

A Comparison of Methods for Estimating the Causal Effect of a Treatment in Randomized

Clinical Trials Subject to Noncompliance

Roderick Little¹, Qi Long² and Xihong Lin³

Abstract

We consider the analysis of clinical trials that involve randomization to an active treatment ($T = 1$) or a control treatment ($T = 0$), when the active treatment is subject to all-or-nothing compliance. We compare three approaches to estimating treatment efficacy in this situation: as-treated analysis, per-protocol analysis, and instrumental variable (IV) estimation, where the treatment effect is estimated using the randomization indicator as an instrumental variable. Both model-based and method-of-moment based IV estimators are considered. The assumptions underlying these estimators are assessed, standard errors and mean squared errors of the estimates are compared, and design implications of the three methods are examined. Extensions of the methods to include observed covariates are then discussed, emphasizing the contrasting role of covariates in these extensions. Methods are illustrated on data from the Women Take Pride study, an assessment of behavioral treatments for women with heart disease.

Keywords: as-treated analysis, per-protocol analysis, causal inference, instrumental variables, principal stratification, propensity scores.

¹ Department of Biostatistics, University of Michigan, Ann Arbor MI 48109. Email: rlittle@umich.edu

² Department of Biostatistics, Emory University, Atlanta, GA 30322. Email: qlong@sph.emory.edu

³ Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115. Email: xlin@hsph.harvard.edu

1. Introduction

Randomized clinical trials that compare treatments are straightforward to analyze when all individuals in the study take the assigned treatment, and outcomes are reported without missing data. The analysis and interpretation is complicated when individuals do not comply with their assigned treatments. The gold-standard analysis of such trials in drug approval processes is intention-to-treat (IT), where compliance information is ignored, and individuals are classified in treatment comparisons according to their assigned treatments, regardless of whether the treatment was taken. IT analysis preserves the benefits of randomized allocation of treatments, and it provides valid measures of the effect of treatment assignment, which is sometimes called treatment *effectiveness*. The analysis is less compelling for estimating the effect of treatment *efficacy*, which concerns the pharmacological or behavioral effectiveness of a treatment when it is in fact taken.

Simple approaches to estimating treatment efficacy are as treated (AT) analysis, which compares average treatment effects with participants classified according the treatment actually received, and per-protocol (PP) analysis, which compares the average treatment effects for participants who comply with the assigned treatment. These analyses both classify participants according to received treatment, and hence are direct measures of treatment efficacy, but they are both subject to bias, in that participants who comply with a particular treatment may be a biased sample of participants randomized to that treatment. The selection bias may be reduced by adjustment for covariates, but it remains a concern. Thus current clinical trial practice for estimating efficacy involves an unappealing choice between IT analysis, which is protected from bias by randomization but is really estimating effectiveness rather than efficacy, and PP and AT

analysis, which are directly attempting to measure efficacy but are subject to bias since the randomization is corrupted by treatment noncompliance.

Recent literature has advocated another approach to estimating efficacy, based on treating the randomization as an instrumental variable (IV), in economic parlance. In simple terms the IV estimator corrects the IT estimator for noncompliance, based on certain assumptions about the outcomes for non-compliers under both treatments. This approach has the attraction of yielding a direct estimate of treatment efficacy, and is also protected from selection bias by the randomization. On the other hand, it does require certain assumptions to be valid, and it also yields estimators with potentially high variance, particularly if the treatment compliance rate is low. Model-based versions of the IV estimator have been proposed that are potentially more efficient, although they make stronger distributional assumptions.

The first objective of this paper is to provide a side-by-side comparison of the PP, AT and IV estimators of efficacy, which elucidates the assumptions made by the different methods and discusses their relative efficiencies. Previous papers that advocate the IV estimator and model-based enhancements essentially dismiss the PP and AT approaches because of their potential bias, but the assumptions under which these methods are valid are rarely explicitly articulated, and we think a direct side-by-side comparison of the methods is illuminating. We attempt to make the comparison as transparent and non-mathematical as possible, by focusing on the simplest non-trivial situation and avoiding unnecessarily algebraic formulations.

A second objective is to elucidate and compare the role of covariates in improving the performance of the PP, AT and IV estimators. We show that covariates can reduce bias for PP and AT estimation but not for IV estimation, and can improve precision for all the methods. Finally we outline some extensions of these results to more general settings.

Our analysis provides an instructive non-technical introduction to some important ideas in causal inference, namely the definition of a *causal effect* of an active treatment as the difference in hypothetical outcomes under that treatment and a control treatment (Rubin 1974, 1977, 1978), and the idea of *principal stratification*, where individuals are stratified according to the values of the post-treatment variable under *both* treatments, rather than simply under the treatment actually observed (Frangakis and Rubin, 2002).

2. The Women Take PRIDE Study

Data from the “Women take PRIDE” (WTP) heart disease management study (Janevic et al., 2003) are used to illustrate our methods. In this study, participants are elder women with heart disease and the intervention is a self-regulatory process for identifying and resolving problems in managing their heart conditions (e.g., increasing physical activity). Two versions of the intervention were administered: a Group format, where 6-8 women study the educational material for 2-2.5 hours/week in a group setting; and a Self-Directed format, where each participant studies the same content on an individual basis at home after attending an in-person orientation session. The intervention program consists of 6 weekly classes for the Group format and 6 weekly units for the Self-Directed format. Compliance is defined as attendance of at least once Group class, or, equivalently, completion of at least one Self-Directed weekly unit.

This study adopted a doubly randomized preference trial (DRPT) (Long et al. 2006), with both a completely randomized arm (n=575), where participants were randomized to Group format, Self-Directed form and a control “usual care” format, and a choice arm (n=553) where participants were allowed to choose their treatment. One of the objectives in the WTP study is to assess efficacy of disease-management programs compared to the control format, accounting for noncompliance. The outcomes of interest include indicators of physical, psychological, and

functional health status, measured at baseline and months 4, 12, and 18. For more details on the WTP study, see Janevic et al., 2003.

The purpose of this paper is to provide an in-depth comparison of the assumptions, properties and relationship of the as-treated (AT), per-protocol (PP) and instrumental variable (IV) estimators for treatment efficacy in trials involving an active treatment and a placebo. We hence restrict our attention to the randomized arm, and compare subjects assigned to Group format ($T=1$, $n=190$) with the control format ($T=0$, $n=184$). In particular, we choose the outcome “6 minute walk” at Month 12, which measures the distance in feet an individual can walk in 6 minutes. Baseline 6-minute walk, and demographic characteristics at baseline, namely age and employment status, are chosen as covariates for methods involving covariate adjustment.

3. The Problem

We consider the analysis of clinical trials that involve randomization to an active treatment ($T = 1$) or a control treatment ($T = 0$), when the active treatment is subject to all-or-nothing compliance (Baker 1997). It is assumed that compliance to the control treatment is perfect, so noncompliance is only an issue when assigned the active treatment. The population can then be divided into two groups: *never-takers* or *non-compliers* ($C = 0$), who take the control treatment whether they are assigned to the control or active treatment, and *compliers* who take the treatment they are assigned ($C = 1$). We call the variable C *principal compliance*, since it is a special case of principal stratification, where subjects are classified by the values of a post-treatment variable (here compliance) under both treatments (Frangakis and Rubin 2002). Note that principal compliance is not the same as observed compliance, which depends only on whether a participant complied with the assigned treatment. Specifically, principal compliance equals observed compliance for cases in the treatment group, since by assumption all individuals

would comply with the control treatment. For cases in the control group observed compliance is always one, but principal compliance is unobserved, since we do not know if individuals would have complied if they had been assigned the active treatment. The key feature of principal compliance is that, unlike observed compliance, it is not affected by the treatment assigned, and hence can be used as a stratification variable in treatment comparisons, if the missing data problem can be solved.

As discussed in Section 6, our analysis extends readily to the case where there is a third group, *always-takers*, who take the active treatment whether or not assigned to the control or active treatment. For the moment we assume there are no always takers, as would be the case if individuals assigned to the control treatment cannot obtain access to the active treatment.

