easily seen from the definitions of the ACEs and the FACEs. Under the assumption that $\tau = 0$,

$$\text{ACE}_{tc}(Y) = \text{FACE}_{tc}(Y) = \beta\rho, \tag{57}$$

regardless of whether or not $\mu_c(r)$ is linear, and hence

$$\beta = \frac{\text{ACE}_{tc}(Y)}{\text{ACE}_{tc}(R)} = \frac{\text{FACE}_{tc}(Y)}{\text{FACE}_{tc}(R)}. \tag{58}$$

The two FACEs in (58) may be estimated simply by the treatment-control mean difference in Y_{SR_S} and R_S, as mentioned earlier, so that (58) provides an alternative way to estimate β that does not assume that $\mu_c(r)$ is constant. In Powers and Swinton (1984), (58) was used to estimate β.

4.4. *Deriving a Structural Equations Model*

The ALICE model may be used to derive the structural equations model given in (9) and (10). If we substitute $S(u)$ for s in (40) and $S(u)$ for s and $R_S(u)$ for r in (41), we get the following pair of equations that involve the observables, S, R_S, Y_{SR_S}:

$$R_S(u) = R_c(u) + \rho S(u) \tag{59}$$

and

$$Y_{SR_S}(u) = Y_{c0}(u) + \tau S(u) + \beta R_S(u). \tag{60}$$

Now let

$$\eta_1(u) = R_c(u) - E(R_c)$$

and

$$\eta_2(u) = Y_{c0}(u) - E(Y_{c0}),$$

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

and then define

$$\alpha = E(R_c), \alpha' = E(Y_{c0}).$$

The following equations, which parallel the structural equations model of (9) and (10), follow immediately:

$$R_S(u) = \alpha + \rho S(u) + \eta_1(u), \tag{61}$$

$$Y_{SR_S}(u) = \alpha' + \tau S(u) + \beta R_S(u) + \eta_2(u). \tag{62}$$

It is easy to see from the definition of η_1 and η_2 that by the independence assumption (justified by randomization), S is independent of η_1 and η_2 over U. But R_S is not independent of η_2 in general. In fact, the condition that equation (62) be interpretable as a conditional expectation is exactly that $\mu_c(r)$ be constant. It follows from the standard theory of structural equations models that ordinary least squares estimates of β are biased in general, so that a simple regression analysis of (62) would not lead to an estimate of the causal effect of studying on test scores. Substituting (61) for R_S in (62) yields

$$\begin{aligned} Y_{SR_S}(u) &= \alpha' + \beta\alpha + (\tau + \rho\beta)S(u) + \beta\eta_1(u) + \eta_2(u) \\ &= a'' + (\tau + \rho\beta)S(u) + \eta_3(u). \end{aligned} \tag{63}$$

Equations (61) and (63) constitute the so-called reduced form of the system (61) and (62). Since S is independent of η_1 and η_2, S is also independent of $\eta_3 = \beta\eta_1 + \eta_2$ in (63). Thus, (61) and (63) can be interpreted as regression functions; therefore, in the language of structural equations models, (61) can be used to estimate ρ and (62) can be used to estimate $\tau + \rho\beta$. Assuming $\tau = 0$ now leads to the second estimate of β discussed in section 4.3.

In closing this section, I point out that the ALICE model leads to the structural equations system (61) and (62); but the ALICE model could be wrong, in various, often testable, ways. Freedman (1987) has argued that models like (61) and (62) should be tested before they are used. Rubin's model gives us a framework for doing that testing. But an assumption like $\tau = 0$ is not testable with the data in hand and must be justified on other grounds.

5. DISCUSSION

One purpose of this paper is to show that path analysis and its generalization, structural equations models, do not justify causal interpretations of regression coefficients. Instead, these models simply define

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

certain regression coefficients as causal effects by fiat, using the loose causal terminology of regression analysis. Rubin's model, on the other hand, precisely defines unit-level causal effects, and these, in turn, may be used to deduce causal interpretations of *some* regression coefficients under *some* assumptions. By explicitly separating the causal theory, $R(u, s)$ and $Y(u, s, r)$, from the observed data, $(S(u), R_S(u)$, and $Y_{SR_S}(u))$, Rubin's model provides analysts with a set of tools that engender careful thought about causal theories and their relationship to data.

