STATISTICS IN MEDICINE
Statist. Med. 2002; **21**:3161–3181 (DOI: 10.1002/sim.1267)

Model misspecification sensitivity analysis in estimating causal effects of interventions with non-compliance

Booil Jo*,†

Division of Social Research Methodology, Graduate School of Education & Information Studies, University of California, Los Angeles, CA, U.S.A.

SUMMARY

Randomized trials often face complications in assessing the effect of treatment because of study participants' non-compliance. If compliance type is observed in both the treatment and control conditions, the causal effect of treatment can be estimated for a targeted subpopulation of interest based on compliance type. However, in practice, compliance type is not observed completely. Given this missing compliance information, the complier average causal effect (CACE) estimation approach provides a way to estimate differential effects of treatments by imposing the exclusion restriction for non-compliers. Under the exclusion restriction, the CACE approach estimates the effect of treatment assignment for compliers, but disallows the effect of treatment assignment for non-compliers. The exclusion restriction plays a key role in separating outcome distributions based on compliance type. However, the CACE estimate can be substantially biased if the assumption is violated. This study examines the bias mechanism in the estimation of CACE when the assumption of the exclusion restriction is violated. How covariate information affects the sensitivity of the CACE estimate to violation of the exclusion restriction assumption is also examined. Copyright © 2002 John Wiley & Sons, Ltd.

KEY WORDS: randomized trial; non-compliance; CACE; exclusion restriction; bias mechanism; sensitivity analysis

1. INTRODUCTION

Non-compliance is a common problem in randomized trials involving human participants. Since both the standard ITT (intent-to-treat) analysis and as-treated analysis may provide biased estimates of treatment effects for compliers in the presence of non-compliance, the possibility of estimating treatment effects only for compliers (CACE, complier average causal effect) has been explored [1–5].

* Correspondence to: Booil Jo, Division of Social Research Methodology, Graduate School of Education & Information Studies, University of California, Los Angeles, CA 90095-1521, U.S.A.
† E-mail: booil@ucla.edu

Contract/grant sponsor: NIMH; contract/grant number: P30 MH38330, P50 MH38725
Contract/grant sponsor: NIAAA; contract/grant number: K02 AA 00230-01
Contract/grant sponsor: NIDA; contract/grant number: RO1 DA11796-01A1

Received September 2000
Accepted February 2002

Under a series of statistical assumptions, CACE estimation provides an unbiased treatment effect estimate for compliers. In the estimation of CACE, the exclusion restriction is one of the critical underlying assumptions. This assumption provides the basis for identifiability in CACE models, given that compliance status is not observed completely. Under this assumption, the difference in outcome between the treatment and the control condition is allowed for compliers, but is not allowed for never-takers (individuals who would not receive the treatment regardless of whether it is offered) or for always-takers (individuals who would receive the treatment regardless of whether it is offered). The assumption of the exclusion restriction plays a critical role in simplifying methodological difficulties involved in CACE approaches. However, this assumption can often be unrealistic in practice [6–8].

When the exclusion restriction is violated, the causal effect of treatment not only can be understated, but also can be exaggerated depending on how the assignment of treatment affects non-compliers. Hirano *et al.* [6] demonstrated the impact of the violation of the exclusion restriction in application to a study of the effect of an influenza vaccine in an encouragement design [9]. Their findings showed little evidence of benefit from the vaccine after taking into account the effect of treatment assignment (that is, encouragement) on always-takers. In contrast, positive effect of the vaccine was found when the effect of treatment assignment on always-takers was ignored. In the Job Search Intervention Study [10, 11], shown as an example in this study, the assumption of the exclusion restriction is more likely to be violated for never-takers. That is, never-takers in the treatment condition could be demoralized by failing to take the intervention opportunity. This negative psychological effect would not occur for never-takers in the control condition, since the treatment is not offered. In this situation, the CACE estimate can be understated by ignoring the effect of treatment assignment on never-takers.

