Bounds on Causal Effects in Three-Arm Trials with Noncompliance

Jing Cheng, Dylan S. Small

University of Pennsylvania, Philadelphia, PA 19104.

Summary. This paper considers the analysis of three-arm randomized trials with noncompliance. In these trials, the average causal effects of treatments within principal strata of compliance behavior are of interest for better understanding the effect of the treatment. Unfortunately, even with usual assumptions, the average causal effects of treatments within principal strata are not point-identified. However, the observable data does provide useful information on the bounds of the identification regions of the parameters of interest. Under two sets of assumptions, we derive sharp bounds for the causal effects within principal strata for binary outcomes, and construct confidence intervals to cover the identification regions. The methods are illustrated by an analysis of data from a randomized study of treatments for alcohol dependence.

Keywords: Three-arm randomized trials; noncompliance; causal effect; principal strata; bounds; confidence interval.

1. Introduction

Many randomized trials with human subjects suffer from noncompliance to assigned treatment. There is a large literature on methods of analysis for two-arm trials with noncompliance (e.g., Angrist, Imbens and Rubin, 1996; Imbens and Rubin, 1997; Goetghebeur and Molenberghs, 1996; Small et al., 2006). Three-arm trials are common in practice but not much attention has been paid to analyzing these trials with noncompliance. In this paper, we use the principal stratification approach of Frangakis and Rubin (2002) to define principal strata for a three-arm trial with noncompliance and then derive sharp bounds on causal effects within principal strata under two sets of assumptions. We provide several reasons why these causal effects within principal strata are of interest. Our motivating example is a three-arm trial of treatments for alcohol dependence, and we analyze this trial using our methodology.

The randomized trial of treatments for alcohol dependence that we consider consists of two active treatments and a control. The control denoted by 0 is simple medication management, in which that arm's primary care physicians are responsible for providing usual care. The two active treatments are (A) simple medication management plus Compliance Enhancement Therapy (CET), in which therapists use motivational interviewing techniques, provide education about the treatment and help the patient develop strategies to improve compliance with taking medication and attending clinic visits; and (B) simple medication management plus Cognitive Behavioral Therapy (CBT), in which therapists instruct subjects on how to monitor and cope with situations that put them at high risk for relapse to alcohol use. Compliance with an active treatment was categorized as a binary variable, whether or not the subject attended at least 80% of the scheduled sessions. This categorization is based on the clinicians' understanding of the mechanism by which

the treatment works. Patients randomized to the control did not have access to the CET or CBT therapist, and patients randomized to CET/CBT could not access the CBT/CET therapist. The outcome of interest is whether or not the subject relapsed, i.e., had five or more drinks in a day, on any day in the past month (a relapse is coded as a 0 and lack of relapse is coded as a 1). The trial is described in more detail in Zisserson et al. (2004).

For a two-arm randomized trial with noncompliance in which the control group cannot receive the treatment, Angrist, Imbens and Rubin (1996) provided an analytical approach that estimates the average causal effect of receiving the treatment for those subjects who would comply with the treatment if offered it. A three-arm trial can be analyzed as two two-arm trials, one comparing the control group to treatment A and one comparing the control group to treatment B. The Angrist, Imbens and Rubin approach estimates the average treatment effect for those subjects who would comply with treatment A if offered it (group A) and the average treatment effect for those subjects who would comply with treatment B if offered it (group B). However, additional analytical possibilities are available if the trial is viewed as one three-arm trial rather than two two-arm trials. The principal stratification (Frangakis and Rubin, 2002) of subjects with respect to treatment received divides the subjects into four groups based on their potential compliance behavior under assignment to treatment A and to treatment B (under assignment to control, all subjects are assumed to take no treatment). The four principal strata are those subjects who would comply with both treatments A and B (group $0AB$), those subjects who would comply only with treatment A (group $0A0$), those subjects who would comply only with treatment B (group $00B$) and those subjects who would comply with neither treatment A nor treatment B (group 000). Note that we observe a subject's compliance behavior under at most one treatment so that a subject's principal strata is never fully observed. The subjects who would comply with treatment A if assigned it in a two-arm trial (group A) are a mixture of groups $0A0$ and $0AB$, and the subjects who would comply with treatment B in a two-arm trial (group B) are a mixture of groups $00B$ and $0AB$. Groups for which membership is determined by a subject's full compliance behavior (e.g., groups 000, $0A0$, $00B$ and $0AB$) are called basic principal strata. Groups that combine basic principal strata (e.g., groups A and B) are called coarsened principal strata (Frangakis and Robin, 2002).

We now describe three settings in which the average causal effects of treatments within the basic principal strata $000, 0A0, 00B$ and $0AB$ provide valuable information beyond that of the average causal effects of treatment within the coarsened principal strata A and B. Note that by the average causal effect of treatment for a principal stratum, we mean the average difference in the potential outcomes under treatment and control for members of the principal stratum. For conciseness, we sometimes call the average causal effect of treatment the average treatment effect. One reason that knowing the average treatment effects for groups $0A0$, $00B$ and $0AB$ is of interest beyond that of knowing the average treatment effects for groups A and B is that the former average treatment effects are of more relevance to a subject who is reasonably confident about what her compliance behavior would be if offered the different treatments. Although it is difficult for a subject to know what her compliance behavior would be, a subject often has enough information to regard her compliance behavior as not being exchangeable with the population's. Knowing the average treatment effects within principal strata allows a subject to utilize her information about her expected compliance behavior to better predict her individual treatment effects. For example, suppose a subject is reasonably confident that she would take treatment A if

suggested it and would take treatment B if suggested it because her usual behavior is to follow a doctor's suggestions. Then the average treatment effect of treatment A for group $0AB$ is a better prediction of this subject's treatment effect for taking treatment A than is the average treatment effect of treatment A for group A.

A second reason for being interested in the average causal effects of treatment within basic principal strata is that they are sometimes relevant to clinicians' decisions about the order in which to suggest treatments to patients. For example, consider a three-arm trial in which both treatments A and B are more effective than the control and both treatments will be made available to the general population after the trial. Suppose that a clinician can quickly determine whether a patient complies with a suggested treatment and will suggest the other treatment if the patient does not comply with the first suggested treatment. Suppose also that if the patient does comply with the first suggested treatment, then the clinician plans to stick with it as long as it does not appear harmful and there is reason to believe it may be helping. A clinician might take this approach to not be disruptive and to increase the patient's confidence in the treatment. In addition, suppose that a patient's final outcome would be the same if the doctor initially suggests a given treatment and the patient complies with it as if the doctor initially suggests the other treatment, the patient does not comply with it, then the doctor suggests the given treatment and the patient complies with it. Finally, suppose that compliance behavior is expected to remain the same when the treatments are offered to the general population as in the trial. Under these conditions, the treatment that has a higher effect for the $0AB$ group should be suggested first because only the $0AB$ group's treatment received will be affected by which treatment is suggested first. Choosing which treatment to suggest first by comparing the average treatment effect of treatment A for group A to the average treatment effect of treatment B for group B could produce a suboptimal strategy. For example, suppose we have equal proportions in the four principal strata, the average causal effect of treatment A is 0.7 for group $0A0$ and 0.4 for group $0AB$, and the average causal effect of treatment B is 0.3 for group $00B$ and 0.6 for group $0AB$. Then the average causal effect of treatment A for group A equals $0.5 \times (0.7 + 0.4) = 0.55$, higher than the average causal effect of treatment B for group B which equals $0.5 \times (0.3 + 0.6) = 0.45$. However, for group $0AB$, the average causal effect of treatment B is higher than the average causal effect of treatment A. Although the above setting is idealized, it illustrates the relevance of understanding interactions between treatments and basic principal strata of compliance behavior for certain clinical decisions.

A third reason for being interested in the average treatment effects for groups $0A0$, $00B$ and $0AB$ is that they provide useful information for a planner trying to anticipate what would happen were the treatment(s) to be introduced into general practice in a setting in which compliance patterns are expected to differ from those of the trial. Compliance outside of the trial could be higher because a treatment has been accepted as effective or lower without the encouragement given in the trial (Robins, 1989). Suppose it is being considered whether to introduce treatment A into general practice and it is expected that the compliance rate will be lower without the encouragement given in the trial. Joffe and Brensinger (2003) provide an approach for a two-arm trial for predicting the effects of introducing treatment A into general practice for a situation in which compliance is expected to change from the trial. A three-arm trial provides additional information for sharpening such predictions. For example, it might be expected that subjects who would comply with treatment A if offered it in the trial but who would not comply with treatment A it offered

it in general practice are more likely to be $0A0$ subjects in the trial than $0AB$ subjects in the trial. Then, the average treatment effect of treatment A for the subjects who would comply with treatment A if offered it in general practice when compared to the average treatment effect of treatment A for subjects who would comply with treatment A if offered it in the trial will move closer to the average treatment effect of treatment A for $0AB$ subjects in the trial and further from the average treatment effect of treatment A for $0A0$ subjects in the trial. Note that in this third setting we consider for which estimating the average causal effects within basic principal strata is of interest, we are assuming that compliance behavior differs between the environment of the trial and the environment of the treatment being offered to the general population, whereas in our second setting, we are assuming that compliance behavior remains the same inside and outside of the trial. These are different plausible possibilities for different treatment settings.

