Published in final edited form as:

Psychol Methods. 2008 March ; 13(1): 1–18. doi:10.1037/1082-989X.13.1.1.

Cluster Randomized Trials With Treatment Noncompliance

Booil Jo,

Department of Psychiatry and Behavioral Sciences, Stanford University

Tihomir Asparouhov,

Muthén & Muthén, Los Angeles, California

Bengt O. Muthén,

Graduate School of Education and Information Studies, University of California, Los Angeles

Nicholas S. Ialongo, and

Department of Mental Health, Johns Hopkins University

C. Hendricks Brown

Department of Epidemiology and Biostatistics, University of South Florida

Abstract

Cluster randomized trials (CRTs) have been widely used in field experiments treating a cluster of individuals as the unit of randomization. This study focused particularly on situations where CRTs are accompanied by a common complication, namely, treatment noncompliance or, more generally, intervention nonadherence. In CRTs, compliance may be related not only to individual characteristics but also to the environment of clusters individuals belong to. Therefore, analyses ignoring the connection between compliance and clustering may not provide valid results. Although randomized field experiments often suffer from both noncompliance and clustering of the data, these features have been studied as separate rather than concurrent problems. On the basis of Monte Carlo simulations, this study demonstrated how clustering and noncompliance may affect statistical inferences and how these two complications can be accounted for simultaneously. In particular, the effect of the intervention on individuals who not only were assigned to active intervention but also abided by this intervention assignment (complier average causal effect) was the focus. For estimation of intervention effects considering noncompliance and data clustering, an ML-EM estimation method was employed.

Keywords

cluster randomized trial; compliance intraclass correlation; complier average causal effect; ML-EM estimation; treatment noncompliance

---

Individual-level randomization is not always possible in field experiments for practical or ethical reasons. Therefore, instead of standard randomized controlled trials, cluster randomized trials (CRTs) have been widely used in practice, treating a cluster (or group) of individuals as the unit of randomization. For example, in primary care settings, the randomization unit is often a doctor or a clinic (e.g., Dexter et al., 1998), where a number of patients belong to each doctor. In school settings, the randomization unit is often a

---

Copyright 2008 by the American Psychological Association

Correspondence concerning this article should be addressed to Booil Jo, Interdisciplinary Brain Science Research, SMC 5795, 401 Quarry Road, Stanford, CA 94305. booil@stanford.edu.

Supplemental materials: <http://dx.doi.org/10.1037/1082-989X.13.1.1.supp>

classroom, where a number of students belong to each teacher (e.g., Ialongo et al., 1999). In CRT settings, individuals belonging to the same cluster are likely to resemble one another because of various factors such as common environment (e.g., clinic, classroom) and common deliverer of intervention treatment (e.g., doctor, teacher). If resemblance in outcomes among individuals in each cluster is ignored in analyzing outcomes in CRTs, standard errors are usually underestimated, which results in inflation of a Type I error. That is, one may overestimate the significance of treatment effects or falsely conclude that treatment effects are significant when they are not. Previous studies have shown how design and analysis strategies need to be adjusted for efficient and fair evaluation of treatment effects in CRTs (e.g., Donner & Klar, 1996; Murray, 1998; Raudenbush, 1997; Brown & Liao, 1999). Multilevel analysis techniques developed in various statistical frameworks (e.g., Aitkin & Longford, 1986; Goldstein, 1986; Longford, 1993; McCulloch, 1997; B. O. Muthén & Satorra, 1995; Raudenbush & Bryk, 2002; Raudenbush, Yang, & Yosef, 2000; Stratelli, Laird, & Ware, 1984; Wong & Mason, 1985; Zeger & Karim, 1991) provide relevant tools for analyzing data accounting for nested data structures.

Another common complication in randomized studies is individual variation in exposure to the intervention. When the intervention is delivered to a group or cluster, exposure can vary because individuals can move in or out of the group during the intervention period. For some interventions, subjects who are targets of the intervention may elect not to attend some of group intervention sessions, thereby receiving only a portion of the intervention. In this article, we refer to all these issues as *treatment noncompliance*, but we recognize that other terms, such as *intervention adherence* or *participation*, may more appropriately describe the interactive processes involved in delivery and reception of the intervention (Snyder et al., 2006). Unlike the situation for laboratory experiments or on-site randomized studies, compliance behavior is often hard to control in larger scale field experiments with intensive treatment regimes. Study participants' compliance behavior can be associated with various factors, such as background characteristics and motivation. For example, people who are highly motivated or have a special interest in the treatment will be more likely to comply with the treatment. In estimating intervention effects in this situation, intention-to-treat (ITT) analysis, where randomized groups are compared regardless of compliance status, has been considered a gold standard. As secondary analyses to complement ITT analysis results, as-treated or per-protocol analyses have been widely used if the effect of treatments when actually received is of interest. However, these secondary analysis methods may yield seriously biased estimates of treatment effects. Given the questionable validity of these analyses, statistical methods such as CACE (complier average causal effect) estimation have been developed to better estimate treatment efficacy taking into account noncompliance (e.g., Angrist, Imbens, & Rubin, 1996; Bloom, 1984; Frangakis & Rubin, 2002; Goetghebeur & Molenberghs, 1996; Hirano, Imbens, Rubin, & Zhou, 2000; Imbens & Rubin, 1997; Jo, 2002a; Little & Yau, 1998).

Further statistical challenges arise when CRTs are accompanied by noncompliance. What is interesting with compliance behavior in the CRT setting is that it may be influenced not only by individual characteristics but also by characteristics of the cluster an individual belongs to. For example, in the Johns Hopkins University Preventive Intervention Research Center (JHU PIRC) school intervention trial (Ialongo et al., 1999), which was used as a prototype for the Monte Carlo simulations reported in this study, the unit of randomization was a classroom. In particular, in the Family-School Partnership (FSP) intervention, poor compliance of parents was one of the major complications. Furthermore, parents' compliance with the intervention activities substantially varied depending on the classroom their children belonged to. When compliance was defined as completion of at least two thirds of intervention activities, average compliance rate in each classroom ranged from 5% to 100%. One possible explanation for this phenomenon would be that some teachers

communicated better with parents and/or were more eager to encourage parents. Another possibility is that children in some classrooms were more ready and willing to collaborate with their parents because of their teachers' characteristics and/or other classroom environment (e.g., proportion of aggressive children). Resemblance in compliance behavior in the same cluster unit is a unique, but common, problem that did not receive proper attention until Frangakis, Rubin, and Zhou (2002) demonstrated the possibility of estimating intervention effects accounting for both noncompliance and clustering within a Bayesian analysis framework. However, there is still little recognition among researchers that these two complications are often closely related, creating unique problems that cannot be handled without concurrent consideration of clustering and noncompliance.

This article intends to promote an understanding of data clustering and noncompliance as simultaneous complications and to propose analyses that address both problems. In particular, the effect of intervention treatments on individuals who would abide by the intervention assignment (CACE) is the focus. The difference from the previous study (Frangakis et al., 2002) is that the current study puts its main emphasis on understanding the mechanisms in which simultaneous complications affect the quality of causal effect estimates and how serious the impact is. Furthermore, this study employs a maximum-likelihood (ML) estimation method (built in Mplus Version 4 and higher; L. K. Muthén & Muthén, 1998–2007), which is more accessible to many applied researchers. This article is presented in the following order. First, the CACE estimation method is briefly reviewed. Second, analytical complications from the simultaneous presence of noncompliance and data clustering are defined by extending the conventional notion of intraclass correlation (ICC). Third, various consequences of these complications are demonstrated using Monte Carlo simulation studies. Fourth, a joint modeling of clustered data and compliance within a multilevel mixture analysis framework is presented. Finally, implications and limitations of the study are discussed.

Complier Average Causal Effect

In randomized field experiments, it is not rare to have a situation where a substantial number of study participants fail to receive the assigned treatment. In this case, if the effect of treatment when actually received (i.e., efficacy) is of interest, both as-treated analysis, which considers treatment receipt but ignores treatment assignment status, and per-protocol analysis, which excludes individuals who fail to adequately abide by the assigned treatment protocol, have been commonly applied as ways of adjusting for noncompliance. However, the results of these analyses are not only hard to interpret as causal effects but also subject to substantial bias (Robins & Greenland, 1994; Sheiner & Rubin, 1995). This study employs the CACE estimation method, which is considered a better alternative to as-treated or per-protocol analysis. The key advantage of the CACE estimation method is that causal effect is defined on the basis of individuals' potential outcomes under every treatment assignment status (Angrist et al., 1996; Frangakis & Rubin, 2002). Given that, causal interpretation is possible with clarified assumptions, and sensitivity analysis is feasible if these assumptions are likely to be violated.

In the potential outcomes approach, identification of causal effects (including ITT effect) is built on the basis of two critical assumptions. First, it is assumed that treatment assignment is random. Second, it is assumed that each individual's potential outcomes are uncorrelated with other individuals' treatment assignment status, which has been labeled as stable unit treatment value (SUTVA; Rubin, 1978, 1980, 1990). Additional assumptions that are commonly used to identify CACE were described in Angrist et al. (1996), in line with the potential outcomes approach. Assuming binary treatment assignment (treatment or control) and binary treatment receipt (receive or not) status, they defined four potential compliance

types. Compliers are individuals who receive treatment only if they are assigned to the treatment condition. Never-takers are individuals who do not receive the treatment even if they are assigned to the treatment condition. Defiers are individuals who do the opposite of what they are assigned to do. Always-takers are individuals who always receive the treatment no matter which condition they are assigned to. Among these four potential types of individuals, emphasis is often given to compliers, and the rest are considered noncompliers. The assumption of monotonicity (Imbens & Angrist, 1994) excludes the possibility of having defiers, which is crucial in identifying causal effects of treatment assignment for compliers. The current study considers the two most common compliance types, compliers and never-takers, which is the case when individuals assigned to the control condition do not have access to the treatment (e.g., JHU PIRC trial; Ialongo et al., 1999). Since there is only one type of noncomplier (i.e., never-takers), noncomplier is used in this article to refer to never-taker. This setting satisfies the assumption of monotonicity.

