Identification of Causal Effects Using Instrumental Variables

Joshua D. Angrist; Guido W. Imbens; Donald B. Rubin

Journal of the American Statistical Association, Vol. 91, No. 434 (Jun., 1996), 444-455.

Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28199606%2991%3A434%3C444%3AIOCEUI%3E2.0.CO%3B2-O

Journal of the American Statistical Association is currently published by American Statistical Association.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/astata.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact jstor-info@jstor.org.

http://www.jstor.org/
Thu Apr 8 15:53:57 2004

Identification of Causal Effects Using Instrumental Variables

Joshua D. ANGRIST, Guido W. IMBENS, and Donald B. RUBIN

We outline a framework for causal inference in settings where assignment to a binary treatment is ignorable, but compliance with the assignment is not perfect so that the receipt of treatment is nonignorable. To address the problems associated with comparing subjects by the ignorable assignment—an "intention-to-treat analysis"—we make use of instrumental variables, which have long been used by economists in the context of regression models with constant treatment effects. We show that the instrumental variables (IV) estimand can be embedded within the Rubin Causal Model (RCM) and that under some simple and easily interpretable assumptions, the IV estimand is the average causal effect for a subgroup of units, the compliers. Without these assumptions, the IV estimand is simply the ratio of intention-to-treat causal estimands with no interpretation as an average causal effect. The advantages of embedding the IV approach in the RCM are that it clarifies the nature of critical assumptions needed for a causal interpretation, and moreover allows us to consider sensitivity of the results to deviations from key assumptions in a straightforward manner. We apply our analysis to estimate the effect of veteran status in the Vietnam era on mortality, using the lottery number that assigned priority for the draft as an instrument, and we use our results to investigate the sensitivity of the conclusions to critical assumptions.

KEY WORDS: Compliers; Intention-to-treat analysis; Local average treatment effect; Noncompliance; Nonignorable treatment assignment; Rubin-Causal-Model; Structural equation models.

1. INTRODUCTION

Economists are typically interested in estimating causal effects rather than mere associations between variables. Potentially interesting causal effects include the effects of education on employment and earnings, the effects of employment training programs on subsequent labor market histories, and the effects of a firm's inputs on its output. The dominant approach to making inferences about causal effects in economics over the last four decades is based on *structural equation models,* which rely on the specification of systems of equations with parameters and variables that attempt to capture behavioral relationships and specify the causal links between variables. Goldberger (1972) and Morgan (1990) provided historical perspectives on these models, which date back to Wright (1928, 1934) and Haavelmo (1943, 1944). Inference in structural equation models often exploits the presence of *instrumental variables* (IV). These are variables that are explicitly excluded from some equations and included in others, and therefore correlated with some outcomes only through their effect on other variables.

Rather than relying on structural equation models, causal inference in statistics, going back at least to work by Fisher (1918, 1925) and Neyman (1923) on agricultural experiments, is fundamentally based on the randomized experiment (see also Kempthorne 1952 and Cox 1958). The basic notion in this formulation, which has been extended by Rubin (1974, 1978) to more complicated situations, including observational studies without randomization, is that of *potential outcomes.* The causal effect of a treatment on a single individual or unit of observation is the comparison (e.g., difference) between the value of the outcome if the unit is treated and the value of the outcome if the unit is not treated. The target of estimation, the estimand, is typically the average causal effect, defined as the average difference between treated and untreated outcomes across all units in a population or in some subpopulation (e.g., males or females). For this definition of causality to be applicable to samples with units already exposed to treatments, we must be able to imagine observing outcomes on a unit in circumstances other than those to which the unit was actually exposed. This approach is now widely used in statistics and epidemiology (e.g., Efron and Feldman 1991 and Greenland and Robins 1986), where it is often referred to as the Rubin Causal Model (RCM; Holland [1986]).

In this article we provide a link between these approaches, capitalizing on the strengths of each. Earlier work combining elements of these approaches includes studies by Hearst, Newman, and Hulley (1986), Holland (1988), Permutt and Hebel (1989), Sommer and Zeger (1991), and Imbens and Angrist (1994). We show how the IV estimand can be given a precise and straightforward causal interpretation in the potential outcomes framework, despite nonignorability of treatment received. This interpretation avoids drawbacks of the standard structural equation framework, such as constant effects for all units, and delineates critical assumptions needed for a causal interpretation. The IV approach provides an alternative to a more conventional intention-to-treat analysis, which focuses solely on the average causal effect of assignment on the outcome (Lee, Ellenberg, Hirtz, and Nelson 1991).

As we show in the context of a specific application, our formulation of these assumptions makes it easier for researchers to judge whether or not a causal interpretation of the instrumental variables estimand is plausible. Standard

Joshua D. Angrist is Associate Professor, Department of Economics, Massachusetts Institute of Technology, Cambridge, MA 02139. Guido W. Imbens is Associate Professor, Department of Economics, and Donald B. Rubin is Professor, Department of Statistics, Harvard University, Cambridge, MA 02138. The authors thank Gary Chamberlain, Clifford Clogg, Sander Greenland, Judy Hellerstein, Paul Rosenbaum, the editor, an associate editor and several referees for comments, the National Science Foundation for financial support, and Tom Newman and Norman Hearst for sharing their data.

© 1996 American Statistical Association
Journal of the American Statistical Association
June 1996, Vol. 91, No. 434, Applications and Case Studies

IV procedures rely on judgments regarding the correlation between functional-form-specific disturbances and instruments. In contrast, our approach forces the researcher to consider the effect of exposing units to specific treatments. If it is not possible (or not plausible) to envision the alternative treatments underlying these assumptions, the use of these techniques may well be inappropriate. Moreover, by separating and defining the critical assumptions, our formulation allows for a clear assessment of the consequences of violations of these assumptions through sensitivity analysis under more general models. Our main results are summarized in three propositions: the first provides conditions for a causal interpretation of the IV estimand, and the others reveal the consequences of violations of the critical assumptions.

We develop our presentation in the context of an evaluation of the effect of serving in the military on health outcomes. Data for this study come from the Vietnam era, when priority for conscription was randomly allocated through the draft lottery. For expository purposes, and to be precise without cumbersome notation, we use the simplest possible example: both the "treatment" (i.e., serving in the military or not, denoted by D) and the "assignment" (i.e., draft status, determined by lottery number, denoted by Z) are binary. If compliance with the draft had been perfect, then all those with a low lottery number ($Z = 1$) would have served in the military ($D = 1$), and all those with a high lottery number ($Z = 0$) would not have served ($D = 0$). We assume that we observe values of Z, D, and the health outcome Y for each person. Our basic results, however, are not limited to this case with binary treatment and binary instrument. The approach developed here can be extended to multi-valued treatments and instruments as in Angrist and Imbens (1995) and Angrist, Graddy and Imbens (1995). Moreover, the generalization to cases with covariates is, in principle, immediate by applying our results at distinct values of the covariates. Also, fully principled methods of estimation using likelihood-based or Bayesian techniques can be derived as in Imbens and Rubin (1994a).

In Section 2 we briefly describe the structural equation approach to causal inference in economics. In Section 3 we develop an alternative approach based on the RCM, and the approaches are contrasted in Section 4. In Section 5 we discuss how to evaluate the sensitivity of the IV estimand to two of the critical assumptions presented in Section 3. In Section 6 we apply this approach to our draft lottery example, where we formulate the critical assumptions in the RCM framework and investigate the implications of violations of these assumptions.

2. STRUCTURAL EQUATION MODELS IN ECONOMICS

Following Goldberger (1972), we define structural equation models as "stochastic models in which each equation represents a causal link, rather than a mere empirical association" (p. 979). Such models are widely used in economics, going back to work by Wright (1928, 1934), Schultz (1928), and Haavelmo (1943, 1944).