An intention-to-treat (IT) analysis in this setting computes and compares the average outcome in the treatment and control groups, ignoring information on compliance. The IT analysis is protected from selection bias by the randomization, and it measures treatment *effectiveness*, which is the effect of assigning the treatment without regard to whether or not the treatment is in fact taken. However, the IT analysis arguably does not provide a satisfactory estimate of *efficacy*, that is, the effect of the treatment itself, since treatment non-compliers are counted as treated cases even though they never received the treatment. We focus here on other approaches that use compliance information to estimate treatment efficacy.

Two simple and widely-used approaches to treatment efficacy are *as-treated* analysis (AT), where individuals are classified according to the treatment they actually received; that is, treatment non-compliers are included in the placebo group; and *per-protocol* analysis (PP), where only individuals who comply with the assigned treatment are included in the treatment comparison; that is, treatment non-compliers are excluded from the analysis. As noted in the

introduction, both of these analyses are subject to selection bias; for example in the case of PP analysis, compliers to the control treatment (by assumption, the entire population) may differ systematically from compliers to the active treatment, so the populations being compared are not comparable. An alternative approach known as instrumental variable estimation (IV) has been suggested, namely the IT estimate (the difference in means for treatment and control ignoring compliance) divided by the proportion of compliers in the treatment group (Bloom 1984; Newcombe, 1988; Robins 1989; Sommer and Zeger 1991; Baker and Lindeman, 1994; Angrist, Imbens and Rubin, 1996, henceforth denoted as AIR). As shown in the next section, the IV estimator is not subject to the selection bias noted for the AT and PP estimates, although it does require assumptions to be valid.

Since AT, PP and IV are all estimates of treatment efficacy, it is important to define precisely what we mean by that term. Two strands of the literature can be distinguished. One focuses on the complier-average causal effect (CACE), which is the average treatment effect in the subpopulation of principal compliers, for which $C = 1$ (AIR). An alternative estimand is the average treatment effect (ATE), which is defined as the difference in mean outcome if all individuals had been assigned and complied with the treatment and the mean if all individuals had received the control treatment. The ATE requires that it is reasonable to consider the treatment outcome for non-compliers, in the counterfactual event that they had complied with the treatment. It is somewhat controversial whether this counterfactual event is useful to contemplate, and arguably this varies according to context. For example, noncompliance to a behavioral treatment such as an exercise regime might plausibly be changed by increased motivation, as might occur if evidence that the treatment is successful becomes widely known. On the other hand, if noncompliance to a drug is the result of intolerable side-effects, then

compliance may require a reformulation of the drug to remove the side effects. Arguably this may change the properties of the drug, and estimation of the ATE is consequently more speculative.

The distinction between the CACE and the ATE is not a major issue in our simple setting, since in the absence of covariates, the ATE and CACE reduce to the same estimand under the assumption that the average outcome under the treatment is the same for compliers as for non-compliers if they had in fact complied. In situations where this assumption does not hold, the ATE and CACE differ, but additional information than that assumed here is needed to estimate the difference.

In the case where covariate information is available, the usual additional assumption to identify the ATE is that the average outcome under the treatment is the same for compliers and non-compliers *within strata defined by the covariates*. The CACE and ATE are then the same within strata, but the overall CACE and overall ATE differ in how the stratum-specific estimates are weighted when combining over strata; specifically for the CACE the natural choice is to weight using the covariate distribution of compliers, whereas for the ATE the natural choice is to weight using the covariate distribution of compliers and non-compliers. This difference in weighting is only important if the covariates modify the effect of the treatment and are related to compliance, and we think it is likely to be minor in many applications. We focus on the CACE in the sequel, in order to avoid the need for assumptions about counterfactual conditions.

In the next section we introduce notation that allows us to define the PP, AT and IV estimators explicitly, and clarify the assumptions under which they yield consistent estimates of treatment efficacy. In Section 5 we consider the precision of the three estimates, and implications for allocation of the sample between the treatment and control group. In Section 6 we consider

extensions of the methods to include observed covariates are then discussed, emphasizing the contrasting role of covariates. Finally in Section 7 we discuss briefly generalizations of the problem considered here and make some closing remarks.

4. The AT, PP and IV Estimators and When they are Consistent for the CACE

Let R denote the indicator for the random treatment assignment, with value 1 if an individual is assigned to treatment and 0 if assigned to control. Note that in our notation R is the treatment randomized, and T is the treatment received, so for non-compliers assigned the active treatment $R = 1$ but $T = 0$. Let C be the indicator for principal compliance, with value 1 for individuals who comply with the treatment if assigned, and 0 otherwise. We make the stable unit-treatment value assumption (SUTVA) (Rubin, 1978), which implies that compliance and outcomes for individuals are not affected by the assignments and outcomes of other individuals in the sample (AIR). Table 1A shows a classification of the population by R and C , assuming a proportion \mathbf{a} of the population is assigned to the population and a proportion \mathbf{b} are principal compliers. The entries reflect the fact that R and C are independent, which is justified under the assumption that treatments are randomly assigned. The proportions in square parentheses are unobserved for the sample, since the principal compliance status of individuals in the control group is unknown – we do not know whether they would have complied with the active treatment if assigned to that treatment.

We now introduce an outcome variable Y , and let \mathbf{m}_c denote the mean of Y for the subpopulation with $R = r$ and $C = c$; let \bar{y}_{rc} denote the corresponding sample mean, and n_{rc} the corresponding sample size. The population means are displayed in Table 1B, with square parentheses indicating quantities for which the corresponding sample estimates are not observed.

The observed sample counts and means are shown in Table 1C. The CACE is then the difference in complier means, $\mathbf{d} = \mathbf{m}_{11} - \mathbf{m}_{01}$.

If principal compliance were known for all individuals in the sample, a direct estimate of the CACE would be $\bar{y}_{11} - \bar{y}_{01}$. However, in practice the means in the four cells of Table 1B are not observed, and additional assumptions are needed to yield a CACE estimate.

One possibility is to assume

$$\text{NCEC: } \mathbf{m}_{00} = \mathbf{m}_{01}, \quad (1)$$

which asserts that the mean outcome in the control group is the same for principal compliers and never-takers. We label this assumption “no compliance effect for controls” (NCEC). Condition (1) is implied by the following conditional independence assumption:

$$\text{NCEC*}: [Y \wedge C \mid R = 0], \quad (2)$$

where the symbol \wedge denotes independence. Under NCEC or NCEC*, it is natural to estimate both \mathbf{m}_{00} and \mathbf{m}_{01} by the marginal control mean \bar{y}_{0+} , leading to the PP estimate of the CACE:

$$\hat{\mathbf{d}}_{PP} = \bar{y}_{11} - \bar{y}_{0+}, \quad (3)$$

which includes cases that take their assigned treatments. The problem with this estimator is that the NCEC assumption is generally considered questionable, since compliers and non-compliers may differ on various unobserved characteristics related to the outcome under the control treatment. NCEC can be weakened by adjusting for known covariates that characterize differences between compliers and never-takers, as discussed Section 5.

Note that participants in the subpopulation of never-takers ($C = 0$) are randomly assigned to treatment or control, and in both cases they receive the same (control) treatment. Thus it is often reasonable to assume the means in the first column of Table 1B are equal, that is:

$$\text{ER: } \mathbf{m}_{00} = \mathbf{m}_{10} . \quad (4)$$

This assumption is often called the exclusion restriction (ER) in the literature, a term that originates in the econometric literature (see e.g. AIR). The ER assumption is implied by the following conditional independence assumption:

$$\text{ER*}: [Y \wedge R \mid C = 0] . \quad (5)$$

ER or ER*, though often plausible, are assumptions, since the outcome may be affected by whether treatment or control is assigned even though the resulting treatment remains the same, particularly in trials of behavioral interventions. Under ER or ER*, it is natural to estimate both \mathbf{m}_{00} and \mathbf{m}_{10} by \bar{y}_{10} . The estimate of \mathbf{m}_{01} that yields the marginal mean \bar{y}_{0+} when averaged with \bar{y}_{10} is

$$\hat{\mathbf{m}}_{01} = (n_{0+}\bar{y}_{0+} - n_{00}\bar{y}_{10}) / n_{01} , \quad (6)$$

but this cannot be computed since n_{00} and n_{01} are unobserved. However, the principal compliance rate for controls n_{01} / n_{0+} can be estimated by the principal compliance rate for cases, n_{11} / n_{1+} , since the underlying population rates are the same by randomization. Replacing n_{01} / n_{0+} by n_{11} / n_{1+} in (6) yields $\hat{\mathbf{m}}_{01} = (n_{1+}\bar{y}_{0+} - n_{10}\bar{y}_{10}) / n_{11}$, and the following IV estimate of the CACE:

$$\hat{\mathbf{d}}_{IV} = (\bar{y}_{11} - \hat{\mathbf{m}}_{01}) = (\bar{y}_{1+} - \bar{y}_{0+}) / \hat{\mathbf{b}} , \quad (7)$$

the intention to treat estimator divided by $\hat{\mathbf{b}} = n_{11} / n_{1+}$, the compliance rate for cases. The estimator (7) is commonly termed the instrumental variable (IV) estimate, since it is a special case of IV estimation with the randomization indicator as the instrument.