For the above reasons, the average treatment effects within basic principal strata $0A0$, $00B$ and $0AB$ are of considerable interest. However, these parameters are not point-identified under usual assumptions. By point-identified, we mean that a parameter would be uniquely determined if we could use the sampling process to obtain an unlimited number of observations. For a two-arm trial, the average treatment effect within the principal stratum of compliers (those who would take the treatment if assigned to it and not take the treatment if not assigned to it) is point-identified under an exclusion restriction that randomization has no direct effect and a monotonicity assumption that there are no defiers (i.e., no patients who would take the treatment only when assigned to the control) plus some other usual assumptions such as SUTVA, randomization of treatment assignment and non-zero average causal effect of random assignment on treatment received (Angrist, Imbens and Rubin, 1996). However, similar assumptions for a three-arm trial do not point-identify the average treatment effects for basic principal strata. One difficulty for the three-arm trial is that the monotonicity assumption that there are no defiers does not suffice to point-identify the proportion of patients in strata $0A0$, $00B$, $0AB$ and 000. This difficulty can be overcome by a further monotonicity assumption that there are no $00B$ patients (Monotonicity II in Section 2.2). However, even with this further assumption, the average treatment effects for treatment A for the principal strata $0A0$ and $0AB$ remain not point-identified because there is no way to fully deconvolve the observed outcomes under treatment A between these two groups without further assumptions. Note that point identification breaks down even in the two-arm case when defiers are allowed (Imbens and Angrist, 1994).

Although the treatment effects within basic principal strata are not point-identified under usual assumptions, they are partially identified, meaning that if we could use the sampling process to obtain an unlimited number of observations, we could place these parameters in a set valued identification region, where the set is a strict subset of the parameter space for at least some probability distributions of the observed data under the sampling process. When parameters of interest are only partially identified under certain assumptions, finding the upper and lower bounds of the identification region of the parameter establishes a domain of consensus among researchers who may hold disparate beliefs about what other assumptions are appropriate (Joffe, 2001; Manski, 2003). Bounds on parameters (in particular, bounds on probabilities) are the basis for some general theories for statistical reasoning in the face of uncertainty and imprecision (Shafer, 1982; Walley, 1991).

For a two-arm trial, bounds on the average treatment effects among the whole popu-

lation and among those who would accept treatment if offered it have been developed by Robins (1989), Manski (1990), and Balke and Pearl (1997). Balke and Pearl (1997) provide the tightest possible bounds on the average treatment effect using linear programming. Joffe (2001) shows how post-treatment covariates can be used to sharpen bounds on the overall and direct effect of a treatment. Zhang and Rubin (2003) develop bounds for causal effects within principal strata in a two-arm trial with censoring by death. In deriving our bounds for a three-arm trial with binary outcomes, we use some of the analytical ideas in Zhang and Rubin (2003).

Although our paper focuses on three-arm trials, our general framework and approach also applies to the analysis of more general multi-arm trials with noncompliance. We discuss extensions to general multi-arm trials in Section 7. However, the efficient computational scheme we have developed for computing bounds on causal effects in three-arm trials does not straightforwardly extend to more general multi-arm trials. Developing efficient and reliable computational procedures for general multi-arm trials requires further research.

Our paper is organized as follows. We introduce notation, assumptions and principal stratification in Section 2. We derive large sample bounds for the average causal effects within basic principal strata under two sets of assumptions in Section 3. In Section 4, we provide confidence intervals that have asymptotically correct coverage for the identification regions. In Section 5, we provide a means of checking the plausibility of the exclusion restriction and Monotonicity II assumptions based on the sample data. In Section 6, we illustrate our methods by analyzing data from the randomized study of treatments for alcohol dependence described in the introduction. In Section 7, we discuss how our methods for three-arm trials with binary compliance can be used to derive bounds on average causal effects within principal strata for general multi-arm trials. Section 8 provides discussion.

2. Framework

In this section, we first give the notation and assumptions used in this paper, and then define principal strata for a three-arm trial.

2.1. Notation

Consider a three-arm trial with Control (0), Treatment A (A) and Treatment B (B).

For all notation, vectors denote variables or indices for all N subjects, whereas scalars denote variables or indices for subject i. To reduce the complexity of the notation, the index i is suppressed in the scalars. We let $\mathbf{Z}$ be the N-dimensional vector of randomization assignments for all subjects, with individual element Z, where $Z = z \in \{0, A, B\}$ is the randomization assignment for subject i. $Z = 0$ if subject i is assigned Control, $Z = A$ for Treatment A, and $Z = B$ for Treatment B. We let $\mathbf{D_Z}$ be the N-dimensional vector of potential treatment-received under randomization assignment $\mathbf{z}$ with element $D_\mathbf{Z}$, where $D_\mathbf{Z} = d \in \{0, A, B\}$ according to whether a person would take the treatment 0, A, or B under randomization assignment $\mathbf{z}$. Let $\mathbf{Y_{z,\,d}}$ be the N-dimensional vector of potential responses under randomization assignment $\mathbf{z}$ and treatment-received $\mathbf{d}$, where $Y_{\mathbf{z},\,\mathbf{d}}$ is the potential response for subject i with the vector of randomization assignments $\mathbf{z}$ and the

vector of treatment-received $\mathbf{d}$.

$Y_{\mathbf{z},\,\mathbf{d}}$ and $D_{\mathbf{z}}$ are "potential" response and treatment-received in the sense that we can only observe one version of them. We let Y and D be the observed outcome and treatment-received variables, respectively.

2.2. Assumptions

In this subsection we use some of the assumptions in Angrist, Imbens and Rubin (1996).

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

a. If $z = z'$, then $D_{\mathbf{z}} = D_{\mathbf{z}'}$ for subject i, where $\mathbf{z}$ and $\mathbf{z}'$ are two different vectors of randomization assignments, and z and z' are the corresponding randomization assignments for subject i under $\mathbf{z}$ and $\mathbf{z}'$.

b. If $z = z'$ and $d = d'$, then $Y_{\mathbf{z},\,\mathbf{d}} = Y_{\mathbf{z}',\,\mathbf{d}'}$ for subject i, where $\mathbf{d}$ and $\mathbf{d}'$ are two different vectors of treatment-received, and d and d' are the corresponding treatment-received for subject i under $\mathbf{d}$ and $\mathbf{d}'$.

The SUTVA assumption allows us to write $Y_{\mathbf{z},\,\mathbf{d}}$ and $D_{\mathbf{z}}$ as $Y_{z,d}$ and D_z respectively for subject i, where $Y_{z,d}$ is the potential outcome with randomization assignment z and treatment-received d for subject i, and D_z is potential treatment-received with randomization assignment z for subject i.

Assumption 2: Random Assignment

For all N subjects, the treatment assignment Z is random: $Pr(\mathbf{Z} = \mathbf{c}) = Pr(\mathbf{Z} = \mathbf{c}')$ for all $\mathbf{c}$ and $\mathbf{c}'$ such that $l^T\mathbf{c} = l^T\mathbf{c}'$, where l is the N-dimensional column vector with all elements equal to one.

The random assignment assumption implies independence between assignment to treatment arm and pretreatment variables including potential outcomes, potential treatment received, and baseline covariates.

Assumption 3: Exclusion Restriction

Because of SUTVA, we can express this assumption in scalars. The assumption is that for subject i, $Y_{z,d} = Y_{z',d}$ for all z, z' and d. In words the exclusion restriction assumes that the randomization assignment affects outcomes only through its effect on treatment received.

This assumption allows us to define potential outcomes $Y_{z,d}$ as a function of d alone. That is, $Y_d = Y_{z,d} = Y_{z',d}$ for all z, z' and d.

Assumption 4: Monotonicity I

$P(D_0 = 0) = 1$ and $P(D_A = B) = P(D_B = A) = 0$. That is, a person in the Control group will not take any treatment, and a person assigned to treatment A or B has no access to treatment B or A respectively. This assumption is satisfied in the example study on alcohol dependence because of the study design.

Assumption 5: Monotonicity II

$P(D_A = A|D_B = B) = 1$. That is, if a person complies with treatment B, then he must comply with treatment A also. This assumption is plausible when A is a treatment that

is similar to treatment B but has fewer or the same side effects for every subject. In the example study on alcohol dependence, the Monotonicity II Assumption is plausible because CET requires attending less sessions than CBT. In Section 5, we show that this assumption imposes testable restrictions on the probability distribution of observable outcomes (Y, D, Z). In Section 6, we will assess the confidence that the distribution of (Y, D, Z) in the alcohol study satisfies the restrictions based on the sample data. Note that although Assumption 5 imposes testable restrictions on the probability distribution of observable outcomes, Assumption 5 cannot be consistently tested (i.e., there is no test of Assumption 5 for which the power converges to one for all fixed alternatives).

In Section 3, Assumptions $1 - 4$ will be used to derive bounds for the average causal effects of the treatment within basic principal strata. Bounds that incorporate the additional Assumption 5 will also be presented and the consequences of the additional assumption on bounds will be examined.

2.3. Principal Strata

Frangakis and Rubin (2002) define a basic principal stratification as a stratification of units by their potential values for a post-randomization variable under the set of randomization assignments being compared. Under Assumptions $1 - 4$, the subjects can be classified into four basic principal strata using actual treatment-received as the post-randomization variable.

