Psychological Methods
2002, Vol. 7, No. 2, 178–193

1082-989X/02/$5.00 DOI: 10.1037//1082-989X.7.2.178

Statistical Power in Randomized Intervention Studies With Noncompliance

Booil Jo
University of California, Los Angeles

This study examined various factors that affect statistical power in randomized intervention studies with noncompliance. On the basis of Monte Carlo simulations, this study demonstrates how statistical power changes depending on compliance rate, study design, outcome distributions, and covariate information. It also examines how these factors influence power in different methods of estimating intervention effects. Intent-to-treat analysis and complier average causal effect estimation are compared as 2 alternative ways of estimating intervention effects under noncompliance. The results of this investigation provide practical implications in designing and evaluating intervention studies taking into account noncompliance.

In randomized field experiments, noncompliance (nonadherence) can be a major threat to obtaining statistical power to detect intervention effects. Noncompliance occurs when study participants do not follow the randomized assignment, and it has several forms (Angrist, Imbens, & Rubin, 1996). The most common form is represented by individuals who are assigned to the experimental group and fail to show up and receive the treatment (never-takers). Other, less frequent forms are represented by individuals assigned to the control group who nevertheless find ways to receive the treatment targeted for the experimental group (always-takers), or by individuals who do the opposite of what they are asked to do (defiers). Among these types, this study focuses on never-takers as the most common type of noncompliers.

In laboratory experimental studies, it is relatively easy to ensure compliance with a randomized assignment. Under these well-controlled circumstances, noncompliance is unlikely to occur because a small number of individuals are closely supervised. In nonlaboratory environments, some researchers limit noncompliance using on-site randomization. Researchers invite all recruited individuals to an initial meeting and then assign them to the treatment or the control condition on-site. The intervention (usually a single session) is given on-site right after assignment.

Although effective in reducing noncompliance rates, laboratory experiments and on-site randomization and treatment are not always possible in practice. First, these strategies are not easy to apply when managing a large number of study participants. Second, intervention trials often deliver an intensive treatment regime. Thus, administering the entire treatment on one occasion is not possible. Third, it is not easy to motivate individuals to show up at the intervention site without informing them whether they are getting treatment or not getting treatment. Although laboratory experiments and on-site randomization are ideal, they are often the exception rather than the rule in intervention studies.

Two studies that used randomization, but neither on-site nor under laboratory conditions, are discussed here. The Job Search Intervention Study (JOBS II; Little & Yau, 1998; Vinokur, Price, & Schul, 1995; Vinokur & Schul, 1997) was a randomized field ex-

This study was supported by National Institute on Alcohol Abuse and Alcoholism Grant K02 AA 00230 (Bengt Muthén, Principal Investigator), National Institute of Mental Health Grant P50 MH38725 (Philip Leaf, Principal Investigator), National Institute of Mental Health Grant P30 MH38330 (Richard Price, Principal Investigator), and National Institute on Drug Abuse Grant R01 DA11796 (Nicholas Ialongo, Principal Investigator).

I thank Bengt Muthén for suggesting the study of this topic and Hendricks Brown for valuable comments. I also thank Nicholas Ialongo and Amiram Vinokur for helpful advice and for providing data that were used as the prototype to formulate realistic settings in the simulation studies.

Correspondence concerning this article should be addressed to Booil Jo, Social Research Methodology Division, Graduate School of Education and Information Studies, University of California, Los Angeles, California 90095-1521. E-mail: booil@ucla.edu

periment intended to prevent poor mental health and to promote high-quality reemployment among unemployed workers. In JOBS II, neither laboratory experiment nor on-site randomization was an option because of intensive treatment protocols and the large scale of the study. Recruited individuals were randomized either to the control or to the intervention condition. Individuals assigned to the intervention group were invited to a series of training sessions. The problem of noncompliance arose in this situation because a substantial proportion of individuals who were assigned to the intervention condition did not show up to the intervention. For example, even though randomization was carried out among individuals who agreed to participate in the study, their motivation could have changed by the time of intervention. In addition, some individuals might have found a job quickly, leaving no need to attend the intervention sessions.

The Johns Hopkins University Preventive Intervention Research Center Study (JHU PIRC; Ialongo et al., 1999) was designed to improve academic achievement and to reduce early behavioral problems of schoolchildren. As in JOBS II, laboratory and on-site settings were not applicable in the JHU PIRC study because of the long duration and intensity of the intervention protocols. First-grade children and their parents were randomly assigned to the family-school partnership intervention condition or to the control condition. In the intervention condition, parents were asked to do a series of take-home activities related to literacy and mathematics. If parents did not carry out these activities, their children did not receive the intended intervention. In this setting, parents complied with the intervention by actually implementing the intervention. The advantage of this setting (self-administration) was that there was no need for parents to show up at the intervention site, which eliminated obstacles such as finding transportation to the site and being unable to choose a convenient time. The disadvantage of this setting was that parents had to sacrifice other activities and motivate themselves to administer the intervention in the absence of supervision. Therefore, despite the flexibility, parents could still be prone to noncompliance.

Noncompliance affects planning and evaluation of randomized intervention trials. In planning a new intervention trial, researchers need to consider the possibility of noncompliance to better estimate statistical power and sample size. For example, estimated sample size on the basis of results from a small laboratory experiment may not yield the expected statistical power when applied to a large-scale community trial, where noncompliance is more likely to occur. Recent developments in statistical methods provide options for researchers to estimate intervention effects with noncompliance. Depending on the purpose of the intervention, one can decide whether (a) to estimate the overall intervention effect for the entire sample or (b) to estimate the intervention effect for only those who comply with the treatment. Intent-to-treat (ITT) analysis is a widely accepted method for the first purpose (a), whereas complier average causal effect (CACE) estimation is a relatively new method that serves the second purpose (b). However, the way in which noncompliance affects statistical power has not been explicitly studied in either context.

On the basis of Monte Carlo simulations, this study examines how noncompliance affects statistical power in various settings of intervention trials. The degree of receipt of treatment can vary among study participants who actually receive the treatment (i.e., there could be partial compliers); however, in this article, discussion is limited to situations in which participants receive either all or none of the intervention treatment. On the basis of dichotomous treatment-receipt status, this study demonstrates how statistical power changes depending on compliance rate, study design (e.g., balanced or unbalanced), outcome distributions, and covariate information. This study also examined how these factors influence power in different methods of estimating intervention effects. ITT analysis and CACE estimation are compared as two alternative methods for estimating intervention effects under noncompliance. The results of this study provide useful answers to several questions that arise in intervention trials with noncompliance: particularly, how to estimate sample size and power, what affects power, how to improve power, and how to reduce cost.

Estimation of Intervention Effects With Noncompliance

The randomization procedure of assigning individuals to an experimental group and a control group is intended to create equivalent groups that can later be compared to determine the effects of treatment. However, if noncompliance occurs (i.e., study participants do not follow the randomized assignment), there is a serious risk that the groups may not be equivalent, and the comparison between them after the treatment may confound the effect of treatment with the initial

difference in their composition. For example, a significant proportion of those randomly assigned to the experimental group may not show up to receive the treatment because of lack of motivation. In this case, if the data are analyzed without including individuals who did not show up (per-protocol analysis), the resulting treatment-effect estimate may not represent the causal effect of the treatment because the experimental group would be composed of individuals who are more highly motivated than those in the control group. If the data are analyzed treating individuals who did not show up as a part of the control group (as-treated analysis), the resulting treatment-effect estimate will not represent the causal effect of the treatment because the experimental and the control groups lose comparability on the basis of random assignment.

In this article I focus on two methods of estimating the causal effects of treatment assignment (ITT analysis and CACE estimation) and do not discuss methods that may not yield causal-effect estimates (e.g., per-protocol analysis and as-treated analysis). Statistical assumptions required to make causal inferences are clearly addressed in Rubin's causal model (RCM; Holland, 1986; Rubin, 1974, 1978, 1980) approach. In RCM, the possibility of statistical causal inference is built on the basis of the effect of treatment at the individual level. The RCM begins with the consideration of an ideal case in which each individual could be separately given the treatment and the control condition at the same time. In this hypothetical situation, the equivalence between the two conditions is guaranteed in the absence of treatment, and each individual's potential outcome can be directly compared between the two conditions. However, in practice, one individual cannot be exposed to different treatment conditions at the same time (see Holland, 1986, for more discussion). Given the two assumptions below, unbiased estimation of the average causal effect can be achieved.