Suppose now we assume NCEC and ER simultaneously:

$$\text{NCEC} + \text{ER: } \mathbf{m}_{00} = \mathbf{m}_{01} = \mathbf{m}_{10} , \quad (8)$$

or the corresponding conditional independence assumptions NCEC*+ER*. The natural estimate of the control mean is $\bar{y}_0 = (n_{0+}\bar{y}_{0+} + n_{10}\bar{y}_{10})/(n_{0+} + n_{10})$, pooling the data in the three cells that receive the control treatment $T = 0$. The resulting estimate of the CACE is AT estimate

$$\hat{d}_{AT} = \bar{y}_{11} - \bar{y}_0, \quad (9)$$

which classifies all cases according to the treatment they received.

To summarize, the NCEC assumption leads to \hat{d}_{pp} , the ER assumption leads to \hat{d}_{IV} and the NCEC and ER assumptions combined lead to \hat{d}_{AT} . The choice between the estimators rests largely on the perceived validity of their underlying assumptions, although the precision of the estimates may also play a role.

5. The Precision of the Alternative Estimators, and Design Implications

5.1. Comparison of the AT, PP and IV Estimators

For simplicity, we assume that the within-cell variance of Y in each of the cells in Table 1 is s^2 ; also since the ER assumption is often plausible, we compare biases and variances of the estimates under that assumption. Under these conditions, the large-sample biases of the ER, PP and AT estimates are

$$B(\hat{d}_{IV}) = 0, \quad (10)$$

$$B(\hat{d}_{pp}) = (1 - b)\Delta s, \quad (11)$$

$$B(\hat{d}_{AT}) = \frac{1-b}{1-ab}\Delta s, \quad (12)$$

where $\Delta = (m_{01} - m_{00})/s$, which is zero under the NCEC assumption. The large sample variances of the three estimators are

$$Var(\hat{d}_{IV}) = \frac{s^2}{na(1-a)b^2} (1 + b(1-b)\Delta^2), \quad (13)$$

$$Var(\hat{\mathbf{d}}_{PP}) = \frac{\mathbf{s}^2}{na(1-a)b} (\mathbf{ab} + 1 - \mathbf{a} + \mathbf{ab}^2(1-b)\Delta^2), \quad (14)$$

$$Var(\hat{\mathbf{d}}_{AT}) = \frac{\mathbf{s}^2}{n} \left(\frac{1}{1-\mathbf{ab}} + \frac{1}{\mathbf{ab}} + \frac{(1-\mathbf{a})\mathbf{b}(1-\mathbf{b})\Delta^2}{(1-\mathbf{ab})^4} (1-2\mathbf{ab} + \mathbf{ab}^2) \right), \quad (15)$$

. It can be shown with some algebra that $\mathbf{s}_{IV}^2 \geq \mathbf{s}_{PP}^2 \geq \mathbf{s}_{AT}^2$.

IV is markedly less efficient than PP and AT for small values of \mathbf{b} . This is illustrated in Figure 1, which plots the asymptotic relative efficiencies $Var(\hat{\mathbf{d}}_{PP})/Var(\hat{\mathbf{d}}_{IV})$ and $Var(\hat{\mathbf{d}}_{AT})/Var(\hat{\mathbf{d}}_{IV})$ of IV compared to PP and AT against \mathbf{b} , when $\mathbf{m}_{b0} = \mathbf{m}_0 = 0$, $\mathbf{m}_1 = 2$, $\mathbf{s}^2 = 1$, $\mathbf{a} = 0.5$. Figures 2 and 3 show the mean squared error ($MSE = B^2(\hat{\mathbf{d}}) + Var(\hat{\mathbf{d}})$) of the three estimators for various choices of \mathbf{a}, \mathbf{b} and Δ , which are measured in units of \mathbf{s}^2/n . If we assume both ER and NCEC, that is, $\Delta = 0$, $\hat{\mathbf{d}}_{AT}$ is preferred to $\hat{\mathbf{d}}_{IV}$. If we assume ER but do not assume NCEC, that is, $\Delta \neq 0$, $\hat{\mathbf{d}}_{AT}$ might still be preferred to $\hat{\mathbf{d}}_{IV}$, if the bias is small and compensated by a large reduction in variance. (For given bias, bias increasingly dominates variance as the sample size increases, so in large samples $\hat{\mathbf{d}}_{IV}$ is preferred to both $\hat{\mathbf{d}}_{AT}$ and $\hat{\mathbf{d}}_{PP}$.)

If one of either NCEC or ER is assumed true, the other assumption can be tested empirically by comparing $\bar{y}_{10} - \bar{y}_{0+}$ with zero; since IV and AT both assume ER, one might increase the efficiency of the CACE estimate by choosing AT over IV if this test is not rejected, or the difference $\bar{y}_{10} - \bar{y}_{0+}$ is “small”. This approach has most appeal when the compliance rate is low, since in this case $\hat{\mathbf{d}}_{IV}$ has substantially higher variance and the power of the test may be reasonable; when compliance is high the power of the test is very modest. An indirect approach to checking the ER assumption using covariates is discussed in Section 6 below.

What fraction \mathbf{a} of cases should be assigned to the treatment group for optimal efficiency? Differentiating (13) with respect to \mathbf{a} , the variance of the IV estimate is minimized when $\mathbf{a} = 0.5$, that is, an equal allocation of treated and control cases. One might think that given noncompliance, more cases should be assigned to the treatment group, but an equal allocation is optimal for the IV estimator under our variance assumptions. On the other hand, the variance of the per-protocol estimate is minimized when $\mathbf{a} = \mathbf{a}_{pp} = 1 / \left(1 + \sqrt{\mathbf{b}(1 + \Delta^2 \mathbf{b}(1 - \mathbf{b}))} \right)$. When $\Delta = 0$, $\mathbf{a}_{pp} = 1 / (1 + \sqrt{\mathbf{b}})$, which does assign more cases to the treatment group when there is noncompliance.

5.2 CACE Estimates for the Women Take Pride Study

We illustrate the AP, PP and IV estimates with data from the Women Take Pride (WTP) study. As discussed in Section 2, we restrict attention to randomized subjects, and compare women assigned to Group behavioral intervention ($R=1$) with the Control “usual care” treatment ($R=0$).

Table 2 shows the observed counts and means for the outcome “6 minute walk” taken at month 12, measuring the distance in feet an individual can walk in 6 minutes. These analyses exclude 69 of the 190 cases in the intervention group and 62 cases in the control group who drop out before month 12. For the current analysis we assume that drop out is random within each treatment group – the analysis in Section 7.2 relaxes this assumption by allowing dependence of the dropout on baseline covariates. In the treatment group $\hat{\mathbf{b}} = 105/121 = 86\%$ complied with treatment, where compliance is defined here as completion of at least one of the treatment modules. Table 2 also shows the estimated cell means from the (a) PP, (b) IV and (c) AT methods; the estimated means in italics are equated because of the model assumptions. For

example, the mean for compliers in the control group equals the mean for compliers in the treatment group by ER, and is estimated to be 694.12; the mean for compliers in the control group equals the mean for non-compliers in the control group by NCEC and is estimated to be 748.90. The last row of the lower panel of Table 2 shows the corresponding estimates of the CACE, and associated standard errors, computed using Eqs. (13)-(15). Note that the estimates of the CACE from the three methods are somewhat different, and the PP and AT estimates are statistically significant at the 0.05 alpha level. The standard errors are ordered as described above, with relative small differences since the compliance rate is quite high in this application. The t-test for the combined NCEC + ER assumptions compares the mean for controls (748.90) with the mean for treatment non-compliers (694.12), statistically not significant even though the difference in means is quite substantial. This illustrates the low power of the test when the compliance rate is high.

