Bounds on Treatment Effects From Studies With Imperfect Compliance

Alexander BALKE and Judea PEARL

This article establishes nonparametric formulas that can be used to bound the average treatment effect in experimental studies in which treatment assignment is random but subject compliance is imperfect. The bounds provided are the tightest possible, given the distribution of assignments, treatments, and responses. The formulas show that even with high rates of noncompliance, experimental data can yield useful and sometimes accurate information on the average effect of a treatment on the population.

KEY WORDS: Causal models; Latent variables; Linear programming; Noncompliance.

1. INTRODUCTION

Consider an experimental study in which random assignment has occurred but compliance is not perfect (i.e., the treatment received differs from that assigned). It is well known that under such conditions, a bias may be introduced. Subjects who did not comply with the assignment may be precisely those who would have responded adversely (positively) to the treatment; therefore, the actual effect of the treatment, when applied uniformly to the population, might be substantially less (more) effective than the study reveals.

In an attempt to avert this bias, analysts sometimes resort to parametric models that make restrictive commitments to a particular mode of interaction between compliance and response (Efron and Feldman 1991). Angrist, Imbens, and Rubin (1996) have identified a set of assumptions under which a nonparametric correction formula, called "instrumental variables," is valid for certain subpopulations. Because these subpopulations cannot be identified from empirical observation alone, the need remains to devise alternative, assumption-free formulas for assessing the effect of treatment over the population as a whole. Manski (1990) and Robins (1989) have derived such bounds, but did not make full use of the information available in the data. In this article we provide sharp (i.e., the tightest possible) bounds on the average treatment effect.

2. PROBLEM FORMULATION

The canonical partial-compliance setting can be graphically modeled as shown in Figure 1. We assume that Z, D, and Y are observed dichotomous variables, where Z represents the (randomized) treatment assignment, D is the treatment actually received, and Y is the observed response. U represents specific characteristics of an individual subject—namely, all factors, both observed and unobserved, that influence the way in which a subject's outcome Y may depend on the treatment D. The experimental study is modeled as a two-step process: (1) treatment selection and (2) treatment administration. In the first step, each subject is allowed to select a treatment in accordance with the following factors: the assignment (Z), basic physiological characteristics (U), and, possibly, initial reactions to the treatment or placebo. (Such reactions are not shown explicitly in the graph, because they merely modify the influence of Z and U on D, and the diagram makes no assumption as to the nature of this influence.) Once the treatment D is selected, the treatment administration step begins, during which subjects are assumed to remain within their selected treatment arms until the outcome Y is recorded; back and forth switching between placebo and active groups is not allowed at this stage.

Given this two-stage process, the second assumption is that the assignment (Z) per se does not alter any physiological characteristics (U) that determine how an individual would react to any given treatment. This assumption, which Angrist et al. (1996) termed "exclusion restriction" and Manski (1990) called "set-level restriction," is represented in the causal diagram of Figure 1 by the absence of a direct link from Z to Y or from Z to U; all paths between Z and Y go through D. (A fuller account of the statistical and causal implications of structural diagrams, and their relation to Rubin's model of counterfactuals [Holland 1988], was given in Pearl 1995a.)

To facilitate the notation, we let z, d, and y represent the values taken by the variables Z, D, and Y, with the following interpretation: $z \in \{z_0, z_1\}$, z_1 asserts that treatment has been assigned (z_0, its negation); $d \in \{d_0, d_1\}$; d_1 asserts that treatment has been administered (d_0, its negation); and $y \in \{y_0, y_1\}$, y_1 asserts a positive observed response (y_0, its negation). Multivalued or continuous outcomes can be easily accommodated in the model using the event $Y \leq y$ as a (dichotomous) outcome variable. Extension to continuous treatments is discussed in Section 3. The domain of U remains unspecified and may in general combine the spaces of several random variables, both discrete and continuous.