A structural equation model for the problem of inferring the effect of veteran status on a health outcome is the *dummy endogenous variable model* (see, e.g., Maddala 1983; Bowden and Turkington 1984; Heckman and Robb 1985). For person i, let Y_i be the observed health outcome, let D_i be the observed treatment (i.e., veteran status), and let Z_i be the observed draft status. A standard dummy endogenous variables model for this problem would have the form

$$Y_i = \beta_0 + \beta_1 \cdot D_i + \varepsilon_i, \tag{1}$$

$$D_i^* = \alpha_0 + \alpha_1 \cdot Z_i + \nu_i \tag{2}$$

and

$$D_i = \begin{cases} 1 & \text{if } D_i^* > 0, \\ 0 & \text{if } D_i^* \leq 0. \end{cases} \tag{3}$$

In this model β_1 represents the causal effect of D on Y. Although simple, this model is typical of the econometric approach to discrete choice (in this case, the choice to serve in the military or not). The latent index formulation involving D_i^* originates in the notion that compliance is a choice determined by comparison of the expected utility of serving and not serving. We note that this dummy endogenous variables model shares many features with the classical simultaneous equations model (Haavelmo 1943): an underlying linear structure, constant coefficients, and a reliance on error terms to characterize omitted variables.

The first assumption typically invoked to identify β_1 is that Z_i is uncorrelated with the disturbances ε_i and ν_i:

$$E[Z_i \cdot \varepsilon_i] = 0, \qquad E[Z_i \cdot \nu_i] = 0. \tag{4}$$

The assumption that the correlation between ε and Z_i is zero and the absence of Z in Equation (1) captures the notion that any effect of Z on Y must be through an effect of Z on D. This is a key assumption in econometric applications of instrumental variables. A second assumption is that the covariance between the treatment D_i and assignment Z_i differs from zero; that is,

$$\text{cov}(D_i, Z_i) \neq 0, \tag{5}$$

which can be interpreted as requiring that α_1 differ from zero. If Z_i satisfies these two assumptions, then it is considered an IV in this model. In general D_i, the *endogenous regressor* in econometric terminology, is potentially correlated with ε_i because the two disturbances ε_i and ν_i are potentially correlated. This implies that the receipt of treatment D_i is not *ignorable* (Rubin 1978) and, in econometric terminology, not *exogenous.*

For this simple example, the IV estimator is defined as the ratio of sample covariances (Durbin 1954)

$$\hat{\beta}_1^{\text{IV}} = \widehat{\text{cov}}(Y_i, Z_i)/\widehat{\text{cov}}(D_i, Z_i)$$

$$= \frac{\sum_{i=1}^N Y_i Z_i / \sum_{i=1}^N Z_i - \sum_{i=1}^N Y_i(1-Z_i)/\sum_{i=1}^N (1-Z_i)}{\sum_{i=1}^N D_i Z_i / \sum_{i=1}^N Z_i - \sum_{i=1}^N D_i(1-Z_i)/\sum_{i=1}^N (1-Z_i)}, \tag{6}$$

where the last equality follows from the binary nature of the instrument.

Structural equation models such as Equations (1)–(3) have not found widespread use among statisticians. One reason is the sensitivity of these models to critical assumptions (see Little 1985) and their apparent inability to reproduce experimental results (see Lalonde 1986). Another reason is the fact that critical assumptions are cast in terms of disturbances from incompletely specified regression functions (i.e., ε_i and ν_i), rather than in terms of intrinsically meaningful and potentially observable variables. Typically the researcher does not have a firm idea what these disturbances really represent, and therefore it is difficult to draw realistic conclusions or communicate results based on their properties. The focus of this article is on the causal interpretation of the limit of the estimator in Equation (6); that is, the *IV estimand,* using the potential outcomes framework, and on the formulation of the critical assumptions in a more transparent manner to make these models more accessible to statisticians.

3. CAUSAL ESTIMANDS WITH INSTRUMENTAL VARIABLES

In this section, we set out an alternative framework for a causal interpretation of the IV estimand based on potential outcomes. First, we discuss the RCM approach to analyzing the causal effects of assignment on treatment received and on the outcome of interest (the intention-to-treat effects). We then define the causal effect of interest, that of treatment received on the outcome, in terms of potential outcomes. Finally, we show how the IV estimand links the two average intention-to-treat effects to a subpopulation average of the causal effect of interest.

3.1 The Rubin Causal Model

As before, $Z_i = 1$ implies that person i has a low lottery number (i.e., would potentially get called to serve in the military), whereas $Z_i = 0$ indicates that person i has a high lottery number (i.e., would not get called to serve in the military). The subsequent notation for D and Y is somewhat different from that in Section 2 because of the need to represent potential outcomes. Let $\mathbf{Z}$ be the N-dimensional vector of assignments with ith element Z_i, and let $D_i(\mathbf{Z})$ be an indicator for whether person i would serve given the randomly allocated vector of draft assignments $\mathbf{Z}$. In a world of perfect compliance with the draft, $D_i(\mathbf{Z})$ would equal Z_i for all i; that is, those with low lottery numbers would actually serve and none of those with high lottery numbers would serve. In practice, $D_i(\mathbf{Z})$ can differ from Z_i for various reasons: individuals may volunteer for military service, they may avoid the draft, or they may be deferred for medical or family reasons.

Similar to the definition of $D_i(\mathbf{Z})$, we define $Y_i(\mathbf{Z}, \mathbf{D})$ to be the response for person i given the vector of service indicators $\mathbf{D}$ and the vector of draft priorities $\mathbf{Z}$; $\mathbf{Y}(\mathbf{Z}, \mathbf{D})$ is the N vector with ith element $Y_i(\mathbf{Z}, \mathbf{D})$. We refer to $D_i(\mathbf{Z})$ and $Y_i(\mathbf{Z}, \mathbf{D})$ as "potential outcomes." The concept of potential outcomes used here can be viewed as analogous to Neyman's (1923) notion of "potential yields" in randomized agricultural experiments, as extended by Rubin (1974, 1978, 1990, 1991) to observational studies where the potential outcomes are partially revealed by a general treatment assignment mechanism, to situations with possible variation of treatments and with possible interference between units, and to Bayesian and likelihood inference where the potential outcomes and assignment have a joint probability distribution. As originally formulated, the potential outcomes $D_i(\mathbf{Z})$ and $Y_i(\mathbf{Z}, \mathbf{D})$ are fixed but unknown values partially observed through the assignment of treatments to units. Differences in these potential outcomes due to assigned and received treatments will be revealed by analyzing data obtained by randomly assigning $\mathbf{Z}$ in the finite population of N units under study. Our initial goal is to provide inferences solely about this finite population.

In evaluation research, some assumptions about how units interact and the variety of possible treatments are required. Our notation has already restricted both Z and D to have only two levels; that is, there is no partial compliance. Here we follow the convention in statistics and medical research by assuming no interference between units.

Assumption 1: Stable Unit Treatment Value Assumption (SUTVA) (Rubin 1978, 1980, 1990).

a. If $Z_i = Z_i'$, then $D_i(\mathbf{Z}) = D_i(\mathbf{Z}')$.
b. If $Z_i = Z_i'$ and $D_i = D_i'$, then $Y_i(\mathbf{Z}, \mathbf{D}) = Y_i(\mathbf{Z}', \mathbf{D}')$.

SUTVA implies that potential outcomes for each person i are unrelated to the treatment status of other individuals. This assumption allows us to write $Y_i(\mathbf{Z}, \mathbf{D})$ and $D_i(\mathbf{Z})$ as $Y_i(Z_i, D_i)$ and $D_i(Z_i)$ respectively. SUTVA is an important limitation, and situations where this assumption is not plausible cannot be analyzed using the simple techniques outlined here, although generalizations of these techniques can be formulated with SUTVA replaced by other assumptions.

Given the set of potential outcomes, we can define the causal effects of Z on D and on Y in the standard fashion (Rubin, 1974).

Definition 1: Causal Effects of Z on D and Z on Y.
The causal effect for individual i of Z on D is $D_i(1) - D_i(0)$. The causal effect of Z on Y is $Y_i(1, D_i(1)) - Y_i(0, D_i(0))$.

In the context of a clinical trial with imperfect compliance these are the intention-to-treat effects, and we adopt this jargon here.

Although Bayesian or likelihood-based inference is straightforward if treatment assignment is ignorable, even if not completely random (Rubin 1978), we assume random assignment here to avoid tangential issues.