6. Model-Based Estimation

We have noted that the potentially high variance of the IV estimator when the compliance rate is low. The precision of $\hat{\mathbf{d}}_{IV}$ can be improved by seeking a more efficient estimator. Technically all the estimators considered so far can be viewed as method-of-moment estimators under the various assumptions. Another approach to inference is maximum likelihood (ML) estimation based on a model for the joint distribution of Y , R and C . In particular, suppose Y given R , C is assumed normal with mean \mathbf{m}_c and variance \mathbf{s}^2 . The loglikelihood of the observed data, assuming independent observations, is:

$$\begin{aligned} \ell(\{\mathbf{m}_c\}, \mathbf{b}, \mathbf{s}^2 \mid data) = & -0.5n \log \mathbf{s}^2 - \sum_{i:r_i=1} 0.5(y_i - \mathbf{m}_{1,c_i})^2 / \mathbf{s}^2 \\ & + \sum_{i:r_i=0} \log \left[\mathbf{p}_i \exp(-0.5(y_i - \mathbf{m}_{01})^2 / \mathbf{s}^2) + (1 - \mathbf{p}_i) \exp(-0.5(y_i - \mathbf{m}_{00})^2 / \mathbf{s}^2) \right], \end{aligned}$$

where \mathbf{p}_i is the compliance rate for subject i , and is assumed equal to \mathbf{b} with no covariates. Without constraints on $\{\mathbf{m}_{cr}\}$, the means in this loglikelihood are not identified, but they are identified under the ER and/or NCEC assumptions. It is easily shown that $\hat{\mathbf{d}}_{pp}$ is ML for this model under the NCEC assumption (1), and $\hat{\mathbf{d}}_{AT}$ is ML under the NCEC and ER assumptions (8). However $\hat{\mathbf{d}}_{IV}$ is not the ML estimate under the ER assumption (4). The ML estimate (say $\hat{\mathbf{d}}_{MLER}$) does not have an explicit form, but is quite easily computed using the EM algorithm, treating the compliance indicators for the controls as missing data (AIR).

The estimate $\hat{\mathbf{d}}_{MLER}$ is more efficient than $\hat{\mathbf{d}}_{IV}$ (Imbens and Rubin 1996), but makes stronger distributional assumptions; in particular, we conjecture that it is sensitive to violations of the assumption of a constant variance for compliers and non-compliers when $R=0$. While $\hat{\mathbf{d}}_{MLER}$ and $\hat{\mathbf{d}}_{IV}$ differ under the normal model, the ML estimate equals $\hat{\mathbf{d}}_{IV}$ for binary outcomes Y with a Bernoulli distribution, providing the resulting means in Table 1B, which are estimated probabilities for a binary Y , all lie between zero and one (Baker and Lindeman 1994).

7. Methods That Include Covariate Information

7.1. CACE Estimation with Covariates.

We now consider how covariates X measured for compliers and non-compliers can be used to improve the performance of the AT, PP and IV estimates. Related questions are the properties of a good covariate, and approaches to dimension reduction when a large number of covariates are available.

For AT and PP, the covariates can be adjusted by a regression model of Y on X and a dummy variable T for the received treatment. All cases are included for AT analyses, and non-compliers assigned the treatment are excluded for PP analysis. As in covariate adjustments in

observational settings, inclusion of X can reduce selection bias from noncompliance and increase precision. Concerning bias, a correctly-specified regression adjustment with covariates X weakens NCEC* in (2) to

$$\text{NCEC}^*(X): [Y \wedge C | X, R = 0], \quad (16)$$

that is, to an assumption that there is no compliance effect for controls within strata defined by X . Covariates that are predictive of the outcome can also increase precision by reducing the residual variance \mathbf{s}^2 of the regression of Y on X and T .

A strategy for robust modeling when there are a number of covariates is to stratify on a coarsened function $c(X)$ of the covariates X . To limit bias, a function is sought such that if (16) is true, then it remains valid with X replaced by $c(X)$. A standard application of the propensity score theory of Rosenbaum and Rubin (1983) implies that the coarsest function $c(X)$ with this property is the propensity to comply, $p(X) = p(C = 1 | X)$. This score can be estimated by regressing the compliance indicator C on the covariates using the cases assigned to treatment; control cases are excluded since C is not known for them. Note that the same propensity score applies for both AT and PP, because the NCEC assumption is shared by both methods. AT also requires the ER assumption, but that is not addressed by the propensity adjustment.

Covariates do not play a role in bias reduction for the IV method. To see this, suppose ER* is assumed conditional on X , that is

$$\text{ER}^*(X): [Y \wedge R | X, C = 0].$$

Then in terms of densities,

$$\begin{aligned}
p(Y | R=1, C=0) &= \int p(Y | X, R=1, C=0) p(X | R=1, C=0) dX \\
&= \int p(Y | X, R=0, C=0) p(X | R=1, C=0) dX \\
&= \int p(Y | X, R=0, C=0) p(X | R=0, C=0) dX \\
&= p(Y | R=0, C=0),
\end{aligned}$$

where the first and last equalities are by definition, the second is implied by $ER^*(X)$, and the third equality is true by the randomization of treatments. Thus $ER^*(X)$ implies ER^* , so the latter is not weakened by conditioning on X , and the covariates do not play a role in bias reduction.

Covariates can be used to increase the precision of $\hat{\mathbf{d}}_{IV}$, however. We first consider the modification of $\hat{\mathbf{d}}_{IV}$ for a single categorical X with J categories. Denote the IV estimator within the stratum $X = j$ as $\hat{\mathbf{d}}_{IV,j} = (\bar{y}_{1+j} - \bar{y}_{0+j}) / \hat{\mathbf{b}}_j$, with the subscript j denoting stratum. Let p_j be the proportion of cases in stratum j , estimated from the pooled sample. Then $p_j \hat{\mathbf{b}}_j / \sum_{k=1}^J p_k \hat{\mathbf{b}}_k$ estimates the proportion of compliers in stratum j . Weighting $\hat{\mathbf{d}}_{IV,j}$ by this proportion and summing over strata yields the stratified IV estimator

$$\hat{\mathbf{d}}_{IV|X} = \sum_{j=1}^J p_j (\bar{y}_{1+j} - \bar{y}_{0+j}) / \sum_{j=1}^J p_j \hat{\mathbf{b}}_j. \quad (17)$$

The numerator of $\bar{y}_{IV|X}$ is a stratified form of the IT estimator, and its precision is improved by stratifying on covariates that are predictive of the outcome. The denominator of $\bar{y}_{IV|X}$ is a stratified estimator of the overall compliance rate, and its precision is improved by stratifying on the covariates that are predictive of compliance. We conjecture that the former of these two components has the greater potential for variance reduction. A natural generalization of $\bar{y}_{IV|X}$ for a set of categorical and/or continuous X 's is

$$\hat{\mathbf{d}}_{IV|X} = \sum_i (\hat{y}_{1i} - \hat{y}_{0i}) / \sum_i \hat{\mathbf{b}}_i, \quad (18)$$

where the summation is over all individuals i in the sample; \hat{y}_{ri} is the predicted outcome for unit i if randomized to treatment r , computed from a regression of Y on X and R ; and \hat{b}_i is the predicted true compliance for unit i , computed from a regression of C on X estimated from the cases assigned to treatment. The latter should be of a form appropriate for a binary outcome, for example logistic or probit regression. We are not aware of discussions of (17) or (18) in the compliance literature.