This study examines the bias mechanism in the estimation of CACE when the assumption of the exclusion restriction is violated. It is demonstrated how the magnitude of bias is affected by compliance rate and the effect of treatment assignment on non-compliers. The study also examines how covariate information affects the sensitivity of the CACE estimate to violation of the exclusion restriction assumption. Simulation studies demonstrate that bias in the CACE estimate due to model misspecification can be reduced substantially by including covariates that are good predictors of compliance. To demonstrate CACE estimation when the assumption of the exclusion restriction is potentially violated, the Job Search Intervention Study for unemployed workers is presented as an example. Maximum likelihood estimation using the EM algorithm is employed in the estimation of CACE in the study.

2. DATA: THE JOB SEARCH INTERVENTION STUDY

The Job Search Intervention Study (JOBS II) is a randomized field experiment intended to prevent poor mental health and to promote high-quality re-employment. The outcome measure that will be the focus in the current study is depression of study participants. Among the mental health problems associated with job loss, depressive symptoms are the most commonly reported [12, 13]. The experimental condition consisted of five half-day training sessions, which included the application of problem-solving and decision-making processes, inoculation against setbacks, provision of social support and positive regard from the trainers, and learning and practising job search skills. The control condition consisted of a booklet briefly describing

job search methods and tips. Although the effect of receiving the booklet was expected to be very small, the booklet was also mailed to treatment condition individuals, including people who did not show up at the intervention seminars, to prevent possible differences in outcome between the treatment and control conditions due to the booklet.

The problem of non-compliance arises in JOBS II because a substantial proportion of individuals who were assigned to the intervention condition did not show up to the intervention. Among study participants assigned to the intervention condition, 55 per cent attended at least one session. Among attendees, 82 per cent attended four or five sessions (mean = 4.3 sessions, median = 5 sessions). A previous study [14] using the instrumental variable approach [2], which normally yields a very similar outcome to CACE analysis, showed that the choice of threshold made little difference in the JOBS Intervention Study, possibly because individuals assigned to the intervention condition either attended most sessions or did not attend any. Little and Yau [5] also employed the binary compliance approach in their CACE analysis of JOBS II using the ML-EM method. In line with previous analyses of JOBS II, the current study also defines compliance as having attended at least one out of five total sessions. However, note that sensitivity of the CACE estimate to the choice of threshold may vary in other situations and needs to be carefully examined.

Given the rate of non-compliance in JOBS II, one can expect that treatment efficacy can be better estimated for compliers if CACE estimation is applied instead of standard ITT analysis. However, special attention is needed in applying the CACE estimation method to JOBS II, because the assumption of the exclusion restriction is questionable. One possible explanation for this phenomenon is that never-takers in the treatment condition become demoralized by failing to take the intervention opportunity. Considering that the major outcome analysed in this study is depression mainly caused by unemployment, this interpretation seems plausible in JOBS II. However, this explanation is not definitive and may not apply to different situations. As in JOBS II, blinding or placebo-control is hard to implement in most randomized field experiments. Although the exclusion restriction assumption is always questionable in this situation, little has been studied about how treatment assignment influences study participants who decide not to comply with the treatment.

3. CACE ESTIMATION UNDER THE ASSUMPTION OF THE EXCLUSION RESTRICTION

Assume the simplest experimental setting where there is only one outcome measure (Y), treatment assignment (Z) is binary (1 = treatment, 0 = control), and the treatment received (D) has only two levels (1 = received, 0 = not received). The behaviour types (C_i) of the subjects based on combinations of Z and D can be classified into four categories based on Rubin's causal model approach, where the possibility of statistical causal inference is built at the individual level [15–18]. Angrist *et al.* [1] labelled the four categories as complier, never-taker, defier and always-taker. Let $D_i(1)$ denote the potential treatment receipt status for individual i when assigned to the treatment condition, and $D_i(0)$ denote the potential treatment receipt status for individual i when assigned to the control condition. Compliers are subjects who do what they are assigned to do ($D_i(1) = 1$ and $D_i(0) = 0$). Never-takers are subjects who do not receive the treatment even if they are assigned to the treatment condition ($D_i(1) = 0$ and $D_i(0) = 0$). Defiers are the subjects who do the opposite of what they are assigned to do

($D_i(1)=0$ and $D_i(0)=1$). Always-takers are the subjects who always receive the treatment, no matter which condition they are assigned to ($D_i(1)=1$ and $D_i(0)=1$).