Assumption 2: Random Assignment.
The treatment assignment Z_i is random:

$$\Pr(\mathbf{Z} = \mathbf{c}) = \Pr(\mathbf{Z} = \mathbf{c}')$$

for all $\mathbf{c}$ and $\mathbf{c}'$ such that $\iota^T\mathbf{c} = \iota^T\mathbf{c}'$, where ι is the N-dimensional column vector with all elements equal to one.

Given SUTVA and random assignment, unbiased estimators for the average intention-to-treat effects can be obtained by taking the difference of sample averages of Y and D classified by the value of Z; that is, by treatment-control mean differences. This has been well known since at least Neyman (1923). Formally, the unbiased estimator for the average causal effect of Z on Y can be written as

$$\frac{\sum_i Y_i Z_i}{\sum_i Z_i} - \frac{\sum_i Y_i(1-Z_i)}{\sum_i (1-Z_i)} = \frac{(1/N)\sum_{i=1}^N Y_i Z_i - (1/N)\sum_{i=1}^N Y_i \cdot (1/N)\sum_{i=1}^N Z_i}{(1/N)\sum_{i=1}^N Z_i Z_i - (1/N)\sum_{i=1}^N Z_i \cdot (1/N)\sum_{i=1}^N Z_i}, \tag{7}$$

and for the average causal effect of Z on D the unbiased estimator can be written as

$$\frac{\sum_i D_i Z_i}{\sum_i Z_i} - \frac{\sum_i D_i(1-Z_i)}{\sum_i (1-Z_i)} = \frac{(1/N)\sum_{i=1}^N D_i Z_i - (1/N)\sum_{i=1}^N Y_i \cdot (1/N)\sum_{i=1}^N Z_i}{(1/N)\sum_{i=1}^N Z_i Z_i - (1/N)\sum_{i=1}^N Z_i \cdot (1/N)\sum_{i=1}^N Z_i}. \tag{8}$$

The ratio of (7) and (8) equals the conventional instrumental variables estimator (6). The limit of the IV estimator (i.e., the IV estimand), therefore equals the ratio of average intention-to-treat effects.

3.2 Instrumental Variables

The critical feature of the problem of evaluating a treatment under imperfect compliance is that even if assignment Z_i is random or ignorable, the actual receipt of treatment D_i is typically nonignorable. Therefore the difference of outcome averages by treatment received does not provide an unbiased or even consistent estimate of the average causal effect of D on Y. In fact, we require additional assumptions just to define the causal effect of D on Y in a meaningful way. The following assumption requires the treatment assignment to be unrelated to potential outcomes once treatment received is taken into account.

Assumption 3: Exclusion Restriction.
$\mathbf{Y}(\mathbf{Z},\mathbf{D}) = \mathbf{Y}(\mathbf{Z}',\mathbf{D})$ for all $\mathbf{Z},\mathbf{Z}'$ and for all $\mathbf{D}$.

This assumption implies that $Y_i(1,d) = Y_i(0,d)$ for $d = 0,1$. It captures the notion underlying instrumental variables procedures that any effect of Z on Y must be via an effect of Z on D. Because the exclusion restriction relates quantities that can never be jointly observed, (i.e., $Y_i(0,d)$ and $Y_i(1,d)$), it is not directly verifiable from the data at hand although it has testable implications when combined with Assumptions 1 and 2. Imbens and Rubin (1994b) discussed a weaker version of the exclusion restriction that impose restrictions only on outcomes that can potentially be observed (i.e., $Y_i(z, D_i(z))$.)

By virtue of Assumption 3, we can now define potential outcomes $\mathbf{Y}(\mathbf{Z},\mathbf{D})$ as a function of $\mathbf{D}$ alone:

$$\mathbf{Y}(\mathbf{D}) = \mathbf{Y}(\mathbf{Z},\mathbf{D}) = \mathbf{Y}(\mathbf{Z}',\mathbf{D}) \quad \forall\, \mathbf{Z},\mathbf{Z}' \quad \text{and} \quad \forall\, \mathbf{D},$$

and then by Assumption 1 we can write $Y_i(D_i)$ instead of $Y_i(\mathbf{Z},\mathbf{D})$.

We now have notation for the causal effects of interest.

Definition 2: Causal Effects of D on Y.
The causal effect of D on Y for person i is $Y_i(1) - Y_i(0)$.

Although we can never observe any of these causal effects, for people with $D_i(0) \neq D_i(1)$ we can observe either one of its terms through appropriate choice of Z_i. We therefore focus on average causal effects in groups of people who can be induced to change treatments. Inferences about such average causal effects are made using changes in treatment status induced by treatment assignment, provided the assignment does affect the treatment.

At this point we introduce a compact notation to denote averages over the entire population or subpopulations. Let $E[g]$ denote the average over the population of N units of any function $g(\cdot)$ of $Z_i, D_i(1), D_i(0), Y_i(0,0), Y_i(0,1), Y_i(1,0)$, or $Y_i(1,1)$. Similarly, the average of $g(\cdot)$ over the subpopulation defined by some fixed value h_0 of some function $h(\cdot)$ will be denoted by $E[g|h(\cdot) = h_0]$. Finally, the relative size of the subpopulation satisfying $h(\cdot) = h_0$ is written as $P[h(\cdot) = h_0] = E[1_{h(\cdot)=h_0}]$, where $1_{\{\cdot\}}$ is the indicator function. We emphasize that this notation simply reflects averages and frequencies in a finite population or subpopulation.

The next assumption requires Z to have some effect on the average probability of treatment.

Assumption 4: Nonzero Average Causal Effect of Z on D.
The average causal effect of Z on $D, E[D_i(1) - D_i(0)]$ is not equal to zero.

The final assumption that we make, originally formulated by Imbens and Angrist (1994), says that there is no one who does the opposite of his assignment, no matter what the assignment.

Assumption 5: Monotonicity (Imbens and Angrist 1994).
$D_i(1) \geq D_i(0)$ for all $i = 1,\ldots,N$.

We refer to the combination of Assumptions 4 and 5, implying that $D_i(1) \geq D_i(0)$ with inequality for at least one unit as strong monotonicity.

Assumptions 1–5 lead to our formal definition of an instrument in the RCM.

Definition 3: Instrumental Variable for the Causal Effect of D on Y.

A variable Z is an instrumental variable for the causal effect of D on Y if: its average effect on D is nonzero, it satisfies

Table 1. Causal Effect of Z on Y, $Y_i(1, D_i(1)) - Y_i(0, D_i(0))$, for the Population of Units Classified by $D_i(0)$ and $D_i(1)$

		$D_i(0)$	
		0	1
$D_i(1)$	0	$Y_i(1, 0) - Y_i(0, 0) = 0$ Never-taker	$Y_i(1, 0) - Y_i(0, 1) = -(Y_i(1) - Y_i(0))$ Defier
	1	$Y_i(1, 1) - Y_i(0, 0) = Y_i(1) - Y_i(0)$ Complier	$Y_i(1, 1) - Y_i(0, 1) = 0$ Always-taker

the exclusion restriction and the monotonicity assumption, it is randomly (or ignorably) assigned, and SUTVA holds (i.e., if Assumptions 1–5 hold).

3.3 Interpreting the Instrumental Variables Estimand

SUTVA and the exclusion restriction are sufficient to establish a fundamental relationship between the intention-to-treat effects of Z on Y and D and the causal effect of D on Y at the unit level:

$$\begin{aligned} & Y_i(1, D_i(1)) - Y_i(0, D_i(0)) \\ &= Y_i(D_i(1)) - Y_i(D_i(0)) \\ &= [Y_i(1) \cdot D_i(1) + Y_i(0) \cdot (1 - D_i(1))] \\ &\quad - [Y_i(1) \cdot D_i(0) + Y_i(0) \cdot (1 - D_i(0))] \\ &= (Y_i(1) - Y_i(0)) \cdot (D_i(1) - D_i(0)). \end{aligned} \tag{9}$$

Thus the causal effect of Z on Y for person i is the product of (i) the causal effect of D on Y and (ii) the causal effect of Z on D. We can therefore write the average causal effect of Z on Y as the weighted sum of average causal effects for two subpopulations, both with $D_i(0) \neq D_i(1)$:

$$\begin{aligned} & E[Y_i(1, D_i(1)) - Y_i(0, D_i(0))] \\ &= E[(Y_i(1) - Y_i(0))(D_i(1) - D_i(0))] \\ &= E[(Y_i(1) - Y_i(0))|D_i(1) - D_i(0) = 1] \\ &\qquad \cdot P[D_i(1) - D_i(0) = 1] \\ &\quad - E[(Y_i(1) - Y_i(0))|D_i(1) - D_i(0) = -1] \\ &\qquad \cdot P[D_i(1) - D_i(0) = -1]. \end{aligned} \tag{10}$$

The weights do not sum to 1 but rather to $P[D_i(0) \neq D_i(1)]$.