Assumption 1 (stable unit treatment value; SUTVA): Potential outcomes for each person are unrelated to the treatment status of other individuals in the sample (Rubin, 1978, 1980, 1990). For example, this assumption would be violated if individuals assigned to the control condition have contact with individuals assigned to the treatment condition and become demoralized by perceiving the treatment condition to be more attractive. West, Biesanz, and Pitts (2000) and Winship and Morgan (1999) provided illustrations of cases in which violations of SUTVA are likely to occur.

Assumption 2 (randomization): Treatment assignment is random. Random assignment enables field experimenters to approximate the equivalence between the treatment condition and the control condition prior to delivering the intervention. If randomization is successfully done, the average characteristics (pretreatment covariates) of individuals will be very similar between the treatment condition and the control condition, which provides the basis for fair comparison (i.e., causal inference).

Standard ITT Analysis

ITT analysis is a standard method used to estimate treatment effects in randomized experimental designs. In this method, average outcomes are compared without considering the aspect of receipt of treatment. Therefore, the ITT effect provided by this method can be thought of as the causal effect of treatment assigned but not necessarily the causal effect of treatment received. If the purpose of the trial is to estimate the overall causal effects of treatment assignment, ITT analysis is an appropriate method. For example, if public health researchers want to learn the overall effectiveness of a new vaccination program in reducing the outbreak of influenza in the community, both compliers and noncompliers should be included in the analysis.

ITT effect (i.e., the average causal effect of treatment assignment) can be defined as

$$\text{ITT} = \mu_1 - \mu_0, \tag{1}$$

where μ_1 denotes the mean potential outcome of the population when assigned to the treatment condition ($Z = 1$), and μ_0 denotes the mean potential outcome of the population when assigned to the control condition ($Z = 0$).

Under Assumptions 1 and 2, the unbiased estimate of the ITT effect is

$$\widehat{\text{ITT}} = \bar{y}_1 - \bar{y}_0, \tag{2}$$

where $\bar{y}_1$ is the sample-mean outcome of individuals assigned to the treatment group, and $\bar{y}_0$ is the sample-mean outcome of individuals assigned to the control group.

Estimation of CACE

The CACE estimation method considers both the assignment and the receipt of intervention treatment to estimate intervention efficacy. In this method, average outcomes are compared between the treatment and the control conditions, but only for compliers. Therefore, the CACE estimate can be thought of as the causal-effect estimate of treatment assignment for compliers. If the purpose of the trial is to estimate the effect of treatment when it is actually received, CACE

estimation is an appropriate method. For example, the purpose of the JOBS II and the JHU PIRC trials was to estimate the intervention effect when it was actually received rather than to estimate the overall effect. The CACE estimate is often of interest in medical research (e.g., clinical trials to test the effect of new drugs, procedures, and therapies).

Bloom (1984) introduced the idea of estimating CACE on the basis of an instrumental-variable (IV) approach, in which treatment effect estimates were adjusted by considering the noncompliance rate. More recently, a refined form of the IV approach based on RCM has been proposed (Angrist et al., 1996; Imbens & Angrist, 1994). The CACE estimation shown in this study is based on assumptions addressed by Angrist et al. (1996), which are described in Assumptions 1–5 in this article. Along with the RCM approach, likelihood-based estimation methods accelerated development in this area. Imbens and Rubin (1997) demonstrated CACE estimation through the maximum-likelihood (ML) estimation method using the expectation-maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977; Little & Rubin, 1987; McLachlan & Krishnan, 1997; Tanner, 1996), and a Bayesian approach using the data augmentation algorithm. Little and Yau (1998) incorporated covariates in this framework and applied the ML–EM method.

In line with RCM, individuals can be categorized not only on the basis of a potential treatment assignment, but also on the basis of a second indicator of the treatment receipt status. Assuming that the treatment has only two levels ($D = 1$, received; $D = 0$, not received), researchers can classify potential compliance types (C_i) of individuals into four categories on the basis of combinations of Z ($Z = 1$, assigned to the treatment condition; $Z = 0$, assigned to the control condition) and D. Angrist et al. (1996) labeled the four categories as complier, never-taker, defier, and always-taker. *Compliers* are individuals who would receive the treatment only if they are assigned to the treatment condition. *Never-takers* are individuals who would not receive the treatment even if they are assigned to the treatment condition. *Defiers* are individuals who would do the opposite of what they are assigned to do. *Always-takers* are individuals who would receive the treatment no matter which condition they are assigned to. Among these four potential compliance types, the current study considers only compliers and never-takers, which is the case when individuals assigned to the control condition do not have access to the treatment, as in the JOBS II and the JHU PIRC studies. The assumption of monotonicity (Imbens & Angrist, 1994) excludes the possibility of having defiers.

> Assumption 3 (monotonicity): There are no defiers. This assumption will hold if study participants are prohibited from receiving a different intervention treatment than the one they were assigned. Although defiers are usually considered the least likely type of noncompliers, violation of this assumption can have a serious impact on the CACE estimate.

The average causal effect of treatment assignment for compliers can be expressed in terms of potential outcomes as

$$\text{CACE} = \mu_{1c} - \mu_{0c}, \tag{3}$$

where μ_{1c} denotes the population-mean potential outcome for compliers when assigned to the treatment condition ($Z = 1$), and μ_{0c} denotes the population-mean potential outcome for compliers when assigned to the control condition ($Z = 0$). The problem with this definition, in practice, is that μ_{0c} cannot be directly estimated from the sample statistics. Because individuals assigned to the control condition are not allowed to receive the treatment, their potential compliance behavior cannot be observed. Only the overall mean μ_0 is estimable from the control condition. That is,

$$\mu_0 = \pi_c \, \mu_{0c} + \pi_n \, \mu_{0n}, \tag{4}$$

where μ_{0n} denotes the population-mean potential outcome for noncompliers when assigned to the control condition ($Z = 0$). π_c is the proportion of compliers in the population, and π_n is the proportion of noncompliers in the population ($\pi_c + \pi_n = 1$). Both π_c and π_n can be estimated directly from the sample statistics because the compliance rate can be observed in the treatment condition. On the basis of randomization, it can be assumed that the control condition has the same (potential) compliance rate.

Here, the exclusion restriction plays a critical role in identifying CACE models by assuming no effect of treatment assignment for noncompliers. Because this study assumes no always-takers, Assumption 4 applies only to never-takers.

> Assumption 4 (exclusion restriction): For never-takers and always-takers, the distributions of the potential outcomes are independent of treatment assignment (Angrist et al., 1996). In other words, never-takers and always-takers receive identical treatment regardless of which treatment condition they are assigned to.

The exclusion-restriction assumption can be violated if the treatment is not identical in different treatment conditions or if the treatment assignment has a psychological effect. For example, in the JHU PIRC study, if a parent (i.e., always-taker) assigned to the control condition somehow (e.g., from his or her friend assigned to the intervention condition) obtained a part of the home-learning activity materials, his or her child may have experienced only part of the intervention, which may not have been identical to getting the intervention when assigned to the intervention condition. In the JOBS II study, individuals who were assigned to the treatment condition but failed to attend the intervention seminars (never-takers) could be demoralized by failing to take the intervention opportunity. This negative psychological effect would not occur for never-takers assigned to the control condition, because the treatment was never offered.

If the exclusion-restriction assumption holds, the potential outcomes of noncompliers are the same regardless of treatment assignment (i.e., $\mu_{0n} = \mu_{1n}$). Therefore, Equation 4 can be rewirtten by replacing μ_{0n} with μ_{1n} as

$$\mu_0 = \pi_c\mu_{0c} + \pi_n\mu_{1n}, \tag{5}$$

where μ_{1n} denotes the population-mean potential outcome for noncompliers when assigned to the treatment condition ($Z = 1$).

Then, from Equation 5, μ_{0c} can be expressed using parameters that are estimable on the basis of sample statistics. That is,

$$\mu_{0c} = (\mu_0 - \pi_n\ \mu_{1n})/\pi_c. \tag{6}$$

Finally, Equation 3 can be rewritten by replacing μ_{0c} with the right-hand side of Equation 6. That is,

$$\text{CACE} = \mu_{1c} - (\mu_0 - \pi_n\mu_{1n})/\pi_c = (\mu_1 - \mu_0)/\pi_c. \tag{7}$$

In Equation 7, Assumption 5 excludes the possibility of a zero denominator (i.e., $\pi_c > 0$).