The model analyzed invokes two assumptions of independence:

Alexander Balke is a research associate and Judea Pearl is Professor, Computer Science Department, University of California, Los Angeles, CA 90024. This research was partially supported by Air Force grant AFOSR 90 0136, National Science Foundation grant IRI-9200918, and Northrop-Rockwell Micro grant 93-124. Alexander Balke was supported by the Fannie and John Hertz Foundation. This work benefited from discussions with Joshua Angrist, David Chickering, Thomas Ferguson, David Galles, Guido Imbens, James Robins, and Donald Rubin. The authors thank Bradley Efron for providing the data used in Section 4.2.

Journal of the American Statistical Association
September 1997, Vol. 92, No. 439, Theory and Methods

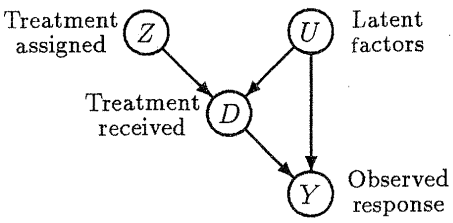

Figure 1. Graphical Representation of Causal Dependencies in a Randomized Clinical Trial With Partial Compliance.

1. For a given individual, the treatment assignment does not influence Y directly, but only through the actual treatment D; that is, $Z \perp\!\!\!\perp Y | \{D, U\}$.
2. Z and U are marginally independent; that is, $Z \perp\!\!\!\perp U$.

This independence is ensured partly through the randomization of Z, which rules out a common cause for both Z and U, and partly through our assumption that physiological factors (U) are not influenced by the assignment (Z).

These two independencies impose on the joint distribution the decomposition

$$P(y, d, z, u) = P(y|d, u)P(d|z, u)P(z)P(u),$$

which of course cannot be observed directly, because U is unmeasurable. (We take the liberty of denoting the prior distribution of U by $P(u)$, even though U may consist of continuous variables.) However, the marginal distribution $P(y, d, z)$, and in particular the conditional distributions

$$P(y, d|z) = \sum_u P(y|d, u)P(d|z, u)P(u) \tag{1}$$

$z \in \{z_0, z_1\}$, are observed, and these observations constrain the factor $P(y|d, u)P(u)$ to produce bounds on treatment effects.

Treatment effects are characterized by a distribution $P(y|\breve{d})$ that represents the probability that Y would have been equal to y, if D were equal to d under a randomized experiment. In general, a value annotated with a *check*(ˇ) will indicate that the corresponding variable has been set to that value by a randomized control. (Angrist et al. [1996] and Holland [1988] denoted this distribution by $P(Y_{D=d})$, but we find the "check" notation more flexible, as it permits one to specify explicitly what is controlled and what is allowed to vary in any given study [Pearl 1995a].) Thus to assess the distribution of Y if the treatment D were applied uniformly to the population, we should calculate

$$P(y|\breve{d}) \triangleq \sum_u P(y|d, u)P(u), \tag{2}$$

where the factors $P(y|d, u)$ and $P(u)$ are the same as those in (1). Similarly, if we are interested in the average *change* in Y due to treatment, then we should compute the average *causal effect*, ACE($D \to Y$) (Holland 1988), given by

$$\begin{aligned} \text{ACE}(D \to Y) &\triangleq P(y_1|\breve{d}_1) - P(y_1|\breve{d}_0) \\ &= \sum_u P(u)[P(y_1|d_1, u) - P(y_1|d_0, u)]. \end{aligned} \tag{3}$$

Our task is then to estimate or bound the expressions in (2) and (3), given the observed probabilities $P(y, d|z_0)$ and $P(y, d|z_1)$, as expressed in (1). This may be accomplished by a procedure that we detailed in earlier work (Balke and Pearl 1994), which is based on linear programming optimization coupled with the fact that the domain of U can be partitioned into 16 equivalence classes, each representing one of four possible mappings from Z to D conjoined with one of four possible mappings from D to Y.

