Interpreting Treatment Effects When Cases Are Institutionalized After Treatment

Daniel F. McCaffrey, Andrew R. Morral, Greg Ridgeway and Beth Ann Griffin

Drug Policy Research Center, The RAND Corporation

ABSTRACT: Drug treatment clients are at high risk for institutionalization, i.e., spending a day or more in a controlled environment where their freedom to use drugs, commit crimes, or engage in risky behavior may be circumscribed. For example, in recent large studies of drug treatment outcomes, more than 40% of participants were institutionalized for a portion of the follow-up period. When longitudinal studies ignore institutionalization at follow-up, outcome measures and treatment effect estimates conflate treatment effects on institutionalization with effects on many of the outcomes of interest. In this paper, we develop a causal modeling framework for evaluating the four standard approaches for addressing this institutionalization confound, and illustrate the effects of each approach using a case study comparing drug use outcomes of youths who enter either residential or outpatient treatment modalities. Common methods provide biased estimates of the treatment effect except under improbable assumptions. In the case study, the effect of residential care ranged from beneficial and significant to detrimental and significant depending on the approach used to account for institutionalization. We discuss the implications of our analysis for longitudinal studies of all populations at high risk for institutionalization.

Keywords: Incarceration; Substance use; Drug treatment; Time at risk; Outcomes evaluations; Adolescents; Treatment modality; Causal models.

Address correspondence concerning this article to: Daniel F. McCaffrey, Ph.D. The RAND Corporation, 4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213, Ph: (412) 683-2300 x4919 Fax: (412) 683-2800 E-mail: daniel_mccaffrey@rand.org. Affiliation and contact information for additional authors: Dr. Andrew Morral, The RAND Corporation, 1200 S. Hayes St., Arlington, VA 22202-5050, Ph: (703) 413-1100 x5119, Fax: (703) 414-4710, E-mail: andrew_morral@rand.org; Dr. Greg Ridgeway, The RAND Corporation, 1776 Main St., Santa Monica, CA 90407-2138 Ph: (310) 393-0411 x7734 Fax: (310) 393-4841 E-mail: gregr@rand.org; Dr. Beth Ann Griffin, The RAND Corporation, 1200 S. Hayes St., Arlington, VA 22202-5050, Ph: (703) 413-1100 x5119, Fax: (703) 414-4710, E-mail: beth_ann_griffin@rand.org;

This research was support by NIDA Grants R01 DA015697, R01 DA016722 and R01 DA017507.

1 Introduction

Drug treatment clients are a population at particularly high risk of institutionalization, defined here as spending a day or more in a controlled environment where the possibility of drug use and criminal activity is substantially diminished (e.g., a jail, prison, hospital, residential treatment or group home setting). This is evident in large samples of drug treatment clients. For instance, in the Drug Abuse Treatment Outcomes Study (Hubbard et al., 1997), 40% of the 2966 clients of U.S. substance abuse treatment programs interviewed 12-months after discharge reported institutionalization for some part of the preceding year (U.S. Dept. of Health and Human Services, National Institute on Drug Abuse, 2004). Among those with any institutionalization, the average number of days institutionalized out of the past 365 was 115 (U.S. Dept. of Health and Human Services, National Institute on Drug Abuse, 2004). Similarly, over 2600 cases (about 52% of the sample) from the National Treatment Improvement Evaluation (NTIES; U.S. Dept. of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Substance Abuse Treatment, 2004; NORC, and RTI, 1997) were institutionalized during the study's 12-month post-treatment evaluation. Of these cases, more than 630 were incarcerated for the entire evaluation period and excluded from analyses.

Because institutionalization limits people's freedoms, it can cause apparent improvement in many of the most important substance abuse treatment outcomes, such as reductions in drug and alcohol use, drug problems, crime, and even psychological problems (Piquero et al., 2001; Webb et al., 2002). Similarly, it can lead to seeming improvements in outcomes such as participation in educational programs or access to health services. However, since institutionalization is often not a positive outcome, these seemingly positive results can lead to misleading inferences about the benefits of alternative treatments.

Consider, for instance, a hypothetical experiment in which drug users randomly assigned to Treatment A are later found to have higher rates of abstinence, but also higher rates of institutionalization than those assigned to Treatment B. This pattern of findings raises the possibility that Treatment A is reducing drug use only by virtue of its effect on institutionalization. This may be unsatisfactory for at least two reasons. First, some types of institutionalization (e.g., prison) may represent a worsening of the clients' conditions, not improvement, at a substantial societal cost. In this case, the positive effect on drug use might actually be a side effect of an otherwise costly negative treatment effect. Second, most stakeholders (clients, payers, referrers) seek treatments that will reduce clients likelihood of using substances when free in the community, rather than while institutionalized. As such, rates of drug use during periods of institutionalization actually obscure the effect of interest

The perspective developed above suggests that for the purpose of understanding most treatment effects (e.g., the differential effects of a single intervention on population subgroups) and providing information needed by stakeholders, estimated rates of drug use (and many other outcomes) should disentangle improvement due to less use among patients when they are free in the community from reductions in use due to institutionalization. For example, we might estimate rates that would be expected if every client was at risk (i.e., not institutionalized) for the entire period of observation. Stated another way, we should try to answer the question, "What would the effects of treatment be if none of its recipients had been institutionalized?"

Figure 1 demonstrates the problem in a common graphical model of direct and indirect effects (MacKinnon et al., 2000). The figure shows that treatment affects institutionalization which in turn affects substance use or other outcomes and also shows that treatment has a direct effect on outcomes. It is the direct link from treatment to outcomes which is of primary interest to stakeholders and others. For clarity, the figure excludes external factors such as deviance that can influence treatment outcomes and institutionalization. However, it illustrates the important point that treatment affects institutionalization, thereby complicating the evaluation of the effectiveness of treatment.

******* Figure 1 About Here *******

This paper makes precise the notions of the effect represented by the direct link from treatment to outcomes, and it discusses estimation of this effect. Because institutionalization occurs post-treatment, estimating such an effect can be challenging. The paper also discusses the four most common of these approaches and identifies the often tacit assumptions required for each approach to recover effects of interest. The methods considered are 1) ignoring institutionalization; 2) combining the outcome of interest and institutionalization into a single measure that is used to assess treatment effects; 3) dropping institutionalized cases from the study without any adjustment for the censoring of the population this creates; and 4) controlling for institutionalization with statistical models such as linear regression, path or structural equation models. This paper's focus on providing technical details of statistical models and analytic methods may make it of greatest interest to methodologists; however, the issues surrounding inferences about treatment effectiveness in the presence of institutionalization during the evaluation should be of concern to all treatment researchers.

The next section describes the dataset used to illustrate the effects of different approaches to the institutionalization confound. Section 3 develops a general statistical model for causal effect analyses given post-treatment confounders such as institutionalization, and uses this model to highlight the assumptions and limitations of each of the most common approaches to addressing these confounds. To aid readers unfamiliar with the potential outcomes framework used in the causal models throughout Section 3, verbal descriptions of the statistical model are presented first, followed by formal notation, which is used for precision. The paper closes with a discussion of strategies for investigating treatment effects in the presence of post-treatment confounders.

2 Empirical Example: The Effects of Treatment Modality on Adolescent Outcomes

To demonstrate the effects of different approaches to addressing institutionalization in outcomes analyses, we use a study of the effects of treatment modality (residential versus outpatient) on the 12-month substance use outcomes for adolescents who participated in the Adolescent Treatment Models (ATM) study, fielded between 1998 and 2002 by the Substance Abuse and Mental Health Services Administration, Center for Substance Abuse Treatment. The ATM study collected treatment admission and 12-month outcomes data for new admissions to 10 community-based treatment programs in the United States, including six residential programs and four outpatient programs (for details of the ATM study and treatment programs, see Stevens and Morral, 2003).

The sample used in the present analyses includes all new admissions to the 10 programs in the main ATM analytic dataset, which was produced in March of 2002. Of these 1,384 cases, 1,256 (90.75%) completed a 12-month follow-up survey and provided data on the outcomes of interest. Only cases with follow-up data are included in the analyses presented below.

For purposes of the illustrations presented in this report, we examine the effects of treatment modality on the Substance Frequency Index (SFI), a widely used scale from the Global Appraisal of Individual Needs (GAIN; Dennis, 1999), the survey instrument used at every site for baseline and 12-month outcome assessments. The SFI averages responses to a series of questions on the frequency of recent drug use, intoxication, and drug problems in the 90 days prior to the 12 month follow-up.[1] It is scaled so that higher values indicate greater substance use and more drug problems. Days of institutionalization at follow-up is assessed with the GAIN's MAXCE variable, which calculates the maximum number of days–in the past 90–which the respondent reports being in any of several different types of controlled environments (e.g., inpatient psychiatric or medical hospitals, residential treatment facilities, juvenile halls or other criminal justice detention facilities, etc.).

Because the ATM is an observational or quasi-experimental study, there were observable differences in the pretreatment characteristics of youths entering residential and outpatient care. Given these pretreatment differences, any observed differences in treatment group outcomes could result either from differential effectiveness of the treatment modalities, or because of differences in how hard their respective populations are to treat. In order to isolate just the treatment effects of interest, we must compare treatments on equivalent cases. Thus, for the case study we compare the effectiveness of the two modalities on cases like those in the ATM sample who entered the residential modality. We achieve this comparison by developing case weights for the outpatient sample that upweight outpatients with pretreatment characteristics similar to those of youths entering the residential modality and downweight outpatient cases that are dissimilar to the residential cases (McCaffrey et al., 2000). Using an optimization algorithm, we chose weights that make the two groups effectively equivalent in terms of the distributions of 86 pretreatment GAIN measures of the American Society of Addiction Medicine (ASAM) patient placement criteria (Mee-Lee et al., 2001).[2] Table 4 in the appendix **Reference Supplemental Material** provides details.

Thus, in the remainder of the analyses reported in this paper, we compare the unweighted residential

[1] In this illustration we multiplied SFI by 90 to make it scale with use rather than use per day.

[2] We chose the weights so the weighted cumulative distribution of the outpatient group matched the unweighted cumulative distribution of the residential group as measured by the Kolmogorov-Smirnov statistic.

sample (n=770) to the weighted outpatient sample (n=486, effective sample size = 125, design effect due to weighting of 3.9), a comparison designed to examine whether the residential modality produces better 12-month outcomes than outpatient care for cases with pretreatment characteristics like those of clients who usually enter residential care. Table 1 provides weighted descriptive statistics for SFI and institutionalization at follow-up by treatment modality.

******* Table 1 About Here *******

3 The Institutionalization Confound

3.1 Causal Effects of Treatment in the Presence of Institutionalization

To understand the impact of institutionalization on estimators of treatment effects, we need a precise definition of a treatment effect that is consistent with the needs of stakeholders and the inferences they make from common estimators. That is, we need a precise definition of the direct link from treatment to outcomes in Figure 1. We start by considering the treatment effect in the simple case without any institutionalization. In this case, we are interested in the change in a youth's outcome following treatment compared to the outcome the youth would have had if given the control condition. Thus, we need to consider how a youth would behave following treatment, which we refer to as the potential outcome for treatment, and we need to consider how this youth would behave following the control condition, which we refer to as the potential outcome for control. The change due to treatment or the "treatment effect" is the difference between these two potential outcomes. The model that uses potential outcomes to define treatment effects, the Neyman-Rubin causal model (Holland, 1986; Pearl, 1996) serves as the basis for our definition of a causal treatment effect in the presence of institutionalization.