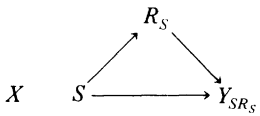
One example of such "careful thought" is the care that must be exercised in identifying variables in complex path models that can truly play the role of both effect and cause. Such variables must measure the amount of exposure of units to a cause and be themselves influenced by another cause. I used the encouragement design to focus my analysis precisely because it is a clearly interpretable example of indirect causation in this sense. Many causal models in the literature are not careful about this point; at best, they merely measure the association between variables rather than produce estimates of causal effects.

I am a little reluctant to place as much emphasis as I have on the ALICE model of section 4, because I do not wish to appear to endorse it as the only way to analyze data from encouragement designs. (For example, it would be inappropriate if "studying" were measured simply as a studied/didn't study dichotomy.) I simply put the ALICE model forward as a basic case from which deductions are easily made and which has interesting consequences for path analysis models. It is a model that captures some of the complexity of encouragement designs and self-selected treatments. Heterogeneity, nonlinearity, and nonadditivity can be added to the ALICE model in various ways to add complexity when that is necessary for proper analyses.

The two alternative ways to estimate the effect of studying on test scores given in section 4.3 were discussed simply to illustrate how the causal model must be used to produce estimates of causal effects. In the studying example, at least, the untestable assumption that $\tau = 0$ is more believable (and understandable) than the untestable assumption that leads to the usual path analytic estimate of the causal effect, i.e. $\delta = 0$. Furthermore, though the structural equations model of (61) and (62) can be used to obtain the ratio estimate of β in section 4.3, there is no way to justify these equations except through the ALICE model, or its generalizations.

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

FIGURE 8.



The assumption of random assignment of the encouragement condition is an important starting place, but there may be applications in which this is impossible or implausible, and we need to consider the corresponding observational study in which S is not independent of $\{R_s\}$ and $\{Y_{sr}\}$. An interesting generalization is to replace randomization by a strong ignorability type of condition given a covariate, $X(u)$ (see the appendix). For example, suppose that there is a covariate X such that given X, S is conditionally independent of $\{R_s\}$ and $\{Y_{sr}\}$. Now, all the equations of section 4 will hold conditionally given X, and the calculation of the FACEs is replaced by the corresponding *covariate-adjusted* FACEs, i.e. the C-FACEs (see the appendix). How might we wish to represent such a system in terms of path diagrams? In Holland (1986*b*), I suggested that an arrow connect two variables only if one indicated a cause (like S or R_S) and the other measured a response (like R_S or Y_{SR_S}). A covariate does not have the status of a causal indicator or of a response, so it ought not be involved with the arrows, according to such a view. Hence, the four-variable system of $(X(u), S(u), R_S(u), Y_{SR_S}(u))$ could be represented as in Figure 8, in which X precedes S and R, to indicate that it is not affected by either causal variable. However, it might be useful to indicate the conditional independence of S and all the variables $\{R_s\}$ and $\{Y_{sr}\}$ given X. This should be done, not in an empirical path diagram like Figure 8, but in a path diagram for a causal theory, like Figure 4. Conditional independence also plays an important role in structural equations models with latent variables, but that is a subject worthy of another paper. The importance of conditional independence suggests the use of two types of arrows in path diagrams: e.g., solid arrows to indicate causal relations and something like dashed arrows to indicate conditional independence. In the complex causal models of the current literature, such distinctions are not made and all arrows indicate causality. This is a mistake and leads to careless and casual causal talk. I hope that my illustration of how Rubin's model can be used to give precision to causal modeling will stimulate similar analyses of more complex causal models.