Figure 1 illustrates the considered randomized trial setting, where individuals are randomly assigned either to the treatment ($Z = 1$) or to the control ($Z = 0$) condition. In this setting, individuals assigned to the treatment condition choose to (or not to) receive treatment, whereas individuals assigned to the control condition do not have this choice (i.e., no one receives treatment). Therefore, in the treatment condition, it is possible to identify each individual's compliance type (C_i) on the basis of his or her observed treatment receipt status. That is, a person is a complier ($C_i = c$) if she or he receives treatment and a noncomplier ($C_i = n$) if she or he does not. However, in the control condition, we cannot distinguish compliers from noncompliers on the basis of observed treatment receipt status, leaving each individual's compliance type unknown (or latent). If compliance type is observed in both the treatment and control conditions, the causal effect of treatment can be estimated for any potential types of individuals defined by Angrist et al. (1996). However, in practice, compliance type is not observed completely, as in the randomized trial setting we consider (where compliance type is unobserved among individuals assigned to the control condition), which complicates the estimation of causal effects given compliance types.

Table 1 shows key parameters involved in CACE estimation and corresponding sample statistics. Since compliance status is observed among individuals assigned to the treatment condition, as illustrated in Figure 1, compliance-specific parameters under the treatment condition have corresponding observed sample statistics. That is, μ_{1n} is the mean potential outcome of noncompliers when assigned to the treatment condition, and μ_{1c} is the mean potential outcome of compliers when assigned to the treatment condition. Corresponding sample statistics $\bar{y}_{1n}$ and $\bar{y}_{1c}$ are approximately unbiased estimates of μ_{1n} and μ_{1c} , where $\bar{y}_{1n}$ is the sample mean outcome of noncompliers assigned to the treatment condition and $\bar{y}_{1c}$ is the sample mean outcome of compliers assigned to the treatment condition. The overall mean potential outcome of the population when assigned to the treatment condition is μ_1 . The parameter μ_1 has the corresponding sample statistic $\bar{y}_1$, which is the overall sample mean outcome of individuals assigned to the treatment condition. Similarly, μ_{0n} is the mean potential outcome of noncompliers when assigned to the control condition, and μ_{0c} is the mean potential outcome of compliers when assigned to the control condition. However, their corresponding sample statistics are not available because of unknown compliance status of individuals assigned to the control condition. The overall mean potential outcome of the population when assigned to the control condition is μ_0 . The parameter μ_0 has the corresponding sample statistic $\bar{y}_0$, which is the overall sample mean outcome of individuals assigned to the control condition. The compliance rate in the population is π_c , and its corresponding sample statistic is p_c , which is the sample proportion of compliers in the treatment group. On the basis of random assignment of treatment conditions, it is assumed that this proportion is the same across the treatment and control conditions.

The distribution of μ_0 can be seen as a mixture (i.e., combination) of the complier (μ_{0c}) and the noncomplier (μ_{0n}) distributions, where π_c determines the mixing proportion (i.e., the ratio between compliers and noncompliers). That is, $\mu_0 = (1 - \pi_c)\mu_{0n} + \pi_c\mu_{0c}$. In general, mixture analysis (McLachlan & Peel, 2000; Titterton, Smith, & Makov, 1985) involves estimation of subpopulation-specific model parameters without information on the number of classes and mixing proportions. In the CACE estimation shown in this article, the number of subpopulations is predetermined (i.e., there are only two subpopulations—compliers and noncompliers), and the mixing proportion is observed from the data (i.e., $\hat{\pi}_c = p_c$). Furthermore, identifying assumptions such as monotonicity and the exclusion restriction are employed to ensure identification of the model (i.e., there can be only one solution). Therefore, the CACE estimation employed in this article can be considered a special case of mixture analysis without its exploratory feature.

Along with the assumption of monotonicity, the exclusion restriction assumption (Angrist et al., 1996; Hirano et al., 2000; Imbens & Rubin, 1997; Jo, 2002a, 2002b) plays a critical role in identifying CACE. Under the exclusion restriction, the effect of treatment assignment is allowed for compliers but disallowed for always-takers and never-takers (in the current setting, it applies only to never-takers). In other words, noncompliers' potential outcomes do not vary depending on treatment assignment status, which directly implies that $\mu_{1n} = \mu_{0n}$ in the current setting. This assumption has been widely used in practice, although its plausibility is often questioned when applied to experiments that do not employ blinding. For example, in the JHU PIRC trial, noncompliers may partially receive the treatment if assigned to the intervention condition, but they cannot receive the treatment at all if assigned to the control condition. Therefore, there is a possibility that noncompliers may have better outcomes when assigned to the intervention condition than when assigned to the control condition. Another possibility is that poorly complying families may be demoralized by failing to comply with the treatment and this may have a negative impact on their children, resulting in less desirable outcomes when assigned to the intervention condition than when assigned to the control condition. For general discussion on underlying assumptions of CACE estimation, see, for example, Angrist et al. (1996), Jo (2002c), Little and Yau (1998), and West and Sagarin (2000).

Assuming two compliance types (compliers and noncompliers) and the exclusion restriction, a continuous outcome Y for individual i ($i = 1, 2, 3, \dots, n$) in a randomized trial without data clustering can be expressed as

$$Y_i = \alpha_n n_i + \alpha_c c_i + \gamma_c c_i Z_i + \varepsilon_{ni} n_i + \varepsilon_{ci} c_i, \quad (1)$$

where $c_i = 0$ and $n_i = 1$ if individual i is a noncomplier and $c_i = 1$ and $n_i = 0$ if individual i is a complier. The treatment assignment status $Z_i = 0$ if individual i is assigned to the control condition, and $Z_i = 1$ if assigned to the treatment condition. The mean potential outcome when $Z = 0$ is α_n (i.e., μ_{0n}) for noncompliers and α_c (i.e., μ_{0c}) for compliers. The average effect of treatment assignment for compliers is γ_c (CACE). A normally distributed residual for noncompliers is ε_{ni} , which has zero mean and the variance σ_n^2 . A normally distributed residual for compliers is ε_{ci} , which has zero mean and the variance σ_c^2 .

In the absence of covariates that predict compliance, the proportions of compliers and noncompliers can be expressed in the empty logistic regression as

$$\begin{aligned} P(c_i=1) &= \pi_i, \\ P(c_i=0) &= 1 - \pi_i, \\ \text{logit}(\pi_i) &= \beta_0, \end{aligned} \quad (2)$$

where π_i is the probability of being a complier for individual i and β_0 is the logit intercept. The proportion of compliers in the population is π_c .

On the basis of Equations 1 and 2, three directly estimable population means (left side of the equations) can be expressed in terms of model parameters (right side of the equations) as

$$\mu_{1n} = \alpha_n, \quad (3)$$

$$\mu_{1c} = \alpha_c + \gamma_c, \quad (4)$$

$$\mu_0 = (1 - \pi_c)\alpha_n + \pi_c\alpha_c, \quad (5)$$

where μ_{1n} , μ_{1c} , and μ_0 have corresponding sample statistics. Under the assumption of the exclusion restriction, the effect of treatment assignment is disallowed for never-takers (i.e., $\mu_{1n} = \mu_{0n}$). Therefore, α_n (i.e., μ_{0n}) is directly identified as μ_{1n} , as shown in Equation 3.

Then, from Equations 3 and 5, α_c can be identified as

$$\alpha_c = \frac{\mu_0 - (1 - \pi_c)\mu_{1n}}{\pi_c}. \quad (6)$$

From Equations 4 and 6, γ_c (CACE) can be identified as

$$\gamma_c = \mu_{1c} - \frac{\mu_0 - (1 - \pi_c)\mu_{1n}}{\pi_c} = \frac{\mu_{1c} - \mu_0}{\pi_c}. \quad (7)$$

Under the assumptions of random assignment, SUTVA, monotonicity, and exclusion restriction, a large-sample-based estimator of CACE is formulated as (Angrist et al., 1996)

$$\gamma_c = \bar{y}_{1c} - \frac{\bar{y}_0 - (1 - p_c)\bar{y}_{1n}}{p_c} = \frac{\bar{y}_1 - \bar{y}_0}{p_c}. \quad (8)$$

CACE in the CRT Setting

Let us assume a CRT where clusters of individuals are randomly assigned either to the treatment or to the control condition. Again, some study participants do not comply with the given treatment. Individual i ($i = 1, 2, 3, \dots, m_j$) now belongs to cluster j ($j = 1, 2, 3, \dots, G$). The expression in Equation 1 is modified as

$$Y_{ij} = \alpha_n n_{ij} + \alpha_c c_{ij} + \gamma_c c_{ij} Z_j + \varepsilon_{nbj} n_{ij} + \varepsilon_{nwi} n_{ij} + \varepsilon_{cbj} c_{ij} + \varepsilon_{cwi} c_{ij}, \quad (9)$$

where Z_j denotes the cluster-level randomization status. The macro-unit residuals ε_{nbj} (noncompliers) and ε_{cbj} (compliers) represent cluster-specific effects given Z and are assumed to be normally distributed with zero mean and the between-cluster variances σ_{nb}^2 (noncompliers) and σ_{cb}^2 (compliers). The micro-unit residuals ε_{nwij} (noncompliers) and ε_{cwij} (compliers) are assumed to be normally distributed with zero mean and the within-cluster variance σ_{nw}^2 (noncompliers) and σ_{cw}^2 (compliers), which are equal across clusters. The total residual variance is the sum of the between- and within-cluster variances.

The logistic regression in Equation 2 is modified as

$$\begin{aligned} P(c_{ij}=1) &= \pi_{ij}, \\ P(c_{ij}=0) &= 1 - \pi_{ij}, \\ \text{logit}(\pi_{ij}) &= \beta_0 + \xi_j, \end{aligned} \tag{10}$$

where the between-cluster residual ξ_j has zero mean and a variance of ψ_b^2 . The logit value varies across clusters ($\beta_0 + \xi_j$), meaning that the proportion of compliers differs across clusters.