Equation (10) does not use monotonicity. The monotonicity assumption requires that $D_i(1) - D_i(0)$ equals either zero or one, so that the average causal effect of Z on Y equals the product of the average causal effect of D on Y for persons with $D_i(0) = 0$ and $D_i(1) = 1$ and their proportion in the population:

$$\begin{aligned} & E[Y_i(D_i(1), 1) - Y_i(D_i(0), 0)] \\ &= E[(Y_i(1) - Y_i(0))|D_i(1) - D_i(0) = 1] \\ &\qquad \cdot P[D_i(1) - D_i(0) = 1]. \end{aligned} \tag{11}$$

This establishes the relationship between the IV estimand and the causal effect of D on Y, which we summarize as a formal proposition.

Proposition 1: Causal Interpretation of the IV Estimand. Given Assumptions 1, 3, 4, and 5, the instrumental variables estimand is

$$\begin{aligned} & \frac{E[Y_i(D_i(1), 1) - Y_i(D_i(0), 0)]}{E[D_i(1) - D_i(0)]} \\ &\qquad = E[(Y_i(1) - Y_i(0))|D_i(1) - D_i(0) = 1]. \end{aligned} \tag{12}$$

We call this the Local Average Treatment Effect (LATE). This result follows directly from (11) combined with two facts: first, that the monotonicity assumption implies that $E[D_i(1) - D_i(0)]$ equals $P[D_i(1) - D_i(0) = 1]$, and second, that $E[D_i(1) - D_i(0]$ differs from zero.

Table 1 helps interpret this result. The four values of $(D_i(0), D_i(1))$ in this two-by-two table generate three distinct values of $D_i(1) - D_i(0)$. Individuals with $D_i(1) - D_i(0) = 1$ (bottom left) are induced to take the treatment by assignment to the treatment, and the causal effect of Z on Y is $Y_i(1) - Y_i(0)$ for individuals of this type, whom we refer to as *compliers.* A value of $D_i(1) - D_i(0) = 0$ (diagonal elements) implies that individual i does not change treatment status with the assigned treatment; the causal effect of Z on Y is zero for such individuals by the exclusion restriction. If $D_i(0) = D_i(1) = 0$, the individual is referred to as a *never-taker,* or in our application, a draft avoider; whereas if $D_i(0) = D_i(1) = 1$, the individual is an *always-taker,* or, in our application, a volunteer. Finally, individuals with $D_i(1) - D_i(0) = -1$ (top right) do the opposite of their assignment; they are induced to avoid the treatment by assignment to it, and induced to take the treatment by assignment to the control group. We call such individuals *defiers,* as suggested by Balke and Pearl (1993) in a comment on an earlier version of this paper (Angrist, Imbens, and Rubin 1993). The causal effect of Z on Y for these individuals is $Y_i(0) - Y_i(1)$. Finally, we refer to never-takers, always-takers, and defiers jointly as *noncompliers.* Note that these labels—compliers, defiers, never-takers, always-takers, and noncompliers—are simply definitions given SUTVA in this experiment and are not assumptions about individual behavior.

By virtue of the exclusion restriction, the two subpopulations corresponding to the two diagonal elements of Table 1 are characterized by a zero causal effect of Z on Y. By virtue of the monotonicity assumption there are no defiers,

and the group corresponding to the top-right element in the table is empty. Finally, by virtue of Assumption 4, the proportion of the population in the cell corresponding to compliers differs from zero and is equal to the average causal effect of Z on D. Combined, these assumptions imply that the average causal effect of Z on Y is proportional to the average causal effect of D on Y for compliers. This is the result in Proposition 1.

Because we can estimate the two intention-to-treat estimands by virtue of random assignment, we can also estimate their ratio; that is, the IV estimand. The ratio of the usual unbiased estimators for the intention-to-treat estimands given in (7) and (8) is equal to the standard instrumental variables estimator for binary instruments given in (6). This estimator does not exploit all the implications of the model developed in this section. In Imbens and Rubin (1994a,b) we discuss implications of this model for estimation.

Finally, it is important to note that (under our assumptions) we cannot generally identify the specific members of the group of compliers, defined by $D_i(0) = 0, D_i(1) = 1$, for whom we can identify the average treatment effect. Thus, the local average treatment effect (i.e., the average causal effect for compliers) is not the average treatment effect for either the entire population or for a subpopulation identifiable from observed values. Stronger assumptions are needed for the identification of average causal effects for subpopulations identifiable from observed data. One assumption that achieves this is random assignment to a control group denied treatment, so that $D_i(0) = 0$ for all i (Zelen 1979; Angrist and Imbens 1991). For examples of other such assumptions see Heckman (1990), Robins and Tsiatis (1991), Efron and Feldman (1991), and Manski (1994).

4. COMPARING THE STRUCTURAL EQUATION AND POTENTIAL OUTCOMES FRAMEWORK

In Section 2 we described a structural equation model for the effect of military service on a health outcome using an indicator of draft eligibility as an instrument. Here we contrast that framework with the approach developed for the same problem in Section 3. In particular, we compare the formulation and clarity of the assumptions in each case. This comparison is useful because several authors have attributed the absence of structural equation methods in statistics to the manner in which such models are commonly formulated. For example, in his discussion of the connection between structural equation methods and path analysis, Goldberger (1972) quoted Moran (1961): "The main reason why Sewall Wright's method of path coefficients is often found difficult to understand is that expositions of the theory do not make clear what assumptions are made" (p. 988). Similarly, Holland (1988) writes, "it is not always evident how to verify assumptions made about [regression disturbances]. For example, why should [they] be independent of $[Z]$ ... when the very definition of [the disturbances] involves $[Z]$" (p. 458).

4.1 The Exclusion Restriction and Ignorable Treatment Assignment

The econometric version of these assumptions requires that the disturbances in the response equation (1) and the participation equation (2) be uncorrelated with, or independent of, the assignment Z. In Imbens and Angrist (1994) this assumption is formulated in a framework using potential outcomes indexed only against the level of the treatment D. The framework we develop here separates this requirement into two assumptions about potentially observable quantities: the exclusion restriction, which says nothing about the treatment assignment mechanism, and ignorable treatment assignment, which says nothing about possible direct effects of assignment.

First, the exclusion restriction requires that the instrument have no effect on the outcome except through D. Thus to verify this assumption, the researcher must consider, at the unit level, the effect of changing the value of the instrument while holding the value of the treatment fixed. To clarify the distinction between this formulation and the econometric formulation, consider the four subpopulations defined by the values of $D_i(0)$ and $D_i(1)$ in Table 1. Someone with $D_i(0) = D_i(1) = 1$ would always serve in the military with a low or high draft lottery number. It seems reasonable to assume that for such a person, the draft lottery number has no effect on health outcome. Next, consider someone with $D_i(0) = D_i(1) = 0$, who would have managed to avoid military service with a high or low lottery number. For someone exempted from military service for medical reasons, it seems plausible that there was no effect of the draft lottery number. But a draftee who managed to avoid military service by staying in school or moving abroad could experience an effect of Z on future life outcomes that would violate the exclusion restriction. For both these groups of noncompliers, the exclusion restriction requires the researcher to consider a difference in outcomes that were potentially observable, even though after the population was randomly allocated to treatment and control groups, only one of the outcomes was actually observed. In fact, if one could identify compliers and noncompliers, then it would be possible to test the exclusion restriction by comparing average outcomes for noncompliers by assignment status.