An alternative to (17) or (18) is to compute ML or Bayes estimates of the CACE given covariates X , using a full model for the distribution of Y and C , given R and X , and treating the compliance indicators in the control group as missing covariates (Imbens and Rubin, 1997a; Little and Yau 1998). The model can be specified in terms of a compliance model for C given R and X indexed by parameters \mathbf{q}_C , and an outcome model for Y given C , R and X indexed by parameters \mathbf{q}_Y . The loglikelihood takes the form

$$\begin{aligned} \ell(Y, C_{\text{obs}} | R, X_Y, X_C; \mathbf{q}_Y, \mathbf{q}_C) = & \\ \sum_{R_i=1} \left(\log \left(f(Y_i | R_i, X_{Y_i}, C_i; \mathbf{q}_Y) \right) + C_i \log \left(\mathbf{p}_i(X_{C_i}; \mathbf{q}_C) \right) + (1 - C_i) \log \left(1 - \mathbf{p}_i(X_{C_i}; \mathbf{q}_C) \right) \right) & \quad (19) \\ + \sum_{R_i=0} \left[\log \left(\mathbf{p}_i(X_{C_i}; \mathbf{q}_C) f(Y_i | R_i, X_{Y_i}, C_i = 1; \mathbf{q}_Y) \right) + \left(1 - \mathbf{p}_i(X_{C_i}; \mathbf{q}_C) \right) f(Y_i | R_i, X_{Y_i}, C_i = 0; \mathbf{q}_Y) \right] & \end{aligned}$$

In these models, the compliance model should exclude the main effect of R and interactions including R , because of the assumption of randomization of treatments. The outcome model should exclude effects of the form RU , where U is the identity or a set of some or all the covariates, because these effects concern the effects of compliance for controls, which are assumed zero under ER. The modeling approach yields gains of efficiency over the IV approach (Imbens and Rubin 1997a, 1997b), but may be more vulnerable to model misspecification; more simulations comparing the methods would be of interest.

We noted in Section 4 that the data do not provide evidence of the validity of ER or NCEC. With covariates, the same comment applies within strata defined by the covariates. However, if the covariates are good predictors of compliance, the relationship between outcome and predicted compliance can be assessed in the control group, and lack of evidence of a relationship might be construed as indirect evidence in favor of the NCEC assumption. Specifically, transform the covariates X into the propensity score $\hat{p}(X)$ and covariates Z orthogonal to $\hat{p}(X)$, and regress Y on $\hat{p}(X)$ and Z in the control group. If the coefficient of $\hat{p}(X)$ in this regression is small, this provides some justification for the NCEC assumption, suggesting that PP or AT analysis may be reasonable options. On the other hand if the regression coefficient of $\hat{p}(X)$ is large, estimates like \hat{d}_{IV} or \hat{d}_{MLER} that do not require the NCEC assumption may be preferable.

7.2 Women Take Pride Data Analysis with Covariates

We now extend our analyses in Section 5.2 to include adjustment for baseline covariates. We considered, in addition to 6-minute walk at baseline, the covariates previously included in Janevic, et al. (2003), including age, employment and Symptom Impact Profile (SIP) physical score, which measures a subject's physical functioning. The ML estimate of the CACE for the data without adjusting for covariates is $\hat{d}_{MLER} = 109.78$, close to the IV estimate of $\hat{d}_{IV} = 108.76$ in Table 2. It has a slightly lower standard error (61.28 vs. 65.53). Table 3 shows the results of fitting a model of form (19) to the data from the WTP study. Block (1) shows the coefficients for the outcome model, with the CACE being the Compliance*Treatment interaction. Block (2) shows coefficients from the compliance model. The covariate-adjusted the ML estimate of the CACE from this model is 97.20 with a reduced standard error, namely 43.73, indicating some improvement in precision from adjusting for the covariates. In addition, women with higher

baseline measure tend to have higher 6-min walk measure at month 12, and older women tend to have lower 6-min walk measure at month 12. Table 4 summarizes the IV, MLN-ER, PP and AT estimates both with and without covariate adjustments. PP and AT estimates with covariate adjustments are obtained using linear regression adjusting for age, baseline 6-minute walk measurement, and employment status. Table 4 shows that covariate adjustment generally improves precision and reduces differences between the methods. The large gain in efficiency may be due to the significant effects of age and baseline measurements on the outcome of interest. The results suggest that compliers performed better under treatment rather than under control with respect to this outcome.

8. Discussion.

We have compared a variety of methods for estimating efficacy in randomized trials for a control and active treatment, when there is all-or-nothing compliance in the treatment arm. This work applies directly to two active treatments A and B, when compliance is perfect for treatment A, and noncompliance to treatment B means taking treatment A. In practice, the choice of methods depends on various factors, effect sizes relative to between-subject variability of the outcome measure, sample size and differences in characteristics of compliers and non-compliers. The choice also depends on the plausibility of the different modeling assumptions and the trade-off between efficiency and robustness. If NCEC or NCEC+ER given a set of covariates can be believed, and regressions on the covariates are correctly specified, then AT can be dramatically more efficient than IV. On the other hand, the IV estimate of the CACE under ER is robust against misspecification of our regression model or false belief in NCEC. Thus, it may be wise to compute and compare all the estimates to assess sensitivity of answers to the choice of method.

Many generalizations and extensions of these methods are possible. A relatively straightforward generalization is to add to the never-takers ($C = 0$) and compliers ($C = 1$) a subpopulation of always-takers (say $C = 2$), who obtain the active treatment whether assigned to treatment or control. This adds an additional column to Table 1. The CACE is still defined as the treatment effect in the subpopulation of compliers, the middle column of the tables, and a second restriction on the means is required to identify the parameters. The PP estimate assumes $\mathbf{m}_{00} = \mathbf{m}_{01}$ (NCEC, as before) and also $\mathbf{m}_{11} = \mathbf{m}_{12}$, that is, the mean outcome under treatment is the same for compliers and always takers. The ER assumption in IT analysis yields two restrictions, $\mathbf{m}_{00} = \mathbf{m}_{10}$ (as before) and $\mathbf{m}_{02} = \mathbf{m}_{12}$, that is, the treatment mean outcome for always-takers is the same regardless of the treatment assigned). The AT analysis assumes $\mathbf{m}_{00} = \mathbf{m}_{01} = \mathbf{m}_{10}$ and $\mathbf{m}_{11} = \mathbf{m}_{12} = \mathbf{m}_{02}$, leading to pooling the means in the (00, 01, 10) and (11, 12, 02) cells. The propensity to comply discussed in Section 3 is replaced by two propensities, $\Pr(C = 1 | X, R = 1, C = 1 \text{ or } 0)$ and $\Pr(C = 1 | X, R = 0, C = 1 \text{ or } 2)$. An interesting new possibility is hybrid models that mix the NCEC and ER assumptions. For example one might assume $\mathbf{m}_{00} = \mathbf{m}_{01}$ (NCEC) and $\mathbf{m}_{20} = \mathbf{m}_{21}$ (ER for always-takers only).

A further extension is to allow the possibility of “defiers” who take the opposite of the treatment assigned (AIR). However this group requires additional assumptions to identify the parameters, and since defiance is incompatible with a consistent preference for one or other of the treatments, the assumption of “no defiers” is commonly made in practice.

Other extensions that require more restrictions to identify the parameters include the case where partial compliance is modeled (Goetghebeur & Molenberghs 1996), or there are more than two treatments, such as two active treatments and a control treatment that is assumed to apply to non-compliers. CACE estimation in trials involving a control group and more than one treatment

groups is more complicated, and consists of more principal compliance categories and involves more complicated identifiability assumptions. The results will be reported elsewhere.

Another extension is to consider joint models for noncompliance and missing data – for simplicity we confined our analyses of 12 month WTP data to completers. For some approaches to this issue see Frangakis and Rubin (1999) and Peng, Little and Raghunathan (2004).

Acknowledgements. This research was supported by grant R01CA76404 from the National Cancer Institute. We thank Noreen Clark for kindly providing the data from the Women Take Pride Study that illustrate the methods.

References

- Angrist, J.D., Imbens, G.W., & Rubin, D.B. (1996). Identification of causal effects using instrumental variables (with discussion and rejoinder). *Journal of the American Statistical Association*, 91, 444-472.
- Baker S. G. (1997). Compliance, all-or-none. In Kotz S, Read CR, Banks DL (eds.). *The Encyclopedia of Statistical Science*, Update Volume 1, 134-138. New York: John Wiley and Sons.
- Baker, S.G., & Lindeman, K.S. (1994). The paired availability design: a proposal for evaluating epidural analgesia during labor. *Statistics in Medicine*, 13: 2269-2278.
- Bloom, H.S. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation Reviews*, 8: 225-246.
- Frangakis, C. E. and Rubin, D. B. (1999). Addressing complications of intent-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika*, 86, 365-379.