$0AB = \{i|(D_0, D_A, D_B)' = (0, A, B)'\}$, the subjects who would comply with the assigned treatment under all three arms. Let $\pi_{0AB} = P(i \in 0AB)$;

$0A0 = \{i|(D_0, D_A, D_B)' = (0, A, 0)'\}$, the subjects who would comply with the assigned treatment under both the control arm and the treatment A arm but would not comply under the treatment B arm. Let $\pi_{0A0} = P(i \in 0A0)$;

$00B = \{i|(D_0, D_A, D_B)' = (0, 0, B)'\}$, the subjects who would comply with the assigned treatment under both the control arm and the treatment B arm but would not comply under the treatment A arm. Let $\pi_{00B} = P(i \in 00B)$. Under Assumption 5 (Monotonicity II), the principal stratum $00B$ is empty;

$000 = \{i|(D_0, D_A, D_B)' = (0, 0, 0)'\}$, the subjects who would not take any treatment under all three arms. Let $\pi_{000} = P(i \in 000)$.

Because membership in a basic principal stratum is not affected by treatment assignment, a comparison of average potential outcomes under two different treatment assignments for a basic principal stratum is a causal effect. Similar to potential outcomes, these basic principal strata cannot be directly observed. However, the observable strata of the observed treatment assignment and the observed treatment-received can provide information on these basic principal strata. Table 1 shows that each observable stratum is a mixture of certain unobservable basic principal strata in the three-arm trial.

Table 1. The relation of observed groups and potential basic principal strata

Z	D	Principal strata
A	A	$0AB$, $0A0$
A	0	$00B$, 000
B	B	$0AB$, $00B$
B	0	$0A0$, 000
0	0	$0AB$, $0A0$, $00B$, 000

3. Bounds

Under Assumptions 1−4, the average causal effects of treatment within basic principal strata - $E(Y_A-Y_0|0AB)$, $E(Y_A-Y_0|0A0)$, $E(Y_B-Y_0|0AB)$, $E(Y_B-Y_0|00B)$, and $E(Y_B-Y_A|0AB)$ $(= E(Y_B - Y_0|0AB) - E(Y_A - Y_0|0AB)$) - are not point identified based on knowledge of the joint distribution of the observables (Y, D, Z). However, the average causal effects of treatments within basic principal strata are partially identified in the sense that knowledge of the distribution of (Y, D, Z) can narrow the range in which these average causal effects of treatments can possibly lie. We first derive bounds for the average causal effects of treatments within basic principal strata for binary outcomes under Assumptions $1 - 4$. We then derive bounds based on the additional Assumption 5 and examine the consequence of this additional assumption on the bounds.

In this section, we derive "large sample" bounds that assume that the population conditional probabilities $P(Y = 1|Z = z, D = d)$ and $P(D = d|Z = z)$ are known. In Section 4, we provide confidence intervals for the bounds that reflect the sampling uncertainty in our estimates of these conditional probabilities.

3.1. Bounds under Assumptions $1 - 4$

Under Assumptions $1 - 4$, we use three steps to obtain sharp bounds for a three-arm trial.

I Based on the relations between the observed (D, Z) strata and the unobserved principal strata, we determine bounds for the proportions in basic principal strata, $\pi_{0AB}, \pi_{0A0}, \pi_{00B}$, and π_{000}.

II Based on the relations between the outcomes in the (D, Z) strata and the potential outcomes within basic principal strata, we find bounds for average potential outcomes within basic principal strata given proportions in basic principal strata.

III We find the bounds for the average causal effects of treatments within basic principal strata using the bounds on proportions in basic principal strata from Step I and the bounds for average potential outcomes in each basic principal stratum given the proportions in the basic principal strata from Step II.

Step I: Bounds for the proportions in basic principal strata

To find bounds for the proportions in the basic principal strata based on the proportions in the observable (D, Z) strata, we express the relations between the proportions in the basic principal strata and the proportions in the observable (D, Z) strata, the fact that the proportions in each basic principal stratum must lie between 0 and 1, and then minimize/maximize the proportions in the basic principal strata given the proportions in the observable (D, Z) strata. From Table 1, we know that each observed stratum of (D, Z) is a mixture of certain unobserved basic principal strata. We use $p_{d|z}$ to denote $P(D = d|Z = z)$ and have the following equations:

$$\begin{aligned}
p_{A|A} &= \pi_{0AB} + \pi_{0A0} && (1)\\
p_{0|A} &= \pi_{00B} + \pi_{000} && (2)\\
p_{B|B} &= \pi_{0AB} + \pi_{00B} && (3)\\
p_{0|B} &= \pi_{0A0} + \pi_{000} && (4)\\
p_{0|0} &= \pi_{0AB} + \pi_{0A0} + \pi_{00B} + \pi_{000} = 1 && (5)
\end{aligned}$$

Furthermore, we have

$$0 \leq \pi_{0AB}, \pi_{0A0}, \pi_{00B}, \pi_{000} \leq 1 \qquad (6)$$

Minimization and maximization of $\pi_{0AB}, \pi_{0A0}, \pi_{00B}, \pi_{000}$ subject to the constraints (1) - (6) define a linear programming problem. The solution to the linear programming problem is

$$\begin{aligned}
max\{0, p_{A|A} - p_{0|B}\} \leq \pi_{0AB} \leq min\{p_{A|A}, p_{B|B}\}\\
max\{0, p_{A|A} - p_{B|B}\} \leq \pi_{0A0} \leq min\{p_{A|A}, p_{0|B}\}\\
max\{0, p_{B|B} - p_{A|A}\} \leq \pi_{00B} \leq min\{p_{B|B}, p_{0|A}\}\\
max\{0, p_{0|B} - p_{A|A}\} \leq \pi_{000} \leq min\{p_{0|A}, p_{0|B}\}
\end{aligned}$$

Step II: Bounds for the potential outcomes given proportions in basic principal strata

In this step, we derive bounds on the average potential outcomes within basic principal strata for known proportions in basic principal strata. From Table 1, we know that the observed outcomes of $Y|Z, D$ are a mixture of potential outcomes from basic principal strata:

$$\begin{aligned}
E(Y|Z = A, D = A) &= \frac{\pi_{0AB}}{\pi_{0AB} + \pi_{0A0}} E(Y_A|0AB) + \frac{\pi_{0A0}}{\pi_{0AB} + \pi_{0A0}} E(Y_A|0A0) && (7)\\
E(Y|Z = B, D = B) &= \frac{\pi_{0AB}}{\pi_{0AB} + \pi_{00B}} E(Y_B|0AB) + \frac{\pi_{00B}}{\pi_{0AB} + \pi_{00B}} E(Y_B|00B) && (8)\\
E(Y|Z = A, D = 0) &= \frac{\pi_{00B}}{\pi_{00B} + \pi_{000}} E(Y_0|00B) + \frac{\pi_{000}}{\pi_{00B} + \pi_{000}} E(Y_0|000) && (9)\\
E(Y|Z = B, D = 0) &= \frac{\pi_{0A0}}{\pi_{0A0} + \pi_{000}} E(Y_0|0A0) + \frac{\pi_{000}}{\pi_{0A0} + \pi_{000}} E(Y_0|000) && (10)
\end{aligned}$$

$$E(Y|Z=0,D=0) = \pi_{0AB}E(Y_0|0AB)+\pi_{0A0}E(Y_0|0A0)+\pi_{00B}E(Y_0|00B) + \pi_{000}E(Y_0|000) \quad (11)$$

Note that for binary outcomes, the expectations in the above equations equal the corresponding probabilities of a favorable outcome occurring, so we have

$$0 \le E(Y_A|0AB), E(Y_A|0A0), E(Y_B|0AB), E(Y_B|00B), E(Y_0|0AB), E(Y_0|0A0), E(Y_0|00B), E(Y_0|000) \le 1 \quad (12)$$

To derive bounds for average potential outcomes within basic principal strata given the proportions in basic principal strata, we use Lemma 1.

LEMMA 1. *Let h be a mixture of two Bernoulli distributions f and g, $h = \alpha f + (1-\alpha)g$, where the mixing proportion α is known, and let P_1, P_2 and P_3 be the probabilities of a positive outcome under f, g and h respectively. Then,*
$max(0, 1-\frac{1-P_3}{\alpha}) \le P_1 \le min(1, \frac{P_3}{\alpha})$
$max(0, 1-\frac{1-P_3}{1-\alpha}) \le P_2 \le min(1, \frac{P_3}{1-\alpha})$

The proof of Lemma 1 is a straightforward solution to the linear programming problem of minimizing/maximizing P_1 and P_2 subject to the constraints $P_3 = \alpha P_1 + (1-\alpha)P_2$, $0 \le P_1 \le 1$, $0 \le P_2 \le 1$, $0 \le P_3 \le 1$.