3. RESULTS

Let the conditional distribution $P(y, d|z)$ over the observed variables be denoted as follows:

$$\begin{aligned} p_{00.0} &= P(y_0, d_0|z_0) \quad p_{00.1} = P(y_0, d_0|z_1) \\ p_{01.0} &= P(y_0, d_1|z_0) \quad p_{01.1} = P(y_0, d_1|z_1) \\ p_{10.0} &= P(y_1, d_0|z_0) \quad p_{10.1} = P(y_1, d_0|z_1) \\ p_{11.0} &= P(y_1, d_1|z_0) \quad p_{11.1} = P(y_1, d_1|z_1) \end{aligned}$$

Optimization of (2) subject to the equality constraints given in (1) defines a linear programming problem that yields a closed-form solution by enumerating all vertices of the constraint polygon of the dual problem. This procedure leads to the following bounds:

$$\max \left\{ \begin{array}{c} p_{10.1} \\ p_{10.0} \\ p_{10.0} + p_{11.0} - p_{00.1} - p_{11.1} \\ p_{01.0} + p_{10.0} - p_{00.1} - p_{01.1} \end{array} \right\} \leq P(y_1|\breve{d}_0)$$

$$\leq \min \left\{ \begin{array}{c} 1 - p_{00.1} \\ 1 - p_{00.0} \\ p_{01.0} + p_{10.0} + p_{10.1} + p_{11.1} \\ p_{10.0} + p_{11.0} + p_{01.1} + p_{10.1} \end{array} \right\}$$

and

$$\max \left\{ \begin{array}{c} p_{11.0} \\ p_{11.1} \\ -p_{00.0} - p_{01.0} + p_{00.1} + p_{11.1} \\ -p_{01.0} - p_{10.0} + p_{10.1} + p_{11.1} \end{array} \right\} \leq P(y_1|\breve{d}_1)$$

$$\leq \min \left\{ \begin{array}{c} 1 - p_{01.1} \\ 1 - p_{01.0} \\ p_{00.0} + p_{11.0} + p_{10.1} + p_{11.1} \\ p_{10.0} + p_{11.0} + p_{00.1} + p_{11.1} \end{array} \right\}.$$

In addition, if we optimize the difference of the two terms in (3) by the same linear programming technique, then we find that the expressions for the upper and lower bounds on the average causal effect ACE($D \to Y$) are equal to the difference of the corresponding bounds on the individual terms; that is, the lower bound on ACE($D \to Y$) is equal to $P(y_1|\breve{d}_1)$'s lower bound less $P(y_1|\breve{d}_0)$'s upper bound, and the upper bound on ACE($D \to Y$) is equal to $P(y_1|\breve{d}_1)$'s upper bound less $P(y_1|\breve{d}_0)$'s lower bound. The resulting formulas are

$$\text{ACE}(D \to Y)$$

$$\geq \max \left\{ \begin{array}{c} p_{00.0} + p_{11.1} - 1 \\ p_{00.1} + p_{11.1} - 1 \\ p_{11.0} + p_{00.1} - 1 \\ p_{00.0} + p_{11.0} - 1 \\ 2p_{00.0} + p_{11.0} + p_{10.1} + p_{11.1} - 2 \\ p_{00.0} + 2p_{11.0} + p_{00.1} + p_{01.1} - 2 \\ p_{10.0} + p_{11.0} + 2p_{00.1} + p_{11.1} - 2 \\ p_{00.0} + p_{01.0} + p_{00.1} + 2p_{11.1} - 2 \end{array} \right\} \quad (4)$$

and

$$\text{ACE}(D \to Y)$$

$$\geq \min \left\{ \begin{array}{c} 1 - p_{10.0} - p_{01.1} \\ 1 - p_{01.0} - p_{10.1} \\ 1 - p_{01.0} - p_{10.0} \\ 1 - p_{01.1} - p_{10.1} \\ 2 - 2p_{01.0} - p_{10.0} - p_{10.1} - p_{11.1} \\ 2 - p_{01.0} - 2p_{10.0} - p_{00.1} - p_{01.1} \\ 2 - p_{10.0} - p_{11.0} - 2p_{01.1} - p_{10.1} \\ 2 - p_{00.0} - p_{01.0} - p_{01.1} - 2p_{10.1} \end{array} \right\}. \quad (5)$$

These bounds represent substantial improvement over those derived by Robins (1989) and Manski (1990), which correspond to the four upper terms in both (4) and (5). One can show that the width of the bounds in (4) and (5) cannot exceed the rate of noncompliance, $P(d_1|z_0) + P(d_0|z_1)$, and may in some cases collapse to a point estimate, even when as many as 50% of subjects switch over to unassigned treatments (Pearl 1995b). Precise determination of treatment effects is feasible whenever (a) the percentage of subjects complying with assignment z_0 is the same as those complying with z_1 and (b) in at least one treatment arm d, y, and z are perfectly correlated.