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

APPENDIX: A BRIEF REVIEW OF RUBIN'S MODEL FOR EXPERIMENTS AND OBSERVATIONAL STUDIES

Discussions of Rubin's model similar to the one here may be found in Holland (1986*a*, *b*). Uses of the model in a variety of significant applications appear in Rubin (1974, 1977, 1978), Holland and Rubin (1983, 1987), Rosenbaum and Rubin (1983*a*, *b*; 1984*a*, *b*; 1985*a*, *b*), Rosenbaum (1984*a*, *b*, *c*; 1987), and Holland (1988). In the simplest case, the logical elements of Rubin's model form a quadruple (U, K, S, Y) where U is a population of units, K is a set of causes or treatments to which each one of the units in U may be exposed, $S(u) = s$ if s is the cause in K to which u is actually exposed, and $Y(u, s) =$ the value of the response that would be observed if unit $u \in U$ were exposed to cause $s \in K$.

The meaning of $Y(u, s)$ needs some explanation. The response variable, Y, depends both on the unit *and* on the cause or treatment to which the unit is exposed. The idea is that if u were exposed to $t \in K$, then we would observe the response value $Y(u, t)$; but if u were exposed to $c \in K$, then we would observe the response value $Y(u, c)$. The requirement that Y be a function on *pairs* (u, s) means that $Y(u, s)$ represents the measurement of some property of u after u is exposed to cause $s \in K$. This has the important consequence of forcing the things that are called causes in K to be *potentially exposable* to any unit in U. This restriction on the notion of cause is of fundamental importance because it prevents us from interpreting a variety of associations as causal: e.g., associations between sex and income or between race and crime. This is discussed more extensively in Holland (1986*b*, sect. 7) and in Holland (1988). The function Y is called the *response function*. In the references cited at the beginning of this section, a subscript notation is used for $Y(u, s)$; i.e.,

$$Y_s(u) = Y(u, s).$$

The subscript notation is convenient, and I will use it when appropriate.

The mapping, S, is the *causal indicator* or assignment rule because it indicates the cause to which each unit is exposed.

The elements of the quadruple (U, K, S, Y) are the *primitives* of Rubin's model, and they serve as the undefined terms. All other concepts are defined in terms of these primitives.

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

The most basic quantity in need of definition is the *observed response* on each unit $u \in U$. This is given by

$$Y_S(u) = Y(u, S(u)).$$

The value $Y_S(u)$ is the value of Y that is actually observed for unit u. The observed data for unit u, in the simplest case, is the pair

$$(S(u), Y_S(u)),$$

where $S(u)$ is the cause or treatment in K to which u is actually exposed and $Y_S(u)$ is the observed value of the response, Y. It is important to distinguish $Y_S(u)$ from $Y(u, s)$: $Y_S(u)$ is the response that is actually observed on unit u, and $Y(u, s)$ is a potentially observed value that is actually observed only if $S(u) = s$.

Note that in the subscript notation, the observed response, $Y_S(u)$, is $Y_{S(u)}(u)$, so that in the usual probabilistic sense, Y_s has a "fixed" subscript and Y_S has a "random" subscript.

In Rubin's model, *causes* are taken as *undefined* elements of the theory, and *effects* are *defined* in terms of the elements of the model.

Definition. The unit-level causal effect of cause $t \in K$ relative to cause $c \in K$ (as measured by Y) is the difference

$$Y(u, t) - Y(u, c) = T_{tc}(u).$$

Hence, the *causal effect*, $T_{tc}(u)$, is the *increase* in the value of $Y(u, t)$, which is what would be observed if u were exposed to t, over that of $Y(u, c)$, which is what would be observed if u were exposed to c. Glymour (1986) points out that in Rubin's model, effects are defined *counterfactually*; i.e., their definitions include sentences of the form "If A were the case then B would be the case," in which A could be false. It should also be noted that $T_{tc}(u)$ is defined *relatively* (i.e., the effect of one cause or treatment is *always* relative to another cause) and at the level of individual units.

The fundamental problem of causal inference. The most vexing problem in causal inference is that it is impossible to simultaneously observe both $Y(u, t)$ and $Y(u, c)$ for two distinct causes t and c; therefore, the causal effect, $T_{tc}(u)$, is *never* directly observable. Rubin's model makes this explicit by separating the *observed data* (S, Y_S) from the function Y. A *causal model* or a *causal theory* is a specification or partial specification of the values of the function Y. Causal inference consists of combining

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

(a) a causal theory, (b) assumptions about data collection, and (c) the observed data to draw conclusions about causal parameters. Many techniques of experimental science are aimed at overcoming the fundamental problem of causal inference by assuming plausible causal theories and then combining them appropriately with data. Some examples of such causal theories are given in the next several paragraphs.

