NBER TECHNICAL PAPER SERIES

IDENTIFICATION OF CAUSAL EFFECTS USING INSTRUMENTAL VARIABLES

J.D. Angrist

G.W. Imbens

D.B. Rubin

Technical Paper No. 136

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 1993

The authors thank the National Science Foundation for financial support, and Tom Newman and Norman Hearst for sharing their data. This paper is part of NBER's research program in Labor Studies. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Technical Paper #136
June 1993

IDENTIFICATION OF CAUSAL EFFECTS USING INSTRUMENTAL VARIABLES

ABSTRACT

In this paper we outline a framework for causal inference in settings where random assignment has taken place, but compliance is not perfect, i.e. the treatment received is nonignorable. In an attempt to avoid the bias associated with simply comparing subjects by the randomized treatment assignment, i.e. 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 this technique can be fitted into the Rubin Causal Model and be used for causal inference without assuming constant treatment effects. The advantages of embedding this approach in the Rubin Causal Model are that it makes the nature of the identifying assumptions more transparent, and that it allows us to consider sensitivity of the results to deviations from these assumptions in a straightforward manner.

J.D. Angrist
Department of Economics
Hebrew University
Mt Scopus, Jerusalem
91905, ISRAEL
and NBER

G.W. Imbens
Department of Economics
Harvard University
Cambridge, MA 02138
and NBER

D.B. Rubin
Department of Statistics
Harvard University
Cambridge, MA 02138

1. INTRODUCTION

World War Two veterans live longer and earn more than non–veterans of the same age. Does this mean that military service in World War Two improved the health and earnings capacity of veterans? Probably not. The difference in civilian mortality between World War Two veterans and nonveterans of the same generation is a classic example of the confounding impact of selection on estimates of treatment effects. The fact that World War Two veterans have lower mortality than nonveterans of the same age is usually attributed to the military screening process (Jablon and Seltzer, 1975) and not to any beneficial effects of wartime service. Moreover, surviving World War II veterans may have been intrinsically healthier than non–surviving soldiers even before service and this too could lead to a possibly erroneous conclusion that military service is good for one's health.

The importance of such bias in observational studies is widely appreciated in statistics, epidemiology and in the medical and social sciences. The task of designing procedures to evaluate treatments for AIDS and other life–threatening conditions has recently helped focus attention on the possibility of similar biases even in randomized clinical trials (Ellenberg, Finkelstein, and Schoenfeld, 1992). An early example of selection bias in clinical trials is the 1966–69 Coronary Drug Project (CDP, Coronary Drug Project Research Group, 1980) in which a range of drugs was tested for efficacy in reducing serum cholesterol. A comparison of treatment and control groups in the CDP shows little difference in outcomes, suggesting the drugs were ineffective. Yet many "treated" patients failed to take their medication as instructed; reasons presumably included negative side effects of the treatment. A comparison of "treated" patients who took the experimental drugs as instructed with the "treated"

patients who did not, initially suggested beneficial effects of cholesterol-reducing medications. Those drug-takers in the CDP control group who actually took the placebo, however, improved essentially as much as did drug-takers in the treatment group. It therefore seems unlikely that simply comparing "treated" patients who took experimental drugs with those who did not tells us much about the ability of these drugs to reduce cholesterol.

Such examples have motivated researchers in a variety of fields to search for new experimental designs and statistical adjustment techniques. In addition to the Ellenberg, Finkelstein, and Schoenfeld paper on AIDS Trials, recent discussions of non-compliance in clinical trials include papers by Efron and Feldman (1991), Gail (1985), Robins and Tsiatis (1991), Lavori (1992) and Robins (1992). The problem of selection bias in observational studies has been discussed for many years in articles and books from many fields. Examples include Cochran (1965) and Rubin (1974) in statistics, Heckman and Robb (1985) and Manski (1992) in economics, Kleinbaum, Kupper, and Morgenstern (1982) and Robins (1989) in epidemiology and medicine, and Nathan (1988) in public policy research. These and many dozens of related studies wrestle with one of the most important issues in applied statistics: how to draw causal inferences in anything other than a controlled laboratory setting.

The purpose of this paper is to describe a framework for causal inference that provides potential solutions to the problem of selection bias both in observational studies and experiments (e.g., clinical trials) with imperfect compliance. The distinguishing feature of our discussion of causality is that we attempt to unify two alternative, and occasionally competing, methodological traditions. We begin with an informal discussion of these traditions, one from statistics, evolving from work on randomized experiments, and one from economics,

evolving from models for individual behavior.

2. Causal inference in statistics and econometrics

Statisticians working on evaluation have been strongly influenced by an approach commonly referred to as Rubin's Causal Model (Rubin 1974, 1977, 1978; Holland 1986). Historical precursors include Fisher (1918), Neyman (1923), Kempthorne (1952), Cox (1958) and others discussed in Rubin (1992). This approach uses the notion of a "potential outcome" to define causality. The causal effect of treatment on a single individual or unit of observation is the 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 of differences in 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. In some cases, such as when the circumstances have been or are to be manipulated by an experimenter (e.g., a randomized drug evaluation, or a proposed change in welfare benefits), this mental exercise is straightforward. In other examples, such as economic studies of the relationship between gender and wages, it may be more difficult to fit the data into this framework. Many advocates of Rubin's model (see Holland 1986) claim that cases that are not fit into this framework are studying only associational, not causal, relationships.

Partly for historical reasons, econometricians (Haavelmo, 1943; Goldberger, 1972; Heck-

man 1992) have been inclined to use a more "structural" or equation-based approach to causality than have statisticians. Of course, this is an over-simplification and many economists use variations on the Rubin Causal Model as a framework for inference. Lewis (1986), for example, discusses union wage effects in terms of individuals with pairs of wages. one for the unionized status and one for the non-unionized status, only one of which is observable. The predominant approach to empirical work in econometrics, however, continues to be structural.

In econometrics (and sometimes in sociology, psychology and educational research). causal modeling typically begins with a set of theoretical relationships, as expressed in a system of mathematical (usually linear) equations. The classic example of this approach is the supply and demand paradigm used by economists to describe how markets determine prices and quantities. The stochastic representation of a market is as a system of simultaneous equations with prices and quantities (the endogenous variables) jointly determined.[2]

In a market model, one typically assumes the existence of supply and demand functions. which describe what quantities producers will supply and consumers will demand as a function of prices and other variables. The problem of causal inference in market models is then reduced to the estimation of the effect of prices on demand (or supply) when we observe only prices and quantities as theoretically generated by points of intersection of demand and supply functions at different times or locations. Economists think of supply and demand parameters as causal because they can be useful for predicting the impact of major market

[2]Sociologists retain the term "path analysis" from the pioneering work of Sewall Wright (1934) for a similar concept.

interventions (i.e. treatments) such as changes in rates of taxation.[3]

The econometric approach to evaluation research – such as evaluating the impact of military service – typically also begins with a behavioral or structural model of the relationship between treatment status and outcomes. For example, the choice to enter the military may be modeled as being based on a comparison of the utilities of alternative options. The most important component of almost all econometric evaluation research, however, is the notion that particular observed variables are related to the outcomes of interest solely by virtue of their relationship with treatment status (or prices in the case of market models). These particular observed variables, which are often intrinsically uninteresting, are then said to be usable as instruments. The Instrumental Variables (IV) estimation technique originated in research on simultaneous equations models by Wright (1928, 1934) and Reiersol (1941), who appears to have coined the term "instrumental variables" (Bowden and Turkington, 1984).