By Lemma 1 and equations (7), (8) and (12), we obtain bounds for the average potential outcomes within basic principal strata, $E(Y_A|0AB)$, $E(Y_A|0A0)$, $E(Y_B|0AB)$ and $E(Y_B|00B)$, given the proportions in each basic principal stratum. Furthermore, by equations (1)-(5), we can express π_{0A0}, π_{00B}, π_{000} in terms of π_{0AB} and the conditional distribution of $D|Z$. We use $q_{1|zd}$ to denote $P(Y=1|Z=z, D=d)$. Thus, we have the following bounds given π_{0AB} and the joint distribution of (Y, D, Z).

$$\begin{aligned} E(Y_d|0AB,\pi_{0AB}) &\in (minE(Y_d|0AB,\pi_{0AB}), maxE(Y_d|0AB,\pi_{0AB})) \\ &= (max\{0, 1-\frac{1-q_{1|zd}}{\frac{\pi_{0AB}}{p_{d|z}}}\}, min\{1, \frac{q_{1|zd}}{\frac{\pi_{0AB}}{p_{d|z}}}\}), \\ \text{where } z=d &= A \text{ or } B; \end{aligned}$$

$$\begin{aligned} E(Y_A|0A0,\pi_{0AB}) &\in (minE(Y_A|0A0,\pi_{0AB}), maxE(Y_A|0A0,\pi_{0AB})) \\ &= (max\{0, 1-\frac{1-q_{1|AA}}{1-\frac{\pi_{0AB}}{p_{A|A}}}\}, min\{1, \frac{q_{1|AA}}{1-\frac{\pi_{0AB}}{p_{A|A}}}\}); \end{aligned}$$

$$\begin{aligned} E(Y_B|00B,\pi_{0AB}) &\in (minE(Y_B|00B,\pi_{0AB}), maxE(Y_B|00B,\pi_{0AB})) \\ &= (max\{0, 1-\frac{1-q_{1|BB}}{1-\frac{\pi_{0AB}}{p_{B|B}}}\}, min\{1, \frac{q_{1|BB}}{1-\frac{\pi_{0AB}}{p_{B|B}}}\}). \end{aligned}$$

Similar but more complicated algebra provides bounds for average potential outcomes under the control ($E(Y_0|0AB)$, $E(Y_0|0A0)$ and $E(Y_0|00B)$) given π_{0AB} and the distribution of (Y, D, Z). The details are provided in Appendix A.

Step III: Bounds for the average causal effects within basic principal strata

In this final step, we combine the bounds from Steps I and II to construct bounds for the average causal effects within basic principal strata. From Step II, we have obtained bounds for average potential outcomes within basic principal strata given π_{0AB}. Then, given π_{0AB}, the average causal effects within basic principal strata can be bounded based on the difference of the corresponding bounds on the average potential outcomes within basic principal strata. For example, for basic principal stratum $0AB$, given π_{0AB}, the average causal effect $E(Y_A - Y_0|0AB, \pi_{0AB})$ can be no less than $E(Y_A|0AB, \pi_{0AB})$'s lower bound minus $E(Y_0|0AB, \pi_{0AB})$'s upper bound, and no greater than $E(Y_A|0AB, \pi_{0AB})$'s upper bound minus $E(Y_0|0AB, \pi_{0AB})$'s lower bound. That is, given π_{0AB}, $E(Y_A - Y_0|0AB, \pi_{0AB})$ must fall in the interval

$$\begin{aligned}&(minE(Y_A|0AB, \pi_{0AB}) - maxE(Y_0|0AB, \pi_{0AB}),\\ &maxE(Y_A|0AB, \pi_{0AB}) - minE(Y_0|0AB, \pi_{0AB})).\end{aligned}$$

However, under Assumptions $1-4$, π_{0AB} is not point-identified but only bounded by Step I, $\pi_{0AB} \in I$, $I = (max(0, p_{A|A} - p_{0|B}), min(p_{A|A}, p_{B|B}))$. Thus, $E(Y_A - Y_0|0AB)$ must fall into

$$\begin{aligned}&(\min_{\pi_{0AB}\in I}[minE(Y_A|0AB, \pi_{0AB}) - maxE(Y_0|0AB, \pi_{0AB})],\\ &\max_{\pi_{0AB}\in I}[maxE(Y_A|0AB, \pi_{0AB}) - minE(Y_0|0AB, \pi_{0AB})])\end{aligned} \tag{13}$$

The minima and maxima of average potential outcomes that appear in (13) are given in Step II. The bounds in (13) are computed using a grid search over $\pi_{0AB} \in I$. Similarly, expressions (13) hold for $E(Y_A - Y_0|0A0)$, $E(Y_B - Y_0|0AB)$ and $E(Y_B - Y_0|00B)$.

3.2. Bounds under Assumptions $1-4$ *and* 5 *(Monotonicity II)*

In this section, we derive bounds that use Assumption 5 in addition to the Assumptions $1-4$ used in the bounds of Section 3.1. In a three-arm trial, Assumption 5 (Monotonicity II) asserts that basic principal stratum $00B$ is empty. This means that the stratum $(D = 0, Z = A)$ belongs only to the basic principal stratum 000 and the stratum $(D = B, Z = B)$ belongs only to the basic principal stratum $0AB$. This additional information on the relationship between the strata of (D, Z) and the basic principal strata has two beneficial consequences: it point-identifies the proportions within each basic principal stratum and it provides additional information about the distribution of potential outcomes within basic principal strata based on the distribution of $Y|D, Z$. The proportions in each basic principal stratum are point-identified by the joint distribution of (D, Z):

$$\begin{aligned}\pi_{0AB} &= p_{B|B}\\ \pi_{0A0} &= p_{A|A} - p_{B|B}\\ \pi_{000} &= p_{0|A}\end{aligned}$$

Following the same reasoning in Step II on the relationship between the distribution of $Y|D,Z$ and the distribution of potential outcomes within basic principal strata but using the fact that the basic principal stratum $00B$ is empty, we have the following:

$$\begin{aligned}E(Y_A|0AB) &\in (max\{0, 1-\frac{1-q_{1|AA}}{\frac{\pi_{0AB}}{p_{A|A}}}\}, min\{1, \frac{q_{1|AA}}{\frac{\pi_{0AB}}{p_{A|A}}}\})\\ E(Y_A|0A0) &\in (max\{0, 1-\frac{1-q_{1|AA}}{1-\frac{\pi_{0AB}}{p_{A|A}}}\}, min\{1, \frac{q_{1|AA}}{1-\frac{\pi_{0AB}}{p_{A|A}}}\})\\ E(Y_B|0AB) &= q_{1|BB}\\ E(Y_0|000) &= q_{1|A0}\\ E(Y_0|0AB) &= min\left(\begin{array}{c}1\\ max\{0, \frac{q_{1|00}-p_{0|B}q_{1|B0}}{p_{B|B}}\}\end{array}\right)\\ E(Y_0|0A0) &= min\left(\begin{array}{c}1\\ max\{0, \frac{p_{0|B}q_{1|B0}-p_{0|A}q_{1|A0}}{p_{A|A}-p_{B|B}}\}\end{array}\right)\end{aligned}$$

Thus, in a three-arm trial under Assumption 5 (in addition to Assumptions $1-4$), $E(Y_B - Y_0|0AB)$ is point-identified. $E(Y_A - Y_0|0AB)$ and $E(Y_A - Y_0|0A0)$ remain only partially identified under Assumptions $1-5$ but the bounds are narrowed from those under only Assumptions $1-4$ because π_{0AB} is known. The bounds under Assumptions $1-5$ are provided in Appendix B.

4. Confidence intervals for the average causal effects within basic principal strata

In Section 3, we have shown that for three-arm trials the average causal effects of the treatments within the basic principal strata are not point-identified but bounded. When deriving the bounds in Section 3, we assumed that the distribution of (Y, D, Z) was known. In practice, there is sampling uncertainty in the distribution of (Y, D, Z) and the lower and upper bounds need to be estimated. Confidence intervals (CIs) are of interest when making inference about the bounds.

For example, from Section 3.2, in a three-arm trial under Assumptions $1-5$

$$E(Y_A - Y_0|0AB) \in (max\left(\begin{array}{c}0\\ 1-\frac{1-q_{1|AA}}{\frac{\pi_{0AB}}{p_{A|A}}}\end{array}\right) - min\left(\begin{array}{c}1\\ max\{0, \frac{q_{1|00}-p_{0|B}q_{1|B0}}{p_{B|B}}\}\end{array}\right),$$

$$min\left(\begin{array}{c}1\\ \frac{q_{1|AA}}{\frac{\pi_{0AB}}{p_{A|A}}}\end{array}\right) - min\left(\begin{array}{c}1\\ max\{0, \frac{q_{1|00}-p_{0|B}q_{1|B0}}{p_{B|B}}\}\end{array}\right)).$$

which is a function of conditional probabilities and can be estimated by substituting the sample values of the conditional probabilities. We would then like to have a CI that describes our uncertainty about these bounds.

Suppose (L_n, U_n) are estimates of the bounds (L, U) on the identification region (the range of possible values for a parameter given the true probability distribution of observable outcomes) of a partially identified population parameter. We are interested in a CI which asymptotically covers the identification region with fixed probability. One way to form such a CI is to find a one-sided CI (L_n^l, ∞) for the lower bound L with coverage probability $P(L_n^l \leq L) = 1 - \frac{\alpha}{2}$ and a one-sided CI $(-\infty, U_n^u)$ for the upper bound U with coverage probability $P(U_n^u \geq U) = 1 - \frac{\alpha}{2}$. By the Bonferroni inequality, $P(L_n^l \leq L, U_n^u \geq U) \geq 1 - \alpha$. Thus, (L_n^l, U_n^u) has at least $(1-\alpha)$ coverage probability. We call the above method for forming a CI for the identification region the Bonferroni method. This method is potentially conservative because it does not take into account the joint distribution of (L_n, U_n).