Among these four types of subjects, the CACE approach focuses on the estimation of causal effect of treatment assignment for compliers. The following assumptions 1 to 5 are critical in estimating CACE:

Assumption 1 (*randomization*). Treatment assignment is random.
Assumption 2 (*stable unit treatment value*, *SUTVA*). Potential outcomes for each person are unrelated to the treatment status of other individuals [17–19].
Assumption 3 (*exclusion restriction*). For never-takers and always-takers, the distributions of the potential outcomes are independent of the treatment assignment [1]. That is, $Y_i(0, D_i(0)) = Y_i(1, D_i(1))$ for units with $D_i(0)=D_i(1)=0$ or $D_i(0)=D_i(1)=1$.
Assumption 4 (*monotonicity*). There are no defiers [3].
Assumption 5 (*non-zero average causal effect of Z on D*). The average causal effect of Z on D is not equal to zero [1].

In addition, the current study also assumes that there are no always-takers (assumption 6) based on the JOBS II example. In JOBS II, neither defier nor always-taker was a likely compliance option, since study participants were prohibited from receiving a different intervention condition than the one that they were assigned to. However, unlike monotonicity, the assumption of having no always-takers is not critical in estimating the CACE, and can be relaxed depending on the situation.

Assumption 6. There are no always-takers.

Under assumptions 4 and 6, the possible compliance behaviour types (C_i) can be reduced to

$$C_i = \begin{cases} c \text{ (complier)} & \text{if } D_i(1)=1 \text{ and } D_i(0)=0 \\ n \text{ (never-taker)} & \text{if } D_i(1)=0 \text{ and } D_i(0)=0 \end{cases}$$

Let $C(t)=\{i \mid C_i=t\}$ for $t \in \{c, n\}$. The differential average causal effect of treatment assignment based on compliance type can be defined as

$$\text{ITT}_t = \sum_{i \in C(t)} [Y_i(1, D_i(1)) - Y_i(0, D_i(0))]/N_t \tag{1}$$

where $Y_i(1, D_i(1))$ denotes the potential outcome for individual i with treatment receipt status D_i when $Z_i=1$, and $Y_i(0, D_i(0))$ denotes the potential outcome for individual i with treatment receipt status D_i when $Z_i=0$. N_t is the number of individuals of compliance type t.

The causal effect of treatment assignment cannot be estimated for individual i, since two potential outcomes (that is, $Y_i(1, D_i(1))$ $Y_i(0, D_i(0))$) cannot be jointly observed. However, the causal effect of treatment assignment can be estimated at the average level. The average causal effect of treatment assignment for compliers ($\text{ITT}_c = \text{CACE}$) can be defined as

$$\text{CACE} = \mu_{1c} - \mu_{0c} \tag{2}$$

where μ_{1c} denotes population mean potential outcome for compliers if $Z=1$, and μ_{0c} denotes population mean potential outcome for compliers if $Z=0$.

This study assumes that the average causal effect of treatment assignment (ITT_t) does not vary across different values of covariates. Under the exclusion restriction, assumption 7 is not critical in the estimation of CACE and can be relaxed depending on situations and research questions.

Assumption 7 (*additivity*). The average causal effect of treatment assignment is constant regardless of varying values of covariates.

The current study employs a maximum likelihood estimation approach, which is known to be often more efficient than the traditional IV (instrumental variable) approach in the estimation of CACE [4, 5]. Given that compliance type C_i cannot be observed in the control condition, the observed-data likelihood function assuming a normally distributed outcome is

$$L(\theta \mid \text{data}) \propto \prod_{i \in \{Z_i=1, D_i=0\}} \pi_{\text{n}} f(y_i \mid \mu_{1\text{n}}, \sigma^2) \times \prod_{i \in \{Z_i=1, D_i=1\}} \pi_{\text{c}} f(y_i \mid \mu_{1\text{c}}, \sigma^2) \times \prod_{i \in \{Z_i=0, D_i=0\}} [\pi_{\text{n}} f(y_i \mid \mu_{0\text{n}}, \sigma^2) + \pi_{\text{c}} f(y_i \mid \mu_{0\text{c}}, \sigma^2)] \tag{3}$$