This and other results regarding bounds on treatment effects in partial compliance studies are elaborated in earlier work (Balke 1995; Balke and Pearl 1993). In particular, it is shown that the basic structural assumptions underlying randomized-assignment experiments, although not directly testable, imply testable restrictions on the observed distributions. By requiring that no upper bound be less than the corresponding lower bound, we obtain

$$\begin{aligned} P(y_0, d_0|z_0) + P(y_1, d_0|z_1) &\leq 1, \\ P(y_0, d_1|z_0) + P(y_1, d_1|z_1) &\leq 1, \\ P(y_1, d_0|z_0) + P(y_0, d_0|z_1) &\leq 1, \end{aligned}$$

and

$$P(y_1, d_1|z_0) + P(y_0, d_1|z_1) \leq 1. \quad (6)$$

If any of these inequalities is violated, then the investigator can deduce that either the assignments were not properly randomized or the assignment exerted some direct influence on subjects' responses. These inequalities, when generalized to multivalued variables, assume the simple form

$$\max_d \sum_y \max_z P(y, d|z) \leq 1,$$

which was called the *instrumental inequality* in earlier work (Pearl 1994).

The instrumental inequality can be further tightened if additional assumptions are made about subjects' behaviors; for example, that no individual would consistently act contrarian to his or her assignment, or, mathematically, that for all u we have

$$P(d_1|z_1, u) \geq P(d_1|z_0, u).$$

Under this assumption, which Angrist et al. (1996) called monotonicity, the inequalities in (6) can be tightened (Balke and Pearl 1993) to give

$$P(y, d_1|z_1) \geq P(y, d_1|z_0)$$

and

$$P(y, d_0|z_0) \geq P(y, d_0|z_1) \quad (7)$$

for all $y \in \{y_0, y_1\}$. The monotonicity assumption can sometimes be verified (or enforced) empirically; for example, by making sure that no subject in the placebo group gains access to active treatment. In such cases, (7) provides more stringent tests for the model assumptions. However, in cases where monotonicity cannot be ensured, violation of the inequalities in (7) may mean that randomization (of Z) was imperfect, that Z has a direct effect on Y, or that contrarian subjects were present.

It can also be shown (Balke and Pearl 1993) that when monotonicity holds, the bounds in (4) and (5) reduce to those derived by Manski (1996) and Robins (1989) (the first four entries in (4) and (5)), and the width coincides precisely with the rate of noncompliance, $P(d_1|z_0) + P(d_0|z_1)$.

Finally, the method of causal analysis we outline here permits one to evaluate a wide variety of counterfactual probabilities; for example, the probability that a given individual would have recovered had he or she not been assigned treatment (z_0), when in actuality he or she has been assigned the treatment (z_1), taken the treatment (d_1), and not recovered (y_0). This intricate probability can be bounded by analyzing the causal effect of the assignment in the subpopulation characterized by $\{z_1, d_1, y_0\}$. We detailed a general method for obtaining such bounds in earlier work (Balke and Pearl 1994).

It is possible to extend this analysis to studies in which treatment may take on more than two values by simply reformulating the linear programming problem over a multivalued variable D. However, this method becomes computationally expensive, because the number of equivalence classes in the U domain increases exponentially with the cardinality of D. Alternatively, using the same linear programming techniques as in the case of dichotomous treatment, one can derive bounds on the difference in causal efficacy of any two treatment levels, say d_0 and d_1, while allowing subjects receiving treatment levels other than d_0 and d_1 (denoted by d_m) to exhibit arbitrary behavior. Remarkably, the bounds derived in this way, letting $d \in \{d_0, d_1, d_m\}$, are expressed identically to (4) and (5), though no assumptions whatsoever have been made about the composition of

d_m or the relation of any values in d_m to Y (Balke 1995). These bounds represent the worst case (least informative) behavior of subjects in the d_m category and are implicitly affected by the size of the d_m category through the equality $P(d_0|z) + P(d_1|z) + P(d_m|z) = 1$.