Defining Complications in CRTs With Noncompliance

Under the assumption of SUTVA, each individual's potential outcomes are uncorrelated with other individuals' treatment assignment status. SUTVA is a critical assumption that makes identification of causal treatment effects possible. However, in randomized trials that deal with groups of individuals who are highly likely to interact with one another, SUTVA is unlikely to hold. Cluster-level randomization plays a critical role in minimizing the impact of interaction (contamination or interference) among individuals by assigning individuals who are most likely to interact with one another (i.e., individuals in the same cluster unit) to the same treatment condition (for further discussion on SUTVA, see, e.g., Sobel, 2006). For example, in the JHU PIRC trial, since the unit of randomization is a classroom, significant contact or interaction among individuals assigned to different treatment conditions is unlikely, although interaction in the same cluster is highly likely. This setting makes the violation of SUTVA a more manageable problem. Whereas contamination across different treatment conditions is a very difficult problem to deal with, interaction within clusters is a manageable problem that can be handled statistically by considering resemblance among individuals with the same cluster membership in the analysis.

In CRTs, inflation of variance is usually expected because of similarity among individuals in the same cluster. If data are appropriately analyzed considering inflation of variance, the resulting statistical power is usually lower than that in trials with individual-level randomization. If data are analyzed ignoring inflation of variance, the resulting Type I error rate will be incorrectly inflated. ICC has been commonly used to gauge possible inflation of variance in CRTs.

Conventionally, ICC represents the level of resemblance among individuals belonging to the same cluster in terms of outcomes. However, ICC defined in this way may not well represent resemblance among individuals in CRTs accompanied by noncompliance. That is, individuals in the same cluster are likely to show resemblance not only in terms of outcomes but also in terms of compliance. Furthermore, the level of resemblance in outcomes may vary across different compliance types. In this case, the conventional ICC may not be informative in evaluating the impact of clustering on Type I error rates. Given that, the

definition of ICC is extended in this article to properly reflect situations where both clustering and noncompliance are present.

Outcome Intraclass Correlation

In the presence of noncompliance, outcome ICC can differ across compliance types. For example, in the FSP intervention condition in the JHU PIRC trial, when compliance is defined as completion of at least two thirds of intervention activities, the ICC estimate among noncompliers is much higher (0.25) than that of compliers (0.05) in terms of the shy behavior outcome at Grade 2. In other words, children's shy behavior outcome was more sensitive to teacher or classroom environment when parents complied poorly with the intervention activities, which may be interpreted as an indicator of low-level parental involvement in general. The conventional overall ICC (0.13) does not reflect this heterogeneity in resemblance across compliers and noncompliers, although it still represents the overall level of resemblance. Given that, we define outcome ICCs conditioning on compliance status.

From Equation 9, the ICC coefficient in outcome Y for noncompliers given Z is

$$ICC_{Y_n} = \frac{\sigma_{nb}^2}{\sigma_{nb}^2 + \sigma_{nw}^2}, \quad (11)$$

where σ_{nb}^2 denotes the between-cluster variance and σ_{nw}^2 denotes the within-cluster variance for noncompliers given Z . The total variance is the sum of the between- and within-cluster variances ($\sigma_n^2 = \sigma_{nb}^2 + \sigma_{nw}^2$).

The ICC coefficient in outcome Y for compliers given Z is

$$ICC_{Y_c} = \frac{\sigma_{cb}^2}{\sigma_{cb}^2 + \sigma_{cw}^2}, \quad (12)$$

where σ_{cb}^2 denotes the between-cluster variance and σ_{cw}^2 denotes the within-cluster variance for compliers given Z . The total variance is the sum of the between- and within-cluster variances ($\sigma_c^2 = \sigma_{cb}^2 + \sigma_{cw}^2$).

Compliance Intraclass Correlation

In CRTs, not only outcomes but also compliance can be similar among individuals in the same cluster. Consequently, the compliance rate may vary across different clusters. There are several ways to present heterogeneity across clusters in proportions (Agresti, 2002; Commenges & Jacqmin, 1994; Haldane, 1940; McCullagh & Nelder, 1989; McKelvey & Zavoina, 1975; Snijders & Bosker, 1999). Assuming the logistic distribution for the within-cluster residuals, the ICC coefficient in compliance can be defined from Equation 10 as

$$ICC_c = \frac{\psi_b^2}{\psi_b^2 + \pi^2/3}, \quad (13)$$

where ψ_b^2 is the between-cluster variance (i.e., variance of ξ_j) and $\pi^2/3$ is the variance for the within-cluster residual in the logistic distribution. This definition has been chosen in this

article because it can be directly extended (Snijders & Bosker, 1999) to define residual ICC (i.e., ICC controlling for covariate effects). ICC_C represents the degree of resemblance in compliance among individuals belonging to the same cluster. For example, in the FSP intervention condition in the JHU PIRC trial, the ICC_C estimate is 0.37, which reflects a substantial variation in the compliance rate across classrooms (compliance rate ranged from 5% to 100%).

Consequences of Simultaneous Complications

Since dealing with either clustering or noncompliance is already quite complex, data are often analyzed considering only one complication. In this article, it is demonstrated how ICC_C , ICC_{Y_n} , and ICC_{Y_c} are connected to variance misestimation when noncompliance is considered but clustering is ignored in the analysis (i.e., single-level CACE estimation). When evaluating efficacy is of interest, one may still want to conduct a single-level CACE analysis despite possible ICCs, given that considering any of ICC_C , ICC_{Y_n} , and ICC_{Y_c} (or conventional overall ICC_Y) involves multilevel mixture analyses that are computationally demanding. Not covered by this study is the analysis where clustering is considered but noncompliance is ignored (i.e., multilevel ITT analysis), which is commonly used in analyzing data from randomized trials to estimate an overall effect of treatment assignment (Brown et al., 2007). In this analysis option, the between-cluster compliance variance cannot be modeled, and therefore, it is captured in the estimation as a part of the between-cluster outcome variance, resulting in substantially decreased power to detect ITT effect. In the current article, we focus on how ignoring clustering affects the quality of CACE estimation and how considering both clustering and noncompliance improves its quality. How ignoring noncompliance affects the quality of ITT analysis and how considering both clustering and noncompliance improves its quality is dealt with in another manuscript (Jo, Asparouhov, & Muthén, in press).

ICCs shown in Equations 11, 12, and 13 raise some new questions that used to be irrelevant in the CRT setting until we started considering noncompliance, such as (a) whether ICC_C alone has any impact on variance misestimation (inflation of the Type I error rate), (b) whether different combinations of ICC_{Y_n} and ICC_{Y_c} differently influence variance misestimation, and (c) whether the impact of ICC_{Y_n} and ICC_{Y_c} varies depending on the level of ICC_C . These speculations, if they turn out to be true, will provide compelling reasons for the simultaneous consideration of clustering and noncompliance in the analysis. This section explores various settings to examine which of these interactions between noncompliance and clustering have actual impact on variance misestimation and in what conditions the impacts are more substantial. Monte Carlo simulations are employed for this purpose, since it is not straightforward to analytically derive possible inflation of the Type I error rate given missing compliance information and mixture distributions of different compliance types.

Data Generation

The Monte Carlo simulation results presented in this study are based on 500 replications. The size of each cluster (m) is either 20 or 40, and the total number of clusters (G) is 100 (50 in the control and 50 in the treatment condition). Although simulation settings are mostly based on the JHU PIRC FSP school intervention trial, a larger number of clusters (100 in this study compared with 18 in the JHU PIRC trial) are employed to avoid another source of variance misestimation and to focus on variance misestimation due only to ICCs. The true ratio of the treatment and control groups is 50%:50%, and the true compliance rate is 50% in all simulation settings.

Outcome ICC values are decided on the basis of the JHU PIRC trial ($ICC_{Y_n} = 0.25$ and $ICC_{Y_c} = 0.05$ for the shy behavior outcome at Grade 2). To examine the impact of different

outcome ICC compositions, four different combinations of outcome ICC values are considered. They are (a) when neither noncompliers nor compliers have any outcome ICC ($ICC_{Y_n} = 0.0$ and $ICC_{Y_c} = 0.0$), (b) when only noncompliers have a substantial outcome ICC ($ICC_{Y_n} = 0.2$ and $ICC_{Y_c} = 0.0$), (c) when only compliers have a substantial outcome ICC ($ICC_{Y_n} = 0.0$ and $ICC_{Y_c} = 0.2$), and (d) when noncompliers and compliers have the same moderate level of ICC ($ICC_{Y_n} = 0.1$ and $ICC_{Y_c} = 0.1$). The first setting ($ICC_{Y_n} = ICC_{Y_c} = 0.0$) is considered to examine whether there is a pure impact of ICC_C in the absence of outcome ICC.

The true compliance ICC value ranges from 0.0 to 1.0 (in the JHU PIRC study, ICC_C is about 0.37 in the intervention condition). The zero ICC_C indicates that compliance behavior is independent of the clusters individuals belong to. This setting is considered to examine whether different compositions of ICC_{Y_n} and ICC_{Y_c} have different impact on variance inflation and the subsequent variance misestimation in the absence of ICC_C . The perfect ICC_C (i.e., 1.0) is the other extreme situation, where every individual in the same cluster shows the same compliance behavior. This could be a possible scenario depending on how compliance is decided and how intervention treatments are delivered. For example, if a teacher or a doctor, who represents the unit of randomization, delivers the intervention and if study participants do not have much room for independent decision on compliance, it is likely that compliance is decided at the cluster level.

Another key component of the simulation settings is the distributional distance between compliers and noncompliers. Given unknown compliance status among control-condition individuals, precision of the CACE estimate depends on how well the mixtures of distributions are separated. Therefore, the distance between the two groups normally improves the estimation quality (i.e., the farther apart the distributions, the better the precision; Jo, 2002c). In the CRT setting, however, having a farther distance does not necessarily have a positive impact on variance estimation. To represent the distance between the two distributions, three conditions are considered. Given Z , noncompliers and complier outcome means are (a) 0.0 SD s apart, (b) 0.5 SD s apart, or (c) 1.0 SD apart. In the JHU PIRC trial, noncomplier and complier outcome means are approximately 1.0 SD apart on the basis of CACE estimation with and without considering data clustering.