To illustrate the IV method as an approach to causal inference we return to the example of veteran status. Vietnam veterans do not live as long as non-veterans the same age, but perhaps this is just the reverse of the postulated positive selection observed among World War Two inductees. Vietnam veterans may have been disproportionately drawn from unfortunate groups in society who would have fared poorly even had they not served in the military.

In an attempt to overcome the selection bias inherent in comparisons of outcomes by veteran status, epidemiologists Hearst, Newman, and Hulley (1986) compared the civilian mortality experience of Vietnam-era cohorts subclassified by draft lottery number. Between

[3]Goldberger (1972) notes that the problem of causal inference in market models is clearly articulated in Henry Schultz's (1938) "Theory and Measurement of Demand". Another early reference is Haavelmo (1943).

1970 and 1973, draft lottery numbers from 1 to 365 were randomly assigned to the birthdays of men born between 1950 and 1953. These lottery numbers were used to determine who would be eligible to be drafted. Veteran status itself was never randomly assigned, but men with the lowest numbers were called earliest to begin the screening process for induction into the military. Men with high numbers were never drafted, but they could still volunteer for military service. In fact, most Vietnam veterans were volunteers (although many evidently volunteered to escape military service under less favorable terms as a draftee).

Hearst, Newman, and Hulley found that men with low lottery numbers were slightly more likely to have died between 1974 and 1984 than men with high lottery numbers. Since lottery numbers were randomly assigned, this clearly suggests some detrimental causal effect of low lottery numbers.

An attempt to link the causal effect of low lottery numbers to a causal effect of veteran status can be made as follows. Men with low lottery numbers were about 15 percentage points more likely to have served in the military than men with high numbers. They were also .09 percentage points more likely to have died of civilian causes during the postwar period of 1974–1983. The standard IV argument states that: (1) if the effect of veteran status on mortality is constant, and (2) the only reason lottery numbers are associated with civilian mortality is because they are associated with veteran status, then the effect of veteran status on mortality can be estimated by the IV formula which is the ratio of the difference in mortality by lottery number to the difference in the probability of serving in the military by lottery number. In the above example, the IV formula gives $0.0009/0.15 = 0.0056$ as the increase of the probability of dying of civilian causes due to serving in the military.

Using the IV interpretation, differences in mortality by lottery number are attributed solely to differences in the probability of having served in the military by lottery number. In a re-analysis of the Hearst, Newman, and Hulley mortality data, we show below that applying the IV formula to data on suicide by men from Vietnam-era cohorts leads to an estimated probability of suicide for veterans nearly double that for nonveterans.[4]

In an econometric application inspired by Hearst, Newman, and Hulley's work, Angrist (1990) uses draft lottery numbers to construct IV estimates of the effect of Vietnam-era military service on civilian earnings, which suggests that Vietnam veterans earn 15 percent less than nonveterans the same age up to 10 years after their discharge from the military. It turns out that World War Two veterans were also conscripted on the basis of birthdays, and quarter of birth is the instrument used by Angrist and Krueger (1989) to estimate the effect of World War II military service on civilian earnings. Here too, IV estimates suggest negative effects of military service on earnings.

Examples of IV approaches from fields other than economics include Powers and Swinton (1984), who describe an experiment run by the Educational Testing Service that effectively uses a randomly assigned letter encouraging college applicants to study for the GRE as an instrument in order to relate self-reported hours of study to exam scores; this example is analyzed in the framework of the Rubin Causal Model in Holland (1988). Permutt and Hebel (1989) use a similar encouragement design to construct instrumental variables estimates of the effect of maternal cigarette smoking on birth weight. Finally, Efron and Feldman (1991).

[4]In their original paper, Hearst, Newman, and Hulley make no reference to the IV interpretation of their use of the draft lottery. But they independently derive an IV-type formula for the effect of veteran status on the relative risk of mortality.

and Robins (1989) present related formulas in their discussions of clinical trials compliance.

Although IV estimators appear to give sensible answers to important causal questions in a wide range of examples where treatment is nonignorable, these applications typically rely on a regression framework with simple constant treatment effects. Recent research in econometrics has been directed at more realistic models allowing for heterogeneity in treatment effects (Heckman 1990; Manski, 1992). One difficulty is that in Rubin's Causal Model, the basic, regression-based, IV assumptions that the instrument affects treatment status and be correlated with outcomes solely because of this, are no longer enough to identify a meaningful average treatment effect estimable by the IV formula. Indeed, it is easy to devise examples where the IV formula leads to misleading results.

Here we provide the conditions under which the IV estimand is a causal effect in the sense of the Rubin Causal Model, and investigate sensitivity of this estimand to violations of critical IV assumptions. The IV estimand is, in our approach, an average causal effect for a subpopulation. This subpopulation cannot be identified from the data, unlike the average casual effects that are typically the focus of evaluation research.

3. Causal Estimands with Instrumental Variables.

We use a hypothetical evaluation of the effect of a new drug (D) on survival (Y) in a population of N units to describe the principles of causal inference using instrumental variables. The actual treatment status D is assumed to be beyond the control of the researcher. Instead, we assume that the evaluation is based on the investigators' randomly assigned intention to treat, indicated by the variable Z; $Z_i = 1$ implies that patient i is assigned to

the treatment group, which is to receive the new drug, whereas $Z_i = 0$ indicates that patient i is assigned to the control group, which is to receive a placebo instead of the new drug. Let $\mathbf{Z}$ be the N dimensional vector of assignments with i^{th} element Z_i, and let $D_i(\mathbf{Z})$ to be the indicator for the drug unit i actually received given the vector of assignments $\mathbf{Z}$. In an ideal research environment, $D_i(\mathbf{Z})$ equals Z_i for all i, that is, the treatment received equals the treatment assigned. In practice, $D_i(\mathbf{Z})$ can differ from Z_i for various reasons: individuals may accidentally receive the incorrect drug, or they may attempt to obtain the new drug despite being assigned to the control group, or individuals in the treatment group may fail to take the assigned drug because of side effects, as in the Coronary Drug Project (1980).