When the treatment is continuous, few if any subjects would take on any given level of treatment precisely. However, it is reasonable to assume that there exists a treatment interval around each d within which the subject's outcome is, for all practical purposes, homogeneous. In other words, for every u we have $P(y|d', u) \sim P(y|d'', u)$ for all $d', d'' \in [d - \delta, d + \delta]$. Under this assumption, which obviously becomes more reasonable as δ decreases, it is possible to apply our previous analysis and derive bounds on the average change in treatment effect between any two treatment levels. This is illustrated in the next section.

4. EXAMPLES

4.1 Vitamin A Supplementation

Consider the study of vitamin A supplementation in northern Sumatra described by Sommer et al. (1986) and Sommer and Zeger (1991). In this study, out of 450 villages, 221 were randomly assigned to the control group and the remaining 229 were assigned to the treatment group. In the treatment group, oral doses of vitamin A were administered in the population at 2–3 months and once again at 6 months; because of government policy, the control group was not administered a placebo. At 12 months after the original census, the mortality (y_0) of the population was determined from the time at which the initial dose was administered. Table 1 presents the final subject counts in terms of our partial compliance model notation. Table 2 presents the probability distribution estimated from the counts in Table 1, making the large-sample assumption and taking the sample frequencies as representing $P(y, d|z)$.

By computing the quantities required for (4) and (5), we obtain

$$\begin{aligned}\text{ACE}(D \to Y) &\geq \max\left\{\begin{matrix}-.1946, -.1982, -.9972, -.9936,\\ -.9910, -1.9898, -.2018, -.3928\end{matrix}\right\}\\ &= -.1946\end{aligned}$$

and

$$\begin{aligned}\text{ACE}(D \to Y) &\leq \min\left\{\begin{matrix}.0054, .8028, .0064, .8018,\\ .0102, .0090, .8072, 1.5982\end{matrix}\right\}\\ &= .0054.\end{aligned}$$

Table 1. Count of Children Classified According to Treatment Assigned (z), Treatment Consumed (d), and Mortality Outcome (y).

	z_0		z_1	
N(y,d,z)	y_0	y_1	y_0	y_1
d_0	74	11,514	34	2,385
d_1	0	0	12	9,663

Table 2. Conditional Probability Distribution P(y,d|z) Derived From the Data in Table 1.

	z_0		z_1	
P(y,d|z)	y_0	y_1	y_0	y_1
d_0	.0064	.9936	.0028	.1972
d_1	0	0	0	.7990

Accordingly, we conclude that the average treatment effect lies in the range

$$-.1946 \leq \text{ACE}(D \to Y) \leq .0054,$$

which is rather revealing: Vitamin A supplement, if uniformly administered, is seen as capable of increasing mortality rate by much as 19.46% and is incapable of reducing mortality rate by more than .54%. The intent-to-treat analysis might mislead one to believe that vitamin A supplement has a beneficial effect of $P(y_1|z_1) - P(y_1|z_0) = .0026$, in total oblivion to the danger presented at the lower end of the range. The instrumental variables estimate advocated by Angrist et al. (1996) calculates to .0035, which further exaggerates the illusionary benefits of vitamin A supplement.

The techniques described in earlier work (Balke and Pearl 1994) may also be used to find a population mix that would explain a particular value of the causal effect magnitude. For example, one may wish to investigate the behavioral characteristics, consistent with the observed data, that would support a detrimental effect of $\text{ACE}(D \to Y) = -.1946$ shown possible at the extreme lower end of the range. For the most part, the population under study would have to be composed of two homogeneous groups. In one group, consisting of almost 80% of the population, all subjects would survive regardless of treatment and would comply perfectly with their treatment assignment. In the other group, consisting of almost 20% of the population, subjects would die if (and only if) they take vitamin A supplements, and, not surprisingly, these subjects would refuse vitamin A supplements under the conditions prevailing in the study. The ability to associate a population mix with any ACE value provides a vantage point from which the plausibility of that ACE value can be assessed.