where $\theta = (\pi_{\text{n}}, \pi_{\text{c}}, \mu_{1\text{n}}, \mu_{1\text{c}}, \mu_{0\text{n}}, \mu_{0\text{c}}, \sigma^2)$ is the set of parameters in the model, and $f(y_i \mid \mu, \sigma^2)$ denotes the probability density of a normal distribution with mean μ and variance σ^2. π_{n} is the proportion of never-takers in the population, and π_{c} is the proportion of compliers in the population. $\mu_{1\text{n}}$ denotes population mean potential outcome for never-takers if $Z = 1$, and $\mu_{0\text{n}}$ denotes population mean potential outcome for never-takers if $Z = 0$.

By maximizing the likelihood in equation (3) with respect to the parameters of interest θ, ML estimates are obtained. The unknown compliance status (C) in the control condition is handled as missing data via the EM algorithm [20–23]. π_{c} and $\pi_{\text{n}} (= 1 - \pi_{\text{c}})$ are parameters that determine the distribution of C. 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. In the current study, ML-EM estimation of CACE was carried out by the M*plus* program [24]. 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.

Based on equation (3), three directly estimable population means can be expressed in terms of model parameters as

$$\mu_{1\text{n}} = \alpha_{\text{n}} + \gamma_{\text{n}} \tag{4}$$

$$\mu_{1\text{c}} = \alpha_{\text{c}} + \gamma_{\text{c}} \tag{5}$$

$$\mu_0 = \pi_{\text{n}} \alpha_{\text{n}} + \pi_{\text{c}} \alpha_{\text{c}} \tag{6}$$

where α_{n} corresponds to $\mu_{0\text{n}}$, α_{c} corresponds to $\mu_{0\text{c}}$, and μ_0 is the overall population mean potential outcome if $Z = 0$. γ_{n} represents the average causal effect of treatment assignment for never-takers (ITT_{n}), and γ_{c} represents the average causal effect of treatment assignment for

compliers ($ITT_c = CACE$). Under the assumption of the exclusion restriction, $\gamma_n = 0$. Therefore, α_n is directly identified as μ_{1n} from equation (4).

From equations (4) and (6), α_c can be identified as

$$\alpha_c = \frac{\mu_0 - \pi_n \mu_{1n}}{\pi_c} \tag{7}$$

From equations (5) and (7), γ_c can then be identified as

$$\gamma_c = \mu_{1c} - \frac{\mu_0 - \pi_n \mu_{1n}}{\pi_c} = \frac{\mu_1 - \mu_0}{\pi_c} \tag{8}$$

where assumption 5 excludes the possibility of a zero denominator (that is, $\pi_c > 0$).

Under assumptions 1 to 6, the approximately unbiased estimator of the average causal effect of treatment assignment for compliers can then be defined as

$$\hat{\gamma}_c = \bar{y}_{1c} - \frac{\bar{y}_0 - p_n \bar{y}_{1n}}{p_c} = \frac{\bar{y}_1 - \bar{y}_0}{p_c} \tag{9}$$

where $\bar{y}_{1c}$ is the sample mean outcome of the treatment group compliers, $\bar{y}_0$ is the sample mean outcome of the control group, $\bar{y}_1$ is the sample mean outcome of the treatment group, $\bar{y}_{1n}$ is the sample mean outcome of the treatment group never-takers, and p_c is the proportion of compliers in the treatment condition ($1 - p_c = p_n$).

3.1. Bias mechanism in a misspecified model

It is shown in the previous section that the assumption of the exclusion restriction plays a critical role in providing identifiability in CACE models, given that compliance information is missing in the control condition. However, assuming the exclusion restriction may cause bias in the CACE estimate, if the assumption does not hold. In line with Angrist *et al.* [1], this section examines the bias mechanism in CACE estimation focusing on compliance rate (π_c) and the effect of treatment assignment on non-compliers ($ITT_n = \gamma_n$).