Similar to the definition of $D_i(\mathbf{Z})$, we define $Y_i(\mathbf{Z}, \mathbf{D})$ to be the response for unit i given the vector of treatments $\mathbf{D}$ and the vector of assignments $\mathbf{Z}$; $\mathbf{Y}(\mathbf{Z}, \mathbf{D})$ is the N vector with i^{th} element $Y_i(\mathbf{Z}, \mathbf{D})$. Holland (1988) uses a similar notation in his discussion of encouragement designs with test scores depending on both the randomized "encouragement" and the subsequent self-selected "amount of studying". We refer to $D_i(\mathbf{Z})$ and $Y_i(\mathbf{Z}, \mathbf{D})$ as "potential outcomes." The concept of potential outcomes is analogous to Neyman's (1923) notion of "potential yields" in randomized agricultural experiments, as formally generalized in Rubin's Causal Model to encompass settings without randomization and with possible interference between units and versions of treatments.

At this point in our development, $D_i(\mathbf{Z})$ and $Y_i(\mathbf{D}, \mathbf{Z})$ are non-stochastic, fixed, but unobserved, constants, and $\mathbf{Z}$ is the only variable that can be manipulated. Differences in these 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. This framework differs importantly from the assumptions and motivation underlying econometric research, where both outcomes and treatment assignments are immediately treated as random variables with an iid (independent and identically distributed) structure governed by some hypothetical parameters. The reason we use an "experimentalist" approach is that it allows us to precisely define the quantities we wish to estimate without specifying the mode of inference to be used to obtain estimators and measures of uncertainty.[5]

In evaluation research, some sort of assumption about how treatment units interact is typically required. Here we follow the convention in statistics and medical research by disallowing interference between treatment units:

Assumption 1 (STABLE UNIT TREATMENT VALUE ASSUMPTION (SUTVA) [RUBIN 1982, 1990a])

(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, or the stability assumption, means that potential outcomes for each particular unit i do not depend on the treatment status of other units. This assumption, by disallowing interference between treatment units, allows us to write $Y_i(\mathbf{Z}, \mathbf{D})$ and $D_i(\mathbf{Z})$ as $Y_i(D_i, Z_i)$ 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 below, although generalizations can be formulated with SUTVA replaced by other assumptions. For example,

[5]Modes of causal inference are discussed in detail by Rubin (1990b, 1991).

models of market phenomena are premised on the notation that large numbers of individuals interact to determine prices and quantities. Each unit's potential outcomes (each individual's response to a price change) can be affected by the distribution of income and by prices and quantities in other markets so that SUTVA is unlikely to be satisfied.

Under SUTVA, our framework allows six potential outcomes for each unit: $Y_i(0,0)$, $Y_i(0,1)$, $Y_i(1,0)$, $Y_i(1,1)$, $D_i(0)$ and $D_i(1)$. Because an experimenter can assign $Z_i = 0$ or $Z_i = 1$ to each unit, we can potentially observe $D_i(0)$, $D_i(1)$, $Y_i(0, D_i(0))$ and $Y_i(1, D_i(1))$. But, it is impossible to observe $Y_i(0, 1 - D_i(0))$ or $Y_i(1, 1 - D_i(1))$ because there is no value of Z_i that would lead to the realization of that outcome; that is, we assume here that there is no manipulation by the researcher affecting D_i other than Z_i. The latter two outcomes can be said to be *a priori* counterfactual because no experimental manipulation can reveal them. Among the other potential outcomes, some are observed and others are not. Those that are not observed can be said to be *ex post* counterfactual because while, *ex post* (after manipulation of Z_i) they were not observed, they could have been observed under an alternate manipulation of Z_i.

For example, consider someone with $D_i(0) = D_i(1) = 0$; this person will not be treated no matter whether assigned to treatment or control group. The *a priori* counterfactuals in this case are $Y_i(0,1)$ and $Y_i(1,1)$; if $Z_i = 0$, $Y_i(0,0)$ is observed and $Y_i(1,0)$ is an *ex post* counterfactual, whereas if $Z_i = 1$, $Y_i(0,1)$ is observed and $Y_i(0,0)$ is an *ex post* counterfactual. In the example we discuss this issue in more detail.

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 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 $Y_i(1, D_i(1)) - Y_i(0, D_i(0))$.

These causal effects are the effects of Z, the treatment assignment, or the "intention to treat" effects. Given Assumption 1 and ignorability of treatment assignment, Z, unbiased estimators for the averages of the causal effects over the population of interest can be obtained by taking the difference of sample averages of Y and D classified by the value of Z, i.e. by treatment-control mean differences, as has been well known since at least Neyman (1923).

To define the causal effect of D on Y in a meaningful way requires more work. Holland (1988) defines the causal effects of D on Y conditional on Z, $Y_i(z,1) - Y_i(z,0)$ for $z = 0, 1$ and in some sense skirts the issue of *a priori* counterfactuals by assuming the causal effects of D on Y to be equal for *all* units, even those who cannot be induced to receive treatment D by the random assignment Z. This approach thus requires us to conceptualize the *a priori* counterfactuals representing outcomes under a treatment that could not be induced by the randomization. In contrast, our approach does not require this.

The following assumption requires the treatment assignment, Z, to be unrelated to potential outcomes once treatment received, D, is taken into account.

Assumption 2 (EXCLUSION RESTRICTION OF TREATMENT ASSIGNMENT GIVEN TREATMENT RECEIVED)

$\mathbf{Y(Z, D) = Y(Z', D)}$ *for all* $\mathbf{Z, Z'}$ *and for all observable* $\mathbf{D}$.

This assumption implies that $Y_i(1,d) = Y_i(0,d)$ for $d = D_i(0), D_i(1)$. Thus, $Y_i(0,d) =$

$Y_i(1,d)$ for $d = 0,1$ for units such that $D_i(1) \neq D_i(0)$, and, for all units such that $D_i(1) = D_i(0)$, $Y_i(0, D_i(0)) = Y_i(1, D_i(1))$. It is a restriction, discussed informally in Section 2, that implies that any effect of Z on Y must be via an effect of Z on the treatment received, D. Since it is an restriction on *a priori* counterfactuals, it is not directly testable from the data at hand. Notice that with this assumption we only need to conceptualize $Y_i(z, D_i(1))$ and $Y_i(z, D_i(0))$ for $z = 0,1$. We do not need to conceptualize the unobservable $Y_i(z,d)$ for a value of d that is not potentially observable, i.e. we do not need to conceptualize *a priori* counterfactuals involving unobservable treatments. In contrast, Holland's (1988) application of the Rubin Causal Model to encouragement designs requires us to consider restrictions on $Y_i(z,d)$ for values d of the treatment that are not even potentially observable.

At this point we define the causal parameters of interest. By virtue of Assumption 2, we can define potential outcomes $\mathbf{Y(Z,D)}$ as a function of $\mathbf{D}$ alone:

$$\mathbf{Y(D) = Y(Z,D) = Y(Z',D)} \qquad \text{for all } \mathbf{Z},\ \mathbf{Z'} \text{ and for all observable } \mathbf{D}.$$

Equivalently, $Y_i(D_i(z)) = Y_i(z, D_i(z))$ for $z = 0,1$.