Data were generated according to Equations 9 and 10. Specifically, the true within-cluster variances σ_{nw}^2 and σ_{cw}^2 take values of 1.0, 0.9, and 0.8. The true between-cluster variances σ_{nb}^2 and σ_{cb}^2 take values of 0.0, 0.1, and 0.2 to reflect ICC_{Y_n} and ICC_{Y_c} of 0.0, 0.1 and 0.2 given the total variance of 1.0. The true correlation between the macrolevel outcome residuals (i.e., ε_{nbj} , ε_{cbj}) is zero. In practice, this correlation may vary depending on trials. For example, outcome variation across clusters given treatment assignment may be similar or very different for compliers and noncompliers. The true control-condition noncomplier outcome mean α_n is 1.0, and the true control-condition complier outcome mean α_c takes values of 1.0, 1.5, and 2.0 to reflect the distance between noncompliers and complier means (0.0, 0.5, and 1.0 SD s apart). The true treatment assignment effect for compliers γ_c (i.e., CACE) is 0.6 (effect size of 0.6 on the basis of the total variance), and the true treatment assignment effect for noncompliers is zero (i.e., exclusion restriction holds). The true logit intercept β_0 is zero (i.e., 50% compliance), and the true between-cluster compliance variance ψ_b^2 takes values of 0.00, 0.82, 2.19, 13.15, and 10,000 on the logit scale to reflect ICC_C of 0.0, 0.2, 0.4, 0.8, and 1.0 according to Equation 13. The true covariance between compliance and outcome at the cluster level (i.e., between ξ_j and ε_{nbj} , between ξ_j and ε_{cbj}) is zero. This correlation may increase in some trials, for example, where clusters with higher proportions of compliers tend to have better outcomes given treatment assignment.

CACE Estimation Without Considering Cluster Randomization

Data are analyzed on the basis of Equations 1 and 2, which represent the standard CACE model without considering the fact that randomization was done at the cluster level. Estimation of CACE considering clustering is presented in a later section of this article. As in the data-generation step, the exclusion restriction is assumed.

The current study employs an ML estimation approach. The moment estimator defined in Equation 8 is helpful for intuitive understanding of identification of CACE. However, this estimator does not distinguish CACE in the CRT setting from CACE in the setting where randomization is conducted at the individual level. Therefore, for the actual CACE estimation in this article, an ML method is consistently used to handle situations with and without data clustering.

Given that compliance type cannot be observed in the control condition, the observed-data likelihood function based on treatment assignment ($Z = 1$: treatment condition, $Z = 0$: control condition) and observed treatment receipt status ($D = 1$: received, $D = 0$: not received) is

$$L(\theta|data) \propto \prod_{i \in \{Z_i=1, D_i=0\}} (1 - \pi_c) f(y_i | \mu_{1n}, \sigma_n^2) \times \prod_{i \in \{Z_i=1, D_i=1\}} \pi_c f(y_i | \mu_{1c}, \sigma_c^2) \times \prod_{i \in \{Z_i=0, D_i=0\}} [(1 - \pi_c) f(y_i | \mu_{0n}, \sigma_n^2) + \pi_c f(y_i | \mu_{0c}, \sigma_c^2)],$$

(14)

where $\theta = (\pi_c, \mu_{1n}, \mu_{1c}, \mu_{0n}, \mu_{0c}, \sigma_n^2, \sigma_c^2)$ is the set of parameters in the model, $f(y_i | \mu, \sigma^2)$ denotes the probability density of a normal distribution with mean μ and variance σ^2 , and π_c denotes the proportion of compliers in the population.

By maximizing the likelihood in Equation 14 with respect to the parameters of interest θ , ML estimates are obtained. The unknown compliance status in the control condition is handled as missing data via the EM algorithm (Dempster, Laird, & Rubin, 1977; Little & Rubin, 2002; McLachlan & Krishnan, 1997; Tanner, 1996). The E step computes the expected values of the complete-data sufficient statistics given data y and current parameter estimates θ . The M step computes the complete-data ML estimates with complete-data sufficient statistics replaced by their estimates from the E step. This procedure continues until it reaches optimal status. Parametric standard errors are computed from the information matrix of the ML estimator using both the first- and the second-order derivatives under the assumption of normally distributed outcomes. In the current study, ML-EM estimation of CACE has been carried out by Mplus Version 4.2 (L. K. Muthén & Muthén, 1998–2007).

Impact of Compliance Intraclass Correlations

In CRTs, individuals in the same cluster may resemble each other not only in terms of outcomes but also in terms of compliance behavior. In particular, the Monte Carlo simulation results presented in this section are based on a hypothetical setting, where individuals in the same cluster are similar in terms of compliance behavior but not in terms of outcomes. Though quite unrealistic, this setting is important to consider to examine whether ICC_C alone has any impact on variance misestimation in the absence of outcome ICCs, which is an intriguing question that has not been explicitly considered in CRT data analysis practice.

Figure 2A shows how the coverage of the CACE estimate decreases as ICC_C increases. In the simulations for Figure 2A, the cluster size is 20. The nominal 95% confidence interval coverage rate is 0.95 (or nominal Type I error rate is 0.05). In the current simulation setting, complier and noncomplier outcomes are normally distributed, and the number of clusters is large enough (i.e., 100), minimizing additional factors that may affect coverage rates. Therefore, a coverage rate below 0.95 mainly indicates that standard error estimates do not appropriately reflect variance inflation due to clustering of compliance behavior. In other words, the level of significance of the treatment effect is overstated. When compliers and noncompliers have homogeneous outcome distributions (0.0 SDs apart), the coverage rate stays close to the nominal level irrespective of the change in ICC_C . When complier and noncomplier outcome means are 0.5 SDs apart, the coverage rate begins to be affected especially by high ICC_C s such as 0.8 and 1.0. When complier and noncomplier outcome means are 1.0 SD apart, the coverage rate decreases substantially with moderate to high ICC_C s.

The simulation results shown in Figure 2B are based on the same settings as those of Figure 2A, except that the cluster size is 40 instead of 20. It is well known that the cluster size affects the magnitude of variance inflation due to outcome ICC (Donner, Birkett, & Buck, 1981; Kish, 1965). Panel (b) shows that the same rule applies to compliance intraclass correlation. As the cluster size increases from 20 to 40, the impact of ICC_C becomes more prominent, resulting in substantial deterioration of the coverage rate even with low levels of ICC_C . As in Figure 2A the coverage rate stays close to the nominal level irrespective of the change in ICC_C when compliers and noncompliers have homogeneous distributions.

The interesting phenomenon shown in Figure 2 can be explained by variance inflation in compliance, which consequently leads to variance inflation in parameters that involve outcome and compliance. However, as compliers and noncompliers have more homogeneous outcome distributions, variance inflation in compliance has less impact on variance inflation in those parameters. For example, in Equation 9, compliance status can be thought of as a between-cluster covariate (not fully observed) if $ICC_C = 1.0$. In this case, if compliers and noncompliers are homogeneous given Z (i.e., $\alpha_n - \alpha_c = 0$), variance inflation in compliance does not influence the between-cluster variance of the outcome. However, as α_n and α_c have some distance between them, the between-cluster compliance status actually produces between-cluster variance in the outcome. In analyses ignoring data clustering, this variance inflation cannot be taken into account and consequently causes variance misestimation of key parameters such as CACE.

Impact of Outcome Intraclass Correlations

The Monte Carlo simulation results in this section are presented to show whether different combinations of ICC_{Y_n} and ICC_{Y_c} differently influence variance misestimation and whether the impact of ICC_{Y_n} and ICC_{Y_c} varies depending on the level of ICC_C . As in Figure 2, the distance between compliers and noncompliers is included as one of the key factors that influence variance inflation. In Figure 3, the first outcome ICC pattern (i.e., $ICC_{Y_n} = ICC_{Y_c} = 0$) is included as a reference setting, where only the impact of ICC_C can be observed as in Figure 2. The other three patterns represent the overall outcome ICC of 0.1 and therefore should have the same influence on variance misestimation unless heterogeneity in the outcome ICC across compliers and noncompliers plays a role. The evidence of interaction between ICC_C and ICC_Y can be found when the impact of ICC_Y changes as ICC_C changes.

Figure 3A shows how the coverage of the CACE estimate changes depending on ICC_{Y_n} , ICC_{Y_c} , and ICC_C when compliers and noncompliers have homogeneous distributions. The pure impact of the outcome ICC can be observed when $ICC_C = 0$. It is shown that different patterns of ICC_Y have a similar impact on variance inflation, and the impact of ICC_Y on

variance misestimation increases as ICC_C increases, although the change is not dramatic. As complier and noncomplier outcome means are farther apart from each other, as shown in Figures 3B and 3C, the coverage rate, in general, decreases faster as ICC_C increases. It is also shown that different patterns of ICC_Y show different coverage rates (i.e., have different impact on variance misestimation). It becomes clear in Figure 3C that ICC_Y has the greatest impact on variance inflation when ICC_Y is concentrated among compliers (i.e., $ICC_{Y_n} = 0.0$, $ICC_{Y_c} = 0.2$) and the smallest impact when it is concentrated among noncompliers (i.e., $ICC_{Y_n} = 0.2$, $ICC_{Y_c} = 0.0$), though the overall outcome ICC is the same.

In the simulations reported in Figure 3, data have been generated and analyzed assuming that the correlation between the two between-cluster residuals (ϵ_{nbj} , ϵ_{cbj}) is zero. In practice, these residuals may be correlated. That is, outcome variation across classrooms given treatment assignment may be similar or very different for compliers and noncompliers. As this correlation deviates from zero, the results can be somewhat different from those reported in Figure 3. The coverage of the CACE estimate is somewhat lower when the correlation between ϵ_{nbj} and ϵ_{cbj} is substantial (e.g., 0.5) than when the correlation is zero. However, both settings lead to similar results as complier and noncomplier means have a substantial distance (i.e., 1.0 *SD* apart). In the simulations reported in Figure 3, data have also been generated and analyzed assuming that the correlation between compliance and outcome at the cluster level is zero (i.e., correlation between ξ_j and ϵ_{nbj} is zero, correlation between ξ_j and ϵ_{cbj} is zero). This correlation may increase in some trials, for example, where clusters with higher proportions of compliers tend to have better outcomes given treatment assignment. The coverage of the CACE estimate is slightly higher in the setting where the correlation is substantial as complier and noncomplier means have a substantial distance (for these additional simulation results, see Figures 5 and 6 in the supplemental materials available online; see the URL at the head of this article).