> Assumption 5 (nonzero-average causal effect of Z on D): The average causal effect of Z on D is not equal to zero (Angrist et al., 1996). In other words, the compliance rate cannot be zero. This assumption can be violated if none of the study participants receive the treatment even though it is offered.

Under Assumptions 1 and 5, the unbiased estimate of CACE is

$$\widehat{\text{CACE}} = (\bar{y}_1 - \bar{y}_0)/p_c \tag{8}$$

where $\bar{y}_1$ is the sample-mean outcome of individuals assigned to the treatment group, $\bar{y}_0$ is the sample-mean outcome of individuals assigned to the control group, and p_c is the sample proportion of compliers in the treatment group.

The current study uses an ML estimation approach, which is known to be often more efficient than the IV approach in estimating the CACE (Imbens & Rubin, 1997; Little & Yau, 1998). By maximizing the likelihood in Equation 9 with respect to the parameters of interest θ, ML estimates are obtained. The unknown compliance status is handled as missing data using the EM algorithm (Dempster et al., 1977; Little & Rubin, 1987; McLachlan & Krishnan, 1997; Tanner, 1996). ML-EM procedures for CACE estimation have been previously presented in Little and Yau (1998). In this study, ML-EM estimation of CACE was carried out by the Mplus program (Muthén & Muthén, 1998–2001). Parametric standard errors are computed from the information matrix of the ML estimator using both the first- and the second-order derivatives assuming conditional normality (i.e., conditional on compliance types and covariates). Mplus input codes used in this study can be obtained by contacting Booil Jo, or from the Mplus Web site, http://www.statmodel.com/mplus/examples/jo/.

Given that compliance type cannot be observed in the control condition, the observed-data likelihood function assuming a normally distributed outcome is

$$\begin{aligned} L(\theta|\text{data}) \propto & \prod_{i\in\{Z_i=1,D_i=0\}} \pi_n f(y_i|\mu_{1n}, \sigma_n^2) \\ & \times \prod_{i\in\{Z_i=1,D_i=1\}} \pi_c f(y_i|\mu_{1c}, \sigma_c^2) \\ & \times \prod_{i\in\{Z_i=0,D_i=0\}} [\pi_n f(y_i|\mu_{0n}, \sigma_n^2) \\ & + \pi_c f(y_i|\mu_{0c}, \sigma_c^2)], \end{aligned} \tag{9}$$

where $\theta = (\pi_n, \pi_c, \mu_{1n}, \mu_{1c}, \mu_{0n}, \mu_{0c}, \sigma_n^2, \sigma_c^2)$ is the set of parameters in the model, and $f(y_i|\mu, \sigma^2)$ denotes the probability density of a normal distribution with mean μ and variance σ^2. σ_n^2 denotes variance for never-takers, and σ_c^2 denotes variance for compliers.

If pretreatment covariates are available in the study, the magnitude and the precision of the CACE estimate can be adjusted depending on the influence of covariates on the outcome measure and compliance status.

In Equation 10, two dummy variables are used to represent compliers and noncompliers. That is, $c_i = 0$ and $n_i = 1$ for a noncomplier and $c_i = 1$ and $n_i = 0$ for a complier. A continuous outcome variable Y for

individual i with compliance status c_i and n_i can be expressed as

$$Y_i = \alpha_n n_i + \alpha_c c_i + \gamma_c c_i Z_i + \boldsymbol{\lambda}_n n_i \boldsymbol{x}_i + \boldsymbol{\lambda}_c c_i \boldsymbol{x}_i + \epsilon_{ni} + \epsilon_{ci}, \quad (10)$$

where $\boldsymbol{x}$ is a vector of covariates, α_n is an intercept for noncompliers, and α_c is an intercept for compliers. γ_c represents the average causal effect of treatment assignment for compliers (CACE). $\boldsymbol{\lambda}_n$ represents covariate effects on the outcome for noncompliers, and $\boldsymbol{\lambda}_c$ represents covariate effects on the outcome for compliers. ϵ_{ni} is a normally distributed residual with a zero mean and variance σ_n^2, and ϵ_{ci} is a normally distributed residual with a zero mean and variance σ_c^2.

The logistic regression of compliance on covariates is described as

$$\begin{aligned} P(c_i = 1|\boldsymbol{x}_i) &= \pi_{ci}, \\ P(c_i = 0|\boldsymbol{x}_i) &= 1 - \pi_{ci} = \pi_{ni}, \\ \text{logit}(\pi_{ci}) &= \beta_0 + \boldsymbol{\beta}_1 \boldsymbol{x}_i, \end{aligned} \quad (11)$$

where π_{ci} denotes the probability of being a complier, π_{ni} denotes the probability of being a noncomplier, β_0 represents a logit intercept, and $\boldsymbol{\beta}_1$ is a vector of logit coefficients, which represent the association between compliance and covariates.

Statistical Power and Noncompliance: Simulation Studies

Statistical power is a major element considered in planning intervention trials. In general, power to detect intervention effect decreases under noncompliance. However, the impact of noncompliance on power can be influenced by various factors such as outcome distributions, covariate information, and study design. Furthermore, noncompliance may affect power differently depending on the method used to estimate intervention effects. Focusing on the ITT analysis and CACE estimation methods, it is demonstrated in this section how statistical power changes in various settings of intervention studies.

Given the possible heterogeneity between compliers and noncompliers, it is unclear, in both the ITT analysis and CACE estimation methods, how to estimate sample size and statistical power by applying standard power calculation methods. For example, conventional power calculation methods (e.g., Cohen, 1988), which are based on ITT effect, cannot reflect change in power because of distributional (i.e., means and variances) differences between compliers and noncompliers. Noncompliance also poses a problem in applying standard power calculation methods in the context of CACE estimation. If a compliance type is completely observed, statistical power can be estimated for compliers assuming normality. Otherwise, CACE estimation involves subgroups of individuals with unknown group membership (i.e., finite mixtures: McLachlan & Peel, 2000; Titterington, Smith, & Makov, 1985), which excludes the possibility of applying standard power calculation methods.

The estimation of statistical power in this study is based on empirical results from Monte Carlo simulations (detailed information about the simulation procedures can be obtained by contacting Booil Jo). Although a number of variations can be considered for intervention settings with noncompliance, this study focuses on simple settings that have been frequently discussed in the recent literature. The assumptions required for CACE estimation (Assumptions 1–5) are satisfied in the simulation studies. It is also assumed that participants assigned to the control group do not have access to the intervention treatment and that participants assigned to the intervention group receive either all or none of the intervention treatment.

On the basis of ITT analysis and CACE estimation, the simulation studies demonstrate how statistical power is influenced by various aspects of intervention trials. First, the simulation studies show how the compliance rate affects power. Second, the studies show how the use of balanced or unbalanced design affects power. Third, the studies show how heterogeneity between compliers and noncompliers in outcome distributions affects power. Finally, the studies show how covariate information affects power.

Compliance Rate and Statistical Power

This section demonstrates how the rate of compliance affects the magnitude and precision of intervention-effect estimates and consequently affects the power to detect intervention effects. Noncompliance has a negative impact on power in both the ITT analysis and CACE estimation methods. However, loss in power because of noncompliance is driven by different mechanisms in ITT analysis and CACE estimation.

In ITT analysis, the magnitude of the intervention-effect estimate (effect size) is dependent on the rate of compliance. On the basis of Equations 1 and 7, the relationship between ITT effect and CACE can be described as

$$\text{ITT} = \pi_c \times \text{CACE}, \quad (12)$$

where π_c is the rate of compliance. This systematic relationship shows that the size of the ITT effect estimate will decrease as the rate of compliance decreases. For example, if the CACE estimate is −0.5, the ITT effect estimate will also be approximately −0.5 when there is no noncompliance. However, if the compliance rate is .3, the ITT effect estimate will be approximately −0.15 (0.3 × −0.5), which is only 30% of the magnitude of the intervention effect when there is no noncompliance.