- Frangakis, C.E. and Rubin, D.B. (2002). Principle stratification in causal inference. *Biometrics*, 58, 21-29.
- Goetghebeur, E., & Molenberghs, G. (1996). Causal inference in a placebo-controlled clinical trial with binary outcome and ordered compliance. *Journal of the American Statistical Association*, 91: 928-934
- Imbens, G.W., & Rubin, D.B. (1997a). Estimating outcome distributions for compliers in instrumental variables models. *Review of Economic Studies*, 64: 555-574.
- Imbens, G.W., Rubin, D.B. (1997b). Bayesian inference for causal effects in randomized experiments with noncompliance. *Annals of Statistics*, 25: 305-327.
- Janevic, M.R., Janz, N.K., Dodge, J.A., Lin, X., Pan, W., Sinco, B.R., and Clark, N.M. (2003). The role of choice in health education intervention trials: a review and case study. *Social Science and Medicine*, 56, 7:1581-1594.
- Little, R.J.A. and Yau, L. (1998). Statistical techniques for analyzing data from prevention trials: treatment of No-Shows using Rubin's causal model. *Psychological Methods*, 3, 2, 147-159.
- Long, Q., Little, R. J. and Lin, X. (2006) Causal inference in hybrid intervention trials involving treatment choice. *Journal of the American Statistical Association*, in press.
- Newcombe, R.G. (1988). Explanatory and pragmatic estimates of the treatment effect when deviations from allocated treatment occur. *Statistics in Medicine*, 7: 1179-1186.
- Peng, Y., Little, R.J. and Raghunathan, T. (2004). An extended general location model for causal inferences from data subject to non-compliance and missing values. *Biometrics*, 60, 598-608.
- Robins, J.M. (1989). The control of confounding by intermediate variables. *Statistics in Medicine*, 8: 679-701.

Rosenbaum, P., & Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70: 41-55.

Rubin, D.B. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology*, 66: 688-701.

Rubin, D.B. (1977). Assignment to a treatment group on the basis of a covariate. *Journal of Educational Statistics*, 2: 1-26.

Rubin, D.B. (1978). Bayesian inference for causal effects: the role of randomization, *the annals of statistics*, 6: 34-58.

Sommer, A., & Zeger, S. (1991). On estimating efficacy from clinical trials. *Statistics in Medicine*, 10: 45-52.

Table 1. Classifications by Treatment and Principle Compliance: (A) Population Proportions; (B) Population Mean Outcomes; (C) Observed Means (Sample Counts).

A

		Compliance C		
		0	1	ALL
Randomized Treatment R	0	$[(1-a)(1-b)]$	$[(1-a)b]$	$1-a$
	1	$a(1-b)$	ab	a
	ALL	$[1-b]$	$[b]$	

$[\]$ = not observed

B

		Compliance C		
		0	1	
Randomized Treatment R	0	$[m_{00}]$	$[m_{01}]$	m_{0+}
	1	m_{10}	m_{11}	m_{1+}
		$[m_{+0}]$	$[m_{+1}]$	

$[\]$ = not directly estimable without assumptions

C.

		Compliance C		
		0	1	
Randomized Treatment R	0	?	?	$\bar{y}_{0+} (n_{0+})$
	1	$\bar{y}_{10} (n_{10})$	$\bar{y}_{11} (n_{11})$	$\bar{y}_{1+} (n_{1+})$
		?	?	

Table 2. Women Take Pride Study: (A) Sample Means (Sample sizes) for Outcome 6 Minute Walk in Control and Group Treatment Subgroups, and (B) Predicted Means under PP, IV and AT Models.

A

		Compliance <i>C</i>		
		0	1	
Randomized Treatment <i>R</i>	0	?	?	748.90 (122)
	1	694.12 (16)	866.01 (105)	843.28 (121)

B

Treatment <i>R</i>	IV		PP		AT	
	Compliance		Compliance		Compliance	
	0	1	0	1	0	1
0	694.12	757.25	748.90	748.90	742.55	742.55
1	694.12	866.01	694.12	866.01	742.55	866.01
CACE (SE)		108.76 (65.53)		117.11 (58.97)		123.45 (57.37)

Table 3. Women Take Pride Study: ML estimates of Regression Coefficients of Models for the Outcome 6-Min Walk and Compliance with Covariates

Parameter	Estimate (B)	SE(B)	P-Value
Outcome Model (1)			
Intercept	707.97	278.89	< 0.001
Age	-6.763	3.404	0.032
Baseline 6-min Walk	0.707	0.048	< 0.001
Employment	28.71	55.18	0.59
Treatment Compliance	-29.72	84.87	0.83
Compliance *Treatment	97.20	43.73	0.043
Compliance Model (2)			
Intercept	2.259	0.497	< 0.001
Employment	-0.464	0.792	0.526
Baseline SIP Physical	-0.040	0.041	0.282

Table 4. Women Take Pride Study: IV, MLN-ER, PP and AT estimates for the Outcome 6-Min Walk, with and without covariate adjustment

Covariates		IV	MLN-ER	PP	AT
Unadjusted	Estimate	108.76	109.78	117.11	123.45
	SE	65.53	61.28	58.97	57.37
Adjusted	Estimate	100.05*	97.20	95.58	90.13
	SE	44.25*	43.73	40.30	38.74

* Estimate is computed using formula (18) and its SE is computed using the bootstrap

Figure 1. Asymptotic Relative Efficiency of IV compared to PP and AT.

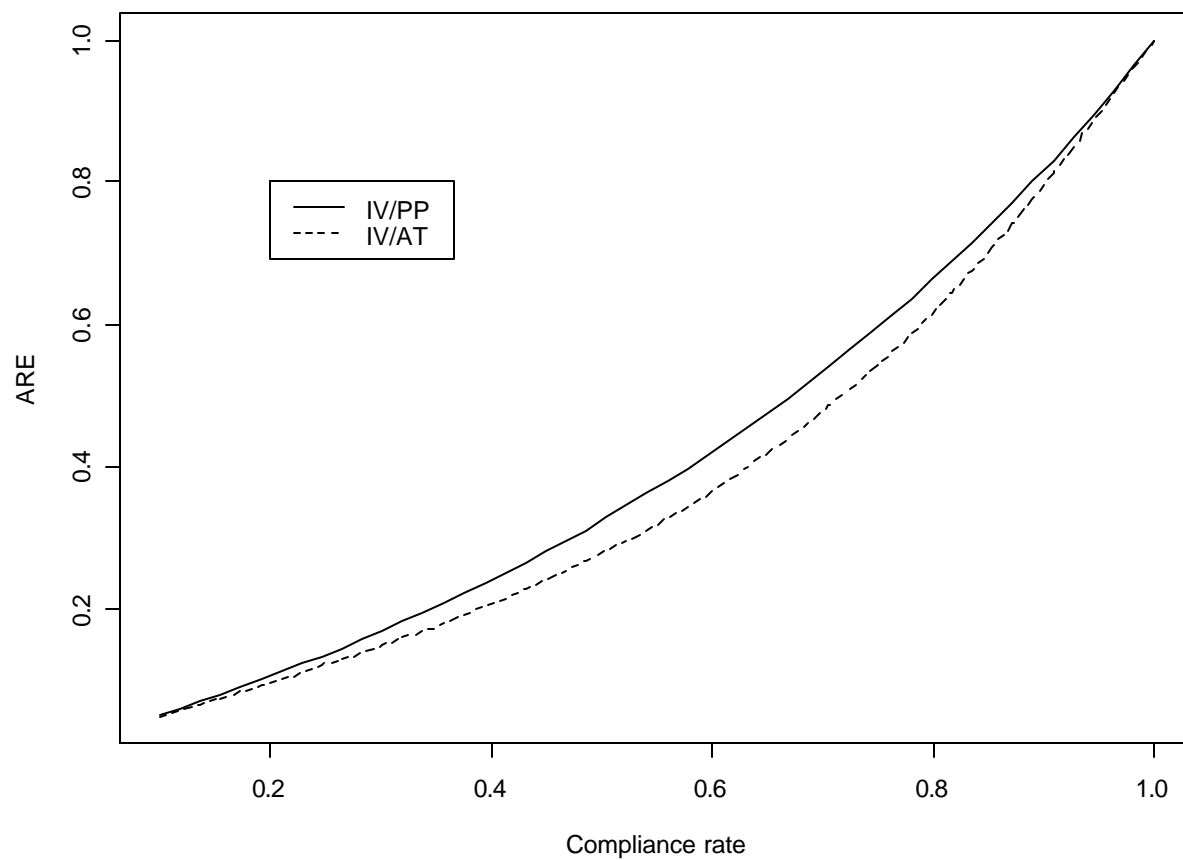


Figure 2. MSE of Estimates for varying population proportions in the treatment arm, a , when compliance rate, $b=0.751$, $n=575$ and treatment effect size $\sqrt{n}\Delta=0,5,10$ clockwise for three graphs, respectively.