4.2 Coronary Primary Prevention Trial

Consider the Lipid Research Clinics Coronary Primary Prevention Trial data. (See Lipid Research Clinic Program 1984 for an extended description of the clinical trial.) A portion of this dataset consisting of 337 subjects was analyzed by Efron and Feldman (1991) using a parametric model; we use the same dataset in our analysis. A population of subjects was assembled, and two preliminary cholesterol measurements were obtained: one prior to a suggested low-cholesterol diet (continuous variable C_{I1}) and one following the diet period (C_{I2}). The initial cholesterol level (C_I) was taken as a weighted average of these two measures: $C_I = .25C_{I1} + .75C_{I2}$. The subjects were randomized into two treatment groups: subjects receiving cholestyramine (z_1) and subjects receiving a placebo (z_0). During several years of treatment, each subject's choles-

terol level was measured multiple times, and the average of these measurements was used as the posttreatment cholesterol level (continuous variable C_F). The compliance of each subject was determined by tracking the quantity of prescribed dosage consumed (continuous variable B). The maximum consumption in the dataset was 101 units.

To apply our analysis to this study, we discretize the continuous data obtained in the Lipid Research Clinic Program (1984) study in the following way:

$$d = \begin{cases} d_0 & \text{if } z = z_0 \text{ or } b = 0 \\ d_1 & \text{if } z = z_1 \text{ and } 87 \le b \le 101 \\ d_m & \text{otherwise} \end{cases} \tag{8}$$

and

$$y = \begin{cases} y_0 & \text{if } c_I - c_F < 38 \\ y_1 & \text{if } c_I - c_F \ge 38. \end{cases} \tag{9}$$

This discretization assumes that each subject's response to treatment is homogeneous between 87 and 101 units of cholestyramine. In addition, (8) reflects the finding that subjects assigned placebo (z_0) did not take cholestyramine—namely, $P(d_1|z_0) = P(d_m|z_0) = 0$. The threshold of 38 in (9) was chosen arbitrarily. Clearly, by varing this threshold over the range of Y, one obtains upper and lower bounds on the entire distribution of the treatment effect, $P(Y \le y|\check{d}_1) - P(Y \le y|\check{d}_0)$.

If the sample data are interpreted according to (8) and (9), then the conditional distribution over (Z, D, Y) results in the distribution given in Table 3. (Here we make the large-sample assumption and take the sample frequencies as representing $P(y, d|z)$.)

By computing the quantities required for (4), we obtain

$$\begin{aligned} \text{ACE}(D \to Y) &\ge \max \left\{ \begin{matrix} .262, -.685, -.976, -.029, \\ .233, -.902, -1.632, -.423 \end{matrix} \right\} \\ &= .262. \end{aligned}$$

Those needed for (5) give us

$$\text{ACE}(D \to Y) \le \min \left\{ \begin{matrix} .868, 1.000, .971, .897, \\ 1.680, 1.815, 1.765, .926 \end{matrix} \right\} = .868.$$

Accordingly, we conclude that the average treatment effect lies in the range

$$.262 \le \text{ACE}(D \to Y) \le .868,$$

which is quite informative. The experimenter can categorically state that when applied uniformly to the population, a cholestyramine dosage of 84–101 units is guaranteed to increase by at least 26.2% the probability of reducing a patient's level of cholesterol by 38 points or more. This guarantee is established despite the fact that 60.6% of the subjects in the treatment group did not comply with their assigned dosage level. For comparison, note that the intent-to-treat analysis in this study gives $P(y_1|z_1) - P(y_1|z_0) = .408$, meaning that enforcing full compliance might result in as much as 46% improvement and no more than 14.6% reduction in the proportion of patients benefiting from the treatment.

Table 3. Conditional Probability Distribution P(y,d|z) for the Lipid Research Clinic Program (1984) Data, made Discrete by (8) and (9).

	z_0		z_1	
$P(y,d\|z)$	y_0	y_1	y_0	y_1
d_0	.971	.029	.024	0
d_m	0	0	.436	.146
d_1	0	0	.103	.291

5. CONCLUSION

In an attempt to avert confounding bias in randomized studies involving noncompliance, analysts usually advocate using "intent-to-treat" analysis, which compares assignment groups regardless of the treatment actually received. Estimates derived by such analysis are free of confounding bias, but decisions based on these estimates require that the experimental conditions perfectly mimic the conditions prevailing in the final implementation of the treatment. In particular, the intent-to-treat analysis is inappropriate when the inducement to receive treatment changes from what it was in the study—for example, when a drug is officially endorsed by a well-meaning authority.