In CACE estimation, precision of intervention-effect estimates is dependent on the rate of compliance. Low precision in the CACE estimate is partly due to subsampling of the entire sample and partly due to missing compliance information. CACE analysis focuses on estimating the intervention effect for compliers only, and the number of compliers decreases as compliance rate decreases, which results in a wider confidence interval for the CACE estimate. Missing compliance information also contributes to a wider confidence interval in the CACE estimate. Given that compliance status is not completely observed, precision of the CACE estimate depends on precision in estimating missing compliance information.

Simulation studies reported in Figure 1 show statistical power in the ITT analysis and CACE estimation methods for various combinations of effect size and compliance rate. To demonstrate the relationship between power and sample size, various sample sizes ranging from 200 to 600 are used. All simulation studies reported in this article are based on 1,000 replications. *Power* is defined as the proportion of replications out of 1,000 replications where the intervention-effect estimate is significantly different from zero ($\alpha = .05$). Equal probability of intervention–control assignment (i.e., balanced design) is assumed, and three conditions of compliance rate are considered (30%, 50%, and 70%).

Without covariates, the continuous outcome *Y,* assuming mixture distributions of compliers and noncompliers with compliance status c_i and n_i, can be described as

$$Y_i = \alpha_n n_i + \alpha_c c_i + \gamma_c c_i Z_i + \epsilon_{ni} + \epsilon_{ci}, \quad (13)$$

which is a simpler form of Equation 10.

In the simulation studies shown in Figure 1, the true control-group mean is 1.5 for both compliers (α_c) and noncompliers (α_n). The true residual variance is 1.0 for both compliers (σ_c^2) and noncompliers (σ_n^2). In line with Cohen (1988), three different conditions of effect size (low, medium, and large) are chosen. True intervention effects for compliers (γ_c = CACE) are −0.2, −0.5, and −0.8, which correspond to 0.2 (small), 0.5 (medium), and 0.8 (large) in terms of effect size. The negative sign of the causal effect reflects the typically desired outcome in intervention research: reduction of a problem behavior. The true noncomplier mean is 1.5 both in the intervention and control conditions assuming the exclusion restriction.

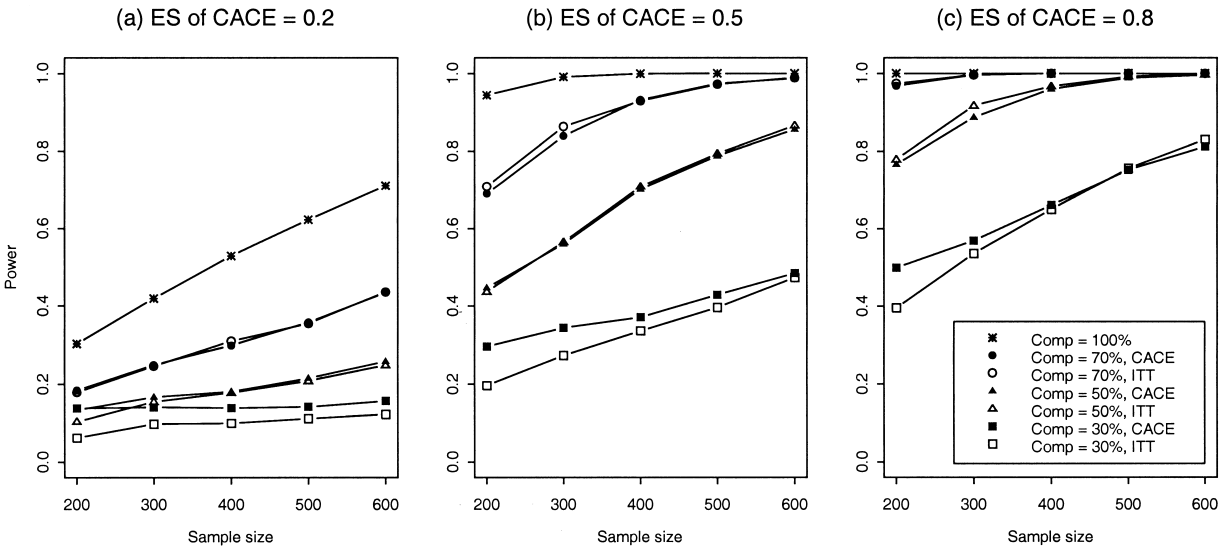

Figure 1. Noncompliance and power to detect intervention effects. The true complier average causal effect (CACE) is −0.2 in Panel a, −0.5 in Panel b, and −0.8 in Panel c, which corresponds to the effect size (ES) of 0.2, 0.5, and 0.8, respectively. No covariates are included. Comp = compliance; ITT = intent-to-treat analysis.

The model described in Equation 13 is used in generating data for the simulation studies, and the same model is used to estimate intervention effects in the CACE estimation method. Equation 14 shows the model assuming a single population, which is used to estimate intervention effects in the ITT analysis:

$$Y_i = \alpha + \gamma Z_i + \epsilon_i, \quad (14)$$

where α represents the overall mean outcome for the control group, and γ is the overall effect of intervention, ITT. ϵ_i is a normally distributed residual with a zero mean and variance σ^2.

Figure 1 illustrates that the compliance rate has a remarkable influence on statistical power to detect intervention effects both in the ITT analysis and CACE estimation methods. To maintain a desired level of statistical power under noncompliance, sample size adjustment is necessary, which usually increases the cost of intervention trials. For example, to reach a power of .8 with an effect size of 0.5, a sample size less than 200 is required if 100% of individuals comply with the intervention. If 50% of individuals comply with the intervention, a sample size greater than 500 is required to reach a power of .8. If the compliance rate is very low (30%), the power to detect an intervention effect is under .5 even with a sample size of 600. Simulation results in Figure 1 clearly illustrate the importance of having a higher rate of compliance.

Figure 1 also shows that ITT analysis and CACE estimation methods have little difference in power, despite the fact that the ITT effect and the CACE estimate have a substantial difference in effect size under noncompliance. If there are no other sources of information than effect size and compliance rate, ITT analysis and CACE estimation methods have little difference in power because there is little information that distinguishes compliers and noncompliers. For the same reason, standard power calculation methods (e.g., Cohen, 1988) and Monte Carlo simulations will provide similar results if compliance rate is taken into account. For example, if one expects a compliance rate of .5 and a complier effect size of 0.5, statistical power and sample size can be estimated when conventional methods are used with an effect size of 0.25 (see Equation 12) and perfect compliance. However, in practice, compliers and noncompliers can be heterogeneous in outcome distributions and auxiliary information, such as from covariates. Conventional power calculation methods do not reflect this heterogeneity, which influences power in both ITT analysis and CACE estimation. Therefore, estimated power and sample size, when conventional methods are used, may deviate from what will be obtained in a new intervention trial. Monte Carlo simulations can be used to better estimate power in the presence of auxiliary information that might reflect heterogeneity between compliers and noncompliers.

Study Design and Statistical Power

Although individuals' compliance behavior is hard to control, the ratio of the intervention and the control condition individuals can be controlled in the design stage. Depending on the situation, the total cost of an intervention study can be somewhat reduced by controlling this ratio. Simulation results reported in Figure 2 show the total cost of an intervention study to reach a power of .8 at various choices of this ratio. The proportion of individuals assigned to the intervention condition ranges from 20% (undersampling for the intervention condition) to 80% (oversampling for the intervention condition). Simulation results are based on a 50% compliance rate and an effect size of 0.5 (true CACE = −0.5). As in Figure 1, the model described in Equation 13 is used in generating data for

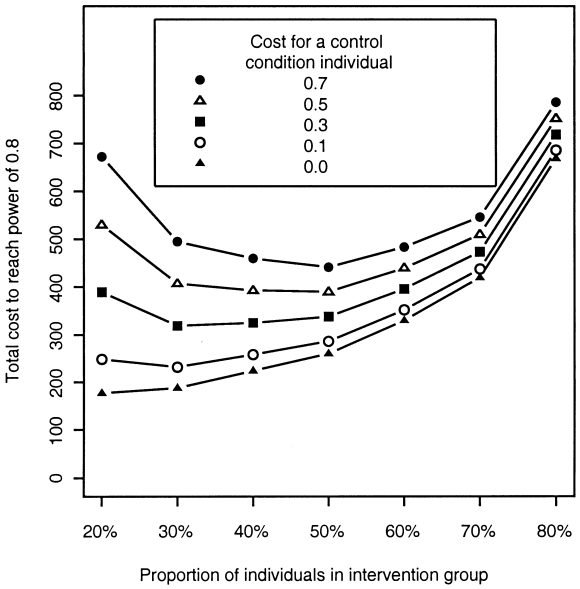

Figure 2. Design and cost of intervention studies when the cost of treating an individual in the intervention condition is the same regardless of compliance status. The results are based on complier average causal effect (CACE) estimation with a desired statistical power of .8. The true compliance rate is .5, and the true CACE is −0.5 (effect size = 0.5). The cost of treating one individual in the intervention condition is 1.0. No covariates are included.

the simulation study, and the same model is used to estimate intervention effects in CACE estimation. The true control group mean is 1.5 for both compliers (α_c) and noncompliers (α_n). The true residual variance is 1.0 for both compliers (σ_c^2) and noncompliers (σ_n^2). The results reported in Figure 2 are based on CACE estimation. Similar results can also be obtained when ITT analysis is applied (not shown).