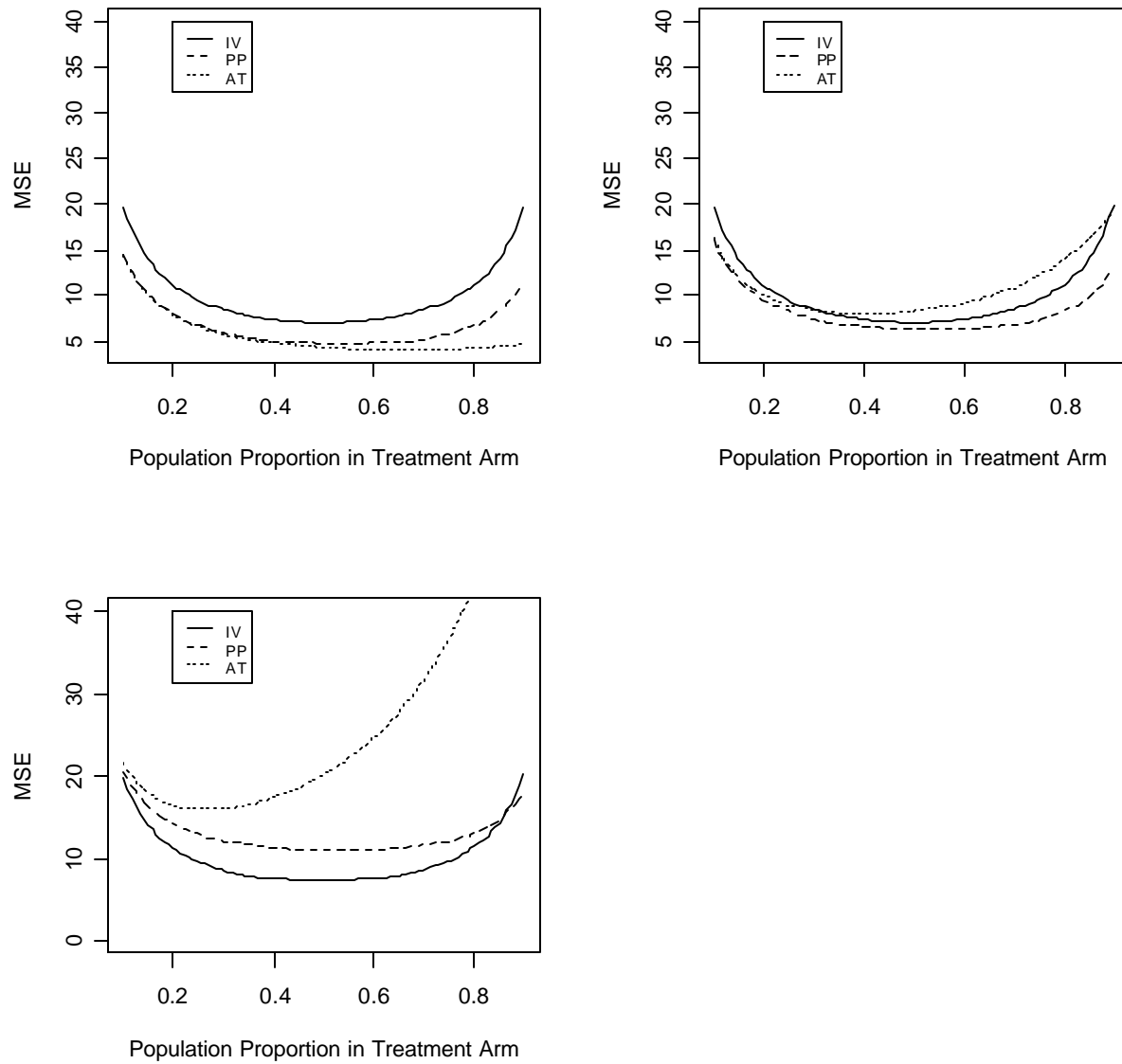
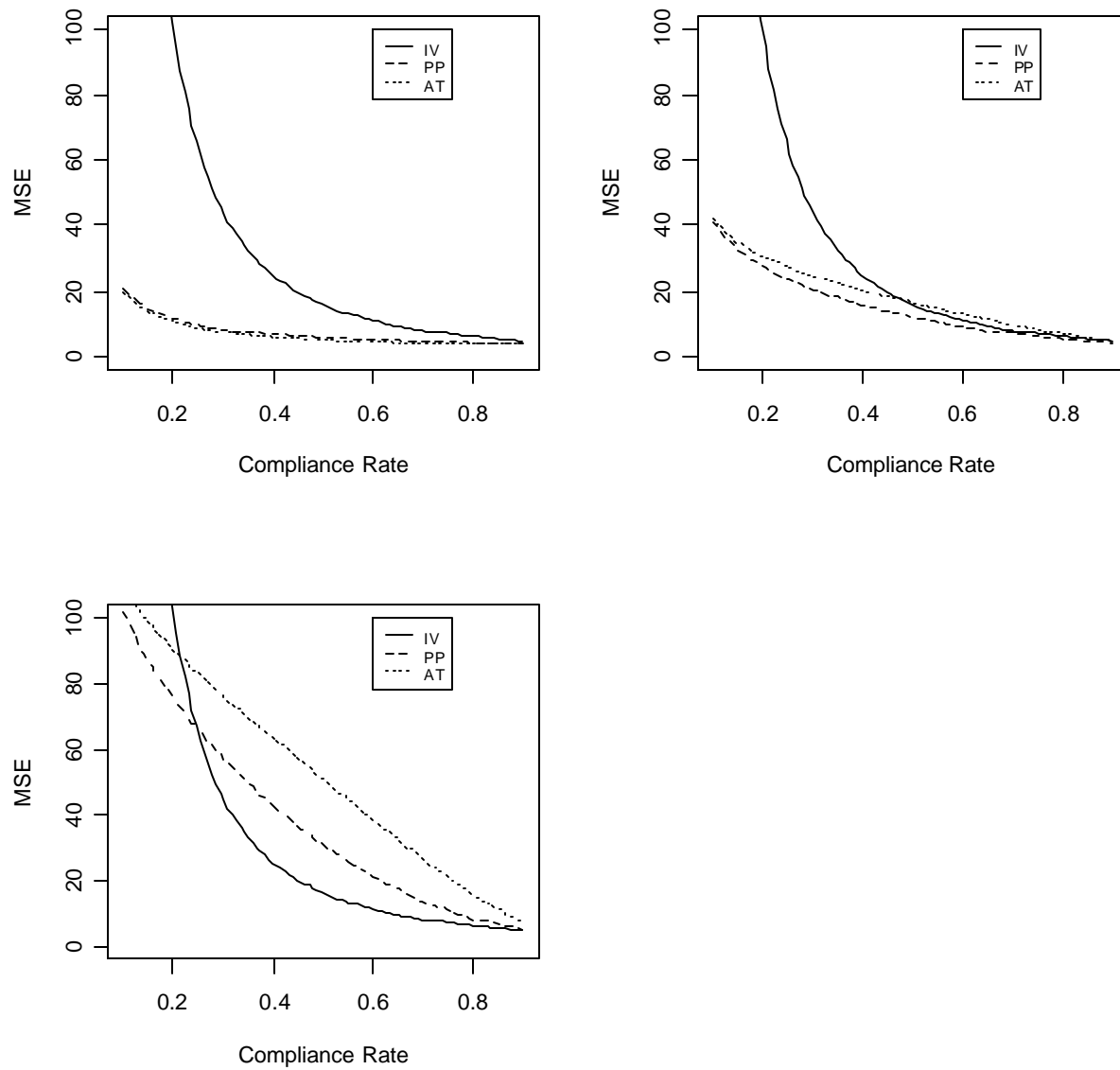


Figure 3. MSE of Estimates for varying compliance rate, b , when the population proportion in the treatment arm $a=0.522$, $n=575$ and treatment effect size $\sqrt{n}\Delta=0,5,10$ clockwise for three graphs, respectively.