In the presence of institutionalization, we need to consider additional potential outcomes; a youth's outcome might depend not only on treatment assignment but also days institutionalized. In the ATM study, for instance, every youth could be viewed as having 182 potential outcomes for his or her SFI score at the 12-month follow-up: one each for every possible number of days institutionalized (0-90) following residential care and one for every possible number of days institutionalized following outpatient care. That is, for residential care we imagine the youth having a set of 91 potential outcomes, one for 0 days institutionalized, one for 1 day institutionalized, etc. Similarly, the youth will have 91 potential outcomes under outpatient care as well. We define treatment effects in terms of the differences between the set of potential outcomes for treatment and the set of potential outcomes for control. For example, a treatment effect of particular interest is the difference between a youth's potential outcome when not institutionalized and receiving residential care and a youth's potential outcome when not institutionalized and receiving outpatient care.

We use the following shorthand notation for these sets of potential outcomes: y denotes the outcome of interest scaled so that high values imply negative outcomes (e.g., greater drug use, higher values of the SFI scale, or lower levels of schooling or employment), t denotes treatment assignment, and z denotes institutionalization. For treatment, $t = 1$, we label the 91 potential outcomes for a youth as $y_1[z]$ where $z = 0, \ldots, 90$, so that $y_1[0]$ is the youth's potential outcome if assigned to treatment and not institutionalized in the 90 days prior to follow-up. Similarly, $y_1[1]$ is the potential outcome following treatment and 1 day of institutionalization prior to the follow-up. There are 89 more potential $y_1[z]$ for $z = 2, 3, \ldots, 90$. The variable z indexes the potential outcomes in the set; it is not restricted to the actual number of days that a youth is observed to be institutionalized. Instead, z_1 denotes a youth's potential days institutionalized following treatment and z_0 denotes potential days institutionalized following control. If a youth receives treatment, we observe z_1 days of institutionalization and we observe the outcome $y_1[z_1]$, which equals the potential outcome for treatment at z_1 days institutionalized. If a youth receives control, we observe z_0 and $y_0[z_0]$.

We assume that treatment assignment is a random variable denoted by T. Given $T = t$, we observe $y_t[z_t]$ for each youth. We let Y_{obs} denote a random variable equal to the observed value of the outcome and Z_{obs} denote random variable equal to the observed value of institutionalization. The variables are random because they are determined by the random treatment assignment variable: $Z_{obs} = z_T$ and $Y_{obs} = y_T[Z_{obs}]$. Given this shorthand notation, we can now provide precise definitions of treatment effects that match the desired inferences of stakeholders.

3.2 Defining Treatment Effects: Complications Due to Suppression and Selection

Institutionalization poses challenges to defining the causal effects of treatments, in part because it can suppress or modify the values observed on many outcomes. For instance, a youth who would otherwise commit a property crime every day might report only 10 days of such crimes in the past 90, if he or she spent 80 of those days in a detention center, jail, residential treatment program or other controlled environment. We refer to the dependence of outcomes on institutionalization as the "dose-response" effect of institutionalization.

Figure 2 presents an example of potential outcomes for a hypothetical youth, which illustrates a dose-response relationship between SFI and institutionalization. The youth has two sets of potential outcomes denoted by two lines. The upper line represents potential outcomes under the control condition, $y_0[z]$, whereas the lower line represents potential outcomes under the treatment condition, $y_1[z]$. Both lines demonstrate dose-response relationships where increased values of institutionalization suppress substance use as measured by SFI.

******* Figure 2 About Here *******

For any given value of z, the difference between the two lines represents the treatment effect at that level of institutionalization. Because the dose-response relationship is not the same for treatment and control (the slope of the control line is steeper than that of the treatment line), the treatment effect will depend on the observed value of z. The figure shows the treatment effect at $z = 0$, which we refer to as the "unsuppressed" treatment effect because it measures the treatment effect on SFI when SFI is not suppressed by institutionalization.[3] Alternative measures, such as the average treatment effect over values of z, might also be of interest in some applications. In this paper, we focus mainly on the unsuppressed effect because it answers the question, "What would the effects of treatment be if none of its recipients had been institutionalized?". Furthermore, the unsuppressed treatment effect is implicit in most inferences drawn from treatment effect estimates including inferences about direct effects made from path

[3] In this paper we use the term "suppression" to refer to the effect of institutionalization on outcomes during the evaluation. This usage is distinct from usage in structural equation modeling, which refers to the effects of suppressor variables biasing statistical estimates of effects in structural models toward zero (MacKinnon et al., 2000). The suppression effects of institutionalization might make treatment effect estimates larger or smaller than the unsuppressed effect.

models or structural equation models that attempt to control for the effects of institutionalization.

While our model allows us to define precisely the treatment effect of interest using potential outcomes, we cannot observe the potential outcome under treatment and no institutionalization and the potential outcome under control and no institutionalization. We can only observe youths under one treatment condition and at whatever level of institutionalization occurs under that condition. Hence, when we use our data to estimate the treatment effect, we need to consider what assumptions about the nature of treatment assignment, institutionalization and substance use are required for that estimate to provide accurate information about the ”unsuppressed” treatment effect. The following paragraphs illustrate ways that, in the absence of such assumptions, institutionalization can affect differences between the treatment control groups and confound estimation of treatment effects.

Estimating treatment effects is challenging because when suppression exists, the average Y_{obs} for control equals the average of $y_0[z_0]$, which is generally not a valid counterfactual for the average of $y_1[z_1]$, unless the distribution of z_0 equals the distribution of z_1, and for both groups observed values for institutionalized cases will not equal $y_t[0]$. Figure 3 depicts this scenario and the bias that can result from the suppression effect of institutionalization. In the figure, there are four hypothetical cases: two who entered treatment (denoted by the dashed lines) and two who entered control (denoted by the solid lines). The figure shows both the $y_0[z]$ and $y_1[z]$ lines for each case, for a total of 8 lines for the four cases. The potential outcomes for control have high values at $z = 0$ and steep slopes; the potential outcomes for treatment are at the bottom of the figure with low values at $z = 0$ and flat slopes. Again the treatment effect is shown by the difference between the two lines for each case. The figure also shows the SFI and institutionalization actually observed for each case (o for control and + for treatment). In this example, the treatment cases are institutionalized for more days than the control cases. Consequently, the difference in the observed means is greater than any true treatment effect including the unsuppressed effect. In situations like Figure 3 when the observed distributions of z_1 and z_0 differ, a model is required to equate these distributions in order to estimate unbiased treatment effects.

******* Figure 3 About Here *******

Modeling to correct for the effects of suppression is complicated by another source of bias caused by institutionalization: selection bias. Selection occurs when youths with different observed values of

Z_{obs} have potential outcomes that are systematically different from those of other cases. For example, youths with higher levels of institutionalization might be youths with the highest potential outcome values (e.g., greatest amount of drug use or lowest levels of schooling or employment) regardless of treatment assignment. In other words, the samples of youths at each observed value of Z_{obs} are not a random subset of the population. Because institutionalization occurs after treatment, cases at any given level of institutionalization might differ on factors that influence outcomes, and their potential outcomes might not be a random subsample of potential outcomes for all the youths in the sample, even though the treatment and control groups were randomly assigned or otherwise equivalent. Because institutionalization occurs post-treatment, balance on these factors before treatment assignment does not guarantee that there will be no selection in the samples with differing values of institutionalization.

Selection can be particularly problematic when it operates differently for the treatment and control groups, because in this situation the control cases at a given level of institutionalization do not provide a valid counterfactual for the treatment cases with the same level of institutionalization. Again, this can occur even in randomized trials where treatment and control groups are matched on pretreatment variables. Figure 4 demonstrates this bias with eight hypothetical cases: four treatment (+) and four control (o). In this example, treatment does not affect outcomes, i.e., $y_1[z] \equiv y_0[z]$ for all youths and all values of z, but treatment does increase institutionalization, yielding a larger average value of z_1 than z_0. There are two kinds of cases, high use (labeled Heavy Users) and low use (labeled Light Users), and the distribution of institutionalization among the two types of cases differs between treatment and control, creating spurious treatment effects. Moreover, as shown by the cases with $z_1 = z_0 = 40$, even comparisons between treatment and control cases with equal institutionalization can result in bias because they involve different kinds of youths from the treatment and control groups.

******* Figure 4 About Here *******

The various approaches discussed below for estimating treatment effects produce valid estimates of the unsuppressed treatment effect only under specific assumptions about both suppression and selection. We now discuss those assumptions.

3.3 Ignoring Institutionalization

Although most longitudinal studies of substance abusers develop elaborate procedures for ensuring institutionalized respondents are included in all follow-up surveys, their institutionalization status at follow-up is rarely reported, or, when it is reported, is not used to correct treatment effect estimates. Many, possibly most, studies of substance abuse treatment could be cited as examples of this claim. Even recent reports of the largest experimental and observational studies report little information about institutionalization of the sample, such as those for the Methamphetamine Treatment Project, Project Match, the National Treatment Outcomes Study, and DATOS-A (Hser et al., 2001; Gossop et al., 2003; Project MATCH Research Group, 1997; Rawson et al., 2004).

In other cases, institutionalization at follow-up is acknowledged and reported, but authors argue that it should not bias treatment effect estimates. For instance, in a study of 447 adolescent probationers referred to a long-term residential substance abuse treatment program or to another probation disposition, Morral et al. (2004) showed that treatment and weighted comparison groups spent equivalent durations institutionalized over the follow-up period, suggesting that bias in treatment effect estimates due to institutionalization should favor neither the treatment nor the comparison conditions. While this argument may be correct when institutionalization is comparable between groups, institutionalization may nevertheless make interpretation of effects difficult, because estimated effects combine outcomes for youths with differing exposure to the risk of drug use.

When institutionalization is ignored, treatment effects are estimated by simple differences between treatment and control group means, and, as demonstrated in Figures 2 to 4, these simple differences fail to recover the unsuppressed treatment effect when either suppression or selection effects occur. Ignoring institutionalization will not bias results when there is no suppression so that all potential outcomes are unrelated to institutionalization ($y_t[z] \equiv y_t$ for all values of z and $t = 0$ or 1). For example, institutionalization might not have a dose-response relationship with measures not sensitive to opportunities to use drugs, such as attitude measures. Similarly, if treatment effects ($y_1[z] - y_0[z]$) do not depend on the level of institutionalization, z for every case in the population of interest, and if there are no treatment effects on institutionalization (z_1=z_0 for every case), then ignoring institutionalization does not bias estimates.

For the case study, the results when institutionalization is ignored are presented in the first two rows of Table 2.[4] Because the SFI measure is highly skewed, we also present the difference in medians (again weighted for the outpatient group). Values are negative and statistically significant for the median indicating that the residential care has a beneficial total effect on substance use frequency. Approximate randomization tests were used to test for differences in the median (Efron and Tibshirani, 1993).

However, as shown in Table 1, institutionalization is very prevalent and differs between modalities in the case study. Fifty-two percent of the residential sample reported some institutionalization at follow-up, compared with 39 percent of the weighted outpatient sample. Thus, the potential outcomes for the observed residential cases are at higher average values of institutionalization than those for outpatient cases. Therefore, the finding that residential cases have lower observed 12-month substance frequency scores than outpatients could occur because: 1) residential care is superior to outpatient care, or 2) residential care is inferior to outpatient care, but this effect is masked by the differential suppression of SFI scores caused by higher rates of institutionalization among residential cases.