If γ_n is not zero, α_n cannot be directly identified as μ_{1n}. Instead, correct specification of α_n from equation (4) is

$$\alpha_n = \mu_{1n} - \gamma_n \tag{10}$$

Equation (10) shows that the estimator of α_n will be biased as much as γ_n, if γ_n is misspecified as zero. For example, if true α_n is 1.0, true μ_{1n} is 1.2 and true γ_n is 0.2, then the estimator of α_n will be forced to be 1.2 if the exclusion restriction is imposed (that is, $\gamma_n = 0$).

The bias in the estimation of α_n is transferred to α_c through equation (6). The correct specification of α_c from equations (4) and (6) is

$$\alpha_c = \frac{\mu_0 - \pi_n \mu_{1n}}{\pi_c} + \frac{\pi_n \gamma_n}{\pi_c} \tag{11}$$

Equation (11) shows that the estimator of α_c will be biased as much as $-\pi_n \gamma_n / \pi_c$, if γ_n is misspecified as zero (compare equations (7) and (11)). For example, if true α_c is 1.5, $\pi_c = \pi_n = 0.5$, and true γ_n is 0.2, then the estimator of α_c will be forced to be 1.3 by imposing the exclusion restriction.

The bias in the estimation of α_n and α_c due to misspecification of γ_n is then transferred to γ_c. The correct specification of γ_c from equations (5) and (11) is

$$\gamma_c = \mu_{1c} - \frac{\mu_0 - \pi_n \mu_{1n}}{\pi_c} - \frac{\pi_n \gamma_n}{\pi_c} \tag{12}$$

Equation (12) shows that the estimator of CACE (γ_c) will be biased as much as $\pi_n\gamma_n/\pi_c$, if γ_n is misspecified as zero. The magnitude of bias is affected by π_n, π_c and γ_n. The bias in the CACE (γ_c) estimate increases if the average causal effect of treatment assignment for never-takers (γ_n) increases and compliance rate (π_c) decreases.

In some situations, γ_n may have the opposite direction to that of CACE. Being assigned to the treatment condition may have a negative psychological impact on never-takers, since they failed to take the given treatment. The directions of γ_n and CACE are opposite in this case, assuming that treatment has a positive impact on compliers. Assume that true CACE is -0.6 and true γ_n is 0.2, which is about one-third of the magnitude of CACE. If $\pi_c = 0.5$, the CACE estimate will be biased as much as 0.2 ($\pi_n\gamma_n/\pi_c = 0.5 \times 0.2/0.5$) by imposing the exclusion restriction. Therefore, the estimated CACE will be around -0.4, which is two-thirds of the magnitude of true CACE. If $\pi_c = 0.3$, the CACE estimate will be biased as much as 0.47 ($\pi_n\gamma_n/\pi_c = 0.7 \times 0.2/0.3$) by imposing the exclusion restriction. In this case, the estimated CACE will be around -0.13, which is less than one-third of the magnitude of true CACE. This implies that if compliance rate is very low, violation of the exclusion restriction assumption can cause a substantial bias in the CACE estimate even when the effect of treatment assignment on never-takers is trivial.

In some situations, γ_n may have the same direction as CACE. Being assigned to the treatment condition may have a positive psychological impact on never-takers, even though they decide not to receive the treatment. The directions of γ_n and CACE are the same in this situation, assuming that treatment has a positive impact on compliers. If true CACE is -0.6, true γ_n is -0.2 and $\pi_c = 0.5$, the CACE estimate will be biased as much as -0.2 ($\pi_n\gamma_n/\pi_c = 0.5 \times -0.2/0.5$) by imposing the exclusion restriction. Therefore, the estimated CACE will be around -0.8, which implies that the CACE estimate can be also exaggerated when the exclusion restriction is violated.

The bias mechanism discussed in this section can be generalized to situations where there are both never-takers and always-takers (see Appendix). However, note that the impact of combined bias on the CACE estimate can be small when there are two types of non-compliers, although the exclusion restriction is substantially violated for both types of non-compliers, because the direction of bias can be different depending on the type of non-compliance.