Implications of Simultaneous Complications

Simulation results reported in Figures 2 and 3 demonstrate that the conventional overall outcome ICC may not properly reflect the impact of clustering on variance misestimation when clustering and noncompliance are simultaneously present. For example, we may predict a mild impact of clustering on the basis of the conventional ICC measure when the actual impact is much more substantial.

In particular, the impact of ICC_C is quite striking, implying the need for checking ICC_C when CRTs are accompanied by treatment noncompliance. Following Equations 10 and 13, ICC_C can be estimated on the basis of a two-level logistic regression analysis using the intervention group data, where compliance type is observed. It is shown that the impact of ICC_C on the coverage rate increases as the distance between complier and noncomplier means increases. Depending on how heterogeneous compliers and noncompliers are, their outcomes may have a substantial distance in the absence of intervention treatments. Whether this distance is large enough (e.g., 0.5–1.0 *SDs* or larger) to connect ICC_C with a substantial bias in variance estimation cannot be observed from the data since compliance types are not known among individuals assigned to the control condition. However, if it is reasonable to assume the exclusion restriction and monotonicity, complier and noncomplier means can be estimated using the standard CACE estimation method. If this distance is substantial and ICC_C is moderate to high, one should be aware that not only ICC_Y but also ICC_C will contribute to variance misestimation, necessitating simultaneous consideration of clustering and noncompliance in the analysis.

The interaction between clustering and noncompliance also has an implication in assessing the impact of ICC_Y . Simulation results show that as the distance between the complier and noncomplier means increases, ICC_Y has a greater impact on variance inflation when ICC_Y is

concentrated among compliers and a smaller impact when it is concentrated among noncompliers, though the overall ICC_Y is the same. The main message here is that it is possible in practice to have different patterns of ICC_Y for different compliance types, and they may affect variance inflation and misestimation differently. As shown in Equations 11 and 12, this information cannot be obtained unless variances are decomposed simultaneously, considering compliance types and clustering. In the analysis accounting for both clustering and noncompliance, different patterns of ICC_Y can be properly modeled, as illustrated in the next section.

In simulation results reported in this article, a large number of clusters are employed (i.e., 100) to avoid another source of variance misestimation and to focus on variance misestimation due only to ICCs. In practice, however, much smaller numbers of clusters are often employed (e.g., 18 in the JHU PIRC trial) in CRTs. With small numbers of clusters, standard errors are likely to be underestimated (i.e., Type I error increases) even with analysis methods that account for clustering. In this situation, simultaneously modeling noncompliance and clustering will at least reduce variance inflation due to the interaction between noncompliance and clustering. However, variance misestimation due to small numbers of clusters still remains as a problem, and therefore, one needs to be aware that estimated standard errors tend to exaggerate the significance of the results. To counter this problem, several methods, such as jackknife sandwich estimates (MacKinnon & White, 1985), adjustment using the t distribution (Thornquist & Anderson, 1992), and adjustment considering the variance of the sandwich estimate (Kauermann & Carroll, 2001), have been suggested. However, little is known about how these methods perform with mixture distributions. Further study is in order to develop reliable methods of adjusting for variance misestimation due to small numbers of clusters in the presence of noncompliance.

The results reported in Figures 2 and 3 are subject to change if the exclusion restriction does not hold. However, including violation of the exclusion restriction as an additional factor results in high dimensional interactions among ICC_C , ICC_Y , and the exclusion restriction. In other words, it is unlikely that consistent conclusions can be reached about the simultaneous impact of these factors. Besides, violation of the exclusion restriction influences not only the standard error but also point estimates of the CACE, further complicating the evaluation of the variance estimation quality. One accessible way of checking sensitivity to violation of the exclusion restriction is to use pretreatment covariates that are predictors of compliance (Jo, 2002a, 2002b). However, this method is not recommended for analyses ignoring clustering because both within- and between-cluster covariates lose predicting power as ICC_C increases. For example, if $ICC_C = 1.0$, within-cluster covariates, which have only within-cluster variances, cannot predict compliance (i.e., logit coefficients are zero), which is a cluster-level variable. Between-cluster covariates cannot predict compliance either because any kind of between-cluster variance cannot be taken into account in the analysis ignoring clustering. However, in the analyses simultaneously considering clustering and noncompliance, both within- and between-cluster covariates can be properly handled as predictors of compliance, alleviating the potential impact of the exclusion restriction violation.

Simultaneous Consideration of Clustering and Noncompliance

Simulation studies shown in the previous section have emphasized the unique portion of variance inflation when CRTs are accompanied by noncompliance. To appropriately reflect this inflation, data clustering and noncompliance need to be considered simultaneously in the analysis. On the basis of Monte Carlo simulations, this section demonstrates the joint analysis that considers both complications. For Mplus codes used for the simulations reported here, see the supplemental materials (available online; see the URL at the head of

this article). Pretreatment covariates are added to data-generation and analysis models employed in the previous section. As clustering of data is considered, between- and within-cluster covariates can be properly handled in the joint analysis given compliance and outcome ICCs. For example, in the JHU PIRC trial, teacher characteristics or classroom environment (e.g., average level of aggression) can be treated as between-cluster or contextual-level variables, whereas students' baseline behavioral measures can be treated as within-cluster or individual-level variables.

Data Generation

In the presence of covariates, the expression in Equation 9 is modified as

$$Y_{ij} = \alpha_n n_{ij} + \alpha_c c_{ij} + \gamma_n n_{ij} Z_j + \gamma_c c_{ij} Z_j + \lambda'_{nb} n_{ij} \mathbf{x}_{bj} + \lambda'_{nw} n_{ij} \mathbf{x}_{wij} + \lambda'_{cb} c_{ij} \mathbf{x}_{bj} + \lambda'_{cw} c_{ij} \mathbf{x}_{wij} + \varepsilon_{nbj} n_{ij} + \varepsilon_{nwij} n_{ij} + \varepsilon_{cbj} c_{ij} + \varepsilon_{cwij} c_{ij}, \quad (15)$$

where $\mathbf{x}_{bj}$ is a vector of between-cluster covariates and $\mathbf{x}_{wij}$ is a vector of within-cluster covariates. The vectors of regression coefficients λ_{nb} and λ_{nw} represent between- and within-cluster covariate effects on Y for noncompliers. The vectors of regression coefficients λ_{cb} and λ_{cw} represent between- and within-cluster covariate effects on Y for compliers. It is assumed that the effect of treatment assignment does not vary across different values of covariates (additivity; Jo, 2002a), and the main effect of treatment assignment for noncompliers γ_n (noncomplier average causal effect [NACE]) is allowed. In other words, the exclusion restriction is not imposed.

In the presence of pretreatment covariates, the probability that individual i in cluster j will comply (π_{ij}) varies depending on the influence of covariates. The logistic regression in Equation 10 is modified as

$$\begin{aligned} P(c_{ij}=1|\mathbf{x}_{bj}, \mathbf{x}_{wij}) &= \pi_{ij}, \\ P(c_{ij}=0|\mathbf{x}_{bj}, \mathbf{x}_{wij}) &= 1 - \pi_{ij}, \\ \text{logit}(\pi_{ij}) &= \beta_0 + \beta'_{1b} \mathbf{x}_{bj} + \beta'_{1w} \mathbf{x}_{wij} + \xi_j, \end{aligned} \quad (16)$$

where the vector of logit coefficients β_{1b} indicates the level of association between compliance and between-cluster covariates and the vector of logit coefficients β_{1w} indicates the level of association between compliance and within-cluster covariates. Nonzero variance of ξ_j (i.e., ψ_b^2) means that the proportion of compliers differs across clusters conditional on these covariates.

The Monte Carlo simulation results presented in this section are based on 500 replications. The size of each cluster (m) is 40, and the total number of clusters (G) is 100. Data have been generated on the basis of Equations 15 and 16. The true within-cluster residual variances σ_{nw}^2 and σ_{cw}^2 are 0.9 and 0.8, respectively. The between-cluster residual variances σ_{nb}^2 (noncompliers) and σ_{cb}^2 are 0.1 and 0.2, reflecting ICC_{Y_n} of 0.1 and ICC_{Y_c} of 0.2 given the total residual variance of 1.0. The control-condition noncomplier mean α_n is 1.0, and the control-condition complier mean α_c is 2.0. The treatment assignment effect for noncompliers γ_n (i.e., NACE) is -0.2 , and the treatment assignment effect for compliers γ_c (i.e., CACE) is 0.6 . One covariate (X_1) with zero mean and within-cluster variance of 1.0 has been generated to represent within-cluster covariates. Another covariate (X_2) with zero mean and between-cluster variance of 1.0 has been generated to represent between-cluster covariates. For these covariates, the within-cluster regression coefficients are -0.1 for noncompliers (λ_{nwX_1}) and -0.2 for compliers (λ_{cwX_1}). The between-cluster regression coefficients are 0.1

for noncompliers (λ_{nbX_2}) and 0.2 for compliers (λ_{cbX_2}). The logit intercept β_0 is zero (50% compliance), and the between-cluster compliance residual variance ψ_b^2 is 2.191 on the logit scale, to reflect ICC_C of 0.4 conditioning on covariates. Both within- and between-cluster logit coefficients (β_{1wX_1} and β_{1bX_2}) are 0.7 (odds ratio of approximately 2.0).

CACE Estimation Without Considering Cluster Randomization

First, data are analyzed ignoring cluster randomization, as in the previous section. That is, the model used for data analysis can be described as

$$Y_i = \alpha_n n_i + \alpha_c c_i + \gamma_n n_i Z_i + \gamma_c c_i Z_i + \lambda'_n n_i \mathbf{x}_i + \lambda'_c c_i \mathbf{x}_i + \varepsilon_{ni} n_i + \varepsilon_{ci} c_i, \quad (17)$$

$$\text{logit}(\pi_i) = \beta_0 + \beta'_1 \mathbf{x}_i, \quad (18)$$

where cluster-level parameters and variances are completely removed from the data-generation model described in Equations 15 and 16. The vector of covariates $\mathbf{x}_i$ consists of the within- (X_1) and the between-cluster (X_2) covariates, and both of them are treated as individual-level variables in the analysis. As in the data-generation model, the exclusion restriction is relaxed (i.e., γ_n is freely estimated), relying on the additivity assumption and covariates that are good predictors of compliance. True values for noncomplier and complier residual variances (σ_n^2 and σ_c^2) are set at 1.0 in Table 2 by combining between- and within-cluster residual variances. For other parameters, true values in the data-generation model have been used, though between- and within-cluster parameters are not distinguished in the analysis (for Mplus codes, see the supplemental materials available online; see the URL at the head of this article).