******* Table 2 About Here *******

If it is reasonable to assume either that institutionalization does not suppress the outcomes or that treatment effects do not depend on the number of days youths were institutionalized, then we can conclude that residential care is indeed superior to outpatient care for reducing substance use. However, Figure 5, which plots the relationship between SFI and institutionalization in our case study sample, strongly suggests that the assumption of no suppression does not hold with the ATM data. There is clearly a dose-response relationship between institutionalization and SFI. While this relationship could result from pernicious selection where youths with the least risk for substance use are institutionalized for the longest periods, this seems unlikely since we typically find a positive relationship between risk for drug use and risk for other problem behaviors that could result in institutionalization.

******* Figure 5 About Here *******

Treatment effect estimates that ignore institutionalization provide unbiased estimates of the combined causal effect treatment has on institutionalization and potential outcomes at each level of institutional-

[4] Recall that throughout this example, we are assuming that pretreatment group differences were eliminated by weighting, so differences in outcomes are unbiased estimates of causal effects

ization. Often this type of effect is referred to as a total treatment effect on outcomes. Studies that report treatment effects without any correction for institutionalization should clarify that either the conditions of no suppression or selection are likely to hold or that inferences should be restricted to the total effect, which may be of limited value to clients and policymakers because it conflates desirable treatment effects on drug use with potentially undesirable effects on institutionalization, and does not distinguish the contribution of either to the final results.

3.4 Combining Outcomes and Institutionalization into a Composite Measure

The problem with the total effect estimated above is that both "good" and "bad" ways to achieve a reduction in substance use are treated equally. A common alternative is to create a composite measure of outcomes identified as good (e.g., no drug use and no institutionalization) and bad (e.g., drug use or institutionalization), and study treatment effects on this composite outcome. For instance, a principal outcome in the Cannabis Youth Treatment study (Dennis et al., 2000) is "recovery," which is defined over a follow-up interval as having few days of institutionalization, no past month substance use, and no symptoms of substance abuse or dependence.

For our illustrative case study, we constructed one such recovery indicator that takes a value of 1 if at follow-up a youth was not institutionalized for more than 15 of the past 90 days and had an SFI score of 0. On this variable, 15.3 percent of youths receiving residential care were in recovery, whereas 18.5 percent of the weighted youths receiving outpatient care met these criterion. The treatment effect is a statistically insignificant (p=0.39) 3.2 percentage point increase in recovery for youths receiving outpatient treatment.

Measures such as recovery have some limitations. These measures only determine whether cases are more likely to have a "good" outcome following treatment or control conditions. They provide no measure of the magnitude of the change in drug use behavior and no direct information about the unsuppressed treatment effect. Similarly, composite measures depend on the definition of a good outcome and that definition can be ambiguous at times or require value judgments not universally accepted. For instance, our recovery measure treats all institutionalization as a negative outcome. In some cases, however, institutionalization might be a positive outcome, such as when clients assigned to a long-term residential treatment program remain in that program 12 months later. In contrast, incarceration in jail or prison settings would ordinarily be considered a negative outcome. The univariate composite recovery defined above does not distinguish between these two types of institutionalization.

As shown in Table 1, 16.5 percent of the ATM residential sample remained in residential treatment at the 12 month follow-up. If we change our composite measure so that youths remaining in residential treatment at follow-up are counted as being in recovery, then 24.8 percent of youths receiving residential care were in recovery compared to 16.5 percent for outpatient care. Changing the definition of recovery dramatically changes the sign of the treatment effect in our case study.

Although this alternative definition of recovery might be more consistent with treatment providers' expectations, other stakeholders might find any long-term institutionalization of an adolescent unfavorable. Still others might argue that residential care following outpatient care is positive, since youths are still part of an evolving regime of treatment. As shown in the ATM example, each change to the definition of a good outcome results in a change to the treatment effect estimate.

Of course, more complex categorical composite measures are also possible. For instance, we could construct an outcome with three levels: not institutionalized and SFI=0, continuing in residential treatment and SFI=0, or any other status. Complex measures like these, however, often do not solve interpretation problems; treatment may reduce some but not all negative outcomes. Moreover, this approach continues to forsake the goal of establishing a unitary treatment effect, such as the unsuppressed effect.

3.5 Dropping Institutionalized Cases

Another common and straightforward approach used to untangle institutionalization from treatment effects drops cases when they have extensive periods of institutionalization. As noted earlier, this was the approach adopted in developing the NTIES analysis dataset, and consequently it has become the de facto method of all analyses using this dataset (e.g., Gerstein and Johnson, 1999; NORC, and RTI, 1997; Zhang et al., 2003).

Estimates of treatment effects from samples restricted to cases with no institutionalization recovers the unsuppressed treatment effect only under the following assumption.

Assumption 1. The distribution of potential control outcomes for cases with zero days of institutionalization under treatment equals the distribution of potential control outcomes for cases with zero days of institutionalization under control.

Using the notation of Section 3.1, Assumption 1 requires $f(y_0[0]|T = 1, Z_{obs} = 0) = f(y_0[0]|T = 0, Z_{obs} = 0)$.[5] That is, the distribution of potential control outcomes at zero institutionalization is the same for treatment and control cases that were not institutionalized. If average treatment effects are of interest, this assumption can be relaxed to $E(y_0[0]|T = 1, Z_{obs} = 0) = E(y_0[0]|T = 0, Z_{obs} = 0)$, the expected value of $y_0[0]$ for treatment cases with no institutionalization equals the expected value of $y_0[0]$ for control cases with no institutionalization.

Often, however, cases with $Z_{obs} = 0$ differ from other cases in terms of risk factors that influence outcomes of interest, so the distribution of potential outcomes for this group differs from other cases. Moreover, because some cases can have $z_1 = 0$ and $z_0 > 0$ (or visa versa), Assumption 1 further requires that if cases with $Z_{obs} = 0$ under treatment differ from the rest of the treatment population in terms of risk factors for outcomes, cases under control must differ in the same way.

When Assumption 1 holds, differences between treatment and control cases with no institutionalization provide unbiased estimates of the unsuppressed treatment effect for cases that would not be institutionalized following the treatment condition.[6] However, the distribution of treatment effects for cases with $z_1 = 0$ or $z_0 = 0$ might not match the distribution of such effects for the entire sample. Hence, Assumption 1 alone does not ensure unbiased estimation for the entire treatment population. To generalize to the entire population, we also need Assumption 2 to hold.

Assumption 2. For both treatment and control, the distribution of unsuppressed treatment effects for cases with zero days institutionalized is the same as would be found for all cases.[7]

[5] We use common notation of $f(Y|X = x)$ to denote the distribution of a variable "Y" conditional the variable "X" equaling the value x and we use $E(Y|X = x)$ to denote the expected value of Y given that $X = x$.

[6] We focus on the effect of "treatment on the treated" (Imbens, 2003) and so we require only that the distributions of potential control outcomes match. If we wanted to estimate treatment effects on the entire population we would need $f(y_0[0]|T = 1, Z_{obs} = 0) = f(y_0[0]|T = 0, Z_{obs} = 0)$ and $f(y_1[0]|T = 1, Z_{obs} = 0) = f(y_1[0]|T = 0, Z_{obs} = 0)$ to hold.

[7] If we are interested only in treatment on the treated then unbiased estimation is possible if Assumption 1 holds and Assumption 2 holds only for the treatment group.

In mathematical notation, Assumption 2 requires $f(y_1[0] - y_0[0]|T = t, Z_{obs} = 0) = f(y_1[0] - y_0[0]|T = t)$, for both t=0 and 1. That is, the distribution of the unsuppressed treatment effects, $y_1[0] - y_0[0]$, for cases with no institutionalization equals the distribution of these effects for all cases. This assumption requires there to be no selection effects. Assumption 2 is violated when the potential outcomes or risk profiles for youths with no institutionalization differ from other youths. For instance, cases that are not institutionalized might be those that are less involved in any illicit activities, including substance use.

Analysts might use pretreatment variables to account for differences in the risk profiles of cases with and without institutionalization, possibly by reweighting cases with no institutionalization to match the entire population in terms of their distributions on pretreatment variables. Such an approach will yield unbiased estimates of the unsuppressed effect for the entire sample, provided Assumptions 1 and 2 hold conditional on the pretreatment variables. However, institutionalization is a post-treatment outcome, which means that if Assumptions 1 and 2 hold conditional on pretreatment variables, then treatment must have had no differential effects on cases in terms of institutionalization and its relationship to outcomes. For example, if cases that respond best to treatment leave residential treatment before 12 months while cases with the greatest potential use remain in treatment, then the cases remaining in treatment are not a random subset of the entire residential population. Moreover, because the treatment influenced which cases where institutionalized, conditioning on pretreatment variables could not capture the differences between the potential outcomes of the cases that were or were not institutionalized. Hence, Assumption 2 would not hold conditional on pretreatment variables.

In the ATM study, youths were much more likely to be institutionalized following residential care. Moreover, as shown in Figure 5, the average SFI outcome of youths with $Z_{obs} = 0$ is substantially lower than the values for youths with a Z_{obs} between 1 and 10 or 20 days. For both treatment modalities, these differences are statistically significant even after controlling for baseline covariates. Given that we assume that individual dose-response curves are constant or decreasing as institutionalization increases, this empirical finding provides strong evidence of selection effects, and, therefore, the violation of Assumption 2, even conditional on pretreatment variables.

To account for violation of Assumptions 1 and 2, Zhang and Rubin (2003) provide bounds for the unsuppressed effect among cases with $Z_{obs} = 0$ that allow for possible selection effects. When applied to the ATM data, the bounds for the effect ranged from about -36 to about 46 for both the mean and the median. This range results from uncertainty about how large the selection effects might be; it does not include any measure of variability due to sampling error, which would make the bounds even wider. Clearly such wide bounds are of little use for making inferences about the relative effectiveness of alternative treatment modalities. The width of the bound depends on the proportion of cases with one or more days of institutionalization, which was very large for the ATM sample. In studies with less institutionalization, the bounds might prove more useful.

3.5.1 Dropping Periods of Institutionalization

An alternative to dropping cases affected by institutionalization is to drop time periods during which cases were institutionalized. In this approach, outcome rates are calculated for the portion of time each case is not institutionalized. For the ATM study, this approach required calculating drug use frequency per day free in the community using $Q_{obs} = Y_{obs}/(90 - Z_{obs})$. We pro-rate this value to 90 days free in the community by multiplying it by 90, so that, under the necessary assumptions, treatment effects on Q equal the unsuppressed effect. Cases institutionalized for 90 days must still be removed from these analyses. Estimates of the effects of residential care on this rate are presented in Table 2. The difference in means is positive suggesting that youths on average have greater rates of use following residential care. The difference in medians is negative but neither value is statistically significant.

Estimation of the unsuppressed treatment effects through an analysis of rates requires the following assumptions.

Assumption 3. A youth's rate of drug use does not depend on institutionalization. In terms of the potential outcomes, the assumption requires that $y_t[z]$ be proportional to $(90 - z)$ for $t = 0$, 1.

When this assumption of constant rate of use holds, then Q_{obs} estimates the rate for all values of z, not just the rate for the observed institutionalization. Values then can be compared across all youths and pro-rated to use at zero days institutionalized. One implication of Assumption 3 is that $y_t[z]$ must converge to zero as z approaches 90.

Without an assumption of a constant rate of use, $y_t[z]/(90 - z)$ depends on z and, for $z \neq 0$, $90 \times y_t[z]/(90 - z) \neq y_t[0]$. Treatment modality means and medians combine differences in the rates per day not institutionalized with differences in the distribution of Z_{obs} for both groups. For example, suppose $y_1[z]/(90 - z) = y_0[z]/(90 - z)$, but both are increasing with z. That is, suppose there is no treatment effect on the dose-response curve ($y_0[z] = y_1[z]$ for all z), but youths' intensity of use increases with fewer days free in the community. Youths enrolled in residential care will have higher average use rates than those enrolled in outpatient care, when youths in residential care have greater levels of institutionalization on average.