3.2. The role of covariates in a misspecified model

It is demonstrated in the previous section that the size of bias in the CACE estimate due to violation of the exclusion restriction is affected by the compliance rate and the effect of treatment assignment on non-compliers. Given that these factors are not easily controllable in practice, the current study focuses on the use of covariate information to reduce the bias due to violation of the exclusion restriction. This section demonstrates how covariate information affects the sensitivity of the CACE estimate to violation of the exclusion restriction.

Let $c_i=0$ and $n_i=1$ if $i\in C(n)$, and $c_i=1$ and $n_i=0$ if $i\in C(c)$. Consider a continuous outcome variable Y for individual i with compliance status c_i and n_i

$$Y_i=\alpha_{\mathrm{n}}n_i+\alpha_{\mathrm{c}}c_i+\gamma_{\mathrm{n}}n_iZ_i+\gamma_{\mathrm{c}}c_iZ_i+\lambda x_i+\varepsilon_i \tag{13}$$

where x is a vector of pretreatment covariates that predict both the outcome measure y and compliance status C. $\alpha_{\mathrm{n}}+\lambda x_i$ represents the potential mean outcome for control condition never-takers with covariates x_i, and $\alpha_{\mathrm{c}}+\lambda x_i$ represents the potential mean outcome for control condition compliers with covariates x_i. For simplicity, it is assumed that covariate effects on outcome is the same for compliers and non-compliers (that is, $\lambda=\lambda_{\mathrm{c}}=\lambda_{\mathrm{n}}$). γ_{n} represents the average causal effect of treatment assignment for never-takers ($\mathrm{ITT_n}$), and γ_{c} represents the average causal effect of treatment assignment for compliers ($\mathrm{ITT_c}=\mathrm{CACE}$). ε_i is a normally distributed residual with zero mean and variance σ^2.

The logistic regression of c on x is described as

$$\begin{aligned} P(c_i=1\,|\,x_i)&=\pi_{\mathrm{c}i}\\ P(c_i=0\,|\,x_i)&=1-\pi_{\mathrm{c}i}=\pi_{\mathrm{n}i}\\ \mathrm{logit}(\pi_{\mathrm{c}i})&=\beta_0+\beta_1x_i \end{aligned} \tag{14}$$

where $\pi_{\mathrm{c}i}$ denotes the probability of being a complier, $\pi_{\mathrm{n}i}$ denotes the probability of being a never-taker, β_0 represents a logit intercept, and β_1 is a vector of logit coefficients. Note that the probability of being a complier ($\pi_{\mathrm{c}i}$) varies depending on x_i (equation (14)), but the average causal effect of treatment assignment (γ_t) does not vary across different values of x_i (equation (13)). Assumption 7 is tested in the JOBS II example through the likelihood ratio test between models with and without imposing the assumption (no significant difference was found in model fit).

Assume that there is only one binary covariate X that predicts both C and y. The observed-data likelihood in equation (3) can be modified as

$$\begin{aligned} L(\theta\,|\,\mathrm{data})\propto&\prod_{i\in\{Z_i=1,D_i=0,X_i=0\}}\pi_{\mathrm{n},X=0}f(y_i\,|\,\mu_{1\mathrm{n},X=0},\sigma^2)\\ &\times\prod_{i\in\{Z_i=1,D_i=0,X_i=1\}}\pi_{\mathrm{n},X=1}f(y_i\,|\,\mu_{1\mathrm{n},X=1},\sigma^2)\\ &\times\prod_{i\in\{Z_i=1,D_i=1,X_i=0\}}\pi_{\mathrm{c},X=0}f(y_i\,|\,\mu_{1\mathrm{c},X=0},\sigma^2)\\ &\times\prod_{i\in\{Z_i=1,D_i=1,X_i=1\}}\pi_{\mathrm{c},X=1}f(y_i\,|\,\mu_{1\mathrm{c},X=1},\sigma^2)\\ &\times\prod_{i\in\{Z_i=0,D_i=0,X_i=0\}}[\pi_{\mathrm{n},X=0}f(y_i\,|\,\mu_{0\mathrm{n},X=0},\sigma^2)+\pi_{\mathrm{c},X=0}f(y_i\,|\,\mu_{0\mathrm{c},X=0},\sigma^2)]\\ &\times\prod_{i\in\{Z_i=0,D_i=0,X_i=1\}}[\pi_{\mathrm{n},X=1}f(y_i\,|\,\mu_{0\mathrm{n},X=1},\sigma^2)+\pi_{\mathrm{c},X=1}f(y_i\,|\,\mu_{0\mathrm{c},X=1},\sigma^2)] \end{aligned} \tag{15}$$