Unit homogeneity. In a scientific laboratory, care is exercised to prepare homogeneous samples of material for study. Such care is often taken to make the following partial specification of Y plausible:

$$Y(u, s) = Y(v, s) \quad \text{for all } u, v \in U \text{ and all } s \in K.$$

This means that the responses of *all* units to cause s are the same, i.e., that units respond homogeneously to each cause. In Holland (1986*a*), I called this the assumption of *unit homogeneity*. It is a partial specification of Y because it restricts the values that Y can take on but it does not specify them completely. If one assumes unit homogeneity, then the causal effect, $T_{tc}(u)$, is easily seen to be given by

$$T_{tc}(u) = Y_t(u) - Y_c(v),$$

for any two distinct units u and v in U. In this case, the effect of t (relative to c) is constant and does not depend on the unit under consideration—a case I call constant effect (see below). Unit homogeneity solves the fundamental problem of causal inference by letting us use the data from two units to measure the causal effect on any single unit.

Fisher's null hypothesis. An assumption about Y that has a long history in statistics and that is formally similar to unit homogeneity is Fisher's null hypothesis:

$$Y(u, s) = Y(u, s') \quad \text{for all } u \in U \text{ and all } s, s'.$$

This means that the response of each unit is unaffected by the cause or treatment to which it is exposed. This is also a partial specification of Y and is a causal theory. Fisher's null hypothesis addresses the fundamental problem of causal inference by assuming that once we observe the value of Y for the pair (u, s), we know the value of Y for the pair (u, s') for any other value of $s' \in K$. Under Fisher's null hypothesis,

$$T_{tc}(u) = 0 \quad \text{for all } u \in U \text{ and all } t, c \in K.$$

So far, I have not given any examples in which assumptions about the data collection process matter. At the population level, "data

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

collection" is contained in the causal indicator variable, S, since S describes the cause to which each unit in U is exposed. Suppose we now consider the joint distribution of S with $\{Y_s: s \in K\}$ as u varies over all of U. By using the term *joint distribution*, I do not mean to imply that S or the $\{Y_s\}$ are stochastic. However, we can use the language of probability to describe this joint distribution. For example, $P(S = s)$ is the proportion of units for which $S(u) = s$, and $E(Y_S|S = t)$ is the average value of Y_S for all those units for which $S(u) = t$. This use of probability notation allows us to discuss other, more statistical, approaches to solving the fundamental problem of causal inference in a convenient manner.

The average causal effect (ACE). We may define an important causal parameter, the ACE, as the average value of $T_{tc}(u)$ over U, or

$$\text{ACE}_{tc} = E(T_{tc}).$$

In this notation, $E(T_{tc})$ denotes the average value of $T_{tc}(u)$, over all $u \in U$. But by definition of $T_{tc}(u)$, this is equivalent to the difference

$$\text{ACE}_{tc}(Y) = E(Y_t) - E(Y_c).$$

The ACE is a useful summary of the unit-level causal effects, $T_{tc}(u)$, when $T_{tc}(u)$ varies little as u ranges over U. In some cases, we are interested in average behavior over the population of units, and in such a case, the ACE is useful regardless of how much $T_{tc}(u)$ varies with u.

When we look at data, we can only observe $S(u)$ and $Y_S(u)$ over U; hence, we can observe data only from the joint distribution of S and Y_S (as opposed to S and $\{Y_s: s \in K\}$). For example, the average value of the observed response Y_S among all those units exposed to cause t is

$$E(Y_S|S = t) = E(Y_t|S = t),$$

and the average value of the observed response among all those units exposed to cause c is

$$E(Y_S|S = c) = E(Y_c|S = c).$$

The difference in average responses between those units exposed to t and those units exposed to c is the *prima facie average causal effect* (FACE) and is given by

$$\begin{aligned}\text{FACE}_{tc}(Y) &= E(Y_S|S = t) - E(Y_S|S = c)\\ &= E(Y_t|S = t) - E(Y_c|S = c).\end{aligned}$$