Definition 2 (CAUSAL EFFECTS OF D ON Y GIVEN THE EXCLUSION RESTRICTION) *The causal effect of D on Y for units i with $D_i(0) \neq D_i(1)$ is $Y_i(1) - Y_i(0)$, which is equal to $Y_i(z,1) - Y_i(z',0)$ for all $z, z' = 0,1$.*

We can never observe this causal effect, of course, although we can observe one of its terms, and therefore focus on typical, e.g., average, causal effects across groups of units. We draw inferences about such causal effects using changes in treatment status D induced by treatment assignment Z.

Assumptions 1 and 2 and Definition 2 are sufficient to establish a fundamental relationship between the causal effect of Z on Y and the causal effect of D on Y for units with $D_i(1) \neq D_i(0)$:

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

Under these same assumptions, for units with $D_i(1) = D_i(0)$, the causal effect of Z on Y is zero:

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

which can be viewed as encompassed in Equation (1) by definition, even though $Y_i(1) - Y_i(0)$ is not defined for these units.

The interpretation of relation (1) is that, for unit i, the causal effect of Z on Y is the product of the causal effect of D on Y and the causal effect of Z on D. This simple but important relationship has been established given only SUTVA and the exclusion restriction on Z given D. However, this relationship is not enough to guarantee the identification of any meaningful average treatment effect because $D_i(0)$ and $D_i(1)$ can be equal to zero or one, and the causal effect of Z on Y, $Y_i(1, D_i(1)) - Y_i(0, D_i(0))$, can be equal to plus or minus the causal effect of D on Y, $Y_i(1) - Y_i(0)$, or zero. To see this even more clearly, partition the population of units in a 2 by 2 table according to the values of $D(0)$ and $D(1)$.

Table 1 illustrates this partition, and the corresponding treatment effects of Z on Y under Assumptions 1 and 2.

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

The four values of $(D_i(0), D_i(1))$ in the 2 by 2 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 "switch-in" to the treatment by the assignment to the treatment, and the causal effect of Z on Y is $Y_i(1) - Y_i(0)$. A value of $D_i(1) - D_i(0) = 0$ (diagonal elements) implies the individual does not change treatment status with the assigned treatment: the causal effect of Z on Y is zero for such individuals by Definition 1 and Assumption 2. Finally, individuals with $D_i(1) - D_i(0) = -1$ (top right) do the opposite of their assignment, they are induced to "switch-out" of the treatment by assignment to it; the causal effect of Z on Y in this case is $Y_i(0) - Y_i(1)$.

At this point it is convenient to 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 $h(\cdot)$,

$$h(Z_i, D_i(1), D_i(0), Y_i(0,0), Y_i(0,1), Y_i(1,0), Y_i(1,1)) = h_0,$$

will be denoted by $E[g|h(\cdot) = h_0]$. Finally, the relative size of this subpopulation satisfying $h(\cdot) = h_0$ is written here as $P[h(\cdot) = h_0] = E[1_{h(\cdot)=h_0}]$, where $1_{\{\cdot\}}$ is the indicator function. We emphasize that this notation, $E[\cdot]$, $E[\cdot|\cdot]$, and $P[\cdot]$, does not reflect (conditional) expectations or probabilities with respect to any stochastic distribution, but simply reflects averages and frequencies in a finite population or subpopulation.

Using this notation, from equation (1), we can write the average causal effect of Z on Y as the weighted sum of the average causal effects for two subpopulations with $D_i(0) \neq D_i(1)$:

$$\begin{aligned} &E[Y_i(1, D_i(1)) - Y_i(0, D_i(0))] \\ &\quad = E[(Y_i(1) - Y_i(0))|D_i(1) - D_i(0) = 1] \cdot P[D_i(1) - D_i(0) = 1] \\ &\qquad - E[(Y_i(1) - Y_i(0))|D_i(1) - D_i(0) = -1] \cdot P[D_i(1) - D_i(0) = -1]. \end{aligned} \tag{2}$$

where the weights do not sum up to 1 but to $P[D_i(0) \neq D_i(1)]$. By virtue of Assumption 2, individuals whose treatment status is unaffected by the treatment assignment only affect the average casual effect of Z on Y through their effect on the sum of the weights. But the groups for whom the treatment status is affected by the treatment assignment include both "switchers-in" and "switchers-out." From equation (2) it is clear that it is theoretically possible to have a situation where the actual treatment effect of D on Y, $Y_i(1) - Y_i(0)$, is

positive for all units, but the relative size of the group of "switchers-in" versus that of the "switchers-out" is such that the average effect of Z on Y is zero or even negative. Suppose, for example, the treatment effect equals C for "switchers-in" and $2C$ for "switchers-out". If $P[D_i(1) - D_i(0) = 1] = 1/2$ and $P[D_i(1) - D_i(0) = 1] = 1/4$, then the average causal effect of Z on Y, $E[Y_i(1, D_i(1)) - Y_i(0, D_i(0))]$, is equal to zero.

As originally noted by Imbens and Angrist (1991), Assumption 3 rules out situations of this type by requiring that, among all individuals whose treatment status D is affected by the treatment assignment Z, everyone is either a "switcher-in" or everyone is a "switcher-out":

Assumption 3 (MONOTONICITY OF TREATMENT ASSIGNMENT AND TREATMENT RECEIVED [IMBENS AND ANGRIST, 1991])

Either $D_i(1) \geq D_i(0)$ *for all* $i = 1, ..., N$ *or* $D_i(1) \leq D_i(0)$ *for all* $i = 1, 2, ..., N$.

Permutt and Hebel (1989) discuss a variant of this assumption informally in the context of a program 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 that received no encouragement to stop smoking, would also stop smoking if encouraged to do so in the treatment group; a related discussion appears in Robins (1989).

Without loss of generality, assume that monotonicity holds with the existence of "switchers-in", i.e. $D_i(1) \geq D_i(0)$ for all i. Therefore $D_i(1) - D_i(0)$ equals either zero or one, so that

$$E[Y_i(D_i(1), 1) - Y_i(D_i(0), 0)] \tag{3}$$

$$= E[(Y_i(1) - Y_i(0))|D_i(1) - D_i(0) = 1] \cdot P[D_i(1) - D_i(0) = 1].$$

Table 1 illustrates the way Assumptions 2 and 3 complement each other by ruling out different causal effects. By virtue of Assumption 2, 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 Assumption 3, one of the groups corresponding to the off-diagonal elements is empty. Combined, these assumptions imply that the non-zero average causal effect of Z on Y arises from just one of the four subpopulations indexed by the values of $D(0)$ and $D(1)$. This is what allows one to express causal effects of D on Y in terms of the causal effects of Z on Y and of Z on D.

In order to be able to use (3) to identify an average causal effect of D on Y, we need to divide both sides by $P[D_i(1) - D_i(0) = 1]$ and so we need the following assumption:

Assumption 4 (NON-ZERO 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.

Note that Assumption 3 with the requirement that $D_i(1) \neq D_i(0)$ for at least one unit i implies Assumption 4. Even without Assumption 3, Assumption 4 leads to the following definition of an *instrument*:

Definition 3 (INSTRUMENTAL VARIABLE)

A variable Z is an instrumental variable if it satisfies Assumptions 2 and 4.