For compliers with $D_i(0) = 0, D_i(1) = 1$, the exclusion restriction compares outcomes that cannot be observed: it requires that $Y_i(0, D_i(0)) = Y_i(1, D_i(0))$ and $Y_i(1, D_i(1)) = Y_i(0, D_i(1))$. For this group, the exclusion restriction amounts to attributing the effect of Z on Y to the change in the treatment received D rather than to the change in assignment Z. Such an assumption is not innocuous, and efforts to ensure it form the rationale for blinding, double blinding, and using placebos in clinical trials. Nevertheless, it underlies most experimental evaluations in economics where blinding and placebos are impossible, and is often thought to be reasonable in those cases.

The second element embedded in the assumption of zero correlation between instruments and disturbances in the standard econometric formulation is that of random, or at

least ignorable, treatment assignment Z. This assumption is trivially satisfied if physical randomization took place, as in the application in Section 6 where Z is a function of a lottery number. Our formulation makes clear that randomization of the instrument, though sufficient to allow unbiased estimation of the average treatment effect of Z on Y and of the average treatment effect of Z on D, does not imply that the IV estimand is interpretable as an average causal effect. In most applications of IV, however, the instrument is not randomly assigned, and this assumption must be argued more carefully. Examples include Angrist and Krueger's (1991) use of quarter of birth as an instrument for the effect of schooling on earnings, Card's (1993) use of distance to college as an instrument for the effect of schooling on earnings, and McClellan and Newhouse's (1994) use of relative distance to hospital as an instrument for the effect of catherization on mortality after acute myocardial infarction.

Whereas the exclusion restriction requires the researcher to contemplate the effect of specific treatments on outcomes, the ignorability assumption requires consideration of the assignment mechanism. Violations of these different assumptions can have different sources and consequences. In our view, pooling these assumptions into the single assumption of zero correlation between instruments and disturbances has led to confusion about the essence of the identifying assumptions and hinders assessment and communication of the plausibility of the underlying model.

4.2 The Monotonicity Condition

The monotonicity assumption rules out the existence of defiers, characterized by $D_i(0) = 1$ and $D_i(1) = 0$. Permutt and Hebel (1989) informally discussed a variant of this assumption in a reanalysis of a program designed to induce pregnant women to stop smoking. In that context, the assumption implies that everyone who would stop smoking if they were in the control group, which received no encouragement to stop smoking, would also stop smoking if encouraged to do so by being in the treatment group. Robins (1989) discussed the effect of this assumption on bounds on population average treatment effects. Monotonicity is implied by designs where those assigned to the control group are prevented from receiving the treatment, as in Zelen's (1979) single-consent designs.

Monotonicity has no explicit counterpart in the econometric formulation, but is implicit in the use of an equation with constant parameters for the relation between Z_i and D_i. The model developed in Section 3 suggests that the constant parameter assumption embodied in (2) is much stronger than needed. On the other hand, it is not sufficient to postulate a nonzero covariance between treatment and assignment, as in (5), for the interpretation of the IV estimand as an average of causal effects.

4.3 Reduced Form and Structural Parameters

Reduced-form parameters for the draft lottery application are the coefficients from a regression of Y on Z and D on Z. In our formulation, these are the average intention-to-treat effects under Assumptions 1 and 2. The structural parameter (β_1) is the average effect of the treatment itself on Y for the subpopulation that complies with assignment. The econometric approach does not distinguish between an effect for the entire population and an effect for the subpopulation of compliers. In our view LATE is structural in the Goldberger (1972) sense of representing a causal link, but not necessarily structural in the sense of representing a parameter that is invariant across populations. Despite this potential lack of generalizability, we view LATE as interesting (perhaps in combination with the intention-to-treat estimand) because it is an average of unit level causal effects of the treatment of interest. For example, for a potential recruit, the average effect of actual military service for a specific subpopulation is likely to be of greater interest than the population average effect of draft eligibility.

A similar rationale applies to clinical trials, which are often based on populations that are more homogeneous than, and not representative of, the population that will eventually be subjected to the treatment. The presumption in such cases, and in our analysis, is the average over the subpopulation of those whose behavior can be modified by assignment are likely to be informative about population averages of those who comply in the future, even if there is substantial heterogeneity in individual-level causal effects.

It should be stressed, however, that the assumptions needed for a causal interpretation of the instrumental variables estimand (Assumptions 1 and 3–5) are substantially stronger than those needed for the causal interpretation of the intention-to-treat estimand (Assumption 1). The plausibility of the additional assumptions (i.e., the exclusion restriction and the monotonicity assumption) must be taken into account when facing the choice to report estimates of the intention-to-treat estimands, of the IV estimands, or both.

5. SENSITIVITY OF THE IV ESTIMAND TO CRITICAL ASSUMPTIONS

The assumptions laid out in Section 3 are sufficient conditions for the identification of a meaningful average treatment effect. In this section we discuss the sensitivity of the IV estimand to deviations from the IV assumptions. As this discussion makes clear, violations of these assumptions need not be catastrophic. We focus on Assumptions 3 and 5 because they form the core of the IV approach. Assumption 4 (a nonzero average causal effect of Z on D) is conceptually straightforward and easy to check. Assumptions 1 and 2 are standard in the RCM approach, and sensitivity to particular violations of those assumptions has been previously discussed (e.g., Rosenbaum and Rubin 1983). In general the IV estimand is most likely to be sensitive to violations of the exclusion restriction and the monotonicity assumption when there are few compliers. In Section 6 we illustrate how this sensitivity analysis can be applied.

5.1 Violations of the Exclusion Restriction

First, we consider violations of the exclusion restriction, while maintaining the other assumptions, stability, and

strong monotonicity. If subject i is a noncomplier, that is, $D_i(0) = D_i(1)$, then the causal effect of Z on Y is

$$H_i = Y_i(1, d) - Y_i(0, d), \tag{13}$$

where $d = 0$ if subject i is a never-taker and $d = 1$ if subject i is an always-taker. Under the exclusion restriction, $H_i = 0$ for all noncompliers.

Proposition 2. Given stability and strong monotonicity, but without the exclusion restriction for noncompliers, the IV estimand equals the Local Average Treatment Effect plus a bias term given by (14):

$$\begin{aligned}&\frac{E[Y_i(1, D_i(1)) - Y_i(0, D_i(0))]}{E[D_i(1) - D_i(0)]} \\ &\quad - E[Y_i(1, D_i(1)) - Y_i(0, D_i(0))|i \text{ is a complier}] \\ &= E[H_i|i \text{ is a noncomplier}] \cdot \frac{P[i \text{ is a noncomplier}]}{P[i \text{ is a complier}]}.\end{aligned} \tag{14}$$

The bias of the IV estimand relative to the Local Average Treatment Effect equals the average direct effect of Z on Y for noncompliers multiplied by the odds of being a noncomplier.

When there is a direct effect of assignment on the outcome for noncompliers, it is plausible that there is also a direct effect of assignment on outcome for compliers. Suppose that for each complier, assignment and treatment had additive effects on the outcome Y; that is,

$$Y_i(1, 0) - Y_i(0, 0) = Y_i(1, 1) - Y_i(0, 1),$$

for all compliers. Additivity for compliers allows us to define the causal effect of Z on Y for compliers as $H_i = Y_i(1, d) - Y_i(0, d)$ for $d = 0, 1$ [analogous to the definition for noncompliers given in (13)] and the causal effect of D on Y as $G_i = Y_i(z, 1) - Y_i(z, 0)$. We can then write the IV estimand as

$$\begin{aligned}&\frac{E[Y_i(1, D_i(1)) - Y_i(0, D_i(0))]}{E[D_i(1) - D_i(0)]} \\ &= E[G_i|i \text{ is a complier}] \\ &\quad + \frac{E[H_i]}{P[i \text{ is a complier}]}.\end{aligned} \tag{15}$$

The bias relative to the average causal effect of D on Y for compliers, the second term in (15), can also be written as

$$\begin{aligned}&E[H_i|i \text{ is a complier}] + E[H_i|i \text{ is a noncomplier}] \\ &\qquad \cdot \frac{P[i \text{ is a noncomplier}]}{P[i \text{ is a complier}]}.\end{aligned} \tag{16}$$

The first term in the bias in (16) has nothing to do with noncompliance, but is the bias due to the direct effect of assignment for those who take the treatment. If compliance were perfect, the second term would be zero but the first term of the bias would still be present. The increased bias in the IV estimand due to noncompliance is directly proportional to the product of the average size of the direct effect of Z for noncompliers and the odds of noncompliance given monotonicity. The higher the correlation between the instrument and the treatment status (i.e., the "stronger" the instrument), the smaller the odds of noncompliance, and consequently the less sensitive the IV estimand is to violations of the exclusion assumption.