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

I use FACE and ACE to draw attention to the fact that we can always compute the FACE from data but that it does not necessarily equal the quantity about which we wish to make an inference, i.e. the ACE. The difference between the FACE and the ACE resides in the difference between

$$E(Y_t) \text{ and } E(Y_t|S=t)$$

and between

$$E(Y_c) \text{ and } E(Y_c|S=c).$$

$E(Y_t)$ is the average of Y_t over all of U, whereas $E(Y_t|S=t)$ is the average of Y_t over only those units that are actually exposed to t. The same is true for $E(Y_c)$ and $E(Y_c|S=c)$.

Independence. It is now time to show the effect of randomization on Rubin's model. Suppose S is independent of $\{Y_s: s \in K\}$. When independence holds we have

$$E(Y_t|S=t) = E(Y_t)$$

and

$$E(Y_c|S=c) = E(Y_c).$$

Hence, if S is independent of $\{Y_s: s \in K\}$, the FACE and the ACE are equal; i.e.,

$$\begin{aligned} \text{FACE}_{tc}(Y) &= E(Y_t|S=t) - E(Y_c|S=c) \\ &= E(Y_t) - E(Y_c) \\ &= \text{ACE}_{tc}(Y). \end{aligned}$$

Thus, independence is important because it relates a causal parameter, i.e. the ACE, to an associational parameter, i.e. the FACE, that can be computed or estimated from the observed data, S and Y_S.

Randomization is related to independence in the following way. Independence is an assumption about the data collection process, i.e., about the relationship between S and Y over the population U. Randomization is a physical process that gives *plausibility* to the independence assumption in some important cases. For example, if U were infinite, then the strong law of large numbers coupled with randomization implies that almost every realization of S would be independent of $\{Y_s\}$. Randomization does not *always* make independence plausible; the best example of this is the case of a small

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

population of units. If U contains only two units, then the physical act of randomization does not make the independence assumption plausible, even though it may still be useful in forming the basis of a test of Fisher's null hypothesis.

Constant effect. An important causal theory is the *constant effect* assumption. Constant effect holds when $T_{tc}(u)$ does not depend on u, i.e., when $T_{tc}(u) = \tau_{tc}$ for all u. This is equivalent to

$$Y_t(u) = Y_c(u) + \tau_{tc}.$$

Thus, *constant effect* is the same as *additivity* in the ANOVA sense. I prefer *constant effect*, since it is more descriptive of the causal theory being assumed.

When constant effect holds, it is easy to see that τ_{tc} equals the $\mathrm{ACE}_{tc}(Y)$:

$$\mathrm{ACE}_{tc}(Y) = E(T_{tc}) = \tau_{tc}.$$

What about the FACE?

$$\begin{aligned}\mathrm{FACE}_{tc}(Y) &= E(Y_t|S=t) - E(Y_c|S=c)\\ &= E(Y_c + \tau_{tc}|S=t) - E(Y_c|S=c)\\ &= \tau_{tc} + \{E(Y_c|S=t) - E(Y_c|S=c)\}.\end{aligned}$$

Hence, under the constant effect assumption,

$$\mathrm{FACE}_{tc}(Y) = \mathrm{ACE}_{tc}(Y) + \mathrm{BIAS},$$

where $\mathrm{BIAS} = E(Y_c|S=t) - E(Y_c|S=c)$. The term BIAS involves the "counterfactual conditional expectation," $E(Y_c|S=t)$, which cannot be computed from data because it is the average value of Y_c among all those units that were exposed to t (and for which only the value of Y_t is known). Under independence, $\mathrm{BIAS} = 0$, and as before, the FACE and the ACE are equal.

Introducing other variables into Rubin's model. So far, I have discussed the simplest form of Rubin's model, in which there is only one variable measured on the units—aside from the causal indicator, S. Now suppose there is a second variable, X. In Rubin's model, X is introduced as a second real-valued function on $U \times K$, $X(u, s)$. The fact that X is real-valued is not important; it could be vector-valued. What is important is that we allow for the fact that, in general, X could depend on both u and s. A special class of variables are the *covariates*.