In the ATM dataset, a comparison of the SFI rate with days institutionalized reveals a weak, roughly linear relationship, which along with the assumption of no selection suggests that Assumption 3 might be violated by this data but that bias caused by the violation would tend to be small.

Dropping periods of institutionalization will result in unbiased estimation of the unsuppressed effect only if, in addition to Assumption 3, Assumptions 1 and 2 hold for rates and 90 days institutionalization because this method excludes youths who were institutionalized for 90 days creating the potential for selection bias. For the ATM sample, 16.5 and 7.2 percent for residential and outpatient care respectively were excluded because they were institutionalized for 90 days. Given the potential for selection bias, we might use the bounding method of Zhang and Rubin (2003) to determine a range of possible treatment effects. However, given the number of dropped cases in this analysis, the bounds range from -4 to 8 for the mean and -9 to 4 for the median, so are still too large to provide meaningful inferences about modality effects.

3.6 Controlling for Institutionalization with Statistical Models

Another common approach is to control statistically for the effects of institutionalization during the follow-up period by conditioning on this variable when estimating treatment effects (Piquero et al., 2001; Webb et al., 2002). Webb et al. (2002), for instance, examined treatment outcomes of youths in the CYT study who either were or were not involved with the legal system at program admission. Because these groups would be expected to have substantially different institutionalization rates that could bias drug use and other outcome measures, post-treatment admission institutionalization was included as a “covariate” in their outcomes models.

Path models and structural equation models like those represented by Figure 1 also include institutionalization as a covariate when estimating direct and indirect effects of treatment. However, unlike true covariates, which are measured prior to treatment, institutionalization occurs after treatment and consequently youths with different values of institutionalization can differ in terms of unobservable factors that affect outcomes, even if systematic differences between groups are eliminated through randomization or other means, such as weighting. Conditioning on this post-treatment variable can distort estimates of the treatment effect on other outcomes because conditioning does not necessarily compare like clients; the potential outcomes of cases with Z_{obs} equal to any specific value might differ from other cases. This potential for bias is shown in Figure 6. In this figure, treatment influences an unobserved and unmeasured variable that affects both institutionalization and outcomes. Thus, cases at different values of institutionalization will have different values of the unobserved confounder, which distorts the relationship between institutionalization and outcomes when the unobserved confounder is not controlled. However, because this variable is unobserved, it cannot be included in this model and path or structural equation models will yield biased results because of unaccounted for confounders (MacKinnon et al., 2000).

***** Figure 6 About Here *****

Figure 7 presents another demonstration of the potential for bias when using institutionalization as a covariate. In this hypothetical sample, there are no treatment effects ($y_1[z] = y_0[z] = y[z]$ for all values of z and every case). There are, however, four types of youths with increasing levels of risk represented by increasing values of $y[0]$. In this example, $y[z] = y[0] + rz$ for all groups with $y[0]$ increasing and r constant across groups. As we might expect, youths with greater risk have greater values of Z_{obs} and treatment increases institutionalization so that the values of z_1 are greater than z_0 for youths from the same risk group. This figure shows the dose-response curves for each risk group (solid lines) and the regression lines fit to all eight observations (dashed line) and to treatment and control observations separately (dotted lines). Because of selection, the slope of the regression line does not capture the slope of the true dose-response curves. Moreover, for treatment and control cases with equal values of z, the control cases are always from a lower risk group. Thus, the control observations are always below the regression line and the control group has lower mean outcomes after adjusting for Z_{obs}. Spurious results occur even if separate lines are fit for the treatment and control groups.

***** Figure 7 About Here *****

The first row of Table 3, labeled Model 1, presents treatment effects after controlling for institutionalization as a covariate with the ATM data. We used a standard linear regression model:

$$y_{obs} = \alpha + \beta z_{obs} + \delta T + \epsilon, \tag{1}$$

where T is a 0-1 treatment indicator variable. The `svyglm` function in `R` (R Development Team Core, 2006) provided all the estimates and tests in Table 3 using weighted data and accounting for weighting in the standard errors through linearization methods (Skinner, 1989). By directly modeling the dose-response curve in this way, we find that youths who received residential care would have had lower levels of SFI had they received outpatient care instead, though this effect is not significant. This analysis is equivalent to the standard path model for estimating the direct effect of treatment separate from its possibly suppressed indirect effects through institutionalization (MacKinnon et al., 2000) and under this modeling paradigm δ would provide the direct treatment effect.

***** Table 3 about here *****

The estimated treatment effects are unbiased estimates of the unsuppressed treatment effect only when the statistical model assumptions are correct. In particular, the following assumptions must hold:

Assumption 4. In terms of their potential outcomes, the youths with any observed value of institutionalization must be a random sample of all youths.

That is, there is no selection and the distribution of the potential outcomes does not depend on the observed values of institutionalization:

$$f(y_t[z]|Z_{obs} = z) = f(y_t[z])$$

for all values of z and t. This is a direct extension of Assumption 2 to include all values of z not just zero. The standard assumptions that are made with linear models like Equation 1, such as a constant treatment effect, a constant slope for institutionalization and an error term (ϵ) independent of institutionalization and treatment, ensure no selection. In this model, selection occurs if institutionalization is not independent of the error term as is demonstrated in Figures 6 and 7.

Assumption 5. The functional form of the dose-response curve must be correctly specified by the statistical model.

Assumption 5 requires that the dose-response curve for potential outcomes be specified correctly. Under Assumption 4, the observed dose-response curve equals the dose-response curve for the potential outcomes; the functional form can be studied using exploratory data analysis and the sensitivity of results to the model can be assessed by fitting alternative models. For example, we relaxed the assumption that the slopes of the dose-response curves were the same for the residential and outpatient groups by fitting a model with an interaction between modality and institutionalization. The results for this model were very similar to those of Model 1 and are not presented.

Model 2 tests the sensitivity of the results to Assumption 4 by refitting Model 1 using only youths with nonzero institutionalization, since youths with zero institutionalization appear to be clearly different from the rest of the sample. Using this model, we predicted the expected SFI at zero institutionalization for cases with any institutionalization and find that, for these youths, residential care appears to result in greater substance use: the treatment effect estimate is 3.48 and is significant at the 0.05 level. For youths with $Z_{obs} = 0$, residential care appears to be beneficial as we saw in Table 2, though the effect is not significant. As shown in Table 3, Model 3, which combines effect from the two subsamples, again has a positive sign, indicating lower values of SFI following outpatient care, but the estimate is small and not significant.

It is still possible that selection effects are contributing to these results. For instance, given that youth who enter residential treatment are more likely to be institutionalized, and institutionalized cases differ from other youths, we might expect the institutionalized cases in the residential and outpatient samples to differ.

Further exploration of possible selection effects could be conducted through sensitivity analysis or through the use of parametric selection models. Selection models involve creating models for both outcomes (SFI) and institutionalization, allowing institutionalization to depend on unobservable quantities that also affect outcomes. Selection models usually require an instrumental variable to uniquely determine or identify values of all the parameters of interest (Greene, 2003). An instrumental variable in this context is a variable that affects institutionalization, but does not affect the outcome except through institutionalization. For example, sentencing guidelines might influence institutionalization but might not influence the amount of drugs youths use when free in the community. Guidelines that differed between treatment and control groups could be used as an instrumental variable. Given the stringent requirements for instrumental variables, they are often not found in observational data unless special care is taken in the study design to capitalize on circumstances that create them, by selecting jurisdictions with different sentencing guidelines, for example. For additional details on instrumental variables see Greene (2003).

Nonlinearities in the model can also be used to identify parameters in some situations. We explored parametric selection models for this data using the nonlinearity of the model for institutionalization for identification. The results were unstable, but again found SFI was greater for the residential cases than the outpatient cases, and found a steeper slope in the relationship between SFI institutionalization than we estimated in Model 1, suggesting that selection bias flattens out the relationship between SFI and institutionalization, as we might expect if the cases at greatest risk for substance use are also at greatest risk for the most time institutionalized. However, an instrumental variable would be necessary to provide truly convincing evidence from a selection model.

4 Discussion

Institutionalization at follow-up is particularly common in substance abuse treatment studies, and can present challenging confounds for the calculation of unbiased treatment effects in either experimental or observational studies. Although some amount of institutionalization exists in nearly all long-term follow-up studies, the most common approach to its confounding effects on outcomes has been to ignore them. While the resulting treatment effect estimate does provide a valid causal effect (what we call the total effect), it is difficult to interpret because it conflates the effects of treatment on the outcome of interest (drug use or crime, for instance) with effects on institutionalization, which itself has an effect on such outcomes. Thus, programs with seemingly positive treatment effects may be achieving these either by reducing crime and drug use, or by increasing incarceration, possibilities with quite different implications for clients and other stakeholders.

The paper presented several alternatives to ignoring institutionalization in treatment effect estimation, and extended the Neyman-Rubin model of causal effects to formalize the analysis of these approaches and the assumptions on which they rest. Each alternative includes some type of conditioning on institutionalization in an effort to estimate the unsuppressed treatment effect, that is, the treatment effect that would have been observed had no cases been institutionalized. Conditioning on institutionalization, like conditioning on any post-intervention event (such as death or study attrition), is problematic because institutionalized clients can differ systematically from those who are not institutionalized, and these differences are associated with their expected outcomes. Therefore, conditioning on institutionalization can result in biased estimates. Moreover, because the differences between clients with differing amounts of institutionalization can result from treatment, adjusting for pretreatment variables might not remove bias.

When institutionalization does not occur at random in the population and affects the outcome of interest, bounding methods and parametric selection models are necessary for estimating treatment effects that are not confounded with institutionalization or biased by attempts to correct for it. Nevertheless, as illustrated with our case study, even when the excluded sample is small (for instance when restricted to cases with $Z_{obs} < 90$), bounds can be so wide as to become useless. On the other hand, selection models require additional untestable assumptions and data, such as instrumental variables, that are often not available. Sensitivity analyses can be conducted to explore selection when directly modeling it is not feasible or requires unlikely assumptions.

Our case study illustrated that the common methods for estimating treatment effects in the presence of institutionalization during the evaluation are not robust to violations of the assumptions presented in Section 3. Depending on how we account for institutionalization (including ignoring it), we might find either significant positive or significant negative effects for residential care. This sensitivity to institutionalization results from the strong dose-response relationship between institutionalization and drug use outcomes, and from strong selection into institutionalization observed in this data set. Given high rates of institutionalization that differed across treatment modality and a strong dose-response curve, the total effect estimated by ignoring institutionalization is of little use for evaluating the relative effectiveness of residential and outpatient treatment. The assumptions to support other simple analysis methods (e.g., excluding institutionalized cases) were unlikely to hold, since we found evidence suggesting that youths were selected for institutionalization on characteristics associated with unobserved drug use propensities.

Although all the factors that existed in the ATM sample might not exist in every study, it is reasonable to believe that measures of substance use, criminal activity and related behaviors will all show strong dose-response relationships with institutionalization. Furthermore, institutionalization rates are often high for studies of substance use treatment. Thus, institutionalization is likely to confound treatment effect estimates in studies on populations like those in ATM. Such confounds might exist for both observational studies and randomized experiments (e.g., Deschenes et al., 1995). Moreover, the confound considered in this report may be equally problematic in subgroup comparisons. For instance, if in a study of racial differences in drug use trajectories, one group was found to be at higher risk of incarceration, the institutionalization confound could bias outcome estimates.