The cost of treating each individual in the control condition usually differs from that in the intervention condition. The relationship between the total cost of an intervention trial and the ratio of the intervention and the control group individuals varies depending on the relative cost to treat these individuals. The simulation study shown in Figure 2 considers various costs of treating the control group individuals relative to the costs of treating the intervention group individuals. The total cost to reach a power of .8 is calculated with the assumption that cost to treat each complier in the intervention group is 1.0. The cost of treating each individual in the control group varies from 0.0 to 0.7. In some situations, managing an intervention group individual can be very costly. Therefore, the cost of treating each control group individual is relatively very low (e.g., 0.0). In other situations, the cost of treating each individual in the control group is not dramatically different from that in the intervention group (e.g., 0.7). In Figure 2, it is assumed that the cost of treating an intervention noncomplier is the same as the cost of treating an intervention complier. For example, if the main intervention activity is a small group seminar led by a psychologist and the cost for each psychologist is fixed, the cost of treating each seminar group will not be reduced even though some individuals assigned to that group do not show up for the seminar.

Figure 2 shows that a desired level of power may be obtained at a lower cost if a smaller number of individuals are assigned to the intervention group (unbalanced design), and if the cost of treating an individual assigned to the control group is much lower than the cost of treating an individual assigned to the intervention group. For example, if the cost of treating a control group individual is 0.1, the total cost when the balanced design is used is 286, and the total cost when the unbalanced design (intervention group proportion = 30%) is used is 232. Therefore, the total cost of the unbalanced design is about 19% less than that of the balanced design. If the cost of treating a control group individual is 0.0, the total cost when the balanced design is used is 260, and the total cost when the unbalanced design (intervention group proportion = 20%) is used is 177. In this case, the total cost of the unbalanced design is about 32% less than that of the balanced design.

Depending on intervention treatments, the cost of treating an individual who receives the intervention treatment can be different from the cost of treating an individual who fails or refuses to receive the intervention treatment. Figure 3 investigates the same set of conditions as in Figure 2 except that it is now assumed that the cost of treating an intervention noncomplier is the same as the cost of treating an individual in the control condition, where the cost is limited to collecting preintervention and follow-up information. For example, if a doctor gives an injection of medicine to participants in the intervention group, and because of noncompliance the unused medicine can be saved for later use or can be returned to the manufacturer, the cost for intervention treatment applies only to individuals who actually receive the treatment.

Figure 3 shows that a desirable level of power may be obtained at a lower cost if a smaller number of individuals are assigned to the intervention condition

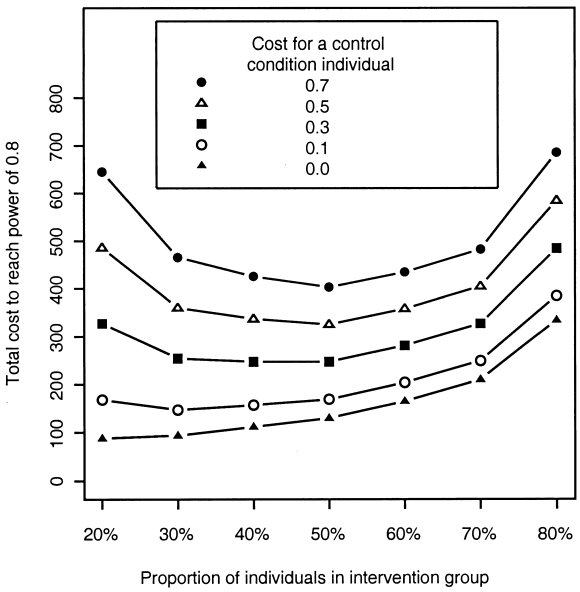

Figure 3. Design and cost of intervention studies when the cost of treating a noncomplier in the intervention condition is the same as the cost of treating a control condition individual. The results are based on complier average causal effect (CACE) estimation with a desired statistical power of .8. The true compliance rate is .5, and the true CACE is –0.5 (effect size = 0.5). The cost of treating one complier in the intervention condition is 1.0. No covariates are included.

and if the cost of treating a control group individual is much lower than the cost of treating an intervention complier. Because the total cost of an intervention is generally lower in this setting, the effect of undersampling intervention group individuals is not as dramatic as in the setting shown in Figure 2. For example, if the cost of treating a control group individual is 0.1, the total cost when the balanced design is used is 169, and the total cost when the unbalanced design (intervention group proportion = 30%) is used is 147. Therefore, the total cost of the unbalanced design is about 13% less (compared to 19% less in Figure 2) than that of the balanced design.

In estimating the cost of intervention trials, this study assumed that the cost for measuring compliance is trivial. However, measuring compliance can be costly in some situations, and the cost of intervention trials may be reduced by measuring compliance behavior for only subgroups of study participants (Frangakis & Baker, 2001). Another design possibility that could be considered is having a run-in period, before the main trial, to select potential compliers (Sheiner & Rubin, 1995). In this setting, only individuals selected as potential compliers are randomly assigned to the treatment conditions in the main trial. This method seeks to maximize the proportion of potential compliers. However, this method is not always applicable for practical and ethical reasons.

Outcome Distributions and Statistical Power

Simulation studies discussed in previous sections assumed a setting in which compliers and noncompliers have minimal difference in terms of outcome distributions (i.e., means and variances). However, individual differences between the complier and noncomplier groups may lead to different distributions even in the absence of treatment. If outcome distributions of compliers and noncompliers are different from those in the simulation setting used in Figure 1, the precision of the ITT effect and CACE estimates will not necessarily remain the same. In addition, the difference in power between the ITT effect and CACE estimates will change because differences between compliers and noncompliers in outcome distributions affect the precision of the ITT effect and CACE estimates differently. This section demonstrates that it is useful to examine how outcome distributions of compliers and noncompliers are different because this will give a better idea about the expected level of statistical power.

Figure 4 demonstrates how statistical power changes in ITT analysis and CACE estimation methods as the distance between mean outcomes of compliers and noncompliers changes. Simulation results reported in Figure 2 are based on a sample size of 300, an effect size of 0.5 (true CACE = −0.5) and a 50% compliance rate. As in Figure 1, the model described in Equation 13 is used in generating data for the simulation studies, and the same model is used to estimate intervention effects in the CACE estimation method. The model described in Equation 14 is used to estimate intervention effects in the ITT analysis. The true residual variance is 1.0 for both compliers and noncompliers. The true control group mean is 1.5 for compliers (α_c). The true noncomplier mean (α_n) has values ranging from 0.0–3.0, which results in the distance between means of compliers and noncompliers ranging from 0.0–1.5 in terms of the pooled standard deviation of 1.0.

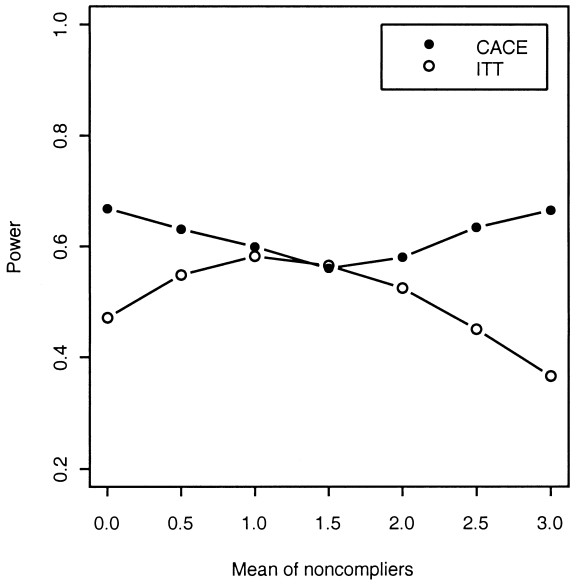

Figure 4. Mean difference and power to detect intervention effects. The true compliance rate is .5, the true complier average causal effect (CACE) is −0.5 (effect size = 0.5), and the sample size is 300. The true α_c is 1.5, and the true σ_c^2 is 1.0. The true α_n varies from 0.0 to 3.0. The true σ_n^2 is 1.0. No covariates are included. ITT = intent-to-treat analysis.