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

Definition. X is a covariate if $X(u, s)$ does not depend on s for any $u \in U$.

In Holland (1986*a*), I used the term *attributes* to refer to *covariates*, but the latter term is preferable because it corresponds to normal experimental usage. Variables measured on units prior to their exposure to treatments are always covariates. Rosenbaum (1984*c*) discusses post-treatment concomitants and their use in statistical adjustments. A post-treatment concomitant is a variable measured *after* the exposure of a unit to the causes in K. For a post-treatment concomitant, the possibility that $X(u, s)$ *does* depend on s cannot be ignored and must be decided. If $X(u, s)$ does depend on s, then X is *not* a covariate in the sense used here.

Observational studies. When the active experimenter is replaced by a passive observer who cannot arrange the values of $S(u)$ to achieve independence, we enter the realm of observational studies. In such studies we are also interested in measuring causal effects; i.e., Rubin's model still applies, but now S is not automatically independent of $\{Y_s\}$. In an observational study, we typically have a covariate, X, and we may check the distribution of X in each exposure group by comparing the values of

$$P(X = x | S = s)$$

across the values of $s \in K$. If there is evidence that $P(X = x | S = s)$ depends on s, then, depending on the nature of X and Y, we may not believe that the independence assumption holds in an observational study. However, we might be willing to entertain a weaker *conditional independence* assumption of the form, given the covariate, X, the variables S, and $\{Y_s: s \in K\}$ are conditionally independent. Combined with the assumption that $P(S = s | X = x) > 0$, the conditional independence assumption is called *strong ignorability* (Rosenbaum and Rubin 1983*a*).

Strong ignorability is the basis for all covariate-adjusted causal effects in observational studies. Covariate adjustments are based on the conditional expectations or regression functions $E(Y_S | S = s, X = x)$, which are used to form the *covariate-adjusted* FACE, i.e., the C-FACE, given by

$$\text{C-FACE}_{tc}(Y) = E\{E(Y_S | S = t, X) - E(Y_S | S = c, X)\}.$$

The C-FACE is like the FACE in that it is generally *not* equal to the

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

ACE, but under conditional independence it is:

$$\begin{aligned}\text{C-FACE}_{tc}(Y) &= E\{E(Y_t|S=t, X) - E(Y_c|S=c, X)\} \\ &= E\{E(Y_t|X) - E(Y_c|X)\} \\ &= E(Y_t) - E(Y_c) \\ &= \text{ACE}_{tc}(Y).\end{aligned}$$

Rubin's model was really developed to address the problem of causal inference in observational studies, and thorough discussions of its application to these types of studies can be found in Rubin (1974, 1977), Holland and Rubin (1983), Rosenbaum and Rubin (1983*a*, *b*; 1984*b*; 1985*a*, *b*), and Rosenbaum (1984*a*, *b*, *c*; 1987).

REFERENCES

Blalock, H. M. 1964. *Causal Inferences in Nonexperimental Research*. Chapel Hill: University of North Carolina Press.

_____. 1971. *Causal Models in the Social Sciences*. Chicago: Aldine.

Campbell, D. T., and J. C. Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand-McNally.

Duncan, O. D. 1966. "Path Analysis: Sociological Examples." *American Journal of Sociology* 72:1–16.

_____. 1975. *Introduction to Structural Equation Models*. New York: Academic Press.

Fisher, R. A. 1926. "The Arrangement of Field Experiments." *Journal of Ministry of Agriculture* 33:503–13.

Freedman, D. A. 1987. "As Others See Us: A Case Study in Path Analysis." *Journal of Educational Statistics* 12:101–28.

Glymour, C. 1986. "Statistics and Metaphysics." Discussion of "Statistics and Causal Inference" by P. W. Holland. *Journal of the American Statistical Association* 81:964–66.

Goldberger, A. S. 1964. *Econometric Theory*. New York: Wiley.

Goldberger, A. S., and O. D. Duncan. 1973. *Structural Equations Models in the Social Sciences*. New York: Seminar Press.

Heise, D. R. 1975. *Causal Analysis*. New York: Wiley.