The framework established in this paper provides a means for analysts to consider the conclusions they draw from estimated treatment effects. It identifies the assumptions that need to be verified and methods for studying them through auxiliary and sensitivity analyses. It also demonstrates the need for additional data that distinguishes the effects of treatment on outcomes from the effects of suppression by institutionalization. For example, data on sentencing guidelines might provide instruments for selection models and repeated follow-up measurements might allow for more elaborate statistical models that could control or greatly reduce the effects of selection. This potential outcomes framework also serves as the groundwork for new statistical methods such as principal stratification (Frangakis and Rubin, 2002), which will potentially provide alternative methods for adjusting for institutionalization at follow-up and give analysts additional tools for studying treatment effects in the future.

References

Dennis, M.L., 1999. Global Appraisal of Individual Needs (GAIN) Manual: Administration, Scoring and Interpretation. Lighthouse Publications, Bloomington, IL.

Dennis, M.L., Babor, T.F., Diamond, G., Donaldson, J., Godley, S.H., Titus, J.C., et al., 2000. The cannabis youth treatment (CYT) experiment: Preliminary findings. Center for Substance Abuse Treatment, Substance Abuse and Mental Health Services Administration, Department of Health and Human Services, Rockville, MD.

Deschenes, E., Turner, S., Petersilia, J., 1995. A dual experiment in intensive community supervision: Minnesota's prison diversion and enhanced supervied release programs. The Prison Journal. 75, 330–356.

Efron, B., Tibshirani, R.J. 1993. An Introduction to the Bootstrap. Chapman & Hall, New York.

Frangakis, C.E, and Rubin, D.B 2002. Principal stratification in causal inference. Biometrics. 58, 21–29.

Gerstein, D.R., Johnson, R.A. 1999. Adolescents and young adults in the national treatment improvement evaluation study. National Evaluation Data Services, Rockville, MD.

Gossop, M., Marsden, J., Stewart, D., Kidd, T., 2003. The national treatment outcome research study (NTORS): 4-5 year follow-up results. Addiction. 98, 291–303.

Greene, W.H. 2003. Econometric Analysis. Prentice Hall, Upper Saddle River, NJ.

Holland, P.W., 1986. Statistics and causal inference. J Am Stat Assoc. 81, 945–960.

Hser, Y. I., Grella, C.E., Hubbard, R.L., Hsieh, S.C., Fletcher, B.W., Brown, B.S.,2001. An evaluation of drug treatments for adolescents in 4 us cities. Arch Gen Psychiatry. 58, 689–695.

Hubbard, R.L., Craddock, S.G., Flynn, P.M., Anderson, J., Etheridge, R., 1997. Overview of 1-year follow-up outcomes in the drug abuse treatment outcome study $DATOS$. Psychol Addict Behav. 11, 261–278.

Imbens, G. (2003). Nonparametric estimation of average treatment effects under exogeneity: A review. Technical Working Paper 294, National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/T0294 December 23, 2005.

MacKinnon, D.P., Krull, J.L., Lockwood, C.M., 2000. Equivalence of the mediation, confounding and suppression effect. Prev Sci. 1, 173–181.

McCaffrey, D.F., Ridgeway, G., Morral, A.R., 2004. Propensity Score Estimation with Boosted Regression for Evaluating Causal Effects in Observational Studies. Psychol Methods. 9, 403–425.

Mee-Lee, D., Shulman, G., Fishman M., Gastfriend, D., Griffith, J. (Eds.), 2001. ASAM Patient Placement Criteria for the Treatment of Substance-Related Disorders (2nd- Revised (ASAM PPC-2R) ed.). American Society of Addiction Medicine, Inc, Chevy Chase, MD.

Morral, A., McCaffrey, D., Ridgeway, G., 2004. Effectiveness of community-based treatment for substance-abusing adolescents: 12-month outcomes of youths entering phoenix academy or alternative probation dispositions. Psychol Addict Behav. 18, 257–268.

NORC, RTI, 1997. National Treatment Improvement Evaluation Survey (NTIES): Final Report. Rockville, MD: Center for Substance Abuse Treatment Substance Abuse and Mental Health Services Administration.

Pearl, J., 1996. Causation, Action and Counterfactuals. In: Shoham, Y. (Ed.), Proceedings of the Sixth Conference of Theoretical Aspects of Rationality and Knowledge. Morgan Kauffman, San Francisco, CA, pp. 57–73.

Piquero, A., Blumstein, A., Brame, R., Haapanen, R., Mulvey, E., Nagin, D., 2001. Assessing the impact of exposure time and incapacitation on longitudinal trajectories of criminal offending. J Adolesc Res. 16, 54–74.

Project MATCH Research Group, 1997. Matching alcoholism treatments to client heterogeneity: Project MATCH posttreatment drinking outcomes. J Stud Alcohol 58, 7–29.

Rawson, R., Marinelli-Casey, A., Anglin, M., Dickow, A., Frazier, Y., C. Gallagher, Galloway, C., Herrell, J., Huber, A., McCann, M., Obert, J., Pennell, S., Reiber, C., Vandersloot, D., Zweben, J., Methamphetamine Treatment Project Corporate Authors (2004). A multi-site comparison of psychosocial approaches for the treatment of methamphetamine dependence. Addiction 99, 708–717.

R Development Team Core, 2006. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org. November 21, 2006.

Skinner, C.J., 1989. Domain means, regression and multivariate analysis In C.J. Skinner, D. Holt, and T.M.F. Smith (Eds.) Analysis of Complex Surveys. New York: John Wiley & Sons.

Stevens, S. Morral, A., (Eds.), 2003. Adolescent Substance Abuse Treatment in the United States: Exemplary Models from a National Evaluation Study. New York: Haworth Press.

U.S. Dept. of Health and Human Services, National Institute on Drug Abuse, 2004. Drug Abuse Treatment Outcome Study (DATOS), 1991-1994: [UNITED STATES] [Computer file] (2nd ICPSR ed.). Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and distributor].

U.S. Dept. of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Substance Abuse Treatment, 2004. National Treatment Improvement Evaluation Study (NTIES), 1992-1997 [Computer file] (3rd ICPSR ed.). Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and distributor].

Webb, C.P.M., Burleson, J.A., Ungemack, J.A., 2002. Treating juvenile offenders for marijuana problems. Addiction 97, 35–45.

Zhang, J., Rubin, D., 2003. Estimation of causal effects via principal stratification when some outcomes are truncated by “death”. Journal of Educational and Behavioral Statistics 28, 353–368.

Zhang, Z., Friedmann, P.D., Gerstein, D. R., 2003. Does retention matter? Treatment duration and improvement in drug use. Addiction 98, 673–684.

Measure	Residential	Outpatient
Mean SFI (Past 90 days)	8.99	9.75
Median SFI (Past 90 days)	1.29	3.78
Mean Days Institutionalized	26.38	14.35
Percent with Zero Days Inst.	48.31	61.52
Percent with 1 to 89 Days Inst.	35.20	32.25
Percent with 90 Days Inst.	16.49	7.23
Mean Days Residential Treatment	10.04	2.79
Percent with Zero Days Res. Tx	83.51	91.82
Percent with 1 to 89 Res. Tx	9.35	6.38
Percent with 90 Days Res. Tx.	7.14	1.80
Mean Days Other Inst.	16.39	11.19
Percent with Zero Days Other Inst.	61.17	66.23
Percent with 1 to 89 Other Inst.	30.00	30.44
Percent with 90 Days Other Inst.	8.83	3.33

Table 1: Descriptive Statistics for the 90 days prior to 12 Month Follow-up for SFI and Institutionalization by Treatment Modality, with Outpatient Case Weighting

				Treatment	
Method	Statistic	Residential	Outpatient	Effect	p-value
SFI, Complete Sample	Mean	8.99	9.75	-0.76	0.54
	Median	1.29	3.78	-2.50	0.02
SFI, Sample with Inst. = 0	Mean	9.80	11.29	-1.49	0.35
	Median	2.36	7.44	-5.08	0.01
SFI Per Day Free in Community,	Mean	13.34	11.49	1.85	0.29
Sample with Inst. <90	Median	3.71	5.39	-1.68	0.37

Table 2: Group Means and Medians and Treatment Effect Estimates for the Substance Use Frequency Index, SFI

	Estimated Mean		Treatment
	Residential	Outpatient	Effect
Model 1	12.14	11.46	0.68
Model 2	19.65	14.17	3.48*
Model 3	13.86	12.78	1.08

Table 3: Model Based Estimates of the Treatment Effect for SFI. Model 1 is a linear regression model with Z_{obs} (observed institutionalization) as a covariate and an indicator for treatment modality; Model 2 is a linear regression model with Z_{obs} as a covariate and an indicator for treatment modality fit only to cases with $Z_{obs} > 0$; Model 3 combines Model 2 with a model for cases with $Z_{obs} = 0$ to estimate the average treatment effect on the entire treated sample. * statistically significant at $p < 0.05$.

Figure Legends

Figure 1: A diagram of treatment effects and institutionalization from the direct and indirect effects perspective. Treatment affects outcomes directly and indirectly through its effect on institutionalization.

Figure 2: Example of suppression of outcomes by institutionalization. A hypothetical youth's potential outcomes decline with institutionalization for both treatment and control. The difference between the two lines at $z = 0$ is the unsuppressed treatment effect.

Figure 3: Example of bias in treatment effect estimate that ignores institutionalization. Potential outcomes for treatment and control are plotted for each of four cases. Potential outcomes for control have high values at $z = 0$ and steep slopes; potential outcomes for treatment are at the bottom of the figure with low values at $z = 0$ and flat slopes. Dashed lines and o denote youths assigned to control. Solid lines and + denote youths assigned to treatment. Institutionalization suppresses outcomes and treatment youths are institutionalized for a longer time than control youths are. Thus, the difference between mean outcomes for control and treatment youths overestimates the unsuppressed treatment effect.

Figure 4: Example of bias due to both suppression and selection. There are no treatment effects, but there are heavy and light users and outcomes decline with institutionalization. Treatment youths (+) are institutionalized longer then control youths (o) for both heavy and light users and a comparison of means does not equal the unsuppressed effect. Note that even among youths with the same amount of institutionalization (40 days), difference in means would not equal treatment effects because the youths are from different groups of users.

Figure 5: Smooth of Mean SFI versus Institutionalization for Institutionalization of One or More Days. Solid dot is Mean for SFI when Institutionalization is Zero Days.

Figure 6: Diagram of potential post-treatment confounding that would create selection bias and bias indirect and direct treatment effect estimates from a linear model. Treatment affects outcomes, institutionalization, and an unobserved and unmeasured confounder which will bias linear model estimates of the direct link from treatment to outcomes.

Figure 7: Example of the failure of linear regression to capture treatment effects in the presence of selection bias. There are no treatment effects and the solid lines represent the outcome of eight hypothetical youths under both treatment and control. There are four types of youths identified by their risk or their

potential outcome at no institutionalization. Youths with greater risk have greater institutionalization and treatment youths (+) are institutionalized for more days than control youths (o). The dashed regression line is biased and does not capture the true dose-response relationship between institutionalization and the outcome. Using linear regression to adjust for institutionalization results in biased estimates because treatment cases are all below the attenuated regression line. Even separate regression lines for the two groups (dotted lines) would not remove the bias.