Figure 4 shows that the difference in power between ITT analysis and CACE estimation methods increases as the distance between mean outcomes of compliers and noncompliers increases. Precision of the CACE estimate benefits from the remoteness between two means, because compliance status can be estimated with a greater precision when compliers are

more distinguishable from noncompliers. In contrast, the remoteness between two means has a negative influence on precision in the ITT effect estimate because the overall variance estimate increases as the distance between means of compliers and noncompliers increases. The relationship between the mean difference and variance is illustrated in Figure 5. If compliers and noncompliers are allowed to have different means, the distance between the two means does not affect variance estimates of compliers and noncompliers. However, if a single distribution is assumed, the overall variance estimate increases as the distance between the two means increases.

Figure 6 demonstrates how statistical power changes in the ITT analysis and CACE estimation methods as the difference between variances of compliers and noncompliers changes. In this simulation setting, true means of compliers and noncompliers with a minimal distance are chosen to emphasize the influence of having different values of variances. The true control group mean is 1.5 for both compliers (α_c) and noncompliers (α_n). The true variance for compliers (σ_c^2) is 1.0. The true variance for noncompliers (σ_n^2) ranges from 0.25–2.00, which is one fourth to twice the size of the true complier variance.

Figure 6 shows that statistical power increases in the CACE estimation method as the size of noncomplier variance becomes smaller than the size of complier variance. When the noncomplier variance is smaller than the complier variance, the difference in size of variances results in a higher precision in estimating compliance, because compliers are more distinguishable from noncompliers in terms of variance. In addition, as the noncomplier variance becomes smaller, the mean outcome for noncompliers is estimated with a higher precision, which results in a higher precision in estimating the mean outcome for compliers and the intervention effect for compliers.

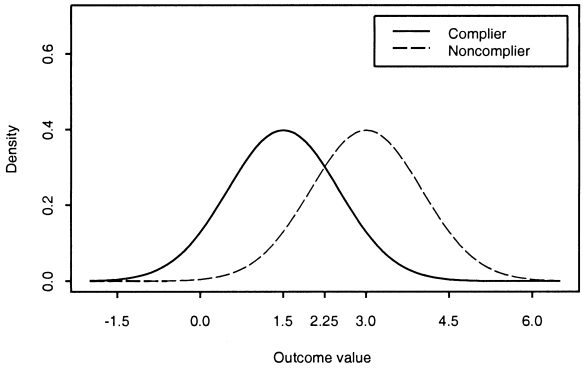

Figure 5. Distributions of compliers and noncompliers with different means.

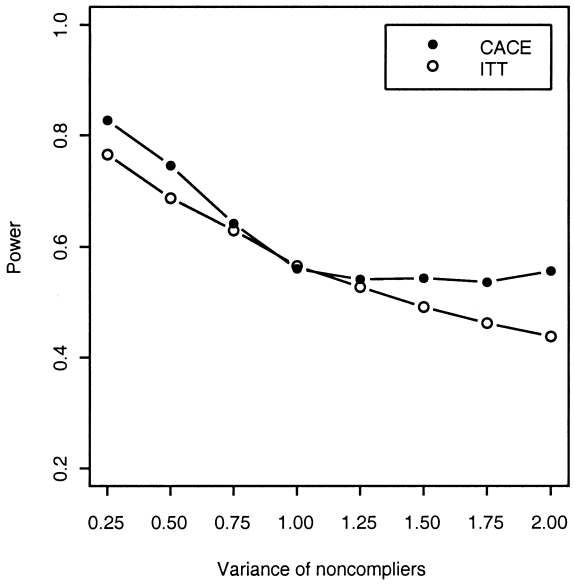

Figure 6. Variance difference and power to detect intervention effects. The true compliance rate is .5, the true complier average causal effect (CACE) is −0.5 (effect size = 0.5), and the sample size is 300. The true α_c is 1.5, and the true σ_n^2 is 1.0. The true α_n is 1.5, and the true σ_n^2 varies from 0.25 to 2.0. No covariates are included. ITT = intent-to-treat analysis.

Figure 6 also shows that statistical power remains stable in the CACE estimation method as the size of noncomplier variance becomes larger than the size of complier variance. When the noncomplier variance is larger than the complier variance, the difference in variances has a positive effect on the precision in estimating compliance. However, as the noncomplier variance becomes larger, the precision in estimating the noncomplier mean decreases, which results in a decreased precision in estimating the complier mean and the complier intervention effect. Because having a larger noncomplier variance relative to the complier variance has both positive and negative influences on precision, statistical power in CACE estimation remains stable. If the noncomplier variance is much larger than the complier variance (e.g., $\sigma_c^2 = 1.0$, $\sigma_n^2 = 4.0$), statistical power in CACE estimation increases (not shown) because having a larger noncomplier variance relative to the complier variance has a more positive than negative influence on precision.

The relationship between power and variance differences is more straightforward in ITT analysis. Figure 6 shows that the size of the noncomplier variance

has a linear effect on statistical power in ITT analysis, given that the true complier variance is invariant. Because the difference between compliers and noncompliers is not considered in ITT analysis, the overall variance estimate increases as the variance of noncompliers increases and decreases as the variance of noncompliers decreases. Therefore, precision of the ITT effect estimate also increases or decreases depending on the size of the overall variance estimate.

Covariate Information and Statistical Power

Although neither the compliance rate nor the outcome distribution is directly under the control of the investigator, the investigator can design the study to collect auxiliary information from pretreatment covariates. Pretreatment covariates contain two sources of information that may affect the precision of intervention effect estimates. One source of information is based on the association between covariates and the outcome measure. The other source of information is based on the association between covariates and compliance. Simulation results reported in Figures 7 and 8 show how this information affects statistical power in ITT analysis and CACE estimation. Simulation studies are based on a sample size of 300, an effect size of 0.5 (true CACE = −0.5) and a 50% compliance rate. The true control group mean is 1.5 for both compliers (α_c) and noncompliers (α_n). The true total variance (pooled across the control and intervention groups) is 1.0 for both compliers and noncompliers. For simplicity, one continuous covariate X is used in the simulation study, where X is normally distributed with a zero mean and the variance of one.

In Figure 7, regression coefficients λ_n and λ_c represent covariate effects on the outcome for noncompliers and compliers (see Equation 10). The true value for λ_n and λ_c is either 0.0 (ES = 0.0) or 0.6 (ES = 0.6). Depending on the λ_n and λ_c values, the true value for residual variances (σ_n^2 and σ_c^2) is either 1.0 or 0.64; therefore, the total variance is 1.0 for both compliers and noncompliers taking into account the variance because of covariate X. In other words, the correlation between Y and X is 0.0 or 0.6 in each compliance group. To emphasize the effect of a covariate associated with outcome, it is assumed that there is no association between the covariate and compliance. The model described in Equation 10 is also used as the model for CACE estimation. The model used for ITT analysis is described as

$$Y_i = \alpha + \gamma Z_i + \lambda X_i + \epsilon_i, \qquad (15)$$

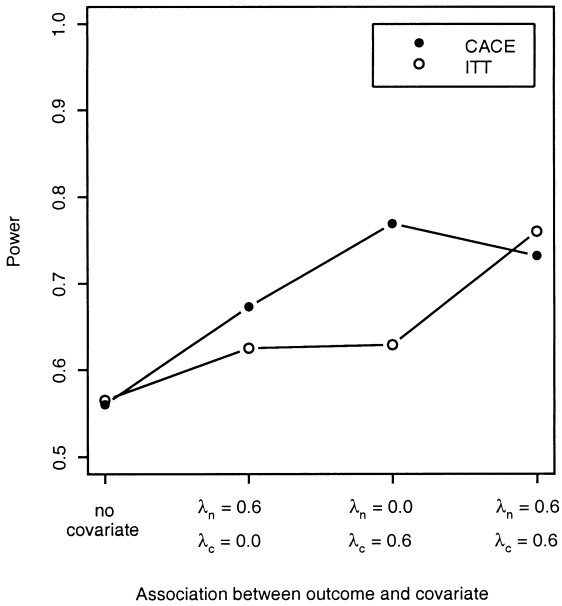

Figure 7. Statistical power and covariate effects on outcome. The true compliance rate is .5, the true complier average causal effect (CACE) is −0.5 (effect size = 0.5), and the sample size is 300. One continuous covariate (M = 0, variance = 1) is included. The regression coefficient (λ_n, λ_c) represents the level of association between outcome and the covariate. ITT = intent-to-treat analysis.