A similar weakness applies to the analysis of Angrist et al. (1996), who derived causal effect formulas for the unobservable subpopulation of "responsive" subjects; that is, subjects who would have changed treatment status if given a different assignment. This subpopulation cannot serve as a basis for policy analysis because it is instrument dependent—individuals who are responsive in the study may not remain responsive in the field, where the incentives for obtaining treatment differ from those used in the study.

In policy evaluation studies, field incentives are normally more compelling than experimental incentives. Hence treatment effectiveness should be assessed by the average causal effect, $E_u[P(y_1|u, d_1) - P(y_1|u, d_0)]$, for which we have provided sharp theoretical bounds. Estimates based solely on intent-to-treat analysis, as well as those based on instrumental variables, can be misleading, as they may lie entirely outside the theoretical bounds. The formulas established in this article provide instrument-independent guarantees for policy analysis and should enable analysts to determine the extent to which efforts to enforce compliance may increase the overall treatment effectiveness.

A topic that should receive considerable attention in future work is the augmentation of the bounds in (4)–(5) with confidence intervals, to account for sample variability. Chickering and Pearl (1996) described a Bayesian method that, using Gibbs sampling, computes the posterior distribution of $\text{ACE}(D \to Y)$ given the data. An alternative approach in this direction is offered by the maximum likelihood ratio test, as applied to the hypothesis H_0: $\text{ACE}(D \to Y) < t$, for arbitrary t, because the maximum likelihood function under H_0 can be computed using linear programming.

[Received October 1993. Revised December 1996.]

REFERENCES

Angrist, J. D., Imbens, G. W. and Rubin, D. B. (1996), "Identification of Causal Effects Using Instrumental Variables" (with comments), *Journal of the American Statistical Association*, 91, 444–472.

Balke, A. (1995), "Probabilistic Counterfactuals: Semantics, Computation, and Applications," Technical Report R-242, University of California, Los Angeles, Computer Science Dept.

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.

———, (1994), "Counterfactual Probabilities: Computational Methods, Bounds and Applications," in *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*, San Francisco: Morgan Kauffman, pp. 46–54.

Chickering, D. M., and Pearl, J. (1996), "A Clinician's Tool for Analyzing Non-Compliance," in *Proceedings of the National Conference on Artificial Intelligence (AAAI-96)*, Boston: Morgan Kauffman, pp. 1269–1276.

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

Holland, P. W. (1988), "Causal Inference, Path Analysis, and Recursive Structural Equations Models," in *Sociological Methodology*, ed. C. Clogg, Washington, DC: American Sociological Association, pp. 449–484.

Lipid Research Clinic Program (1984), "The Lipid Research Clinics Coronary Primary Prevention Trial Results, Parts I and II," *Journal of the American Medical Association*, 251, 351–374.

Manski, C. F. (1990), "Nonparametric Bounds on Treatment Effects," *American Economic Review, Papers and Proceedings*, 80, 319–323.

Pearl, J. (1994), "On the Testability of Causal Models With Latent and Instrumental Variables," in *Uncertainty in Artificial Intelligence 11*, eds. P. Besnard and S. Hanks, San Francisco: Morgan Kaufmann, pp. 435–443.

——— (1995a), "Causal Diagrams for Experimental Research," (with discussion, *Biometrika*, 82, 669–710.

——— (1995b), "Causal Inference From Indirect Experiments," *Artificial Intelligence in Medicine*, 7, 561–582.

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. L. Sechrest, H. Freeman, and A. Mulley, Washington, DC: U.S. Public Health Service, pp. 113–159.

Sommer, A., Tarwotjo, I., Djunaedi, E., West, K. P., Leodin, A. A., Tilden, R., and Mele, I. (1986), "Impact of Vitamin A Supplementation on Childhood Mortality: A Randomized Controlled Community Trial," *The Lancet*, 1169–1173.

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