Table 2 shows the results from the analysis using the CACE model ignoring cluster randomization. As in the CACE estimation presented in the previous section, an ML-EM estimation method is used. It is demonstrated that the average standard error estimates do not capture variance inflation of many parameters. In particular, the average standard error estimate of CACE is about half of the standard deviation of CACE estimates from 500 replications (i.e., empirical standard deviation), indicating poor estimation of the variance of the CACE estimate. As a result, the 95% confidence interval coverage rate of CACE is only 0.662. In other words, the Type I error rate is 0.338, which is more than 6 times the nominal rate. Low coverage rates are also evident among other cluster-level parameters. In the logistic regression of compliance on covariates, bias is observed not only in standard error estimates but also in point estimates. Logit coefficient estimates for both the within- and the between-level covariates (β_{1X_1} and β_{1X_2}) are noticeably biased by ignoring compliance ICC_C . The exclusion restriction is relaxed in the analysis assuming a randomized field trial, where the plausibility of the exclusion restriction is questionable (e.g., in the absence of blinding, it is possible that noncompliers may experience a negative psychological effect of being assigned to the treatment condition, whereas they do not experience this negative effect when assigned to the control condition). In this example, covariates still have reasonable predicting power to support the identification of γ_n (NACE). The results, however, imply that relying on covariate information to relax the exclusion restriction can be risky in CRTs with potentially high ICC_C .

CACE Estimation Considering Cluster Randomization

Simulation results presented in this section are based on the analysis using the same model used for data generation, described in Equations 15 and 16. The path diagram presented in Figure 4 illustrates this model. Here, the compliance status (C) of the control group is latent (unknown), and the compliance status of the treatment group is observed (known). In this path diagram, squares represent observed variables, and circles represent latent (missing) variables. In the within-level, the path from X_1 to C corresponds to the logistic regression of C on a within-cluster covariate. The path from X_1 to Y corresponds to the regression of Y on a within-cluster covariate. The arrows from C to this path and the residual error path indicate that the covariate effect and the residual variance are different depending on compliance status. In the between-level, the path from X_2 to C corresponds to the logistic regression of C on a between-cluster covariate. The error path to C indicates that the proportion of compliers varies across clusters. The path from X_2 to Y corresponds to the regression of Y on a between-cluster covariate. The arrow from C to this path indicates that the covariate effect is different depending on compliance status. The path from Z to Y corresponds to the treatment effect. The arrow from C to this path indicates that the treatment effect is different depending on compliance status. The residual error path to Y indicates the existence of between-cluster outcome variance. The arrow from C to this error path indicates that the between-cluster residual variance varies across compliance types. To estimate this model, we employed a formal multilevel mixture analysis (Asparouhov & Muthén, 2008; B. O. Muthén, 2004). For ML-EM estimation of CACE in this framework, Mplus Version 4.2 (L. K. Muthén & Muthén, 1998–2007) was used (for Mplus codes, see the supplemental material available online; see the URL at the head of this article). Estimation details are given in the Appendix.

Table 3 shows the results from the CACE analysis considering both clustering and noncompliance. Within-cluster residual variances (σ_{nw}^2 , σ_{cw}^2) are now separated from the between-cluster variances, allowing for a better comparison between compliance types. In the logistic regression of compliance, the within-cluster covariate logit coefficient (β_{1wX_1}) estimate is close to the true value with a reasonable coverage rate. It is also shown from the logit intercept (β_0) estimation that compliance type is estimated properly taking into account ICC_C . Fixed effect between-cluster parameters show reasonable coverage rates. In particular, the coverage rate of the CACE is 0.952 (compared with 0.662 in the analysis ignoring data clustering). Noncomplier and complier intercepts (α_n , α_c) also show good point estimates and coverage rates. In the logistic regression of compliance, the average between-cluster covariate logit coefficient (β_{1bX_2}) estimate is close to the true value with a reasonable coverage rate. The exclusion restriction is relaxed in the analysis assuming a randomized field trial, where the plausibility of the exclusion restriction is questionable. Given correctly estimated covariate effects on compliance (β_{1bX_2}), average treatment assignment effects for compliers (CACE) and for noncompliers (NACE) are more likely to be successfully separated in the analysis considering clustering than in the analysis ignoring clustering. Between-cluster residual variances (σ_{nb}^2 , σ_{cb}^2) are separated from the within-cluster residual variances, revealing the fact that between-cluster variance is more concentrated among compliers given covariates and treatment assignment. However, these random effect between-cluster parameters show somewhat low coverage rates. The results imply that, unless large numbers of clusters are available, the level of significance of between-cluster random effects should be interpreted with caution. We expect that the quality of standard error estimates will improve as the number of clusters increases (the number of clusters is currently 50 per compliance type). Finally, in the logistic regression of compliance, between-cluster variation of compliance is captured by the between-cluster residual variance (ψ_b^2), indicating a sizable ICC_C given covariates.

Application to Real Data

The multilevel CACE model described in Equations 15 and 16 can be directly applied to analyze real data. However, ML-EM estimation of this model is computationally demanding, and some parameter restrictions may be necessary to achieve identifiability and convergence. Simpler analyses using only the treatment group or analyses ignoring clustering may provide useful information that can be used to formulate more efficient estimation models.

On the basis of a simple two-level analysis using only the intervention group data, where compliance type is observed, ICC_{Y_n} (from a noncomplier-only analysis) and ICC_{Y_c} (from a complier-only analysis) can be estimated. Similarly, ICC_C can be estimated on the basis of a two-level logistic regression analysis using only the intervention group data. If all these estimates are very small, there is less need to employ the two-level CACE approach. As discussed earlier, depending on how heterogeneous complier and noncompliers are, their outcomes may have a substantial distance in the absence of intervention treatments. Whether this distance is large enough (e.g., 0.5–1.0 *SDs* or larger) to connect ICC_C with a substantial bias in variance estimation cannot be observed from the data since compliance types are not known among individuals assigned to the control condition. However, if it is reasonable to assume the exclusion restriction and monotonicity, complier and noncomplier means can be estimated using the standard CACE estimation method ignoring clustering (both the treatment and control groups are included in the analysis). If this distance is substantial and ICC_C is moderate to high, one should be aware that not only ICC_Y but also ICC_C will contribute to variance misestimation.

These investigations are also very helpful in interpreting CACE estimation results when the multilevel approach is not employed. For example, in the JHU PIRC trial, the proposed ML-EM estimation of the multilevel CACE model is not a recommended option mainly because of the small number of clusters (i.e., 18 clusters total). Along with the simulation results reported in this article, these simpler analyses provide better ideas on the possible impacts of outcome and compliance ICCs when clustering is not taken into account.

In this article, Mplus (L. K. Muthén & Muthén, 1998–2007) has been used for ML-EM estimation of the multilevel CACE model (for annotated Mplus codes for the simpler analyses and the full two-level CACE analysis using real data, see the supplemental materials available online; see the URL at the head of this article). The input data for this analysis need to include the outcome of interest, treatment assignment condition, compliance status, within- and between-cluster covariates, and cluster membership (e.g., which classroom). Details of these variables and the estimation model can also be found in the supplemental materials. For those who are familiar with path diagrams, Figure 4 will be helpful in understanding Mplus codes provided in the supplement.

Conclusions

This study demonstrates the impact of ICCs on variance inflation in the estimation of CACE in diverse CRT settings. The Monte Carlo simulation results show various types of variance inflation that are unique to CRTs accompanied by treatment noncompliance. First, it has been demonstrated that compliance ICC (ICC_C) itself can cause serious variance inflation in trials where cluster membership is likely to influence individuals' compliance behavior. Second, given the same overall outcome ICC, the impact of clustering may differ substantially depending on how the ICC is spread across compliance types. Furthermore, the impact of outcome ICCs (ICC_{Y_n} , ICC_{Y_c}) may vary depending on the level of ICC_C . Whether compliers and noncompliers are similar or different in the control condition turns out to be an important factor that affects variance inflation. The higher the level of heterogeneity

across complier and noncomplier distributions, the larger the impact of compliance and outcome ICCs on variance inflation. The Monte Carlo simulation results indicate that the conventional ICC, which is limited to outcome with no distinction between compliers and noncompliers, may not be a sufficient indicator of how serious variance inflation is in CRTs with noncompliance.

A formal multilevel analysis combined with the mixture analysis has been suggested as a way of dealing with both data clustering and noncompliance. On the basis of the formal multilevel analysis approach, between- and within-cluster-level parameters can be explicitly specified and estimated. On the basis of the mixture model approach, compliance-class-specific parameters, such as CACE, can be estimated, considering mixture distributions of compliers and noncompliers. The interaction between data clustering and noncompliance can be explicitly modeled in the joint analysis that combines the two approaches. The ML-EM estimation of the multilevel mixture models has been implemented in the Mplus program (L. K. Muthén & Muthén, 1998–2007), providing an accessible tool for a wide range of researchers. The instrumental variable approach (Angrist et al., 1996; Bloom, 1984), which is basically the method of moments estimator, can also be used to estimate the CACE. However, the ML-EM method is known to be often more efficient than the instrumental variable approach in estimating the CACE (Imbens & Rubin, 1997; Little & Yau, 1998). Furthermore, likelihood-based estimation methods are essential in our context for elaborate parametric modeling of noncompliance in the CRT context. The current study has limited its investigation to cases without missing outcome data. However, extending the model to accommodate outcome missing data is certainly a possibility in the ML-EM framework, which we plan to explore in the near future.