Horowitz and Manski (2000) develop a CI that takes into account the joint distribution of (L_n, U_n). The interval $(L_n - z_{n\alpha}, U_n + z_{n\alpha})$, where $z_{n\alpha}$ is chosen so that $P(L_n - z_{n\alpha} \leq L, U \leq U_n + z_{n\alpha}) = 1 - \alpha$ asymptotically, has asymptotically $(1 - \alpha)$ probability of containing both the lower and the upper bounds of the identification region. Horowitz and Manski's approach requires the use of the same $z_{n\alpha}$ in the CI $(L_n - z_{n\alpha}, U_n + z_{n\alpha})$; it is not balanced, meaning that $P(L_n - z_{n\alpha} \geq L)$ might not equal $P(U_n + z_{n\alpha} \leq U)$ asymptotically. Also, Horowitz and Manski's approach should be modified when L_n or U_n are equal to the lower or upper bound of the identification region respectively.

The asymptotic Bonferroni and Horowitz-Manski CIs can be obtained by using the delta method or the bootstrap. For the Bonferroni CIs, we use empirical percentile bootstrap CIs to construct $(1 - \frac{\alpha}{2})$ one-sided CIs for L and U respectively. For Horowitz-Manski's method, by repeated bootstrap sampling, the distribution of bootstrap estimates, (L_n^*, U_n^*), conditional on the data can be estimated and used to find $z_{n\alpha}^*$ such that $P^*(L_n^* - z_{n\alpha}^* \leq L_n, U_n \leq U_n^* + z_{n\alpha}^*) = 1 - \alpha$, where P^* is the probability measure induced by bootstrap sampling conditional on the data. Then the bootstrap $(1 - \alpha)$ CI for (L, U) is $(L_n - z_{n\alpha}^*, U_n + z_{n\alpha}^*)$ (Horowitz and Manski, 2000), which has asymptotic coverage probability $(1 - \alpha)$ (Bickel and Freedman, 1981).

Another approach to finding a confidence interval for the identification region is to find a joint 95% confidence region for (L, U) and then take the smallest value for L and the largest value for U in this confidence region. This takes into account the joint distribution of the (L_n, U_n) without the limitations on the form of the confidence interval of Horowitz-Manski. Beran (1988) develops the B method to find simultaneous confidence sets that are balanced and have correct overall coverage probability asymptotically. The B method simultaneous confidence intervals for L and U are $\{L_n^l : L_n - L_n^l \leq \hat{H}_{n,l}^{-1}[\hat{H}_n^{-1}(1 - \alpha)]\}$ and $\{U_n^u : U_n^u - U_n \leq \hat{H}_{n,u}^{-1}[\hat{H}_n^{-1}(1 - \alpha)]\}$ respectively, where $\hat{H}_{n,l}$ and $\hat{H}_{n,u}$ are distributions of $(L_n^* - L_n)$ and $(U_n - U_n^*)$ respectively (where L_n^* and U_n^* are estimates of L_n and U_n from each bootstrap resample respectively), and $\hat{H}_n$ is the distribution of

$max\{\hat{H}_{n,l}(L_n^* - L_n), \hat{H}_{n,u}(U_n - U_n^*)\}$. The B method CI for (L, U) is $(minL_n^l, maxU_n^u)$. Beran (1990) provides the B^2 method as an improvement of the B method that reduces the asymptotic order of imbalance in the B method simultaneous confidence set as well as the asymptotic order of error in overall coverage probability. The B^2 method requires a double bootstrap Monte Carlo algorithm that uses roughly the square of the computer time needed by the B method.

The Bonferroni, Horowitz-Manski and B method CIs are CIs that cover the entire identification region with probability greater than or equal to $(1 - \alpha)$ asymptotically. Perforce, these confidence intervals contain the true parameter with probability greater than or equal to $(1 - \alpha)$ asymptotically and are hence confidence intervals for the true parameter. However, Imbens and Manski (2004) show that by dropping the requirement that a confidence interval contain the identification region with probability greater than or equal to $(1 - \alpha)$ asymptotically, a narrower confidence interval for the true parameter can be constructed. Because our interest is typically in a confidence interval for the true parameter rather than the identification region, extending Imbens-Manski's approach to our setting is of considerable interest. However, doing so is beyond the scope of this paper. In particular, verifying Assumption 1 of Imbens and Manski (2004), or constructing an alterative assumption, requires further research for our setting.

In Section 6, the data from the alcohol study are used to compare bootstrap CIs based on the Bonferroni method, Horowitz and Manski's approach, and the B method, and simulation studies are done to examine the finite sample coverage of these three methods.

5. Checking the Plausibility of the Exclusion Restriction and Monotonicity II Assumptions

All the bounds in this paper are derived under the exclusion restriction assumption (Assumption 3). The bounds in Section 3.2 are derived under the additional Monotonicity II assumption. Neither the exclusion restriction nor the Monotonicity II assumptions are "point-identified". In other words, it cannot be uniquely determined from the probability distribution of observable outcomes (Y, D, Z) whether or not they hold. However, both assumptions imply restrictions on the probability distribution of observable outcomes (Y, D, Z) such that the assumptions cannot hold under certain distributions of (Y, D, Z). In this section, we discuss the restrictions on the distribution of (Y, D, Z) implied by the exclusion restriction and Monotonicity II assumptions and provide a means of assessing the confidence that the distribution of (Y, D, Z) satisfies these restrictions based on the sample data.

Pearl (1995) provides a necessary condition on the probability distribution of (Y, D, Z) for the exclusion restriction assumption to hold. For our setting of three-arm trials with binary outcomes in which one group cannot access the treatment assigned to the other groups, a necessary condition for the exclusion restriction assumption to hold is that the six following inequalities hold:

$$P(Y = 0, D = 0|Z = z) + P(Y = 1, D = 0|Z = \bar{z}) \leq 1, \tag{14}$$

where

$$z = \left\{ \begin{array}{c} 0 \\ A \\ B \end{array} \right\}, \bar{z} = \left\{ \begin{array}{c} \bar{0} \\ \bar{A} \\ \bar{B} \end{array} \right\} = \left\{ \begin{array}{c} A, B \\ 0, B \\ 0, A \end{array} \right\}$$

We now assume that the exclusion restriction assumption holds and derive restrictions on the distribution of (Y, D, Z) for Monotonicity II to be plausible. Under the assumption of the exclusion restriction, the set of probability distributions of observable outcomes (Y, D, Z) can be divided into three subsets: (i) probability distributions of observable outcomes for which Monotonicity II does not hold for any of the corresponding probability distributions of potential outcomes; (ii) probability distributions of observable outcomes for which there are some corresponding distributions of potential outcomes for which Monotonicity II holds as well as some probability distributions of potential outcomes for which Monotonicity II does not hold; and (iii) probability distributions of observable outcomes for which Monotonicity II must hold. The union of the subset (ii) and (iii) of probability distributions of observable outcomes for which Monotonicity II is plausible is the set of probability distributions on (Y, D, Z) satisfying the following constraints:

$$q_{1|AA}, q_{1|A0}, q_{1|BB} \in (0, 1); \tag{15}$$

$$q_{1|B0} \in (\frac{p_{0|A}}{p_{0|B}} q_{1|A0}, \frac{p_{A|A} - p_{B|B}}{p_{0|B}} + \frac{p_{0|A}}{p_{0|B}} q_{1|A0}); \tag{16}$$

$$q_{1|00} \in (p_{0|A} q_{1|A0}, p_{A|A} + p_{0|A} q_{1|A0}). \tag{17}$$

The constraints (15)-(17) are obtained based on the relationship between observed outcomes and potential outcomes and the fact that potential binary outcomes are 0 or 1. Given that (Y, D, Z) satisfies constraints (15)-(17), the subset (iii) of probability distributions of observable outcomes for which Monotonicity II must hold is the set of probability distributions on (Y, D, Z) satisfying a further constraint:

$$min(p_{B|B}, p_{0|A}) = 0, \tag{18}$$

which is obtained based on the fact that $P(\pi_{00B} = 0) = 1$ under Monotonicity II.

The set of probability distributions of observable outcomes (Y, D, Z) can be divided into three subsets: (a) distributions which do not satisfy (15)-(17); (b) distributions which satisfy (15)-(17) but not (18); (c) distributions which satisfy (15)-(18). Under the assumption that the exclusion restriction holds, (a) corresponds to subset (i) above, (b) corresponds to (ii) and (c) corresponds to (iii). Suppose it is observed that the empirical distribution of (Y, D, Z) falls into subset (a). We would like to know how "confident" we should be that the true distribution of (Y, D, Z) falls into subset (a). This is an example of the problem of regions discussed by Efron and Tibshirani (1998). A simple bootstrap procedure for estimating the confidence that the true probability distribution of (Y, D, Z) falls into subset (a) based on a sample is the following:

Table 2. Observed proportions in the alcohol study and the hypothetical study

Observed Proportions	Alcohol Study	Hypothetical Study
$P(D = A\|Z = A)$	0.70	0.95
$P(D = 0\|Z = A)$	0.30	0.05
$P(D = B\|Z = B)$	0.52	0.80
$P(D = 0\|Z = B)$	0.48	0.20
$P(D = 0\|Z = 0)$	1.00	1.00
$P(Y = 1\|Z = A, D = A)$	0.60	0.95
$P(Y = 1\|Z = A, D = 0)$	0.47	0.20
$P(Y = 1\|Z = B, D = B)$	0.87	0.70
$P(Y = 1\|Z = B, D = 0)$	0.71	0.25
$P(Y = 1\|Z = 0, D = 0)$	0.60	0.45

(1) Bootstrap from the empirical distribution of (Y, D, Z);

(2) Count what proportion of the bootstrapped empirical distributions do not satisfy the constraints (15)-(17). This proportion is the estimated confidence that the true probability distribution of observable outcomes falls into subset (a).