5.2 Violations of the Monotonicity Condition

Next we consider violations of the monotonicity assumption. Because we maintain the exclusion restriction, the causal effect of D on Y for person i with $D_i(1) \neq D_i(0)$ is still uniquely defined, and equal to $Y_i(1) - Y_i(0)$.

Proposition 3. Given stability, the exclusion restriction, and a nonzero average causal effect of Z on D, but without the monotonicity assumption, the IV estimand equals the Local Average Treatment Effect plus a bias term given by (17):

$$\begin{aligned}&E[Y_i(1, D_i(1)) - Y_i(0, D_i(0))]/E[D_i(1) - D_i(0)] \\ &\quad - E[Y_i(1) - Y_i(0)|i \text{ is a complier}] \\ &= -\lambda \cdot \{E[Y_i(1) - Y_i(0)|i \text{ is a defier}] \\ &\quad - E[Y_i(1) - Y_i(0)|i \text{ is a complier}]\},\end{aligned} \tag{17}$$

where

$$\lambda = \frac{P(i \text{ is a defier})}{P(i \text{ is a complier}) - P(i \text{ is a defier})}. \tag{18}$$

The bias due to violations of monotonicity is composed of two factors. The first factor, $\lambda = P(i \text{ is a defier})/(P(i \text{ is a complier}) - P(i \text{ is a defier}))$, is related to the proportion of defiers and is equal to zero under the monotonicity assumption. The smaller the proportion of defiers, the smaller the bias will be from violations of the monotonicity assumption. However, because the denominator of this factor is the average causal effect of Z on D, the bias can be large even if there are few defiers, as long as the average causal effect of Z on D is small. Note again that the stronger the instrument, the less sensitive the IV estimand is to violations of the monotonicity assumption. The second factor is the difference in average causal effects of D on Y for the compliers and defiers. If the average causal effects of D on Y are identical for defiers and compliers, violations of the monotonicity assumption generate no bias. The less variation there is in the causal effect of D on Y, the smaller the bias from violations of the monotonicity assumption.

Without monotonicity, the IV estimand can also be written as

$$\begin{aligned}&(1 + \lambda) \cdot E[Y_i(1) - Y_i(0)|i \text{ is a complier}] \\ &\qquad - \lambda \cdot E[Y_i(1) - Y_i(0)|i \text{ is a defier}],\end{aligned}$$

with λ as defined in (18). In this representation, the estimand is still a weighted average of average treatment effects despite the violation of the monotonicity assumption, but the weights are always outside the unit interval because $\lambda > 0$.

6. AN APPLICATION: THE EFFECT OF MILITARY SERVICE ON CIVILIAN MORTALITY

Hearst, Newman, and Hulley (1986) showed that men with low lottery numbers in the Vietnam Era draft lottery (i.e., men with $Z_i = 1$) had elevated mortality after their discharge from the military. The authors attribute this elevated mortality to the detrimental effect of serving in the military during wartime on well-being. Similarly, Angrist (1990) attributed differences in subsequent earnings by lottery number to the effect of serving in the military on earnings. These conclusions are primarily based on the fact that between 1970 and 1973, priority for the draft was randomly assigned in a lottery using dates of birth. Each date of birth in the cohorts at risk of being drafted was assigned a *random sequence number* (RSN) from 1–365. The Selective Service called men for induction by RSN up to a ceiling determined by the defense department. Men born in 1950 were potentially drafted up to RSN 195 in 1970, men born in 1951 were potentially drafted up to RSN 125 in 1971, and men born in 1952 were potentially drafted up to RSN 95 in 1972. We refer the reader to Hearst et al. (1986) and Angrist (1990) for further details on these data and the draft.

In their paper, Hearst et al. focused on the difference in mortality risk by draft status. For example, they compare the number of deaths of men born in 1950 with RSN below 195 to the number of deaths of men born in 1950 with RSN above 195. Our purpose in returning to this example is twofold. First, we discuss the validity of Assumptions 1–5 in this context. Second, we show how the sensitivity of the estimated average treatment effect to violations of the exclusion restriction and the monotonicity assumption can be explored using the results from the previous section.

6.1 Assessment of Assumptions 1–5

The potential outcome in this example, $Y_i(z, d)$, is an indicator variable equal to one if person i would have died between 1974 and 1983 given lottery assignment z and military service indicator d. To distinguish this from mortality during the war period, we refer to Y_i as civilian mortality. For simplicity, we ignore the effect that mortality during the war might have on the size of the population at risk.

For a valid causal interpretation of the IV estimand, we require:

- SUTVA, Assumption 1: The veteran status of any man at risk of being drafted in the lottery was not affected by the draft status of others at risk of being drafted, and, similarly, that the civilian mortality of any such man was not affected by the draft status of others;
- Ignorable Assignment, Assumption 2: Assignment of draft status was random;
- Exclusion restriction, Assumption 3: Civilian mortality risk was not affected by draft status once veteran status is taken into account;
- Nonzero Average Causal Effect of Z on D, Assumption 4: Having a low lottery number increases the average probability of service;
- Monotonicity assumption, Assumption 5: There is no one who would have served if given a high lottery number, but not if given a low lottery number.

Although we believe these assumptions are plausible, a case can be made for violations of most. For example, it has been argued that the fraction of a cohort that served in the military affects the civilian labor market response to veterans (De Tray 1982). If this assertion is true, then the SUTVA assumption very likely does not hold. Another reason for possible violations of SUTVA is that people not drafted may be induced to serve in the military by friends who were drafted.

There is also some evidence that some men with low lottery numbers changed their educational plans so as to retain draft deferments and avoid the conscription (Angrist and Krueger 1992b). If so, then the exclusion restriction could be violated, because draft status may have affected civilian outcomes through channels other than veteran status. We return to this issue in some detail shortly.

Monotonicity would be violated if, for example, someone, who would have volunteered for the Navy when not at risk of being drafted because of a high lottery number, would have chosen to avoid military service altogether when at risk of being drafted because of a low lottery number. It seems unlikely that there were many in the population in this category.

It is clear that the Assumption 4 is satisfied because the likelihood of serving in the military sharply increases with draft status.

Another uncontroversial assumption is the ignorability of treatment assignment, which allows simple unbiased estimation of the average causal effects of Z on D and of Z on Y. Although there is some evidence that the first lottery, which was executed using a poorly designed physical randomization, was not actually random (Fienberg 1971) it nevertheless is almost certainly ignorable. Ignoring this complication and postponing consideration of the possible problems with the exclusion restriction and the monotonicity condition, we forge ahead with the IV approach.

6.2 The Instrumental Variables Estimates

Table 2 presents data and some estimates of the effects of military service on civilian mortality for white men born in 1950 and 1951 by year of birth and draft status. Column 3 shows the number of deaths in both Pennsylvania and California between 1974–1983. Columns 5 and 6 show the average number of civilian deaths and suicides respectively per 1,000, computed as the number of deaths divided by the population at risk estimated using the 1970 census. Column 7 shows the frequency of veteran status, estimated from the 1984 Survey of Income and Program Participation (SIPP). In columns 5–7, the entries in the third pair of rows give the difference in probability of death, suicide, and veteran status between those with low and high lottery numbers (draft eligible or not). The fourth pair of rows in columns 5 and 6 give the ratio of these differences to the difference in the probability of being veteran by draft eligibility. These are the standard IV estimates. An alternative approach to

Table 2. Data on Civilian Mortality for White Men Born in 1950 and 1951

Year	Draft eligibility[a]	Number of deaths[b]	Number of suicides[c]	Probability of death[d]	Probability of suicide	Probability of military service[e]
1950	Yes	2,601	436	.0204	.0034	.3527
				(.0004)	(.0002)	(.0325)
	No	2,169	352	.0195	.0032	.1934
				(.0004)	(.0002)	(.0233)
Difference (Yes minus No)				.0009	.0002	.1593
				(.0006)	(.0002)	(.0401)
IV estimates[f]				.0056	.0013	
				(.0040)	(.0013)	
1951	Yes	1,494	279	.0170	.0032	.2831
				(.0004)	(.0002)	(.0390)
	No	2,823	480	.0168	.0029	.1468
				(.0003)	(.0001)	(.0180)
Difference (Yes minus No)				.0002	.0003	.1362
				(.0005)	(.0002)	(.0429)
IV estimates				.0015	.0022	
				(.0037)	(.0016)	

[a] Determined by lottery number cutoff: RSN 195 for men born in 1950, and RSN 125 for men born in 1951.