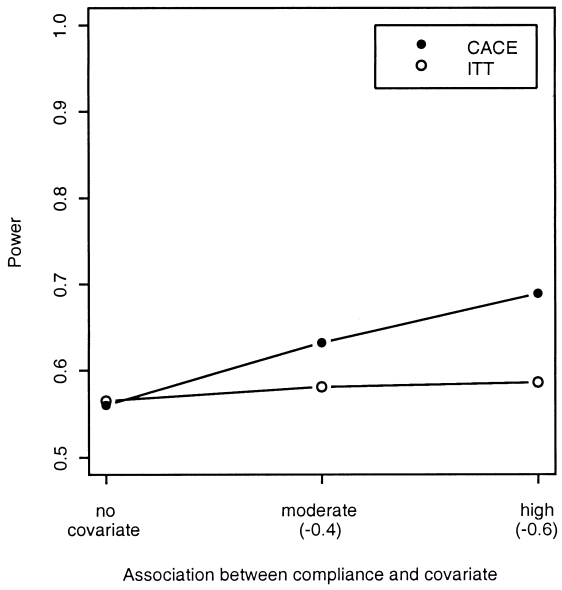

Figure 8. Statistical power and covariate effects on compliance. The true compliance rate is .5, the true complier average causal effect (CACE) is −0.5 (effect size = 0.5), and the sample size is 300. One continuous covariate (M = 0, variance = 1) is included. The biserial correlation between compliance and the covariate is presented in parentheses. ITT = intent-to-treat analysis.

where α is the overall mean outcome for the control group, γ is the overall effect of intervention (ITT effect), λ is the overall effect of covariate X on the outcome, and ϵ_i is a normally distributed residual with a zero mean and variance σ^2.

Figure 7 shows that a strong association between the covariate and outcome generally improves the power to detect intervention effects in ITT analysis and CACE estimation. It also shows that this association affects the precision of ITT effect and CACE estimates differently. In ITT analysis, where covariate effects on the outcome for compliers and noncompliers are not differentiated, precision of the intervention effect estimate benefits the most when the covariate has strong and similar effects on the outcome for compliers and noncompliers (i.e., $\lambda_n = \lambda_c = 0.6$). In CACE estimation, where covariate effects on the outcome for compliers and noncompliers can be differentiated, precision of the intervention effect estimate benefits the most when the covariate has a strong effect on the outcome only for compliers (i.e., $\lambda_n = 0$, $\lambda_c = 0.6$). Improvement in precision is less dramatic when the covariate has strong but similar effects on the outcome for compliers and noncompliers, because compliers and noncompliers are less distinguishable from each other on the basis of the covariate effects on the outcome.

Figure 8 demonstrates how the association between covariate and compliance affects power in ITT analysis and CACE estimation. To emphasize the effect of this association, it is assumed that there is no association between the covariate and outcome. The logistic regression of compliance on the covariate has been described in Equation 11. In the simulation study shown in Figure 8, three different situations are considered: when there is no covariate, when there is a moderate association between the covariate and compliance, and when there is a high association between the covariate and compliance. The true logit coefficients (β_1 in Equation 11) for moderate and high association are −0.69 and −1.20 respectively, which correspond to 0.50 and 0.30 in terms of the odds ratio (approximately −.40 and −.60 in terms of the biserial correlation). Equation 11 is a part of the CACE estimation model. However, this part is not included in the ITT analysis, where compliance information is not used in estimating intervention effects.

Figure 8 shows that a strong association between the covariate and compliance improves the power to detect intervention effects in CACE estimation. In ITT analysis, the association between the covariate and compliance has little impact on the precision of the intervention effect estimate. As a result, the difference in power between ITT analysis and CACE estimation methods increases as the level of association between compliance and covariate increases. As discussed previously in the context of Bayesian inference (Frangakis, Rubin, & Zhou, 1998, in press; Imbens & Rubin, 1997), covariates associated with compliance are unique sources of information that can be used in CACE estimation. Covariate information can often be collected from various sources such as background variables, initial status of the target outcome measure, and other relevant covariates that may reflect the characteristics of compliance behavior.

Summary

This study examines how noncompliance affects statistical power in the ITT analysis and CACE estimation methods. If there are no other sources of information than effect size and compliance rate, ITT analysis and CACE estimation methods have little difference in power, and standard power-calculation methods with adjusted effect size can be used as a crude way of estimating statistical power. In both methods, simulation studies showed that compliance rate has a critical impact on power. To maintain a desired level of statistical power under noncompliance, sample size adjustment is necessary, which usually increases the cost of intervention trials.

Researchers may reduce the cost of intervention trials by controlling the design of randomized trials, depending on where the major concern lies in terms of cost. Simulation studies showed that a desired level of power may be obtained at a lower cost if a smaller number of individuals are assigned to the intervention condition (unbalanced design), and if the cost of managing individuals in the control group is much lower than the cost of managing individuals in the intervention group.

Statistical power can be affected by outcome distributions of compliers and noncompliers. Further, outcome distributions influence power differently in ITT analysis and CACE estimation, which may cause a larger difference in power between the two methods. It was shown that the remoteness between the complier and noncomplier means has a negative effect on power in ITT analysis but has a positive effect on power in CACE estimation. Statistical power in ITT analysis increases as the overall variance decreases and decreases as the overall variance increases. In CACE estimation, heterogeneity between complier

and noncomplier variances has a positive influence on power, whereas the magnitude of these two variances has a negative influence on power.

Along with outcome distributions, pretreatment covariates are another source of information that potentially increases the difference in power between ITT analysis and CACE estimation. Unlike outcome distributions, collecting covariate information is relatively controllable in intervention studies. Including covariates that are highly associated with outcome may increase power in both methods. Precision in ITT analysis benefits the most when covariates have strong and equal effects on the outcome. Precision in CACE estimation benefits the most if covariate effects are stronger among compliers than among noncompliers. Covariates associated with compliance provide unique sources of information that may improve power in CACE estimation. Although this study assumed constant covariate effects, treatment effects may vary depending on covariate values. To avoid bias in CACE estimation, interaction effects should be properly specified. However, including unnecessary interaction parameters may reduce gains in precision due to covariate information.

Limitations

On the basis of Monte Carlo simulations, this study examined several factors that affect statistical power in intervention studies with noncompliance. The study was intended to help researchers to focus on controllable factors that may improve power in future intervention studies, and to evaluate completed studies with fuller understanding. However, this study did not cover many other factors that might affect power depending on specific settings. For example, randomized trials may face not only noncompliance but also other complications, such as clustering and attrition.

In intervention trials, the unit of intervention or the unit of randomization is often a cluster (group) of individuals. In the JOBS II study, job search seminars were given to groups of participants. In this setting, individuals belonging to the same group interacted with each other and shared common features of the treatment. In the JHU PIRC study, the unit of randomization was a classroom. Therefore, students belonging to the same classroom naturally shared the same classroom environment, including the teacher. The resemblance among individuals, due to clustering, needs to be taken into account to avoid overestimating statistical power. Adjusting statistical power is further complicated if clustering is accompanied by noncompliance (Jo, Muthén, Ialongo, & Brown, 2002). This study did not include this situation, which is quite common in intervention trials.