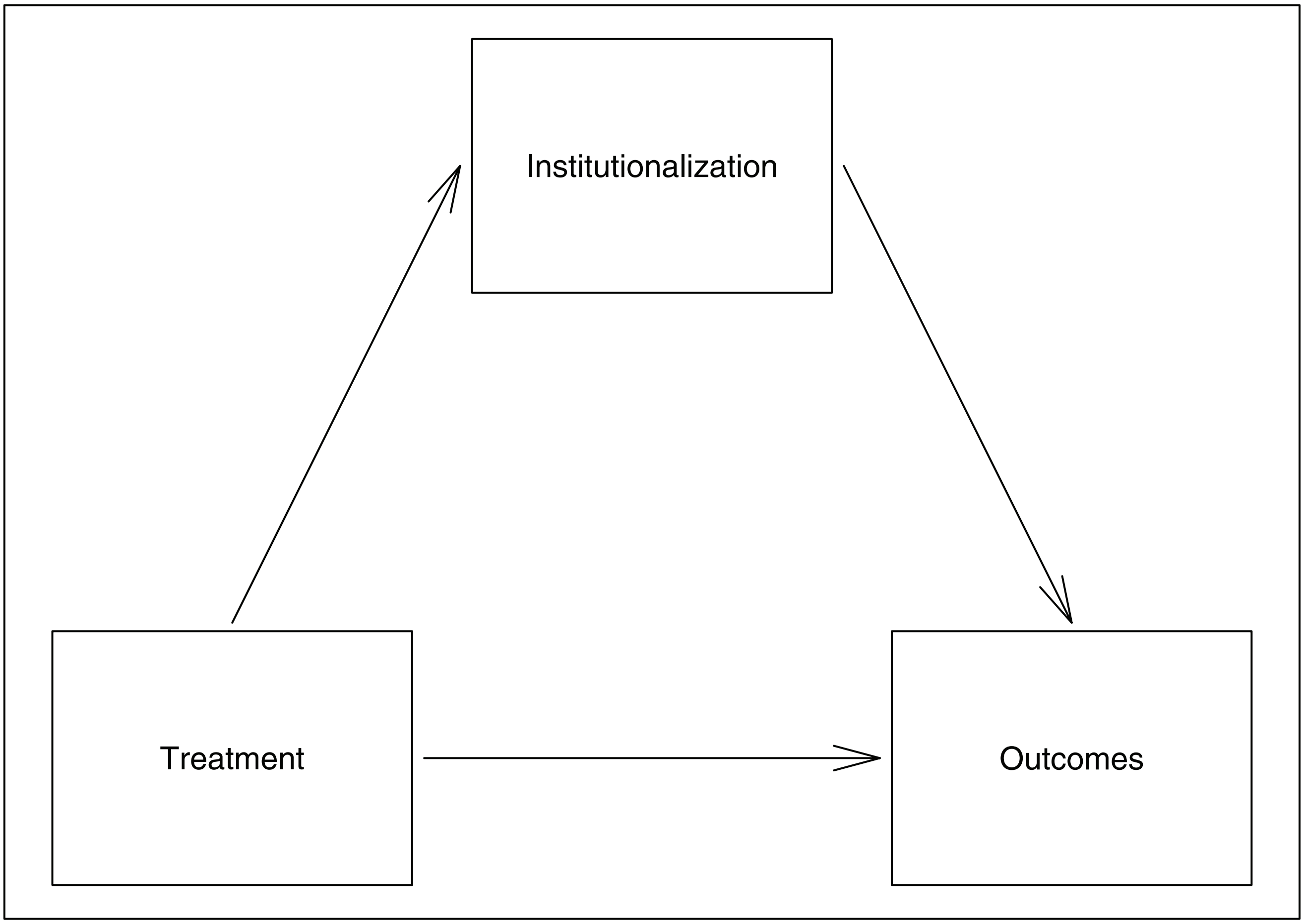

Figure 1:

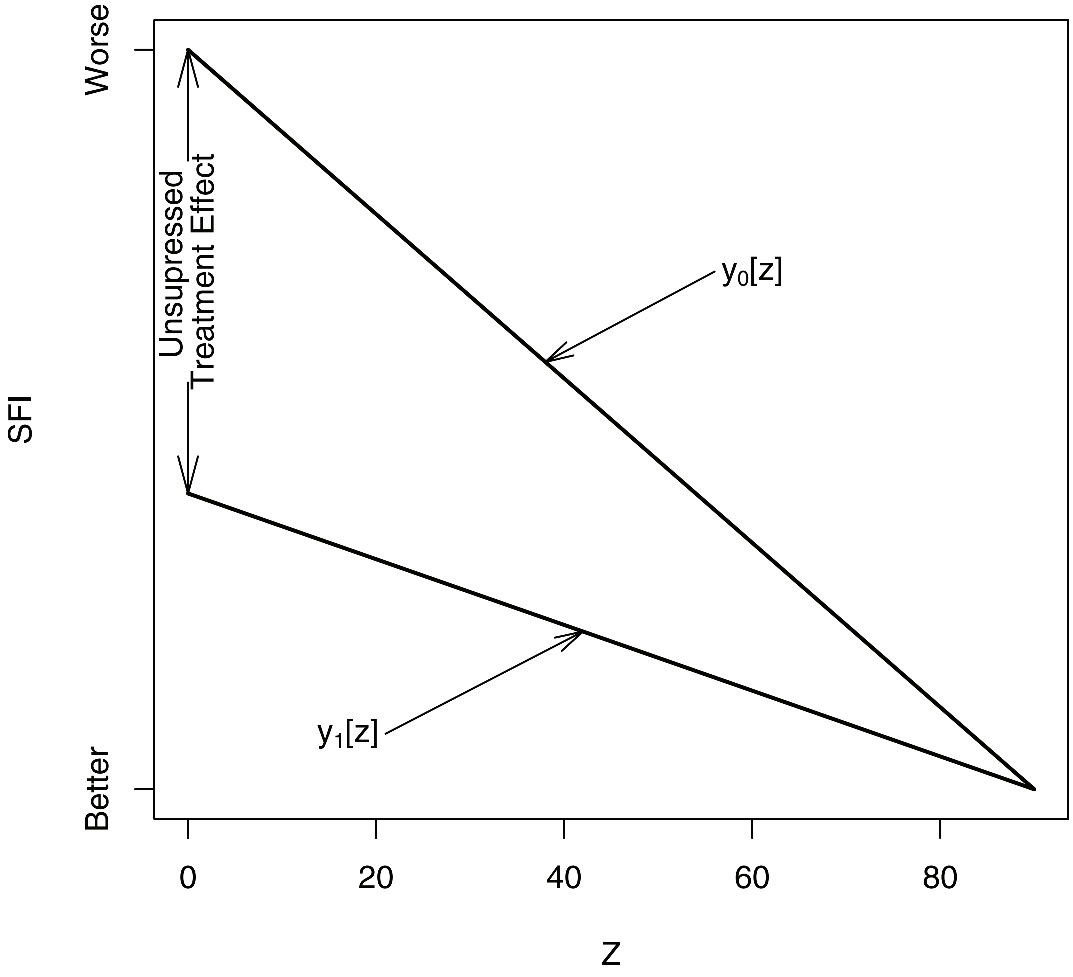
SFI
Worse
Better
Unsupressed
Treatment Effect
$y_0[z]$
$y_1[z]$
0
20
40
60
80
Z

Figure 2:

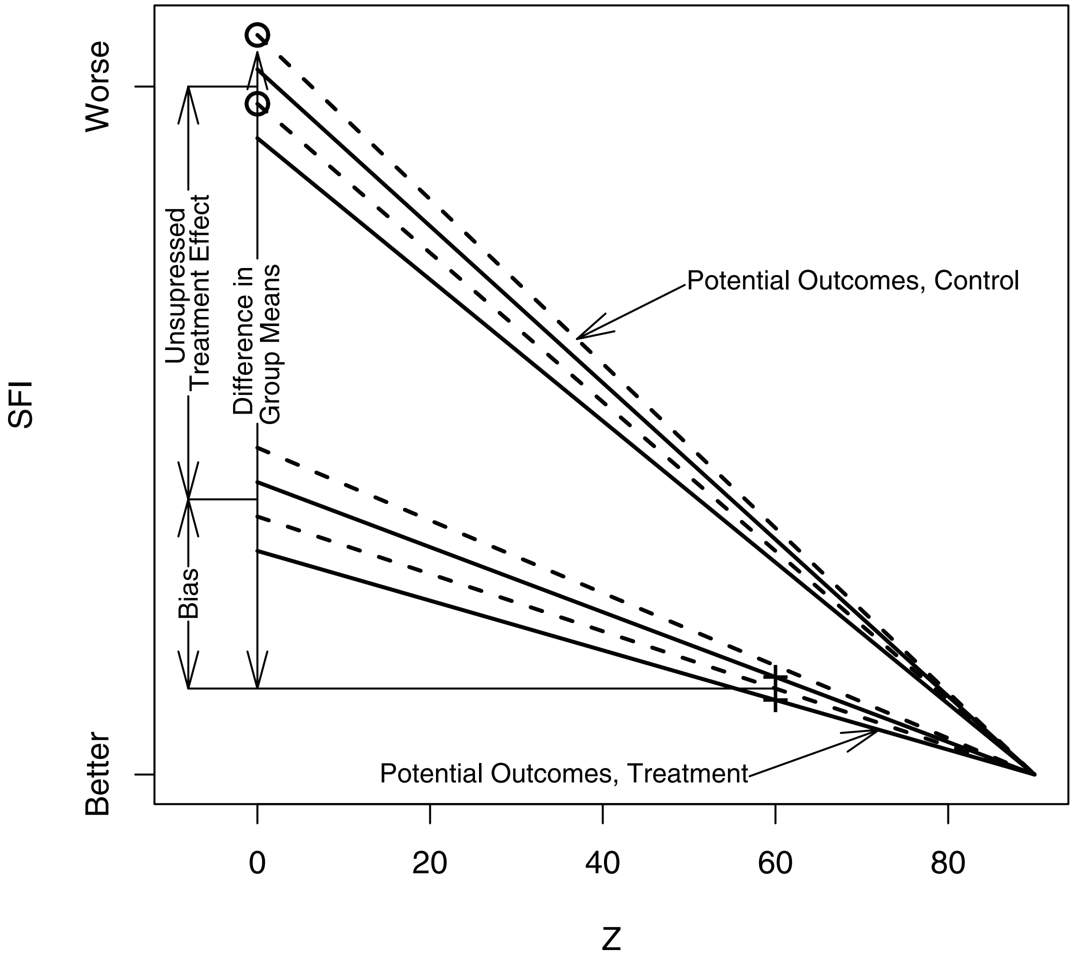
Worse
Better
SFI
Z
0
20
40
60
80
Unsupressed
Treatment Effect
Difference in
Group Means
Bias
Potential Outcomes, Control
Potential Outcomes, Treatment

Figure 3:

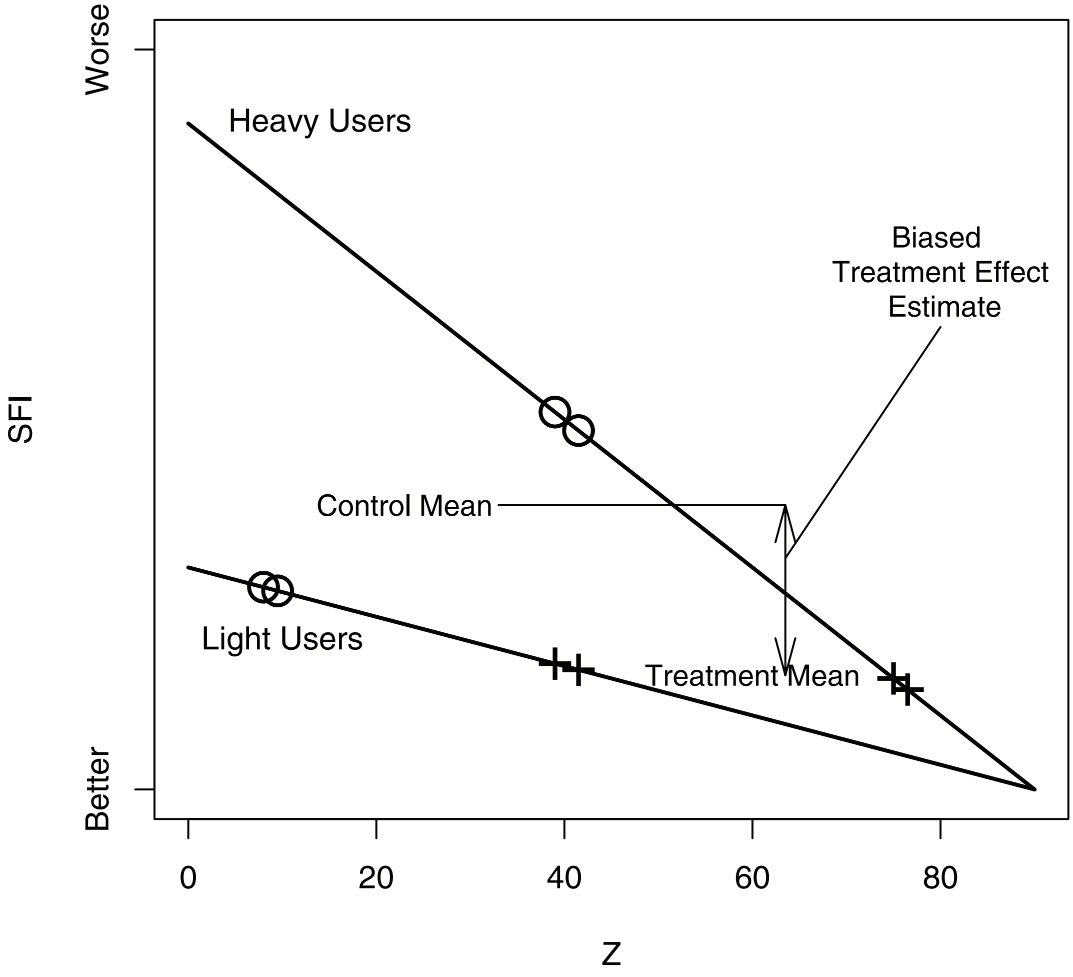
Worse
Better
SFI
Heavy Users
Light Users
Biased
Treatment Effect
Estimate
Control Mean
Treatment Mean
0
20
40
60
80
Z

Figure 4:

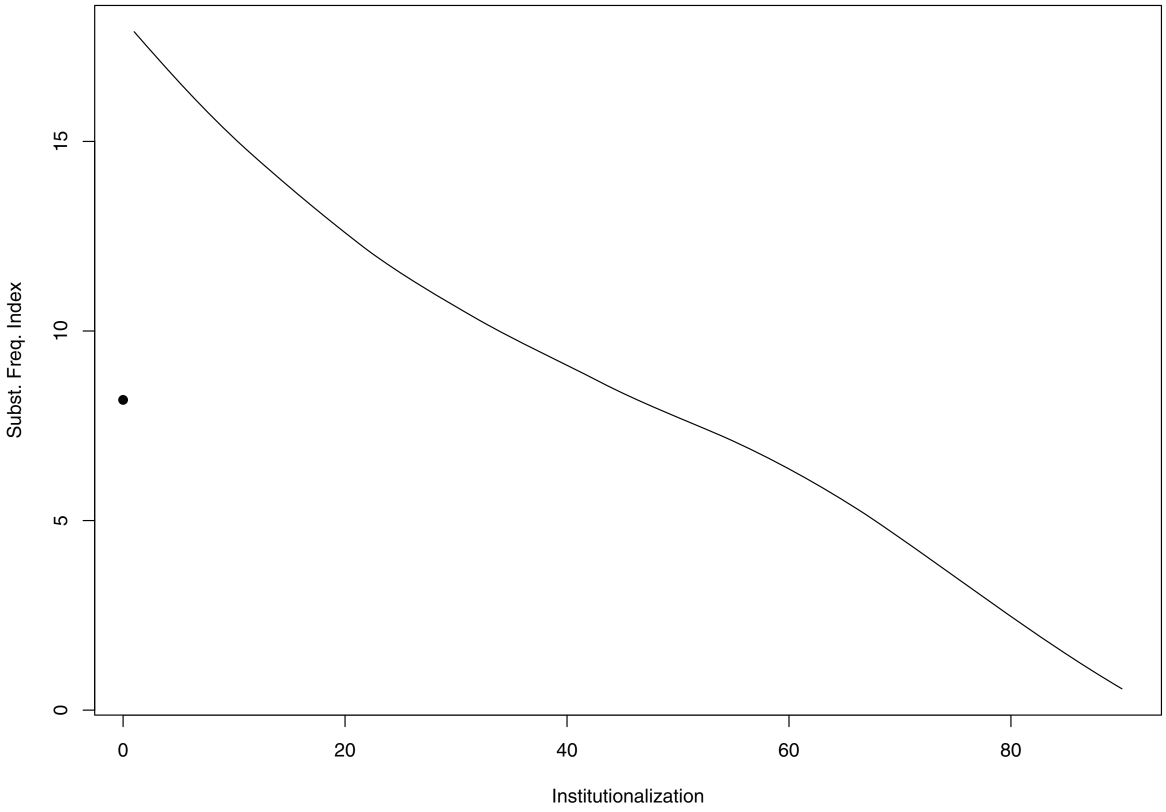

Figure 5:

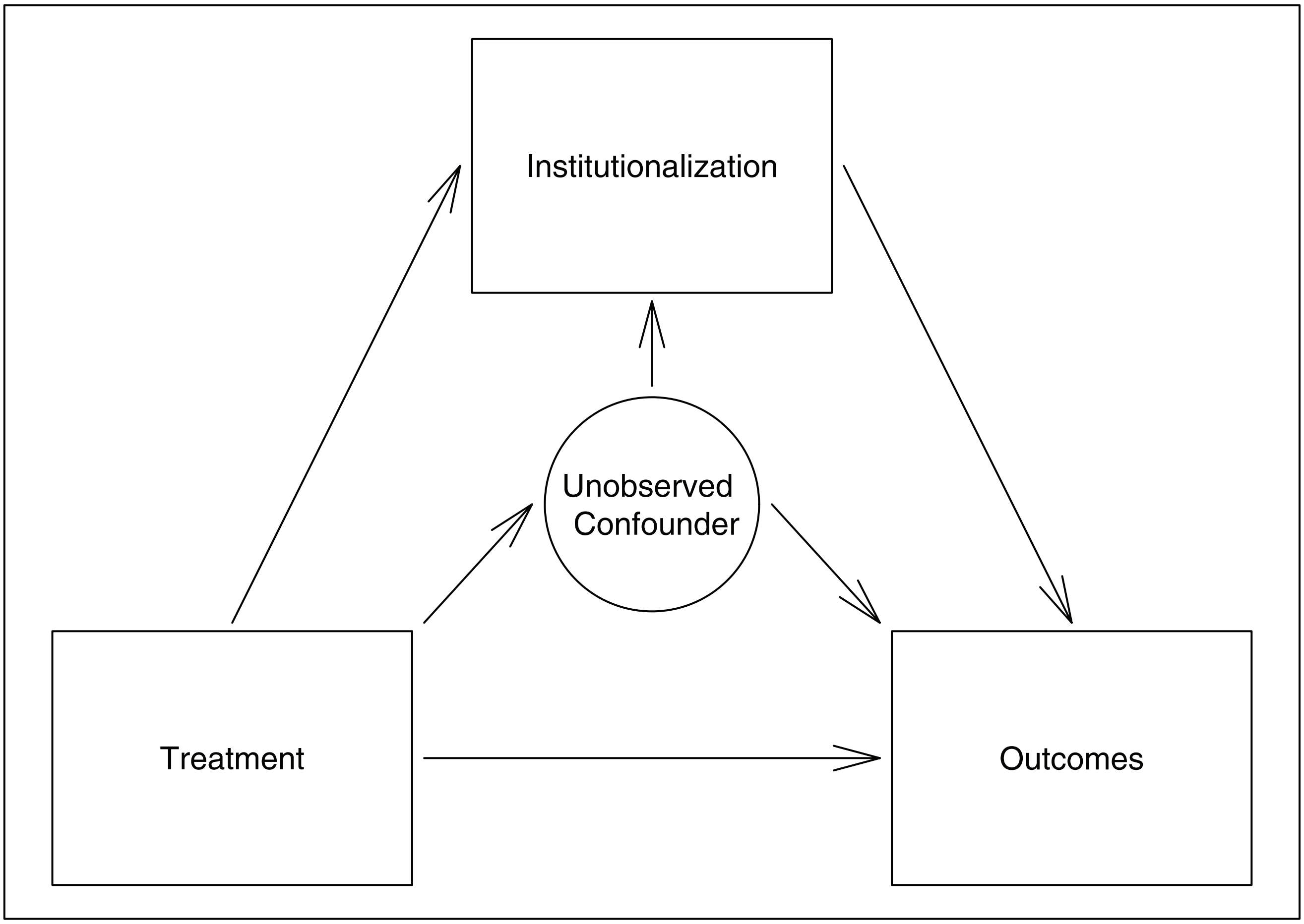
Institutionalization
Unobserved Confounder
Treatment
Outcomes

Figure 6:

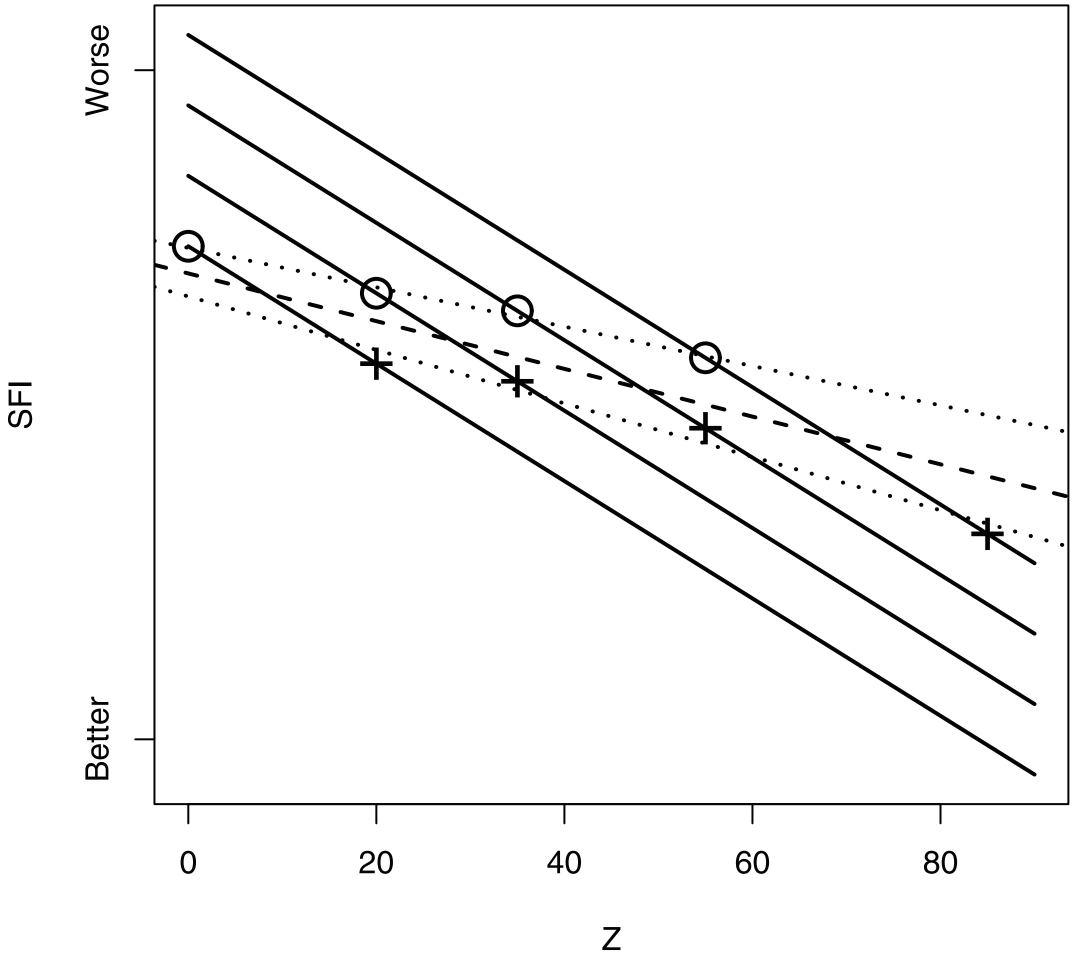

Figure 7:

5 Appendix

	Mean		Std. Dev.		Effect		
Name	Res.	OP	Res.	OP	Size	p	KS
Age	16.09	15.82	1.30	1.09	0.20	0.02	0.08
Gender						0.95	
Male	0.71	0.72	0.45	0.45	-0.01		0.00
Female	0.29	0.28	0.45	0.45	0.01		0.00
Race						0.13	
Am. Indian	0.16	0.07	0.36	0.26	0.23		0.08
Black	0.09	0.09	0.29	0.29	-0.02		0.00
White	0.40	0.49	0.49	0.50	-0.18		0.09
Mexican	0.18	0.16	0.38	0.37	0.03		0.01
Other Race	0.18	0.18	0.38	0.38	-0.01		0.00
Current living status*						0.45	
House/Apt	0.74	0.77	0.44	0.42	-0.07		0.03
Friend/Relative	0.08	0.11	0.28	0.31	-0.09		0.02
Jail/Correctional	0.09	0.07	0.28	0.25	0.06		0.02
Other status	0.09	0.05	0.29	0.22	0.14		0.04
No. people lived with^{a}	4.76	4.05	6.17	3.50	0.12	0.07	0.10
Times received tx for drug/alc	1.98	1.41	7.91	2.51	0.07	0.11	0.04
Prior tx psychological problems	0.43	0.43	0.49	0.49	0.00	1.00	0.00
Lifetime detox admissions	0.32	0.34	1.38	1.56	-0.01	0.94	0.03
Currently in Alc/Drug tx	0.19	0.18	0.39	0.38	0.02	0.81	0.01
Attended AA or other self-help	0.52	0.47	0.50	0.50	0.09	0.37	0.04
Physical Health Tx Index	0.02	0.03	0.03	0.11	-0.43	0.35	0.04
Mental Health Tx Index*	0.02	0.02	0.05	0.07	-0.01	0.90	0.02
Substance Abuse Tx Index	0.06	0.07	0.15	0.16	-0.02	0.84	0.03
Treatment Resistance Index	1.54	1.54	1.09	1.08	0.00	0.99	0.02
Treatment Motivation Index*	2.89	2.92	1.11	1.18	-0.02	0.84	0.02
High resistance	0.20	0.20	0.40	0.40	-0.02	0.86	0.01
Low motivation	0.32	0.33	0.47	0.47	0.00	0.99	0.00
Need for heroin tx (self-reported)*	0.07	0.05	0.25	0.22	0.07	0.61	0.02

Continued on next page

	Mean		Std. Dev.		Effect		
Name	Res.	OP	Res.	OP	Size	p	KS
Mention of Alc as tx need	0.29	0.24	0.45	0.42	0.12	0.19	0.05
Mention of Marijuana as tx need	0.54	0.56	0.50	0.50	-0.04	0.66	0.02
Mention of Cocaine as tx need	0.12	0.10	0.32	0.30	0.05	0.58	0.02
Mention of Opiates as tx need	0.07	0.06	0.26	0.24	0.04	0.79	0.01
Mention of Amph's as tx need	0.11	0.07	0.31	0.26	0.12	0.16	0.04
Mention of Other Drugs as tx need	0.07	0.07	0.26	0.26	-0.01	0.93	0.00
Substance Frequency Index*	0.27	0.25	0.19	0.18	0.11	0.24	0.06
Substance Problem Index^a	11.22	11.09	3.68	3.75	0.03	0.69	0.03
Substance Problem Index^b	5.70	5.50	5.03	4.85	0.04	0.67	0.05
Substance Dependence Index^a	4.74	4.66	2.01	2.03	0.04	0.67	0.03
Substance Dependence Index^b	2.29	2.20	2.40	2.31	0.04	0.67	0.04
Low alc/drug use problem orientation*	0.42	0.44	0.49	0.50	-0.05	0.57	0.03
First high/intoxicated under age 15	0.92	0.92	0.28	0.27	-0.01	0.91	0.00
Prior opiate use	0.36	0.39	0.48	0.49	-0.06	0.55	0.03
Using daily	0.80	0.80	0.40	0.40	0.01	0.90	0.00
Using opioids^d	0.28	0.29	0.45	0.45	0.00	0.97	0.00
Used in past two days	0.14	0.18	0.34	0.39	-0.14	0.13	0.05
Drunk/high most of day^d	33.84	32.22	30.45	28.83	0.05	0.56	0.06
Tobacco Dependence Index^{b,e,*}	50.38	53.23	36.71	35.82	-0.08	0.40	0.06
Substance use despite prior tx	0.24	0.24	0.43	0.43	-0.01	0.93	0.00
Substance most likes						1.00	
Alcohol	0.14	0.14	0.35	0.35	0.01		0.00
Cannabis	0.63	0.62	0.48	0.48	0.00		0.00
Crack or cocaine	0.06	0.07	0.24	0.25	-0.02		0.01
Opiates	0.06	0.06	0.23	0.23	0.00		0.00
Hallucinogens	0.04	0.03	0.18	0.18	0.02		0.00
Amphetamines	0.07	0.08	0.26	0.26	-0.02		0.00
Other Substance	0.01	0.01	0.09	0.09	0.00		0.00
Needle Frequency Index	0.04	0.04	0.13	0.13	0.02	0.87	0.06
Needle Problem Index	0.47	0.48	1.78	1.80	0.00	0.99	0.01
Needle Use^c	0.14	0.08	0.34	0.27	0.16	0.16	0.05

Continued on next page

	Mean		Std. Dev.		Effect		
Name	Res.	OP	Res.	OP	Size	p	KS
Current Withdrawal Index	3.43	3.50	5.06	5.24	-0.01	0.90	0.03
Seizures/current withdrawal symptoms	0.16	0.18	0.37	0.38	-0.05	0.62	0.02
Living Risk Index	7.93	7.77	3.51	3.74	0.04	0.68	0.05
Homeless/Runaway[a]	0.24	0.22	0.43	0.42	0.04	0.69	0.02
Environmental Risk Index	32.64	32.05	10.17	10.13	0.06	0.53	0.05
Recovery Environment Risk Index	0.33	0.34	0.11	0.11	-0.04	0.67	0.07
General Social Support Index	6.80	6.99	1.96	1.95	-0.10	0.30	0.06
Social Risk Index	13.66	13.62	4.81	5.22	0.01	0.94	0.06
Weekly family problems[c]*	0.48	0.40	0.50	0.49	0.17	0.08	0.08
Family Hist. of Substance Use	0.87	0.86	0.34	0.35	0.03	0.75	0.01
Days of victimization[c]	0.34	0.32	0.47	0.47	0.04	0.68	0.02
High Victimization	0.70	0.70	0.46	0.46	-0.01	0.88	0.01
General Victimization Index	5.08	4.82	3.06	2.86	0.08	0.31	0.05
CJ System Index	0.55	0.63	0.47	0.46	-0.17	0.07	0.09
Crime Violence Index	10.40	10.70	6.54	6.00	-0.05	0.63	0.06
Illegal Activities Index*	0.10	0.10	0.05	0.05	-0.05	0.57	0.04
Drug Crime Index	1.04	1.03	0.89	0.85	0.01	0.92	0.02
Current CJ Involvement	0.80	0.80	0.40	0.40	0.00	0.99	0.00
Total arrests[c]	0.79	0.77	1.47	1.08	0.01	0.87	0.03
Controlled Environment Index*	0.30	0.30	0.38	0.36	0.02	0.85	0.04
Institutionalization[c]*	19.18	19.51	25.84	26.63	-0.01	0.90	0.04
Employed	0.35	0.35	0.48	0.48	0.00	0.96	0.00
Employment Problem Index*	1.23	1.19	1.79	1.63	0.03	0.78	0.03
Employment Activity Index	0.24	0.23	0.33	0.33	0.02	0.80	0.04
Training Problem Index*	4.62	4.86	2.36	2.23	-0.10	0.26	0.05
Training Activity Index*	0.48	0.49	0.32	0.32	-0.04	0.69	0.05
Vocational Risk Index	11.04	10.76	6.12	6.08	0.05	0.66	0.04
In school	0.74	0.77	0.44	0.42	-0.06	0.57	0.02
Last grade completed in school	9.10	9.02	1.33	1.34	0.06	0.55	0.03
General Satisfaction Index	13.88	13.76	4.45	4.79	0.03	0.77	0.04
Personal Sources of Stress Index	1.82	1.61	1.45	1.36	0.15	0.09	0.05

Continued on next page

	Mean		Std. Dev.		Effect		
Name	Res.	OP	Res.	OP	Size	p	KS
Other Sources of Stress Index	1.65	1.50	1.69	1.52	0.09	0.28	0.04
Mental distress	0.42	0.42	0.49	0.49	0.00	0.99	0.00
Homicidal-Suicidal Thought Index	0.25	0.23	0.43	0.42	0.03	0.76	0.01
Self Efficacy Index	3.56	3.59	1.52	1.36	-0.02	0.80	0.07
Problem Orientation Index*	2.20	2.05	2.03	1.98	0.07	0.43	0.05
Internal Mental Distress Index	13.04	12.59	8.76	8.88	0.05	0.57	0.04
Emotional Problem Index	0.30	0.33	0.21	0.20	-0.12	0.20	0.09
Behavior Complexity Index*	15.32	15.35	7.64	7.61	0.00	0.96	0.03
High behavioral problems^a	0.38	0.39	0.49	0.49	-0.02	0.84	0.01
Sex Protection Ratio	0.76	0.77	0.37	0.38	-0.02	0.87	0.05
Health Problem Index	0.15	0.14	0.16	0.15	0.02	0.79	0.03
Health Distress Index^a	1.94	1.88	1.11	1.00	0.06	0.47	0.05
Acute Health Problems*	0.37	0.37	0.48	0.48	0.00	0.96	0.00

Table 4: Summary of Covariates Distribution After Weighting for Pretreatment Differences. Effect sizes the ratio of the absolute difference in group means to the residential standard deviation. ^{a}Past year, ^{b}Past Month, ^{c}Past 90 Days, ^{d}Weekly, ^{e}Lifetime,*Used in selection model