[b] From California and Pennsylvania administrative records, all deaths 1974–1983. Data sources and methods documented by Hearst et al. (1986). Note: Sample sizes differ from Hearst et al., because non-U.S.-born are included to match SIPP data in the last column.

[c] The mortality figures are tabulated from the data set analyzed by Hearst et al. (1986).

[d] The estimated population at risk is from the author's tabulation of 1970 census data. Estimates by draft-eligibility status are computed assuming a uniform distribution of lottery numbers. Standard errors are given in parentheses.

[e] These figures are taken from Angrist (1990), table 2, and were tabulated using a special version of the SIPP that has been matched to indicators of draft eligibility. Note that probabilities estimated using the SIPP are for the entire country and do not take account of morality. The impact of mortality on differences in the probability of being a veteran by eligibility status is small enough to have only trivial consequences for the estimation.

[f] The standard errors, following econometric practice (e.g. Imbens and Angrist 1994), were calculated based on a normal approximation to the sampling distribution of the ratio of the difference in estimated probability of death/suicide and the difference in estimated probability of serving. We assume independence of numerator and denominator because they were calculated from different data sets. Pooled estimates show a statistically significant increase in risk at conventional significance levels (e.g., Hearst, Newman, Hulley 1986).

estimating the local average treatment effect, which takes into account the full implications of the assumptions, is provided in Imbens and Rubin (1994b).

As a specific example, consider men born in 1950. Of the men with low lottery numbers ($Z_i = 1$), 35.3% actually served in the military. Of those who had high lottery numbers ($Z_i = 0$), only 19.3% served in the military. Random assignment of draft status suggests that draft status had a causal effect that increased the probability of serving by an estimated 15.9% on average. Similarly, of those with low lottery numbers, 2.04% died between 1974 and 1983, compared to 1.95% of those who had high lottery numbers. The difference of .09% can be interpreted as an estimate of the average causal effect of draft status on civilian mortality. Assuming that these estimated causal effects are population averages, the ratio of these two causal effects of draft status is, under the Assumptions 1–5, the causal effect of military service on civilian mortality for the 15.9% who were induced by the draft to serve in the military. For this group, the average causal effect is .56%, which amounts to approximately a 25% increase in the probability of death (given average mortality rates around 7%). These estimates highlight the fact that the IV estimator does not require observations on individuals; sample averages of outcomes and treatment indicators by values of the instruments are sufficient. In applications like the one discussed here, these moments are drawn from different data sets. (For a detailed discussion of IV estimation with moments from two data sets, see Angrist and Krueger 1992a.)

6.3 Sensitivity to the Exclusion Restriction

Suppose that the exclusion restriction is violated because men with low lottery numbers were more likely to stay in school. A schooling–lottery connection could arise because, for much of the Vietnam period, college and graduate students were exempt from the draft. Although new graduate student deferments were eliminated in 1967 and new undergraduate deferments were eliminated in December 1971, many of the men with low lottery numbers in 1970 and 1971 could have postponed conscription by staying in school. Working with special versions of the March 1979 and March 1981–1985 Current Population Surveys (CPS's), Angrist and Krueger (1992b) showed that men born in 1951 with lottery numbers 1–75 had completed .358 more years of schooling than men with lottery numbers above 150, who were not drafted.

How much bias in estimates of the effect of military service on mortality is this correlation between lottery numbers and schooling likely to generate? Addressing this question requires data on the connection between schooling and mortality. The relationship between socioeconomic variables and mortality is uncertain and the subject of considerable research in epidemiology and social science. (An early study in this area is Kitagawa and Hauser 1973.) For the purposes of illustration, we have taken estimates from Duleep's (1986) study of socioeconomic variables and mortality using men surveyed in the March 1973 CPS and linked to 1973–1978 Social Security data. Estimates presented in Table 1 of Duleep (1986) suggest that married white men 25 years old with 1–3 years of college have

mortality rates roughly .0017 per thousand higher than do men with only high school degrees.

Assume that the excess mortality among men with some college accumulates linearly, so that an additional year of schooling raises mortality by $.0017 \times (1/3) = .00056$. Men with low lottery numbers may have as much as .358 more years of schooling than men with high lottery numbers. Thus an estimate of the mortality difference attributable to the effect of draft status on schooling is $.358 \times .00056 = .00019$, essentially as large as the .0002 observed difference in mortality by draft status for white men born in 1951. Assuming additive causal effects of education and military service on mortality, the bias formula (15) applied to this example is $E[H_i]/(E[D_i(1) - D_i(0)])$, which is estimated by $.00019/.1362 = .0014$ because there is a .1362 difference in the probability being a veteran by draft eligibility status. Thus taking account of this potential bias could eliminate the estimated .0015 impact of veteran status on civilian mortality!

This calculation illustrates the cautions that should accompany the IV estimates. But the extent to which the causal interpretation of the estimates in Table 2 should be discounted in light of these findings is unclear. First, there is no evidence of a schooling–lottery number connection for the 1950 cohort, yet lottery-based estimates of the effects of service are even larger for men born in 1950 than for the 1951 cohort used in the illustration. Second, the schooling–mortality connection is not well determined [the Duleep (1986) estimate used here is not actually significantly different from zero], and this relationship is also subject to sign reversals. For example, although men with some college have higher mortality than high school–only graduates, the Duleep study showed almost no difference between the mortality of high school only graduates and college graduates. Thus, a calculation based solely on graduates would indicate no bias.

6.4 Sensitivity to the Monotonicity Assumption

Without monotonicity, the average causal effect of Z on D estimates the difference between the proportions of compliers and defiers. Table 2 therefore suggests that 15.93% more people are compliers than defiers. Suppose that 5% of the population are defiers. This would imply that about 21% of the population are compliers, and that the multiplier $P[i \text{ is a defier}]/(P[i \text{ is a complier}] - P[i \text{ is a defier}])$ could be as large as .33 rather than zero, as required by monotonicity. Next, suppose that we assume the difference between average treatment effects for compliers and defiers is at most .0041. This number was chosen because the range of IV estimates in Table 2 (.0056 for 1950 and .0015 for 1951) is equal to this amount. This implies that the estimated average treatment effect for compliers could be as small as $.0056 - .33 \times .0041 = .0042$ or as large as $.0056 + .33 \times .0041 = .0070$. To reverse the sign of the average causal effect through violations of the monotonicity assumption would therefore require the presence of an implausibly large group of defiers, or very large differences between average effects for compliers and defiers.

7. CONCLUSION

In this article we have outlined a framework for causal inference in settings where random assignment has taken place, but compliance is not perfect; that is, the treatment received is nonignorable. In an attempt to estimate the effect of receipt of treatment, rather than assignment of treatment as in intention-to-treat analysis, we make use of instrumental variables. This approach has long been used by economists in the context of regression models with constant treatment effects. We show that this technique can be fit into the Rubin Causal Model and used for causal inference without assuming constant treatment effects. The advantages of embedding this approach in the RCM are twofold. First, it makes the nature of the identifying assumptions more transparent. Second, it allows us to consider the sensitivity of results to deviations from these assumptions in a straightforward manner. We hope that the approach outlined in this article serves to make the IV approach more accessible to statisticians, while helping economists understand and interpret the strong assumptions required for a causal interpretation of IV estimates.

[Received June 1993. Revised December 1995.]