This definition requires that Z satisfies an exclusion restriction on outcomes given treatment status, *and* that it affects, on average, the treatment received. It is the combination of these two assumptions that allows one to use Z as an instrument for uncovering causal relations between D and Y.

Rewriting equation (3), using Assumptions 3 and 4, gives

$$E[(Y_i(1) - Y_i(0))|D_i(1) - D_i(0) = 1] = \frac{E[Y_i(D_i(1),1) - Y_i(D_i(0),0)]}{E[D_i(1) - D_i(0)]}, \quad (4)$$

because Assumption 3 implies that $P[D_i(1) - D_i(0) = 1] = E[D_i(1) - D_i(0)]$ and Assumption 4 implies that this is not equal to zero. The ratio on the right-hand side of equation (4) will be called the IV estimand, whether or not Assumptions 1-3 hold:

Definition 4 (IV ESTIMAND)

The IV estimand is the ratio of the average causal effect of Z on Y, $E[Y_i(1, D_i(1)) - Y_i(0, D_i(0))]$, to the average causal effect of Z on D, $E[D_i(1) - D_i(0)]$.

Variations on this estimand has been analyzed extensively in econometrics, (Durbin, 1954, Goldberger, 1972, Heckman and Robb, 1985), and recently in other fields, (Holland, 1988, Permutt and Hebel, 1989, Robins, 1989), often under the assumption that the causal effect of D on Y is constant.

The relation between the IV estimand and causal effects of D on Y under our assumptions is our main theoretical result and we summarize it as a formal proposition:

Proposition 1 (IDENTIFICATION OF CAUSAL ESTIMAND USING INSTRUMENTAL VARIABLES)

Given Assumptions 1-4, the IV estimand is equal to the Local Average Treatment Effect.

$E[Y_i(1) - Y_i(0)|D_i(1) \neq D_i(0)]$

It is important to note that we cannot identify the group, defined by $D_i(1) \neq D_i(0)$, for which we can identify the average treatment effect. The average treatment effect identified here is therefore not the average treatment effect for the entire population or for an

identifiable, observable subpopulation, which is typically the focus of attention in evaluation research. Stronger assumptions are needed for the identification of average causal effects for identifiable observable subpopulations. For examples of such assumptions see Robins (1989). Heckman (1990), Efron and Feldman (1991) and Manski (1992).

4. Sensitivity of the IV Estimand to Critical Assumptions

Before returning to an application, we discuss the sensitivity of the Instrumental Variables (IV) estimand to deviations from Assumptions 2 and 3. We focus on these assumptions because they form the core of the IV approach, and, moreover, sensitivity to the other, more conventional assumptions of ignorability of treatment assignment has been discussed in the context of the conventional Rubin Causal Model (e.g., Rosenbaum and Rubin, 1983). The focus of our discussion is on the consequences of deviations from the IV assumptions on the average causal effect for the subpopulation defined by $D_i(1) \neq D_i(0)$. This local average is, under Assumptions 1–4 equal to the ratio of the average causal effect of Z on Y to the average causal effect of Z on D.

4.1 Violations of the Exclusion Restriction.

First we consider violations of the exclusion restriction, Assumption 2, while maintaining the monotonicity assumption, Assumption 3 (with monotonicity satisfied for "switchers-in" for descriptive convenience), and the other Assumptions 1 and 4. The following assumption allows us to define the causal effect of D on Y in the absence of Assumption 2 by forcing the two conditional causal effects to be equal.

Assumption 5 (ADDITIVE CAUSAL EFFECTS)

For all units i,

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

For units with $D_i(1) = D_i(0)$ this assumption is trivially satisfied. For units with $D_i(1) \neq D_i(0)$ Assumption 5 implies that the effects of Z and D are additive. For notational convenience, define

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

to be the direct causal effect of Z on Y. Assumption 2, the exclusion restriction, implies that $H_i = 0$ for all units, and so the causal effect of D on Y is uniquely defined. Under the weaker Assumption 5, the causal effect of D on Y is still uniquely defined to be

$$G_i = Y_i(z, 1) - Y_i(z, 0) \qquad \text{for } z = 0, 1.$$

for all units with $D_i(1) \neq D_i(0)$. The average causal effect of Z on Y can now be written as

$$E[Y_i(1, D_i(1)) - Y_i(0, D_i(0))]$$

$$= P[D_i(1) \neq D_i(0)] \cdot E[G_i | D_i(1) \neq D_i(0)] + E[H_i].$$

The average causal effect of Z on D is not affected by the violation of Assumption 2. The IV estimand is the ratio of these two average causal effects and is therefore, given Assumptions 1, 3, 4 and 5, equal to

$$\frac{E[Y_i(1, D_i(1)) - Y_i(0, D_i(0))]}{E[D_i(1) - D_i(0)]} = E[G_i | D_i(1) \neq D_i(0)] + \frac{E[H_i]}{E[D_i(1) - D_i(0)]}.$$

The ratio is equal to the sum of two parts: first the average causal effect of D on Y over the subpopulation defined by $D_i(1) \neq D_i(0)$, that is, $E[G_i|D_i(1) \neq D_i(0)]$; and second, the direct effect of Z on Y, $E[H_i]$, divided by the frequency of that subpopulation, which by Assumption 3 is equal to $P[D_i(1) - D_i(0) = 1] = E[D_i(1) - D_i(0)]$.

Consider first the special case where $D_i(1) - D_i(0) = 1$, in other words all units comply with their randomized assignment. In this case, the IV estimand equals the average causal effect of Z on Y, the "intent to treat" estimand, which equals the average causal effect of D on Y, $E[G_i]$, *plus* the average (direct) causal effect of Z on Y, $E[H_i]$. Since with perfect compliance $D_i \neq D_i(0)$ for all units, this can be written as:

$$E[G_i + H_i] = E[G_i|D_i(1) \neq D_i(0)] + E[H_i|D_i(1) \neq D_i(0)] \tag{5}$$

Thus if the assignment has a direct causal effect separate from the effect of treatment, even a randomized experiment with perfect compliance will not estimate the average causal effect of the treatment D but the rather combined effect of D and Z.

This effect is well known, and it is the motivation for standard precautions in experimental design. For example blinding, double blinding, and the use of placebos are designed to help ensure that any effects observed can be attributed to the treatment being studied, D, and not simply the result of the randomized assignment, Z.

Now consider the case of imperfect compliance. The IV estimand can be written as

$$E[G_i|D_i(1) \neq D_i(0)] + \frac{E[H_i]}{E[D_i(1) - D_i(0)]}$$

$$= E[G_i|D_i(1) \neq D_i(0)] + E[H_i|D_i(1) \neq D_i(0)] \tag{6}$$

$$+E[H_i|D_i(1) = D_i(0)] \cdot \frac{P[D_i(1) = D_i(0)]}{P[D_i(1) \neq D_i(0)]}.$$

Thus, comparing (5) and (6), we see that with additive effects, the increased bias in the IV estimand due to non-compliance is the last term, which is directly proportional to the product of (1) the average size of the direct effect of Z for those who cannot be induced to change treatment, and (2) $p/(1-p)$ where p is the proportion of units that cannot be induced to change.