Nonresponse (e.g., attrition, dropout) is another common complication that was not considered in this study. Intervention trials often suffer from both noncompliance and missingness in outcome because of nonresponse. For example, in JOBS II, a substantial number of individuals did not respond to follow-up surveys, and nonresponse rates were different for compliers and noncompliers. Yau and Little (2001) demonstrated the simultaneous modeling of noncompliance and nonresponse using the ML-EM and Bayesian estimation methods, assuming no correlation between noncompliance and nonresponse (i.e., missing at random; Little & Rubin, 1987). Frangakis and Rubin (1999) demonstrated the simultaneous modeling of noncompliance and nonresponse using Bayesian estimation methods, allowing for possible correlation between noncompliance and nonresponse (i.e., missing not at random; Little & Rubin, 1987). Statistical power will be affected not only depending on whether this missingness is taken into account, but also depending on what is assumed for the relationship between noncompliance and nonresponse.

Researchers also need to be aware that the results of CACE estimation shown in this study are based on the five statistical assumptions originally outlined by Angrist et al. (1996). Violating these statistical assumptions has a critical impact on statistical power and needs to be carefully examined. Although violating any of these assumptions will also affect ITT analysis, consequences of violation are more critical in CACE estimation because it claims to target compliers. If compliers and noncompliers are well separated without violating any of these assumptions, CACE estimation will provide an unbiased treatment effect estimate and statistical power for compliers. However, if underlying assumptions are violated, the CACE estimate's validity and power will be challenged.

The exclusion restriction is one of the critical underlying assumptions in the estimation of CACE. The assumption provides the basis for identifiability in CACE models, given that compliance status is not observed completely. However, this assumption can be often unrealistic in practice (Hirano, Imbens, Rubin, & Zhou, 2000; Jo, in press-a, in press-b; West & Sagarin, 2000). For example, in JOBS II, noncompliers in the treatment condition could have been demoralized by failing to take the intervention opportunity.

This negative psychological effect did not occur for never-takers in the control condition because the treatment was not offered. In this situation, the CACE estimate can be understated by ignoring the effect of treatment assignment on noncompliers. As a result, statistical power could be underestimated if the exclusion restriction is assumed.

This study also assumed that participants receive either all or none of the intervention treatment. However, the generalizability of this setting is often limited in reality. Intervention trials often include several treatment sessions or doses, and study participants may choose different levels of exposure. For example, in the JHU PIRC family-school partnership intervention, parents could choose to implement 0 to 66 activities. Previous research (Angrist & Imbens, 1995; Efron & Feldman, 1991; Goetghebeur & Molenberghs, 1996; Holland, 1988) has shown some possibilities for estimating treatment effects taking into account the varying intensity of compliance (or dose). However, applicability of these methods in psychosocial intervention studies has not been fully studied.

This study did not deal with situations involving interaction effects. For example, effects of covariates may vary depending on the treatment assignment condition. Given underlying assumptions such as the exclusion restriction, CACE models do not require constant effect of treatment to remain identifiable. In other words, both the main and the interaction effect can be estimated for compliers. If interaction effects are not specified in the model, the resulting CACE estimate should be interpreted carefully, given possible confounding with interaction effects. To avoid bias in CACE estimation, interaction effects should be properly specified. However, including unnecessary interaction parameters may reduce gains in precision due to covariate information. To minimize a loss in power, CACE models need to be structured in a more parsimonious way by focusing only on plausible interactions. This study investigated covariate effects on statistical power assuming a linear relationship between outcome and covariate and between compliance and covariate. More research is needed to clarify the potential complications in estimating intervention effects when covariates have a nonlinear relationship with outcome and compliance.

References

Angrist, J. D., & Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association, 90,* 431–442.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association, 91,* 444–455.

Bloom, H. S. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation Review, 8,* 225–246.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Dempster, A., Laird, N., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, 39,* 1–38.

Efron, B., & Feldman, D. (1991). Compliance as an explanatory variable in clinical trials. *Journal of the American Statistical Association, 86,* 9–17.

Frangakis, C. E., & Baker, D. B. (2001). Compliance subsampling designs for comparative research: Estimation and optimal planning. *Biometrics, 57,* 889–907.

Frangakis, C. E., & Rubin, D. B. (1999). Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika, 86,* 365–379.

Frangakis, C. E., Rubin, D. B., & Zhou, X. H. (1998). Analyzing clustered encouragement design: The central role of modeling. In *American Statistical Association, Proceedings of the Biometrics Section* (pp. 71–79). Alexandria, VA: American Statistical Association.

Frangakis, C. E., Rubin, D. B., & Zhou, X. H. (in press). Clustered encouragement design with individual noncompliance: Bayesian inference and application to advance directive forms. *Biostatistics.*

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.

Hirano, K., Imbens, G. W., Rubin, D. B., & Zhou, X. H. (2000). Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics, 1,* 69–88.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81,* 945–970.

Holland, P. W. (1988). Causal inference, path analysis, and recursive structural equation models. *Sociological Methodology, 18,* 449–484.

Ialongo, N. S., Werthamer, L., Kellam, S. G., Brown, C. H., Wang, S., & Lin, Y. (1999). Proximal impact of two first-grade preventive interventions on the early risk behaviors for later substance abuse, depression and antiso-

cial behavior. *American Journal of Community Psychology, 27,* 599–642.

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

Imbens, G. W., & Rubin, D. B. (1997). Bayesian inference for causal effects in randomized experiments with noncompliance. *Annals of Statistics, 25,* 305–327.

Jo, B. (in press-a). Estimating intervention effects with noncompliance: Alternative model specifications. *Journal of Educational and Behavioral Statistics.*

Jo, B. (in press-b). Model misspecification sensitivity analysis in estimating causal effects of interventions with noncompliance. *Statistics in Medicine.*

Jo, B., Muthén, B. O., Ialongo, N. S., & Brown, C. H. (2002). *Cluster randomized trials: ML-EM estimation of treatment effects with noncompliance.* Unpublished manuscript.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: Wiley.

Little, R. J. A., & Yau, L. H. Y. (1998). Statistical techniques for analyzing data from prevention trials: Treatment of no-shows using Rubin's causal model. *Psychological Methods, 3,* 147–159.

McLachlan, G. J., & Krishnan, T. (1997). *The EM algorithm and extensions.* New York: Wiley.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models.* New York: Wiley.

Muthén, L. K., & Muthén, B. O. (1998–2001). *Mplus user's guide.* Los Angeles: Muthén & Muthén.

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

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics, 6,* 34–58.

Rubin, D. B. (1980). Discussion of "Randomization analysis of experimental data in the Fisher randomization test" by D. Basu. *Journal of the American Statistical Association, 75,* 591–593.

Rubin, D. B. (1990). Comment on "Neyman (1923) and causal inference in experiments and observational studies." *Statistical Science, 5,* 472–480.

Sheiner, L. B., & Rubin, D. B. (1995). Intention-to-treat analysis and the goals of clinical trials. *Clinical Pharmacology and Therapeutics, 57,* 6–15.

Tanner, M. (1996). *Tools for statistical inference: Methods for the exploration of posterior distributions and likelihood functions.* New York: Springer.

Titterington, D. M., Smith, A. F. M., & Makov, U. E. (1985). *Statistical analysis of finite mixture distributions.* New York: Wiley.

Vinokur, A. D., Price, R. H., & Schul, Y. (1995). Impact of the JOBS intervention on unemployed workers varying in risk for depression. *American Journal of Community Psychology, 23,* 39–74.

Vinokur, A. D., & Schul, Y. (1997). Mastery and inoculation against setbacks as active ingredients in the JOBS intervention for the unemployed. *Journal of Consulting and Clinical Psychology, 65,* 867–877.

West, S. G., Biesanz, J. C., & Pitts, S. C. (2000). Causal inference and generalization in field settings: Experimental and quasi-experimental designs. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 40–84). New York: Cambridge University Press.

West, S. G., & Sagarin, B. J. (2000). Participant selection and loss in randomized experiments. In L. Bickman (Ed.), *Research design* (pp. 117–154). Thousand Oaks, CA: Sage.

Winship, C., & Morgan, S. L. (1999). The estimation of causal effects from observational data. *Annual review of sociology* (Vol. 25, pp. 659–707). Palo Alto, CA: Annual Reviews.

Yau, L. H. Y., & Little, R. J. A. (2001). Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed. *Journal of the American Statistical Association, 96,* 1232–1244.

Received December 14, 2000
Revision received January 29, 2002
Accepted January 29, 2002 ■