Another useful feature of the joint analysis is that it allows a flexible modeling of covariate effects. Considering the complex influence of covariates may help in better understanding intervention mechanisms. As demonstrated in simulation results reported in Tables 2 and 3, the effects of within- and between-cluster covariates can be properly modeled in this joint framework. In general, including covariates that are good predictors of compliance is likely to (a) increase precision in estimating compliance status and therefore increase power to detect CACE, (b) decrease sensitivity of CACE estimates to violation of underlying assumptions such as the exclusion restriction, and (c) increase identifiability of CACE when critical identifying assumptions such as the exclusion restriction are relaxed (see Tables 2 and 3 and related discussions). The same benefits apply to CACE estimation in the CRT setting if clustering is taken into account in the analysis. If not, both within- and between-cluster covariates lose predicting power as ICC_C increases, which undermines all three benefits of having good covariate information. Related to the use of covariate information, another possible way of modeling clustering and noncompliance is to utilize the propensity score method, which heavily depends on covariate information. However, little investigation has been done regarding how this method should be implemented in the context of CACE analysis, in particular, simultaneously considering clustering and noncompliance.

In this article, we have focused on the CACE estimation approach and have shown how the significance of the results can be exaggerated when clustering is not considered in the analysis. Not covered by this study is the ITT analysis, which is commonly used to estimate an overall effectiveness of treatment by comparing groups as randomized. Given that we are not interested in compliance-type-specific treatment assignment effects and that the effect of cluster-level randomization can be taken into account using the standard multilevel analysis techniques, it is unclear whether we need to worry about the effect of treatment noncompliance in estimating the ITT effect in CRTs. A failure to consider both clustering and noncompliance has different consequences in ITT analysis, which was dealt with in a separate study (Jo et al., in press). That study showed that ignoring compliance information

may result in substantially decreased power to detect ITT effect, mainly because of compliance ICC. In other words, ITT analysis may also benefit from the analyses that consider both clustering and noncompliance.

On the basis of incorrectly estimated standard errors, gauging statistical power is meaningless. Therefore, power has not been explicitly discussed in the study, though power is of great concern in planning randomized trials. However, once variance inflation is properly taken into account, as demonstrated in the multilevel mixture analysis that considers both data clustering and noncompliance, statistical power regains its validity. How statistical power varies depending on various settings of CRTs with noncompliance remains as a topic for future study.

Another major limitation of the proposed multilevel mixture analysis is that it tends to yield inaccurate cluster-level variance and standard error estimates when the number of clusters is small (e.g., 18 in the JHU PIRC trial), which is a well-known problem in multilevel modeling. Small numbers of clusters are often employed in psychosocial intervention trials, and therefore, the subsequent variance inflation is an important matter to be resolved. Further investigation is needed in this area to provide methods to improve the quality of standard error estimates given small numbers of clusters. Some possible directions include the use of the bootstrap method, the use of the Bayesian method with strong prior information, the use of an approximate F test, and the use of the sandwich estimator. These strategies have been applied in multilevel modeling. However, little is known about how these methods perform with mixture distributions in the CRT context. On the basis of these developments, we expect to establish some guidelines for analyzing CRT data with noncompliance when faced with small numbers of clusters.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Booil Jo's research was supported by National Institute of Mental Health Grants MH066319, MH066247, and MH40859 and by National Institute on Drug Abuse Grant DA11796. Bengt O. Muthén's research was supported by National Institute on Alcohol Abuse and Alcoholism Grant AA00230. We thank participants of the Prevention Science Methodology Group for helpful feedback.

References

- Agresti, A. Categorical data analysis. New York: Wiley; 2002.
- Aitkin M, Longford N. Statistical modeling issues in school effectiveness studies (with discussion). *Journal of the Royal Statistical Society* 1986;149(A):1–43.
- Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 1996;91:444–455.
- Asparouhov, T.; Muthén, BO. Multilevel mixture models. In: Hancock, GR.; Samuelsen, KM., editors. *Advances in latent variable mixture models*. Greenwich, CT: Information Age Publishing; 2008. p. 27–51.
- Bloom HS. Accounting for no-shows in experimental evaluation designs. *Evaluation Review* 1984;8:225–246.
- Brown CH, Liao J. Principles for designing randomized preventive trials in mental health: An emerging developmental epidemiologic paradigm. *American Journal of Community Psychology* 1999;27:677–714.

- Brown CH, Wang W, Kellam SG, Muthén BO, Petras H, Toyinbo P, et al. Methods for testing theory and evaluating impact in randomized field trials: Intent-to-treat analyses for integrating the perspectives of person, place, and time. 2007 Unpublished manuscript.
- Commenges D, Jacqmin H. The intraclass correlation coefficient: Distribution-free definition and test. *Biometrics* 1994;50:517–526. [PubMed: 8068852]
- Dempster A, Laird N, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society* 1977;39(B):1–38.
- Dexter P, Wolinsky F, Gramelspacher G, Zhou XH, Eckert G, Waisburd M, Tierney W. Effectiveness of computer-generated reminders for increasing discussions about advance directives and completion of advance directives. *Annals of Internal Medicine* 1998;128:102–110. [PubMed: 9441569]
- Donner A, Birkett N, Buck C. Randomization by cluster: Sample size requirements and analysis. *American Journal of Epidemiology* 1981;114:906–914. [PubMed: 7315838]
- Donner A, Klar N. Statistical considerations in the design and analysis of community intervention trials. *Journal of Clinical Epidemiology* 1996;49:435–439. [PubMed: 8621994]
- Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics* 2002;58:21–29. [PubMed: 11890317]
- Frangakis CE, Rubin DB, Zhou XH. Clustered encouragement design with individual noncompliance: Bayesian inference and application to advance directive forms. *Biostatistics* 2002;3:147–164. [PubMed: 12933609]
- Goetghebuer E, Molenberghs G. Causal inference in a placebo-controlled clinical trial with binary outcome and ordered compliance. *Journal of the American Statistical Association* 1996;91:928–934.
- Goldstein H. Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika* 1986;73:43–56.
- Haldane JBS. The mean and variance of X_2 , when used as a test of homogeneity, when expectations are small. *Biometrika* 1940;31:346–355.
- Hirano K, Imbens GW, Rubin DB, Zhou XH. Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics* 2000;1:69–88. [PubMed: 12933526]
- Ialongo NS, Werthamer L, Kellam SG, Brown CH, Wang S, Lin Y. Proximal impact of two first-grade preventive interventions on the early risk behaviors for later substance abuse, depression and antisocial behavior. *American Journal of Community Psychology* 1999;27:599–642. [PubMed: 10676542]
- Imbens GW, Angrist J. Identification and estimation of local average treatment effects. *Econometrica* 1994;62:467–476.
- Imbens GW, Rubin DB. Bayesian inference for causal effects in randomized experiments with non-compliance. *Annals of Statistics* 1997;25:305–327.
- Jo B. Estimating intervention effects with noncompliance: Alternative model specifications. *Journal of Educational and Behavioral Statistics* 2002a;27:385–420.
- Jo B. Model misspecification sensitivity analysis in estimating causal effects of interventions with noncompliance. *Statistics in Medicine* 2002b;21:3161–3181. [PubMed: 12375297]
- Jo B. Statistical power in randomized intervention studies with noncompliance. *Psychological Methods* 2002c;7:178–193. [PubMed: 12090409]
- Jo B, Asparouhov T, Muthén BO. Intention-to-treat analysis in cluster randomized trials with noncompliance. *Statistics in Medicine*. (in press).
- Kauermann G, Carroll RJ. A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association* 2001;96:1387–1396.
- Kish, L. Survey sampling. New York: Wiley; 1965.
- Little, RJA.; Rubin, DB. Statistical analysis with missing data. New York: Wiley; 2002.
- Little RJA, Yau L. Statistical techniques for analyzing data from prevention trials: Treatment of no-shows using Rubin's causal model. *Psychological Methods* 1998;3:147–159.
- Longford, N. Random coefficient models. New York: Oxford University Press; 1993.

- MacKinnon JG, White H. Some heteroscedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics* 1985;29:305–325.
- McCullagh, P.; Nelder, JA. *Generalized linear models*. London: Chapman & Hall; 1989.
- McCulloch CE. Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association* 1997;92:162–170.
- McKelvey RD, Zavoina W. A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology* 1975;4:103–120.
- McLachlan, GJ.; Krishnan, T. *The EM algorithm and extensions*. New York: Wiley; 1997.
- McLachlan, GJ.; Peel, D. *Finite mixture models*. New York: Wiley; 2000.
- Murray, DM. *Design and analysis of group-randomized trials*. New York: Oxford University Press; 1998.
- Muthén, BO. Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In: Kaplan, D., editor. *Handbook of quantitative methodology for the social sciences*. Newbury Park, CA: Sage; 2004. p. 345–368.
- Muthén, BO.; Satorra, A. Complex sample data in structural equation modeling. In: Marsden, PV., editor. *Sociological methodology* 1995. Cambridge, MA: Blackwell; 1995. p. 267–316.
- Muthén, LK.; Muthén, BO. *Mplus user's guide*. Los Angeles: Muthén & Muthén; 1998–2007.
- Raudenbush SW. Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods* 1997;2:173–185.
- Raudenbush, SW.; Bryk, AS. *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage; 2002.
- Raudenbush SW, Yang ML, Yosef M. Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation. *Journal of Computational and Graphical Statistics* 2000;9:141–157.
- Robins JM, Greenland S. Adjusting for differential rates of PCP prophylaxis in high- versus low-dose AZT treatment arms in an AIDS randomized trial. *Journal of the American Statistical Association* 1994;89:737–749.
- Rubin DB. Bayesian inference for causal effects: The role of randomization. *Annals of Statistics* 1978;6:34–58.
- Rubin DB. Discussion of “Randomization analysis of experimental data in the Fisher randomization test” by D. Basu. *Journal of the American Statistical Association* 1980;75:591–593.
- Rubin DB. Comment on “Neyman (1923) and causal inference in experiments and observational studies.”. *Statistical Science* 1990;5:472–480.
- Sheiner LB, Rubin DB. Intention-to-treat analysis and the goals of clinical trials. *Clinical Pharmacology and Therapeutics* 1995;57:6–15. [PubMed: 7828382]
- Snijders, TAB.; Bosker, RJ. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage; 1999.
- Snyder J, Reid J, Stoolmiller M, Howe G, Brown CH, Dagne GA, et al. Measurement systems for randomized intervention trials: The role of behavioral observation in studying treatment mediators and short-term outcomes. *Prevention Science* 2006;7:43–56. [PubMed: 16572301]
- Sobel ME. What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference. *Journal of the American Statistical Association* 2006;101:1398–1407.
- Stratelli R, Laird N, Ware JH. Random-effects models for serial observations with binary response. *Biometrics* 1984;40:961–971. [PubMed: 6534418]
- Tanner, M. *Tools for statistical inference: Methods for the exploration of posterior distributions and likelihood functions*. New York: Springer; 1996.
- Thornquist, MD.; Anderson, GL. Small sample properties of generalized estimating equations in group-randomized designs with Gaussian response. Paper presented at the annual meeting of American Public Health Association; Washington, DC. 1992 Oct.
- Titterton, DM.; Smith, AFM.; Makov, UE. *Statistical analysis of finite mixture distributions*. New York: Wiley; 1985.
- West, SG.; Sagarin, BJ. Participant selection and loss in randomized experiments. In: Bickman, L., editor. *Research design*. Thousand Oaks, CA: Sage; 2000. p. 117–154.