The higher the correlation between the instrument and the treatment status, i.e., the "stronger" the instrument, the smaller p, the less sensitive the IV estimand is to violations of the exclusion assumption. Also, the smaller the direct effect of Z on Y, H_i, for those who do not change treatment as a result of a change in assignment (those with $D_i(0) = D_i(1)$), the less affected the IV estimand is relative to the "intent to treat" estimand under random assignment with perfect compliance.

4.2 Violations of the Monotonicity Condition

Next we consider violations of the monotonicity assumption, Assumption 3. Since we maintain Assumption 2, the causal effect of D on Y for unit i with $D_i(1) \neq D_i(0)$ is still uniquely defined, and equal to $Y_i(1) - Y_i(0)$. Again we investigate the average causal effect of Z on Y:

$$E[Y_i(1, D_i(1)) - Y_i(0, D_i(0))]$$

$$= E[(Y_i(1) - Y_i(0))|D_i(1) - D_i(0) = 1] \cdot P[D_i(1) - D_i(0) = 1]$$

$$- E[(Y_i(1) - Y_i(0))|D_i(1) - D_i(0) = -1] \cdot P[D_i(1) - D_i(0) = -1].$$

This average causal effect now consists of two terms. The IV estimand can be written

$$
\begin{aligned}
&E[Y_i(1, D_i(1)) - Y_i(0, D_i(0))/E[D_i(1) - D_i(0)] \\
&\quad = (1 - \lambda) \cdot E[Y_i(1) - Y_i(0)|D_i(1) > D_i(0)] + \lambda \cdot E[Y_i(1) - Y_i(0)|D_i(1) < D_i(0)] \\
&\quad = E[Y_i(1) - Y_i(0)|D_i(1) > D_i(0)] \qquad (7) \\
&\qquad +\lambda \cdot \Big[E[Y_i(1) - Y_i(0)|D_i(1) < D_i(0)] - E[Y_i(1) - Y_i(0)|D_i(1) > D_i(0)]\Big],
\end{aligned}
$$

where $\lambda = -P[D_i(1) - D_i(0) = -1]/E[D_i(1) - D_i(0)] < 0$. In the first representation, the estimand is a weighted average of average treatment effects, but the weights are always outside the unit interval. The latter term in the second representation is the bias, and it is composed of two factors. The first factor, λ, is related to the frequency of units with $D_i(1) < D_i(0)$; λ is equal to zero under Assumption 3 because $P[D_i(1) - D_i(0) = -1] = 0$. Because λ is equal to the frequency of "switchers-out" divided by the average causal effect of Z on D, equal to $E[D_i(1) - D_i(0)] = P[D_i(1) - D_i(0) = 1] - P[D_i(1) - D_i(0) = -1]$, it can be large even if $P[D_i(1) - D_i(0) = -1]$ is small, as long as the average causal effect of Z on D is small. The bias term resulting from the violation of the monotonicity assumption is also proportional to the difference in average causal effects for the "switchers-in" ($D_i(1) > D_i(0)$), and the "switchers-out" ($D_i(1) < D_i(0)$). If the average causal effects of D on Y are identical for switchers-in and switchers-out, violations of Assumption 3 generate no bias. A constant treatment effect of D on Y for units with $D_i(1) \neq D_i(0)$ is sufficient but not necessary for this. In general, the closer λ is to zero, and the less variation there is in the causal effect of D on Y, the smaller the bias will be from violating Assumption 3. Note that again the average

causal effect of Z on D is in the denominator of the bias. The "weaker" the instrument, or the further away the instrument from a "true" experiment with full compliance, the worse the consequences of any violation of Assumption 3.

Finally, note that the assumptions laid out in Section 3 are *sufficient* conditions for the identification of a meaningful average treatment effect. As this discussion makes clear violations of these assumptions need not be catastrophic. However, the IV estimand is very sensitive to violations of Assumptions 2 and 3 when the average causal effect of the treatment assignment Z on the treatment status D is low. In the next section we illustrate how this sensitivity analysis can be applied.

5. An Application: The effect of military service on civilian mortality.

Hearst, Newman and Hulley (1986) showed that men who were at high risk of being drafted in the Vietnam Era draft lottery had elevated mortality risk after their discharge from the military. The authors attribute this elevated risk to the higher probability of military service faced by draft–eligible men. This assumption is plausible because between 1970 and 1973, the risk of being drafted was randomly assigned in a lottery based on 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 called up to RSN 195, men born in 1951 were called up to RSN 125, and men born in 1952 were called up to RSN 95.

In their paper, Hearst, Newman and Hulley focus on the difference in mortality risk by

draft–eligibility 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. This procedure can be used to provide a valid estimate of the effects of military service on mortality if draft–eligibility is correlated with civilian mortality solely by virtue of its correlation with veteran status. Our purpose in returning to this example is threefold. First, we use the example to illustrate the distinction between *a priori* and *ex post* counterfactual outcomes. Second, we discuss the validity of Assumptions 1–4 in this context. Third, we show how the sensitivity of the estimated average treatment effect to violations of Assumption 2 can be explored using the results from the previous section.

To apply the analysis of Section 3, let the instrument Z_i be a binary indicator of draft eligibility, equal to one if individual i was assigned an RSN less than the draft cutoff. Let $D_i(z)$ indicate whether someone would have served in the military when draft eligible ($z = 1$) or not ($z = 0$). It is perfectly reasonable to postulate the existence of both of these treatment assignments even though one assignment is *ex post* (after Z is realized) counterfactual. The reason both potential values of $D_i(z)$ are well–defined is that Z_i was randomly set by the federal government and could easily have been set in a different way.

The potential outcome in this example is an indicator variable, $Y_i(z, d)$, equal to one if person i died between 1974 and 1983. To distinguish this from mortality during the war period we will refer to this as civilian mortality. For simplicity we ignore the effect that different mortality during the war might have had on mortality after the war. Is it reasonable to postulate a complete set of potential outcomes for all values of d and z? Consider someone with $D_i(0) = D_i(1) = 0$ – this man would not have served in the military

whether draft eligible or not. For such a man, $Y_i(0,0)$ and $Y_i(1,0)$ are clearly well defined for the same reason that $D_i(0)$ and $D_i(1)$ are well-defined. But the definitions of $Y_i(0,1)$ and $Y_i(1,1)$ can problematic. These outcome variables represent what would have happened if such a man had served in the military, despite the fact that no manipulation of draft lottery numbers could have revealed these outcomes. Of course, some other intervention or manipulation might have been able to cause this man to serve in the military, but that would be a different experiment from the one we have available, the draft lottery. However, the IV estimand does not involve these *a priori* counterfactuals because it is an average over the subpopulation that can be induced to serve by a change in lottery number.