An analogous procedure can be used to estimate the confidence that the probability distribution of (Y, D, Z) falls into subsets (b) or (c) when the empirical distribution of (Y, D, Z) falls into subsets (b) or (c) respectively. Also an analogous procedure can be used to estimate the confidence that the probability distribution of (Y, D, Z) does or does not satisfy the constraint (14) that is necessary for the exclusion restriction assumption to hold. Efron and Tibshirani (1998) provide some refinements on this simple bootstrap procedure that improve the accuracy of the estimated confidence.

6. Application

In this section, we will use data from the trial of treatments for alcohol dependence discussed in the introduction to illustrate the methods developed in this paper to construct the bounds and confidence intervals for the average causal effects within principal strata under different assumptions. †

In the alcohol study, we have 141 subjects and observe the proportions shown in the first column of Table 2.

For the alcohol study, the empirical distribution of (Y, D, Z) satisfies the constraint (14), indicating that the exclusion restriction assumption is plausible, but it does not satisfy the constraints (15)-(17). Using the bootstrap procedure in Section 5, 65% of 1000 bootstrapped distributions do not satisfy the constraints (15)-(17), so we are fairly confident that Monotonicity II does not hold for the alcohol study. Under Assumptions $1 - 4$, the bounds are estimated using the results of Section 3, and the bootstrap 95% Bonferroni,

†R codes to compute the bounds and CIs with the methods developed in this paper are available from the authors.

Table 3. The estimated bounds and 95% Bonferroni, Horowitz-Manski and B method CIs for the average causal effects within principal strata for the alcohol study

Causal effect		$1-4$
$E(Y_A - Y_0 \mid 0AB)$	Estimated Bounds	$(-0.61, 0.67)$
	95% Bonferroni CIs	$(-1, 1)$
	95% Horowitz-Manski CIs	$(-1, 1)$
	95% B method CIs	$(-1, 1)$
$E(Y_A - Y_0 \mid 0A0)$	Estimated Bounds	$(-1, 0.40)$
	95% Bonferroni CIs	$(-1, 0.79)$
	95% Horowitz-Manski CIs	NA
	95% B method CIs	$(-1, 0.81)$
$E(Y_B - Y_0 \mid 0AB)$	Estimated Bounds	$(0.11, 0.67)$
	95% Bonferroni CIs	$(-0.85, 1)$
	95% Horowitz-Manski CIs	$(-0.44, 1)$
	95% B method CIs	$(-0.43, 1)$
$E(Y_B - Y_0 \mid 00B)$	Estimated Bounds	$(-0.64, 1)$
	95% Bonferroni CIs	$(-1, 1)$
	95% Horowitz-Manski CIs	NA
	95% B method CIs	$(-1, 1)$

Horowitz-Manski, and B method CIs are obtained based on the corresponding methods introduced in Section 4. Because of the binary outcome, the confidence limits for the bounds which are estimated to be -1 or 1 are defined as -1 or 1 respectively. The results are shown in Table 3.

For the alcohol study, the estimated bounds and 95% CIs for all principal strata are wide under Assumptions $1-4$. The 95% CIs for treatment A or B for principal strata $0AB$, $0A0$ and $00B$ all contain 0, indicating that there is not strong evidence that treatment A (CET) or treatment B (CBT) has a beneficial causal effect on alcohol relapse. The CIs for treatment A and treatment B overlap for principal stratum $0AB$, so the data provide no strong evidence about which treatment is better for principal stratum $0AB$. Because the Horowitz-Manski method requires the use of the same $z_{n\alpha}$ in the CI $(L_n - z_{n\alpha}, U_n + z_{n\alpha})$, it is not applied to the bounds with -1 or 1. The CIs based on different methods have similar length except that the Horowitz-Manski CI and B method CI for $E(Y_B - Y_0 \mid 0AB)$ are shorter than the corresponding Bonferroni CI.

To illustrate that the bounds can be highly informative in certain cases, we consider some hypothetical data. Suppose we have a three-arm trial with binary outcome Y equal to one if the treatment is successful, $n = 1200$, and observe the proportions shown in the second column of Table 2. The empirical distribution of (Y, D, Z) satisfies the constraint (14), indicating that the exclusion restriction assumption is plausible, and it satisfies the constraints (15)-(17). Because 95% of 1000 bootstrapped distributions satisfy the constraints (15)-(17), we are confident that Monotonicity II is plausible for the hypothetical study. Note that although the data does not suggest Monotonicity II is implausible, whether Monotonicity II is in fact a reasonable assumption depends on the nature of the actual treatments. The bounds and bootstrap 95% Bonferroni, Horowitz-Manski, and B method CIs for the

Table 4. The estimated bounds and 95% Bonferroni, Horowitz-Manski and B method CIs for the average causal effects within principal strata for the hypothetical data

Causal effect		$1-4$	$1-4$ and 5
$E(Y_A - Y_0 \mid 0AB)$	Estimated Bounds	$(0.41, 0.51)$	$(0.44, 0.50)$
	95% Bonferroni CIs	$(0.33, 0.58)$	$(0.36, 0.57)$
	95% Horowitz-Manski CIs	$(0.34, 0.58)$	$(0.37, 0.57)$
	95% B method CIs	$(0.33, 0.58)$	$(0.37, 0.57)$
$E(Y_A - Y_0 \mid 0A0)$	Estimated Bounds	$(0.39, 0.79)$	$(0.42, 0.73)$
	95% Bonferroni CIs	$(0.13, 0.89)$	$(0.18, 0.87)$
	95% Horowitz-Manski CIs	$(0.22, 0.96)$	$(0.23, 0.92)$
	95% B method CIs	$(0.21, 0.91)$	$(0.21, 0.87)$
$E(Y_B - Y_0 \mid 0AB)$	Estimated Bounds	$(0.16, 0.23)$	0.20
	95% Bonferroni CIs	$(0.06, 0.32)$	$(0.11, 0.29)$
	95% Horowitz-Manski CIs	$(0.06, 0.32)$	NA
	95% B method CIs	$(0.07, 0.32)$	NA
$E(Y_B - Y_0 \mid 00B)$	Estimated Bounds	$(-1, 1)$	Undefined
	95% Bonferroni CIs	$(-1, 1)$	NA
	95% Horowitz-Manski CIs	$(-1, 1)$	NA
	95% B method CIs	$(-1, 1)$	NA

average causal effects within principal strata are computed under Assumptions $1-4$ and $1-5$ shown in Table 4.

In the hypothetical study most of the bounds are informative in terms of not including zero. The addition of Assumption 5 helps to narrow the bounds. The CIs based on different methods have similar length in this example. For both sets of assumptions, the 95% CIs for treatment A for both principal strata $0AB$ and $0A0$ and for treatment B for principal stratum $0AB$ do not contain 0, indicating that there is strong evidence that treatment A has a beneficial causal effect for both strata $0AB$ and $0A0$, and treatment B has a beneficial causal effect for stratum $0AB$. For principal stratum $0AB$, under both sets of assumptions, the lower endpoints of the CIs for the treatment A are greater than the upper endpoints of the CIs for treatment B. This provides strong evidence that treatment A is better than treatment B for principal stratum $0AB$.

To estimate the true coverage probabilities of the Bonferroni, Horowitz-Manski, and B method 95% CIs for the bounds, we generated one thousand simulations for each study with the same sample size of the original data using the empirical distribution of the data. For each simulation, we constructed the Bonferroni, Horowitz-Manski, and B method 95% bootstrap CIs. The true coverage probabilities (for the empirical distribution of the data) are estimated by checking how many of the one thousand bootstrap CIs cover the bounds of the identification region of the empirical distribution of the data. The estimated coverage probabilities for the Bonferroni, Horowitz-Manski, and B method are in the range of $0.94-0.96$, $0.91-0.96$, and $0.90-0.96$ respectively for the hypothetical data, and $0.92-0.97$, $0.65-0.92$, and $0.76-0.96$ respectively for the alcohol data, which has a much smaller sample size than the hypothetical data.

7. Extensions to General Multi-Arm Trials

In this section, we consider more general multi-arm trials than three-arm trials. For similar reasons as discussed in the introduction for three arms, the causal effects within basic principal strata are often of interest in general multi-arm trials. As with three-arm trials, our three-step strategy described in Section 3.1 can be used to compute bounds for causal effects within basic principal strata. However, computation of such bounds for trials with more than three arms is more difficult. We discuss this further below; we first describe the structure of the basic principal strata in several types of multi-arm trials.