REFERENCES

Angrist, J. D. (1990), "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence From Social Security Administrative Records," *American Economic Review*, 80, 313–335.

Angrist, J. D., and Imbens, G. W. (1991), "Sources of Identifying Information in Evaluation Models," Technical Working Paper 117, National Bureau of Economic Research.

——— (1995), "Two-Stage Least Squares Estimation of Average Causal Effects in Models With Variable Treatment Intensity," *Journal of the American Statistical Association*, 90, 431–442.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1993), "Identification of Causal Effects Using Instrumental Variables," Technical Working Paper 136, National Bureau of Economic Research.

Angrist, J., Graddy, K., and Imbens, G. W. (1995), "Nonparametric Demand Analysis With an Application to the Demand for Fresh Fish," Technical Working Paper 178 NBER April '95, Massachusetts Institute of Technology.

Angrist, J. D., and Krueger, A. (1991), "Does Compulsory School Attendance Affect Schooling and Earnings?," *Quarterly Journal of Economics*, 106, 979–1014.

——— (1992a), "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables With Moments From Two Samples," *Journal of the American Statistical Association*, 87, 328–336.

——— (1992b), "Estimating the Payoff to Schooling Using the Vietnam-Era Draft Lottery," Working Paper 4067, National Bureau of Economic Research.

Balke, A., and Pearl, J. (1993), "Nonparametric Bounds on Causal Effects From Partial Compliance Data," Technical Report R-199, University of California, Los Angeles, Computer Science Dept.

Bowden, R. J., and Turkington, D. A. (1984), *Instrumental Variables*, Cambridge, U.K.: Cambridge University Press.

Card, D. (1993), "Using Geographic Variation in College Proximity to Estimate the Returns to Schooling," Working Paper 4483, National Bureau of Economic Research.

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

De Tray, D. (1982), "Veteran Status as a Screening Device," *American Economic Review*, 72, 133–142.

Duleep, H. O. (1986), "Measuring the Effect of Income on Adult Mortality Using Longitudinal Administrative Record Data," *Journal of Human Resources*, 21, 238–251.

Durbin, J. (1954), "Errors in Variables," *Review of the International Statistical Institute*, 22, 23–32.

Efron, B., and Feldman, D. (1991), "Compliance as an Explanatory Variable in Clinical Trials," *Journal of the American Statistical Association*, 86, 9–26.

Fienberg, S. (1971), "Randomization and Social Affairs: The 1970 Draft Lottery," *Science*, 171, 255–261.

Fisher, R. A. (1918), "The Causes of Human Variability," *Eugenics Review*, 10, 213–220.

——— (1925), *Statistical Methods for Research Workers*, London: Oliver & Boyd.

Goldberger, A. S. (1972), "Structural Equation Methods in the Social Sciences," *Econometrica*, 40, 979–1001.

Greenland, S., and Robins, J. M. (1986), "Identifiability, Exchangeability, and Epidemiological Confounding," *International Journal of Epidemiology*, 15, 413–419.

Haavelmo, T. (1943), "The Statistical Implications of a System of Simultaneous Equations," *Econometrica*, 11, 1–12.

——— (1944), "The Probability Approach in Econometrics," *Econometrica*, 12 (Supplement), 1–115.

Hearst, N., Newman, T., and Hulley, S. (1986), "Delayed Effects of the Military Draft on Mortality: A Randomized Natural Experiment," *New England Journal of Medicine*, 314, 620–624.

Heckman, J. (1990), "Varieties of Selection Bias," *American Economic Review*, 80, 313–318.

Heckman, J., and Robb, R. (1985), "Alternative Methods for Evaluating the Impact of Interventions," in *Longitudinal Analysis of Labor Market Data*, eds. Heckman and Singer, New York: Cambridge University Press, pp. 156–245.

Holland, P. (1986), "Statistics and Causal Inference," *Journal of the American Statistical Association*, 81, 945–970.

——— (1988), "Causal Inference, Path Analysis, and Recursive Structural Equations Models," (with discussion) in *Sociological Methodology*, ed. Washington, DC: American Sociological Association, pp. 449–493.

Imbens, G. W., and Angrist, J. (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–476.

Imbens, G. W., and Rubin, D. B. (1994a), "Bayesian Inference for Causal Effects in Randomized Experiments With Noncompliance," Harvard University, Dept. of Economics.

——— (1994b), "On the Fragility of Instrumental Variables Estimators," Working Paper 1356, Harvard Institute of Economic Research.

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

Kitagawa, E. M., and Hauser, P. M. (1973), *Differential Mortality in the United States: A Study in Socioeconomic Epidemiology*, Cambridge, MA: Harvard University Press.

Lalonde, R. (1986), "Evaluating the Econometric Evaluations of Training Programs," *American Economic Review*, 76, 604–620.

Lee, Y., Ellenberg, J., Hirtz, D., and Nelson, K. (1991), "Analysis of Clinical Trials by Treatment Actually Received: Is It Really an Option?," *Statistics in Medicine*, 10, 1595–1605.

Little, R. (1985), "A Note About Models for Selectivity Bias," *Econometrica*, 53, 1469–1474.

Maddala, G. S. (1983), *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge, U.K.: Cambridge University Press.

Manski, C. F. (1994), "The Selection Problem," in *Advances in Econometrics*, ed. C. Sims, New York: Cambridge University Press, pp. 143–170.

McClellan, M., and Newhouse, J. (1994), "Marginal Benefits of Medical Treatment Intensity: Acute Myocardial Infarction in the Elderly," mimeo, National Bureau of Economic Research.

Moran, P. (1961), "Path Coefficients Reconsidered," *Australian Journal of Statistics*, 3, 87–93.

Morgan, M. (1990), *The History of Econometric Ideas*, Cambridge, U.K.: Cambridge University Press.

Neyman, J. (1923), "On the Application of Probability Theory to Agricultural Experiments: Essay on Principles, Section 9," translated in *Statistical Science*, 5, 465–480, 1990.

Permutt, T., and Hebel, J. (1989), "Simultaneous-Equation Estimation in a Clinical Trial of the Effect of Smoking on Birth Weight," *Biometrics*, 45, 619–622.

Robins, J. M. (1989), "The Analysis of Randomized and Non-Randomized AIDS Treatment Trials Using a New Approach to Causal Inference in Longitudinal Studies," in *Health Service Research Methodology: A Focus on AIDS*, eds. by L. Sechrest, H. Freeman, and A. Bailey, Washington, DC: U.S. Public Health Service.

Robins, J. M., and Tsiatis, A. A. (1991), "Correcting for Non-Compliance in Randomized Trials Using Rank-Preserving Structural Failure Time Models," *Communications in Statistics, Part A—Theory and Methods*, 20, 2069–2631.

Rosenbaum, P., and Rubin, D. B. (1983), "Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study With Binary Outcome," *Journal of the Royal Statistical Society*, Ser. B, 45, 212–218.

Rubin, D. B. (1974), "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66, 688–701.

——— (1978), "Bayesian Inference for Causal Effects," *The Annals of Statistics*, 6, 34–58.

——— (1980), Comment on "Randomization Analysis of Experimental Data: The Fisher Randomization Test," by D. Basu, *Journal of the American Statistical Association*, 75, 591–593.

——— (1990), Comment: "Neyman (1923) and Causal Inference in Experiments and Observational Studies," *Statistical Science*, 5, 472–480.

——— (1991), "Practical Implications of Modes of Statistical Inference for Causal Effects and the Critical Role of the Assignment Mechanism," *Biometrics*, 47, 1213–1234.

Schultz, H. (1928), *Statistical Laws of Supply and Demand*, Chicago: University of Chicago Press.

Sommer, A., and Zeger, S. (1991), 'On Estimating Efficacy from Clinical Trials," *Statistics in Medicine*, 10, 45–52.

Wright, S. (1928), Appendix to *The Tariff on Animal and Vegetable Oils*, by P. G. Wright, New York: MacMillan.

——— (1934), "The Method of Path Coefficients," *Annals of Mathematical Statistics*, 5, 161–215.

Zelen, M. (1979), "A New Design for Randomized Clinical Trials," *New England Journal of Medicine*, 300, 1242–1245.