Holland, P. W. 1986*a*. "Which Comes First, Cause or Effect?" *New York Statistician* 38:1–6.

_____. 1986*b*. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81:945–70.

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

_____. 1988. "Causal Mechanism or Causal Effect: Which is Best for Statistical Sciences?" Discussion of "Employment Discrimination and Statistical Science" by A. P. Dempster. *Statistical Science* (in press).

Holland, P. W., and D. B. Rubin. 1983. "On Lord's Paradox." Pp. 3–25 in *Principals of Modern Psychological Measurement*, edited by H. Wainer and S. Messick. Hillsdale, NJ: Lawrence Erlbaum.

_____. 1987. "Causal Inference in Retrospective Studies." Technical Report No. 87-73. Princeton: Educational Testing Service Program Statistics Research.

Kenny, D. A. 1979. *Correlation and Causality*. New York: Wiley.

Lewis, D. 1986. *Philosophical Papers*. Vol. 2. Oxford: Oxford University Press.

Neyman, J. 1935. (With K. Iwaszkiewicz and S. Kolodziejczyk) "Statistical Problems in Agricultural Experimentation" (with discussion). *Journal of the Royal Statistical Society*, Suppl., 2:107–80.

Powers, D. E., and S. S. Swinton. 1984. "Effects of Self-Study for Coachable Test Item Types." *Journal of Educational Measurement* 76:266–78.

Robins, J. M. 1984. "A Statistical Method to Control for the Healthy Worker Effect in Intracohort Comparisons." *American Journal of Epidemiology* 120:465.

_____. 1985. "A New Theory of Causality in Observational Survival Studies—Application to the Healthy Worker Effect." *Biometrics* 41:311.

_____. 1986. "A New Approach to Causal Inference in Mortality Studies with a Sustained Exposure Period—Application to the Control of the Healthy Worker Survivor Effect." *Mathematical Modelling* 7:1393–1512.

Rogosa, D. 1987. "Casual Models Do Not Support Scientific Conclusion: A Comment in Support of Freedman." *Journal of Educational Statistics* 12:185–95.

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.

_____. 1984*b*. "The Consequences of Adjustment for a Concomitant Variable that has been Affected by the Treatment." *Journal of the Royal Statistical Society*, ser. A, 147:656–66.

_____. 1984*c*. "Conditional Permutation Tests and the Propensity Score in Observational Studies." *Journal of the American Statistical Association* 79:565–74.

_____. 1987. "The Role of a Second Control Group in an Observational Study" (with discussion). *Statistical Science* 2:292–316.

Rosenbaum, P. R., and D. B. Rubin. 1983*a*. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70:41–55.

_____. 1983*b*. "Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome." *Journal of the Royal Statistical Society*, ser. B, 45:212–18.

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions

_____. 1984*a*. 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.

_____. 1984*b*. "Reducing Bias in Observational Studies using Subclassification on the Propensity Score." *Journal of the American Statistical Association* 79:516–24.

_____. 1985*a*. "Constructing a Control Group using Multivariate Matched Sampling Methods that Incorporate the Propensity Score." *American Statistician* 39:33–38.

_____. 1985*b*. "The Bias Due to Incomplete Matching." *Biometrics* 41:103–16.

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." *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* 75:591–93.

_____. 1986. "Which Ifs Have Causal Answers?" Discussion of "Statistics and Causal Inference" by P. W. Holland. *Journal of the American Statistical Association* 81:961–62.

Saris, W. E., and L. H. Stronkhorst. 1984. *Causal Modeling in Nonexperimental Research*. Amsterdam: Sociometric Research Foundation.

Swinton, S. 1975. "An Encouraging Note." Unpublished manuscript.

Tukey, J. W. 1954. "Causation, Regression, and Path Analysis." Pp. 35–66 in *Statistics and Mathematics in Biology*, edited by O. Kempthorne. Ames, IA: Iowa State College Press.

Wright, S. 1934. "The Method of Path Coefficients." *Annals of Mathematical Statistics* 5:161–215.

This content downloaded on Wed, 13 Feb 2013 18:37:26 PM
All use subject to JSTOR Terms and Conditions