For a k-arm trial with the same structure as a three-arm trial under Assumptions $1-4$, i.e., the control group cannot receive any active treatments and each active treatment arm cannot receive any other active treatments not assigned to it, there will be $(2k-1)$ observed (Z, D) strata and 2^{k-1} basic principal strata. Each (Z, D) stratum will have 2^{k-2} basic principal strata in it except for the $(Z, D) = (0, 0)$ strata which has 2^{k-1} basic principal strata in it.

For a k-arm trial with the same structure as a three-arm trial under Assumptions $1-5$, i.e., Monotonicity II holds, the proportions in each basic principal stratum are identified in Step I, the average potential outcomes for control within all basic principal strata and the average potential outcome for the hardest to comply with treatment within the basic principal stratum that would take all treatments are identified in Step II. Hence the average causal effect for the hardest to comply with treatment within the basic principal stratum that would take all treatments is identified in Step III. All other average causal effects within basic principal strata are not point identified.

In the alcohol study that we considered, compliance to an arm was categorized as a binary variable, whether or not the patient attended at least 80% sessions, based on the clinicians' understanding of how the treatment works. Categorization of compliance into a binary variable is often done in analyses of causal effects when there is noncompliance (e.g., Sommer and Zeger, 1991; Ten Have et al., 2004; Small et al., 2006). Although such categorization is often a reasonable simplification as in the alcohol study that we consider, compliance often actually involves more than two levels. For example, for the alcohol study, compliance could be classified into three levels – none, half or full – based on the percentage of sessions attended. In this case, there are nine basic principal strata (000, $0(\frac{1}{2}A)0$, $0A0$, $00(\frac{1}{2}B)$, $00B$, $0(\frac{1}{2}A)B$, $0AB$, $0A(\frac{1}{2}B)$, $0(\frac{1}{2}A)(\frac{1}{2}B)$) and there are nine basic principal strata in each (Z, D) stratum other than the $(0, 0)$ stratum; for example, the $(Z, D) = (A, \frac{1}{2}A)$ stratum contains the basic principal strata $0(\frac{1}{2}A)0$, $0(\frac{1}{2}A)B$, $0(\frac{1}{2}A)(\frac{1}{2}B)$.

For both settings in which there are more than three arms and settings in which there are more than two levels of compliance, the ratio of basic principal strata to observed (Z, D) strata increases compared to three-arm trials. This makes the identification problems more severe, and potentially reduces the informativeness of the bounds. For complex multi-arm trials, it might be worthwhile to consider parametric models that parameterize the relationships among outcomes in different principal strata. For example, for trials with more than two levels of compliance, Goetghebeur and Molenberghs (1996) discuss the use of parametric models for the association between principal stratum membership and potential outcomes.

Besides the additional identification problems raised by multi-arm trials with more than three arms or more than two levels of compliance, such trials raise computational difficulties for computing bounds. For three-arm trials with two levels of compliance, our three-step strategy described in Section 3.1 enables us to compute bounds by carrying out a one-dimensional grid search in which the function we evaluate at each grid point involves solving simple linear programming problems. We can quickly compute the bounds and this enables bootstrap methods to be used to compute confidence intervals for bounds in a manageable amount of time (e.g., less than 10 seconds for Bonferroni bootstrap CIs for all bounds with 1000 resamplings). For general multi-arm trials, our three-step strategy can still be used to compute the bounds. For Step II, the bounds for average potential outcomes within basic principal strata given the proportions in the basic principal strata can be still be computed using linear programming. However, for Step I, the region of feasible proportions in the basic principal strata can no longer be parameterized as an interval for one parameter as in three-arm trials. Instead, the feasible region must be parameterized as a multi-dimensional convex polytope. Polyhedral computation techniques (Fukuda, 2004) can be used to parameterize the feasible region in Step I. As a consequence of the more complex region of feasible proportions in principal strata, for Step III, we need to carry out a grid search over a multidimensional convex polytope rather than a one-dimensional interval. Another approach for computing bounds in general multi-arm trials besides our three-step strategy is to directly write the bounds as the solution to nonlinear programming problems. However, because the programming problems are not in general convex, solutions to the programming problem may be only local optima. Development of fast and reliable computational techniques for computing bounds for multi-arm trials more complex than three-arm trials with two levels of compliance is a valuable topic for future research.

8. Discussion

For three-arm trials, average causal effects of the treatments within basic principal strata are of interest for several reasons. They provide useful information to patients who have enough information to regard their own compliance behavior as not being exchangeable with the population's and will choose between treatments, to clinicians who do not know patients' compliance behavior and want to determine which treatment to offer first, and to planners who try to anticipate what would happen were the treatments to be introduced into general practice. Even with usual assumptions, when noncompliance is present, the average treatment effects within basic principal strata in three-arm trials are not point-identified. However, we show that the observable data does provide useful information that can narrow the bounds on the identification regions of the average causal effects. We derive sharp bounds for the average causal effects within basic principal strata given the distribution of observables (Y, D, Z) under two sets of assumptions. To account for the sampling uncertainty in the distribution of (Y, D, Z), we develop confidence intervals for the bounds of the identification regions.

When the confidence intervals on the bounds of the identification regions for the average causal effect of treatments A and B for basic principal stratum $0AB$ overlap, making a decision on treatment strategy (whether to offer A or B first) is not straightforward based on the available data. We are working on two approaches that can provide further

information about average causal effects within basic principal strata beyond the bounds presented in this paper. One approach we are working on is to take advantage of additional information, such as covariates that predict principal stratum membership, to provide more decisive results. Another approach we are working on is to develop methods of sensitivity analysis. In this paper, we derive the bounds based on the most extreme relationship between certain potential outcomes which are associated with each observable outcome. If the relationship between certain potential outcomes follows a parametric model, we can do sensitivity analyses on the parameters of interest and then the bounds are the extreme results of these sensitivity analyses. We are working on developing appropriate models for sensitivity analysis. Related work on sensitivity analysis includes Vansteelandt and Goetghebeur (2001) and Vansteelandt et al. (2005).

Our method to derive bounds on average causal effects within basic principal strata for three-arm trials in this paper can be extended to trials with more than three arms or more than two levels of compliance. However, fast and reliable methods for computing the bounds for these trials requires further research. Additionally, the ratio of basic principal strata to observed strata of (Z, D) increases as the number of arms and/or compliance levels increase, making the identification problems more severe and hence making the bounds of average causal effects within basic principal strata less informative. Appropriate further assumptions that can reduce the identification problems when there are multiple arms and more than two levels of compliance is a valuable topic for future research. In this study, we focus on binary outcomes. Extensions to more general outcomes would be worthwhile.

Acknowledgements

This research was supported by a grant R01-MH-61892 "Mixed Models for Discrete Data with Non-compliance". The authors thank Thomas R. Ten Have for valuable discussions and for insightful comments on the paper. The authors also thank Marshall Joffe and Kevin Lynch for valuable comments and thank David Oslin for providing us with the data from the trial of treatments for alcohol dependence and for helpful discussion of the data. The authors are also grateful to the referees and associate editor for insightful comments that improved the paper.

APPENDIX

Appendix A:

CONSTRUCTION OF BOUNDS FOR AVERAGE POTENTIAL OUTCOMES UNDER THE CONTROL GIVEN PROPORTIONS IN PRINCIPAL STRATA WITH ASSUMPTIONS $1-4$:

For constructing bounds for average potential outcomes under the control given proportions in principal strata, it is useful to re-express (9)-(11) as

$$\begin{aligned} E(Y|Z=0,D=0) - p_{0|B}E(Y|Z=B,D=0) & \\ -p_{0|A}E(Y|Z=A,D=0) &= \pi_{0AB}E(Y_0|0AB) - \pi_{000}E(Y_0|000) \quad (19) \\ E(Y|Z=0,D=0) - p_{0|A}E(Y|Z=A,D=0) &= \pi_{0AB}E(Y_0|0AB) + \pi_{0A0}E(Y_0|0A0) \quad (20) \\ E(Y|Z=0,D=0) - p_{0|B}E(Y|Z=B,D=0) &= \pi_{0AB}E(Y_0|0AB) + \pi_{00B}E(Y_0|00B) \quad (21) \end{aligned}$$

By maximizing/minimizing $E(Y_0|0AB)$ with constraints (12) and (19)-(21), or equivalently (9)-(12), using linear programming as in Lemma 1, we have the following bounds for $E(Y_0|0AB)$ given π_{0AB}:

$$\begin{aligned} E(Y_0|0AB,\pi_{0AB}) &\in (minE(Y_0|0AB,\pi_{0AB}), maxE(Y_0|0AB,\pi_{0AB})) \\ &= (max\begin{pmatrix} 0 \\ min\{1, \frac{q_{1|00}-p_{0|B}q_{1|B0}-p_{0|A}q_{1|A0}}{\pi_{0AB}}\} \\ min\{1, \frac{q_{1|00}-p_{0|A}q_{1|A0}-p_{A|A}+\pi_{0AB}}{\pi_{0AB}}\} \\ min\{1, \frac{q_{1|00}-p_{0|B}q_{1|B0}-p_{B|B}+\pi_{0AB}}{\pi_{0AB}}\} \end{pmatrix}, \\ & \quad min\begin{pmatrix} 1 \\ max\{0, \frac{q_{1|00}-p_{0|B}q_{1|B0}-p_{0|A}q_{1|A0}}{\pi_{0AB}} + \frac{p_{0|B}-p_{A|A}+\pi_{0AB}}{\pi_{0AB}}\} \\ max\{0, \frac{q_{1|00}-p_{0|A}q_{1|A0}}{\pi_{0AB}}\} \\ max\{0, \frac{q_{1|00}-p_{0|B}q_{1|B0}}{\pi_{0AB}}\} \end{pmatrix}). \end{aligned}$$