Wong GY, Mason WM. The hierarchical logistic regression model for multilevel analysis. *Journal of the American Statistical Association* 1985;80:513–524.

Zeger SL, Karim MR. Generalized linear models with random effects: A Gibbs sampling approach. *Journal of the American Statistical Association* 1991;86:79–86.

Appendix

ML-EM Estimation of Complier Average Causal Effect in the Multilevel Mixture Model Framework

The observed-data two-level likelihood function is based again on the treatment assignment Z and the observed treatment receipt status D . The likelihood for the j th cluster is

$$L_j(\theta|\text{data}) \propto \int \left(\prod_{i \in \{Z_{ij}=1, D_{ij}=0\}} (1 - \pi_{ij}) f_n(y_{ij}|\text{data}, \varepsilon_{nbj}) \times \prod_{i \in \{Z_i=1, D_i=1\}} \pi_{ij} f_c(y_{ij}|\text{data}, \varepsilon_{cbj}) \times \prod_{i \in \{Z_i=0, D_i=0\}} [(1 - \pi_{ij}) f_n(y_{ij}|\text{data}, \varepsilon_{nbj}) + \pi_{ij} f_c(y_{ij}|\text{data}, \varepsilon_{cbj})] \right) \phi(\varepsilon_{cbj}, \varepsilon_{nbj}, \xi_j) d\varepsilon_{cbj} d\varepsilon_{nbj} d\xi_j$$

where f_n and f_c are the probability density of a normal distribution,

$$\begin{aligned} f_n(y_{ij}|\text{data}, \varepsilon_{nbj}) &= \\ \text{Exp} \left(-\frac{(y_{ij} - \alpha_n - \gamma_n Z_{ij} - \lambda'_{nb} x_{bj} - \lambda'_{nw} x_{wij} - \varepsilon_{nbj})^2}{2\sigma_{nw}^2} \right) / (\sqrt{2\pi} \sigma_{nw}), \\ f_c(y_{ij}|\text{data}, \varepsilon_{cbj}) &= \\ \text{Exp} \left(-\frac{(y_{ij} - \alpha_c - \gamma_c Z_{ij} - \lambda'_{cb} x_{bj} - \lambda'_{cw} x_{wij} - \varepsilon_{cbj})^2}{2\sigma_{cw}^2} \right) / (\sqrt{2\pi} \sigma_{cw}), \end{aligned}$$

π_{ij} is the probability of compliance,

$$\pi_{ij} = \frac{\text{Exp}(\beta_0 + \beta'_{1b} x_{bj} + \beta'_{1w} x_{wij} + \xi_j)}{1 + \text{Exp}(\beta_0 + \beta'_{1b} x_{bj} + \beta'_{1w} x_{wij} + \xi_j)},$$

and $\phi(\varepsilon_{cbj}, \varepsilon_{nbj}, \xi_j)$ is the joint density function for the random effects,

$$\phi(\varepsilon_{cbj}, \varepsilon_{nbj}, \xi_j) = \text{Exp}(-\varepsilon_{cbj}^2/(2\sigma_{cb}^2) - \varepsilon_{nbj}^2/(2\sigma_{nb}^2) - \xi_j^2/(2\psi_b^2)) / ((2\pi)^{1.5} \sigma_{cb} \sigma_{nb} \psi_b).$$

The total likelihood function is

$$L(\theta|\text{data}) = \prod_j L_j(\theta|\text{data}).$$

The parameters in the model are

$$\theta = (\alpha_c, \alpha_n, \sigma_{nw}, \sigma_{cw}, \sigma_{nb}, \sigma_{cb}, \psi_b, \lambda_{nw}, \lambda_{cw}, \lambda_{nb}, \lambda_{cb}, \beta_0, \beta_{1w}, \beta_{1b}).$$

By maximizing the total likelihood function with respect to the parameters of interest θ , maximum-likelihood (ML) estimates are obtained. Numerical integration is used to

approximate the integration over the between-level random effects. The unknown compliance status in the control condition and the between-level random effects are handled as missing data via the EM algorithm. Parametric standard errors are computed from the information matrix of the ML estimator using both the first- and the second-order derivatives under the assumption of normally distributed outcomes. In the current study, ML-EM estimation of complier average causal effect was carried out by Mplus Version 4.2 (L. K. Muthén & Muthén, 1998–2007).

Figure 1.
Treatment noncompliance in randomized trials.

Figure 2.

Impact of ICC_C on variance misestimation when $ICC_Y = 0$ (i.e., $ICC_{Y_n} = ICC_{Y_c} = 0$).

Complier and noncomplier outcome means are 0.0, 0.5, and 1.0 SDs apart given treatment assignment. The y-axis represents the coverage rate for nominal 95% confidence intervals.

A: 20 per cluster. B: 40 per cluster. ICC_C = compliance intraclass correlation; ICC_{Y_n} = outcome intraclass correlation for noncompliers; ICC_{Y_c} = outcome intraclass correlation for compliers.

Figure 3.

Impact of varying combinations of ICC_{Y_n} , ICC_{Y_c} , and ICC_C on variance misestimation. Each cluster consists of 20 individuals. Complier and noncomplier outcome means are (A) 0.0, (B) 0.5, and (C) 1.0 SDs apart given treatment assignment. The y-axis represents the coverage rate for nominal 95% confidence intervals. ICC_C = compliance intraclass correlation; ICC_{Y_n} = outcome intraclass correlation for noncompliers; ICC_{Y_c} = outcome intraclass correlation for compliers.

Figure 4. Path Diagram for the model described in Equations 15 and 16. C = compliance status; X_1 = within-cluster covariates; X_2 = between-cluster covariates; Y = outcome; Z = treatment assignment status.

Table 1

Key Parameters and Corresponding Sample Statistics

Parameter	Description	Sample statistic
μ_1	Mean potential outcome if $Z = 1$	y_1
μ_{1n}	Mean potential outcome of noncompliers if $Z = 1$	y_{1n}
μ_{1c}	Mean potential outcome of compliers if $Z = 1$	y_{1c}
μ_0	Mean potential outcome if $Z = 0$	y_0
μ_{0n}	Mean potential outcome of noncompliers if $Z = 0$	Not available
μ_{0c}	Mean potential outcome of compliers if $Z = 0$	Not available
π_c	Mean compliance probability	p_c

Note. Z is treatment assignment status: 1 = treatment condition, 0 = control condition.

Table 2

Simulation: Estimation of CACE Ignoring Cluster Randomization (100 Clusters, 40 per Cluster)

Parameter	True value	Average estimate	Empirical SD	Average SE	95% CI coverage
γ_n (NACE)	-0.200	-0.212	0.121	0.064	0.700
γ_c (CACE)	0.600	0.608	0.132	0.064	0.662
α_n	1.000	1.010	0.101	0.058	0.744
α_c	2.000	1.994	0.111	0.058	0.684
λ_{nx1}	-0.100	-0.099	0.028	0.027	0.940
λ_{cx1}	-0.200	-0.202	0.028	0.027	0.934
λ_{nx2}	0.100	0.100	0.051	0.028	0.712
λ_{cx2}	0.200	0.204	0.067	0.027	0.594
σ_n^2	1.000	0.996	0.047	0.040	0.918
σ_c^2	1.000	0.989	0.055	0.040	0.838
β_0	0.000	0.004	0.164	0.048	0.452
β_{1x1}	0.700	0.515	0.055	0.048	0.066
β_{1x2}	0.700	0.513	0.166	0.050	0.246

Note. CACE = complier average causal effect; CI = confidence interval; NACE = noncomplier average causal effect.

Table 3

Simulation: Estimation of CACE Considering Cluster Randomization (100 Clusters, 40 per Cluster)

Parameter	True value	Average estimate	Empirical SD	Average SE	95% CI coverage
Within level					
$\lambda_{nw}x_1$	-0.100	-0.099	0.026	0.026	0.952
$\lambda_{cw}x_1$	-0.200	-0.200	0.025	0.025	0.952
σ_{nw}^2	0.900	0.900	0.035	0.036	0.952
σ_{cw}^2	0.800	0.799	0.032	0.032	0.934
β_0	0.000	0.008	0.212	0.210	0.942
$\beta_{lw}x_1$	0.700	0.702	0.058	0.058	0.944
Between level					
γ_n (NACE)	-0.200	-0.204	0.105	0.102	0.946
γ_c (CACE)	0.600	0.603	0.119	0.123	0.952
a_n	1.000	1.002	0.087	0.087	0.942
a_c	2.000	1.999	0.099	0.101	0.942
$\lambda_{nb}x_2$	0.100	0.099	0.048	0.047	0.936
$\lambda_{cb}x_2$	0.200	0.202	0.059	0.058	0.924
σ_{nb}^2	0.100	0.095	0.025	0.025	0.898
σ_{cb}^2	0.200	0.190	0.045	0.041	0.876
$\beta_{lb}x_2$	0.700	0.692	0.208	0.196	0.932
ψ_b^2	2.191	2.162	0.488	0.482	0.930

Note. CACE = complier average causal effect; CI = confidence interval; NACE = noncomplier average causal effect.