Next, consider someone with $D_i(0) = 0$ and $D_i(1) = 1$. This man would not serve in the military if not drafted, and would serve if drafted. For such men we need to conceptualize the set of four potential outcomes, $Y_i(0,0)$, $Y_i(0,1)$, $Y_i(1,0)$ and $Y_i(1,1)$. The two potentially observable outcomes $Y_i(0,0)$ and $Y_i(1,1)$ are straightforward, but we also need to conceptualize the outcome when this man was drafted, but did not serve, despite the fact that he would have served if drafted. Assumption 2 links this outcome to one that is potentially observable, by assuming that, because the random assignment itself has no effect, it equals the outcome when drafted and serving, which is potentially observable.

Next we turn to the validity of Assumptions 1–5. These assumptions comprise SUTVA, an exclusion restriction, and monotonicity in the relationship between instruments and treatments. In other words, we require: (Assumption 1) that the veteran status of any man at risk of being drafted in the lottery is not a function of the lottery number assigned to others. Similarly, the civilian mortality risk of any man is not determined by the lottery number

or veteran status of others; (Assumption 2) civilian mortality risk is not affected by lottery numbers once veteran status is taken into account; (Assumption 3) A man who would serve in the military given that his lottery number made him ineligible for the draft would also have served in the military if his lottery number was below the cutoff ceiling, (Assumption 4) On average, men would have been more likely to serve with a low lottery number than with a high lottery number.

Although we believe these assumptions are plausible, a case can be made for violations of each one. 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 possibly violations of SUTVA is that people with high lottery numbers may be induced to serve in the military by friends who received low lottery numbers. There is also some evidence that men with low lottery numbers changed their educational plans so as to retain draft deferments and avoid the conscription (Angrist and Krueger 1992b), and this means that the exclusion restriction could be violated because draft lottery numbers may have affected civilian outcomes through channels other than veteran status. The third assumption, monotonicity, may be violated if, for example, someone who would have volunteered for the Navy when not at risk of being drafted would have chosen to avoid military service altogether if he received a low lottery number, but it seems unlikely that there were many in the population in this category.

The least controversial assumption required to use the draft lottery to estimate average causal affects is ignorability of treatment assignment, needed for unbiased estimation of the

average causal effects of Z on D and of z on Y. Here too, however, there is evidence that the first lottery, executed using a poorly designed physical randomization, was not actually random (Fienberg 1971), but even so, any correlation between lottery number and "potential" outcomes is essentially impossible. Ignoring this complication and postponing consideration of the potential problems outlined above, we can forge ahead with the instrumental variables approach.

Table 2 presents data and some estimates of the effects of military service on civilian mortality for white men born in 1950 and 1951.

Table 2: Data on civilian mortality for white men born in 1950 and 1951.

Year	Draft Eligibility[6]	No of Deaths[7]	No of Suicides[8]	Prob of Deaths[9]	Prob of Suicides	Prob of Mil Service[10]
1950	Yes	2601	436	0.0204	0.0034	0.3527
	No	2169	352	0.0195	0.0032	0.1934
Difference (Yes minus No)				0.0009	0.0002	0.1593
IV estimates				0.0056	0.0013	
1951	Yes	1494	279	0.0170	0.0032	0.2831
	No	2823	480	0.0168	0.0029	0.1468
Difference (Yes minus No)				0.0002	0.0003	0.1362
IV estimates				0.0015	0.0022	

Column 1 shows the number of deaths in Pennsylvania and California between 1974–83

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

[7]From California and Pennsylvania administrative records, all deaths 1974-1983. Data sources and methods documented in Hearst, Newman and Hulley (1986). NOTE: Sample sizes differ from Hearst, et al., because non-US-born are included

by draft–eligibility.[8] Columns 3 and 4 show the probability of death and suicide respectively, computed as the number of deaths divided by the population at risk estimated using the 1970 census.[9] Column 5 shows the probability of veteran status by year of birth and draft-eligibility status, estimated from the 1984 Survey of Income and Program Participation (SIPP).[10] In columns 3–5 the entry in the third row gives the difference in frequency of death, suicide and veteran status between those who were draft eligible and those who where not. The fourth row in columns 3 and 4 gives the ratio of these differences to the difference in the probability of being veteran by draft eligibility. As an example consider the men born in 1950. Of the men that were draft eligible, 35.3% actually served in the military. Of those that were not draft eligible only 19.3% served in the military. Standard application of the Rubin causal model suggests that the draft had a causal effect that increased the likelihood of serving by an estimated 15.9% on average. Similarly, of those that were draft eligible, 2.04% died between 1974 and 1983, compared with 1.95% of those that were not. The difference of 0.09% can be interpreted as an estimate of the average causal effect of the draft lottery number on civilian mortality. Assuming that these estimated causal effects are the population averages, the ratio of these two causal effects is, under the assumptions made above, the causal effect of military service for the 15.9% who were induced by the draft to serve in the military. For this group, the average causal effect is 0.56%, which amounts to a

[8]The mortality figures are tabulated from the data set analyzed by Hearst, Newman and Hulley (1986)

[9]The estimated population at risk is the author's tabulation of the total number of white men born in 1950 and 1951 in California and Pennsylvania from the 1970 census (IICPSR Study number 18). Estimates by draft–eligibility status are computed assuming a uniform distribution of lottery numbers.

[10]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 mortality. 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.

25% increase in the probability of death for this group (from 1.95% to 2.51%).[11]

To investigate the sensitivity of the estimates to violations of the exclusion restriction, suppose, for example, that Assumption 2 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 graduates student deferments were eliminated in 1967 and new undergraduate deferments were eliminated in December 1971, many of the men at risk in the 1970 and 1971 draft lotteries could, in theory, have postponed conscription by staying in school.

Working with special versions of the March 1979 and March 1981-85 Current Population Surveys, Angrist and Krueger (1992b) show that men born in 1951 with lottery numbers 1-75 have completed .358 more years of schooling than men with higher lottery numbers above 150 who were at no risk of being drafted.[12]

How much bias in estimates of the effect of military service on mortality is this correlation between lottery numbers and schooling likely to generate? To answer 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 sciences.[13] For the purposes of illustration, we have taken estimates from Duleep's (1986) study of socioeconomic variables and mortality using men surveyed in

[11]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 (1992)

[12]The standard error of this estimate is .147. Estimates of the effect of lottery numbers on schooling are variable and imprecise for most of the age cohorts at risk of being drafted. There is evidence of a significant correlation, however, for men born in 1945 and 1946 (at risk in the 1970 lottery) and men born in 1951 (at risk in the 1971 lottery).

[13]An early study in this area is Kitagawa and Hauser 1973.

the March 1973 CPS and linked 1973-78 Social Security data. Estimates presented in Table 1 of Duleep (1986) suggest that white married men 25 years old with 1-3 years of college have mortality rates roughly .0017 *higher* than do high school graduates.[14]

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

This calculation provides useful cautionary background 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)

[14]This figure comes from a regression of mortality on age, income and interactions of age with income.

estimate used here is not actually significantly different from zero), and this relationship is also subject to sign reversals. For example, while men with some college have higher mortality than high school graduates, the Duleep paper shows almost no difference between the mortality of high school and college graduates. Thus, a calculation based solely on graduates would indicate no bias.