Given π_{0AB}, maximizing/minimizing $E(Y_0|0A0)$ with constraints (12) and (19)-(21), or equivalently (9)-(12), is equivalent to maximizing/minimizing $E(Y_0|0A0)$ with constraints (12) and (20) given $E(Y_0|0AB,\pi_{0AB})$ bounded as above. Using linear programming as in Lemma 1 we have

$$\begin{aligned} E(Y_0|0A0,\pi_{0AB}) &\in (minE(Y_0|0A0,\pi_{0AB}), maxE(Y_0|0A0,\pi_{0AB})) \\ &= (max\begin{pmatrix} 0 \\ min\{1, \frac{q_{1|00}-p_{0|A}q_{1|A0}-\pi_{0AB}[maxE(Y_0|0AB,\pi_{0AB})]}{p_{A|A}-\pi_{0AB}}\} \end{pmatrix}, \\ & \quad min\begin{pmatrix} 1 \\ max\{0, \frac{q_{1|00}-p_{0|A}q_{1|A0}-\pi_{0AB}[minE(Y_0|0AB,\pi_{0AB})]}{p_{A|A}-\pi_{0AB}}\} \end{pmatrix}). \end{aligned}$$

Similarly we have

$$\begin{aligned} E(Y_0|00B,\pi_{0AB}) &\in (minE(Y_0|00B,\pi_{0AB}), maxE(Y_0|00B,\pi_{0AB})) \\ &= (max\begin{pmatrix} 0 \\ min\{1, \frac{q_{1|00}-p_{0|B}q_{1|B0}-\pi_{0AB}[maxE(Y_0|0AB,\pi_{0AB})]}{p_{B|B}-\pi_{0AB}}\} \end{pmatrix}, \\ & \quad min\begin{pmatrix} 1 \\ max\{0, \frac{q_{1|00}-p_{0|B}q_{1|B0}-\pi_{0AB}[minE(Y_0|0AB,\pi_{0AB})]}{p_{B|B}-\pi_{0AB}}\} \end{pmatrix}). \end{aligned}$$

Appendix B:

RESULTS ON THE AVERAGE CAUSAL EFFECTS WITHIN PRINCIPAL STRATA UNDER ASSUMPTIONS $1-5$

$$E(Y_B - Y_0|0AB) = q_{1|BB} - min\left(\begin{array}{c} 1 \\ max\{0, \frac{q_{1|00} - p_{0|B} q_{1|B0}}{p_{B|B}}\} \end{array} \right);$$

$$\begin{aligned} E(Y_A - Y_0|0AB) \in & \left(max\left(\begin{array}{c} 0 \\ 1 - \frac{1-q_{1|AA}}{\frac{\pi_{0AB}}{p_{A|A}}} \end{array} \right) - min\left(\begin{array}{c} 1 \\ max\{0, \frac{q_{1|00} - p_{0|B} q_{1|B0}}{p_{B|B}}\} \end{array} \right), \right. \\ & \left. min\left(\begin{array}{c} 1 \\ \frac{q_{1|AA}}{\frac{\pi_{0AB}}{p_{A|A}}} \end{array} \right) - min\left(\begin{array}{c} 1 \\ max\{0, \frac{q_{1|00} - p_{0|B} q_{1|B0}}{p_{B|B}}\} \end{array} \right)\right); \end{aligned}$$

$$\begin{aligned} E(Y_A - Y_0|0A0) \in & \left(max\left(\begin{array}{c} 0 \\ 1 - \frac{1-q_{1|AA}}{1-\frac{\pi_{0AB}}{p_{A|A}}} \end{array} \right) - min\left(\begin{array}{c} 1 \\ max\{0, \frac{p_{0|B} q_{1|B0} - p_{0|A} q_{1|A0}}{p_{A|A} - p_{B|B}}\} \end{array} \right), \right. \\ & \left. min\left(\begin{array}{c} 1 \\ \frac{q_{1|AA}}{1-\frac{\pi_{0AB}}{p_{A|A}}} \end{array} \right) - min\left(\begin{array}{c} 1 \\ max\{0, \frac{p_{0|B} q_{1|B0} - p_{0|A} q_{1|A0}}{p_{A|A} - p_{B|B}}\} \end{array} \right)\right), \end{aligned}$$

where

$$\pi_{0AB} = p_{B|B}.$$

References

Angrist, J. D., Imbens, G. W. and Rubin, D. B. (1996) Identification of causal effects using instrumental variables. *J. Am. Statist. Ass.*, **91**, 444-455.

Balke, A. and Pearl, J. (1997) Bounds on treatment effects from studies with imperfect compliance. *J. Am. Statist. Ass.*, **92**, 1171-1176.

Beran, R. (1988) Balanced simultaneous confidence sets. *J. Am. Statist. Ass.*, 83, 679-697.

——— (1990) Refining bootstrap simultaneous confidence sets. *J. Am. Statist. Ass.*, 85, 417-426.

Bickel, P. J. and Freedman, D. A. (1981) Some asymptotic theory for the bootstrap. *Ann. Statist.*, **9**, 1196-1217.

Efron, B. and Tibshirani, R. (1998) The problem of regions. *Ann. Statist.*, 26, 1687-1718.

Frangakis, C. E. and Rubin, D. B. (2002) Principal stratification in causal inference. *Biometrics*, **58**, 21-29.

Fukuda, K. (2004) Frequently asked questions in polyhedral computation. *http://www.ifor.math.ethz.ch/staff/fukuda/polyfaq/polyfaq.html.*

Goetghebeur, E. and Molenberghs, G. (1996) Causal inference in a placebo-controlled clinical trial with binary outcome and ordered compliance. *J. Am. Statist. Ass.*, 91, 928-934.

Horowitz, J. and Manski, C. (2000) Nonparametric analysis of randomized experiments with missing covariate and outcome data. *J. Am. Statist. Ass.*, **95**, 77-84.

Imbens, G. W. and Angrist, J. D. (1994) Identification and estimation of local average treatment effects. *Econometrica*, 62, 467-476.

Imbens, G. W. and Rubin, D. B. (1997) Bayesian inference for causal effects in randomized experiments with noncompliance. *Ann. Statist.*, 25, 305-327.

Imbens, G. W. and Manski, C. F. (2004) Confidence intervals for partially identified parameters. *Econometrica*, **72**, 1845-1857.

Joffe, M. M. (2001) Using information on realized effects to determine prospective causal effects. *J. R. Statist. Soc. B*, **63**, 759-774.

Joffe, M. M. and Brensinger, C. (2003) Weighting in instrumental variables and G-estimation. *Statist. Med.*, **22**, 1285-1303.

Manski, C. F. (1990) Non-parametric bounds on treatment effects. *Am. Econ. Rev., Papers and Proceedings*, **80**, 319-323.

——— (2003) *Partial identification of probability distributions.* New York: Springer-Verlag.

Pearl, J. (1995) On the testability of causal models with latent and instrumental variables. In *Uncertainty in Artificial Intelligence 11* (eds P. Besnard and S. Hanks) , pp. 435-443. San Francisco: Morgan Kaufmann.

Robins, J. M. (1989) The analsyis 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. Bailey), pp. 113-159. Washington, DC: NCHSR, U.S. Public Health Service.

Rubin, D. B. (1980) Comment on 'Randomization analysis of experimental data: the Fisher randomization test' by D. Basu. *J. Am. Statist. Ass.*, 75, 591-593.

Shafer, G. (1982) Belief functions and parametric models. *J. R. Statist. Soc. B*, **44**, 322-352.

Small, D., Ten Have, T. R., Joffe, M. and Cheng, J. (2006) Random effects logistic models for analyzing efficacy of a longitudinal randomized treatment with non-adherence. *Statist. Med.*, 25, 1981-2007.

Sommer, A. and Zeger, S. L. (1991) On estimating efficacy from clinical trials. *Statist. Med.*, 10, 45-52.

Ten Have, T. R., Elliott, M. R., Joffe, M., Zanutto, E. and Datto, C. (2004) Causal models for randomized physician encouragement trials in treating primary care depression. *J. Am. Statist. Ass.*, 99, 16-25.

Vansteelandt, S. and Goetghebeur, E. (2001) Generalized linear models with incomplete outcomes: the IDE algorithm for estimating ignorance and uncertainty. *J. Comput. Graph. Statist.*, 10, 656-676.

Vansteelandt, S., Goetghebeur, E., Kenward, M. G. and Molenberghs, G. (2006) Ignorance and uncertainty as inferential tools in a sensitivity analysis. *Statist. Sin. (In Press)*.

Walley, P. (1991) *Statistical reasoning with imprecise probabilities*. London: Chapman and Hall.

Zhang, J. L. and Rubin, D. B. (2003) Estimation of causal effects via principal stratification when some outcomes are truncated by death. *J. Educ. Behav. Statist.*, **28**, 353-368.

Zisserson, R. N., Lynch, K. G., Pettinati, H. M., Volpicelli, J. R. and Oslin, D. W. (2004) The effect of social support on alcoholism treatment outcome (manuscript).