Appendix: Large sample variances of IV, PP and AT

All the estimates are of the form $g(\bar{y}_{10}, \bar{y}_{11}, \bar{y}_{0+}, \hat{\mathbf{b}})$, where $(\bar{y}_{10}, \bar{y}_{11}, \bar{y}_{0+}, \hat{\mathbf{b}})$ are asymptotically independent with variances

$$Var(\bar{y}_{10}) = \frac{\mathbf{s}^2}{na(1-\mathbf{b})},$$

$$Var(\bar{y}_{11}) = \frac{\mathbf{s}^2}{na\mathbf{b}},$$

$$Var(\bar{y}_{0+}) = \frac{\mathbf{s}^2(1+\Delta^2\mathbf{b}(1-\mathbf{b}))}{n(1-\mathbf{a})},$$

$$Var(\hat{\mathbf{b}}) = \frac{\mathbf{b}(1-\mathbf{b})}{na}$$

$$\text{Hence } Var(g) = \left(\frac{\partial g}{\partial \bar{y}_{10}} \right)^2 Var(\bar{y}_{10}) + \left(\frac{\partial g}{\partial \bar{y}_{11}} \right)^2 Var(\bar{y}_{11}) + \left(\frac{\partial g}{\partial \bar{y}_{0+}} \right)^2 Var(\bar{y}_{0+}) + \left(\frac{\partial g}{\partial \mathbf{b}} \right)^2 Var(\hat{\mathbf{b}})$$

$$(A) \hat{\mathbf{d}}_{IV} = \bar{y}_{11} + \bar{y}_{10}(1-\hat{\mathbf{b}})/\hat{\mathbf{b}} - \bar{y}_{0+}/\hat{\mathbf{b}}$$

$$\left(\frac{\partial g}{\partial \bar{y}_{10}} \right) = \hat{\mathbf{b}}^{-1} - 1, \left(\frac{\partial g}{\partial \bar{y}_{11}} \right) = 1, \left(\frac{\partial g}{\partial \bar{y}_{0+}} \right) = -\hat{\mathbf{b}}^{-1}, \left(\frac{\partial g}{\partial \hat{\mathbf{b}}} \right) = \frac{\mathbf{m}_0 - \mathbf{m}_{0+}}{\hat{\mathbf{b}}^2} = \frac{\hat{\mathbf{b}}\Delta\mathbf{s}}{\hat{\mathbf{b}}^2}$$

$$Var(\hat{\mathbf{d}}_{IV}) = \frac{\mathbf{s}^2}{n} \left[\frac{(1-\mathbf{b})^2}{\mathbf{b}^2} \frac{1}{\mathbf{a}(1-\mathbf{b})} + \frac{1^2}{\mathbf{a}\mathbf{b}} + \frac{1+\Delta^2\mathbf{b}(1-\mathbf{b})}{\mathbf{b}^2(1-\mathbf{a})} + \frac{\Delta^2\mathbf{b}(1-\mathbf{b})}{\mathbf{b}^2\mathbf{a}} \right]$$

$$\text{So } = \frac{\mathbf{s}^2}{n\mathbf{b}^2\mathbf{a}(1-\mathbf{a})} [(1-\mathbf{b})(1-\mathbf{a}) + \mathbf{b}(1-\mathbf{a}) + \mathbf{a}] + \frac{\Delta^2\mathbf{s}^2(1-\mathbf{b})}{n\mathbf{b}\mathbf{a}(1-\mathbf{a})} [\mathbf{a} + (1-\mathbf{a})]$$

$$= \frac{\mathbf{s}^2}{n\mathbf{b}^2\mathbf{a}(1-\mathbf{a})} [1 + \Delta^2\mathbf{b}(1-\mathbf{b})]$$

$$(B) \hat{\mathbf{d}}_{PP} = \bar{y}_{11} - \bar{y}_{0+}$$

$$\left(\frac{\partial g}{\partial \bar{y}_{11}} \right) = 1, \left(\frac{\partial g}{\partial \bar{y}_{0+}} \right) = -1 \text{ So } Var(\hat{\mathbf{d}}_{PP}) = \frac{\mathbf{s}^2}{n} \left[\frac{1}{\mathbf{a}\mathbf{b}} + \frac{1+\Delta^2\mathbf{b}(1-\mathbf{b})}{(1-\mathbf{a})} \right]$$

$$(C) \hat{\mathbf{d}}_{AT} = \bar{y}_{11} - \frac{(1-\mathbf{a})\bar{y}_{0+} + \mathbf{a}(1-\hat{\mathbf{b}})\bar{y}_{10}}{1-\mathbf{a}\hat{\mathbf{b}}} = \bar{y}_{11} - \bar{y}_{0+} + \frac{\mathbf{a}(1-\hat{\mathbf{b}})(\bar{y}_{10} - \bar{y}_{0+})}{1-\mathbf{a}\hat{\mathbf{b}}}$$

$$\left(\frac{\partial g}{\partial \bar{y}_{10}} \right) = \frac{\mathbf{a}(1-\hat{\mathbf{b}})}{1-\mathbf{a}\hat{\mathbf{b}}}, \left(\frac{\partial g}{\partial \bar{y}_{11}} \right) = 1, \left(\frac{\partial g}{\partial \bar{y}_{0+}} \right) = \frac{1-\mathbf{a}}{1-\mathbf{a}\hat{\mathbf{b}}}, \left(\frac{\partial g}{\partial \hat{\mathbf{b}}} \right) = \frac{-\Delta \mathbf{b} \mathbf{s} \mathbf{a} (1-\mathbf{a})}{(1-\mathbf{a}\hat{\mathbf{b}})^2}$$

So

$$\begin{aligned} Var(\hat{\mathbf{d}}_{AT}) &= \frac{\mathbf{s}^2}{n} \left[\frac{\mathbf{a}^2(1-\hat{\mathbf{b}})^2}{(1-\mathbf{a}\hat{\mathbf{b}})^2} \frac{1}{\mathbf{a}(1-\mathbf{b})} + \frac{1}{\mathbf{a}\hat{\mathbf{b}}} + \frac{(1-\mathbf{a})^2}{(1-\mathbf{a}\hat{\mathbf{b}})^2} \frac{1+\Delta^2 \mathbf{b}(1-\mathbf{b})}{(1-\mathbf{a})} + \frac{\mathbf{b}^2 \Delta^2 \mathbf{a}^2 (1-\mathbf{a})^2}{(1-\mathbf{a}\hat{\mathbf{b}})^4} \frac{\mathbf{b}(1-\mathbf{b})}{\mathbf{a}} \right] \\ &= \frac{\mathbf{s}^2}{n} \left[\frac{\mathbf{a}(1-\hat{\mathbf{b}})}{(1-\mathbf{a}\hat{\mathbf{b}})^2} + \frac{1}{\mathbf{a}\hat{\mathbf{b}}} + \frac{(1-\mathbf{a})}{(1-\mathbf{a}\hat{\mathbf{b}})^2} \right] + \frac{\mathbf{s}^2 \Delta^2 \mathbf{b}(1-\mathbf{b})(1-\mathbf{a})}{n(1-\mathbf{a}\hat{\mathbf{b}})^4} \left[(1-\mathbf{a}\hat{\mathbf{b}})^2 + \mathbf{a}(1-\mathbf{a})\hat{\mathbf{b}}^2 \right] \\ &= \frac{\mathbf{s}^2}{n} \left[\frac{1}{(1-\mathbf{a}\hat{\mathbf{b}})} + \frac{1}{\mathbf{a}\hat{\mathbf{b}}} \right] + \frac{\mathbf{s}^2 \Delta^2 \mathbf{b}(1-\mathbf{b})(1-\mathbf{a})}{n(1-\mathbf{a}\hat{\mathbf{b}})^4} \left[1 - 2\mathbf{a}\hat{\mathbf{b}} + \mathbf{a}\hat{\mathbf{b}}^2 \right] \end{aligned}$$