6. Conclusion

In this paper we have outlined a framework for causal inference in settings where random assignment has taken place, but compliance is not perfect, i.e. the treatment received is nonignorable. In an attempt to avoid the bias associated with simply comparing subjects by the randomized treatment assignment, i.e. 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 this technique can be fitted into the Rubin Causal Model and be used for causal inference without assuming constant treatment effects. The advantages of embedding this approach in the Rubin Causal Model are that it makes the nature of the identifying assumptions more transparent, and that it allows us to consider sensitivity of the results to deviations from these assumptions in a straightforward manner.

References

AMEMIYA, T., (1985), *Advanced Econometrics*, Harvard University Press, Cambridge MA.

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

— , (1991), "Instrumental Estimation of Average Treatment Effects in Econometrics and Epidemiology," National Bureau of Economic Research Technical Working Paper No. 115. November.

— AND A. KRUEGER, (1991), "Does Compulsory School Attendance Affect Schooling and Earnings", *Quarterly Journal of Economics*, 106, 979-1014.

— AND —, (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, June.

— AND —, (1992b), "Estimating the Payoff to Schooling Using the Vietnam-Era Draft Lottery," National Bureau of Economic Research Working Paper No. 4067, May 1992.

BOWDEN, R.J., AND D.A. TURKINGTON, (1984), *Instrumental Variables*, Cambridge: Cambridge University Press.

COCHRAN, (1965), "The Planning of Observational Studies of Human Populations (with Discussion)," *Journal of the Royal Statistical Society, Series A* 128, 234-255.

Coronary Drug Project Research Group (1980), "Influence of Adherence to Treatment and Response of Cholesterol on Mortality in the Coronary Drug Project," *New England Journal of Medicine* 303, 1038-1041.

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

DE TRAY, D., (1982), "Veteran Status as a Screening Device." *American Economic Review* 72. 133-142.

DULEEP, HARRIET ORCUTT, (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 D. FELDMAN, (1991), "Compliance as an Explanatory Variable in Clinical Trials", *Journal of the American Statistical Association*, 86, 9-26.

ELLENBERG, S.S., D.M. FINKELSTEIN, AND D.A. SCHOENFELD, "Statistical Issues Arising in AIDS Clinical Trials," *Journal of the American Statistical Association* 87(418). 562-583.

FIENBERG, S., (1971), "Randomization and Social Affairs: The 1970 Draft Lottery," *Science* 171, January.

FISHER, R., (1918), "The Causes of Human Variability", *Eugenics Review*, Vol 10, 213-220.

GAIL, M., (1985), "Eligibility Exclusions, Losses to Follow-up, Removal of Randomized Patients. and Uncounted Events in Cancer Clinical Trials," *Cancer Treatment Reports*, 69(10). 11-7-1112.

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

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

HEARST, N., NEWMAN, T., AND S. HULLEY, (1986), "Delayed Effects of the Military Draft on Mortality: A Randomized Natural Experiment," *New England Journal of Medicine*, 314

(March 6), 620-624.

HECKMAN, AND R. ROBB, (1985), "Alternative Methods for Evaluating the Impact of Interventions," in J. Heckman and B. Singer, eds., *Longitudinal Analysis of Labor Market Data*. New York: Cambridge University Press.

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

HECKMAN, J. J. (1992), "Haavelmo and the Birth of Modern Econometrics: A Review of The History of Econometric Ideas by Mary Morgan," *Journal of Economic Literature*, 330(2). June.

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." Chapter 13 in: *Sociological Methodology*, Washington: American Sociological Association.

IMBENS, G., AND J. ANGRIST, (1991), "Identification and Estimation of Local Average Treatment Effects", NBER Technical Working Paper no 118, December.

JABLON, AND SELTZER, (1975), "Effect of Selection on Mortality," *American Journal of Epidemiology*, 100, 367-372.

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

KITAGAWA, E.M., AND P.M. HAUSER, (1973), *Differential Mortality in the United States: A Study in Socioeconomic Epidemiology*, Cambridge: Harvard University Press.

KLEINBAUM, D.G., L.L KUPPER, AND H. MORGENSTERN, (1982), *Epidemiological Research*. New York: Van Nostrand Reinhold.

LAVORI, P., (1992), "Clinical Trials in Psychiatry: Should Protocol Deviation Censor Patient

Data", *Neuropsychopharmacology*, Vp; 6, No 1.

LEAMER, E.E., (1988), "Discussion," Chapter 14 in: *Sociological Methodology*. Washington: American Sociological Association.

LEWIS, H. G.,, (1986), *Union Relative Wage Effects*, Chicago, University of Chicago Press.

MANSKI, C. F., (1992), "The Selection Problem," in *Advances in Econometrics*, edited by C. Sims, New York: Cambridge University Press.

NATHAN, RICHARD, (1988), *Social Science in Government*, New York: Basic Books.

NEYMAN, J., (1923), "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9," translated in *Statistical Science*, Vol 5, No 4, 465–480. 1990.

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

POWERS, D. E., AND S. S. SWINTON, (1984), "Effects of Self-Study for Coachable Test Item Types," *Journal of Educational Psychology* 76, 266-78.

REIERSOL, O. (1941), "Confluence Analysis by Means of Lag Moments and Other Methods of Confluence Analysis," *Econometrica* 9, 1-24.

ROBINS, JAMES M. (1989), "The Analysis of Randomized and Non–Randomized AIDS Treatment Trials Using a New Approach to Causal Inference in Longitudinal Studies," Health Service Research Methodology: A Focus on AIDS, edited by L. Sechrest, H. Freeman, and A. Bailey, NCHSR, U.S. Public Health Service.

ROBINS, J. M., AND A. A. TSIATIS, "Correcting for Non–Compliance in Randomized Trials Using Rank–Preserving Structural failure Time Models," *Communications in Statistics - Theory and Methods*, 20(8), 2069-2631.

ROSENBAUM, P., AND D. RUBIN, (1983), "Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome," *Journal of the Royal Statistical Society, Series B*, 45, 212-218.

ROY, A. (1951), "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers* 3, 135-46.

RUBIN, D. (1974), "Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies," *Journal of Educational Psychology*, 66, 688-701.

—, (1977), "Assignment to a Treatment Group on the Basis of a Covariate", *Journal of Educational Statistics*, 2, 1-26.

—, (1978), "Bayesian inference for causal effects", *Annals of Statistics*, 6:34-58.

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

—, (1990b), "Formal Modes of Statistical Inference for Causal Effects", *Journal of Statistical Planning and Inference*, Vol. 25, 279-292.

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

SEXTON, AND J. HEBEL, (1984), "A Clinical Trial of the Change in Maternal Smoking and its Effect on Birth Weight," *Journal of the American Medical Association*, 251(7), (February 17), 911-915.

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.