where the binary covariate X has two values ($X=0$ or $X=1$). The proportions of compliers and never-takers in the population vary across different values of X. The population mean potential outcomes for compliers and never-takers vary across different values of X.

Based on equation (15), the average causal effect of treatment assignment for compliers adjusted for covariate X can be defined as

$$\text{CACE} = \mu_{1c,X=0} - \mu_{0c,X=0} = \mu_{1c,X=1} - \mu_{0c,X=1} \tag{16}$$

where $\mu_{0c,X=0}$ and $\mu_{0c,X=1}$ are not directly estimable quantities.

Based on equations (13) and (15), six directly estimable population means can be expressed in terms of model parameters as

$$\mu_{1n,X=0} = \alpha_n + \gamma_n \tag{17}$$

$$\mu_{1n,X=1} = \alpha_n + \gamma_n + \lambda \tag{18}$$

$$\mu_{1c,X=0} = \alpha_c + \gamma_c \tag{19}$$

$$\mu_{1c,X=1} = \alpha_c + \gamma_c + \lambda \tag{20}$$

$$\mu_{0,X=0} = \pi_{n,X=0}\alpha_n + \pi_{c,X=0}\alpha_c \tag{21}$$

$$\mu_{0,X=1} = \pi_{n,X=1}\alpha_n + \pi_{c,X=1}\alpha_c + \lambda \tag{22}$$

where α_n corresponds to $\mu_{0n,X=0}$, and α_c corresponds to $\mu_{0c,X=0}$. $\mu_{0,X=0}$ is the population mean potential outcome if $Z=0$ and $X=0$, and $\mu_{0,X=1}$ is the population mean potential outcome if $Z=0$ and $X=1$.

Equations (17) to (22) show that there are two different sources of information that contribute to the identification of CACE: the exclusion restriction and the covariate. Although both sources of information affect the CACE estimate simultaneously, how covariate information moderates bias due to violation of the exclusion restriction can be better understood by looking at the two sources of information separately.

First, α_n, α_c and γ_c can be identified based on directly estimable quantities relying on the exclusion restriction assumption, but ignoring the association between compliance and covariate. Let us call these tentative estimators $\hat{\alpha}_n^{t_E}$, $\hat{\alpha}_c^{t_E}$ and $\hat{\gamma}_c^{t_E}$.

According to the exclusion restriction assumption, $\gamma_n = 0$. Therefore, $\hat{\alpha}_n^{t_E}$ can be defined from equation (17) as

$$\hat{\alpha}_n^{t_E} = \hat{\mu}_{1n,X=0} \tag{23}$$

where $\hat{\mu}_{1n,X=0}$ is an ML estimate of $\mu_{1n,X=0}$. If the exclusion restriction holds, $\hat{\alpha}_n^{t_E}$ is an unbiased estimator of α_n.

Based on equation (21), $\hat{\alpha}_c^{t_E}$ is defined as

$$\hat{\alpha}_c^{t_E} = \frac{\hat{\mu}_{0,X=0} - \hat{\pi}_{n,X=0}\hat{\alpha}_n^{t_E}}{\hat{\pi}_{c,X=0}} \tag{24}$$

where $\hat{\mu}_{0,X=0}$, $\hat{\pi}_{n,X=0}$ and $\hat{\pi}_{c,X=0}$ are ML estimates of $\mu_{0,X=0}$, $\pi_{n,X=0}$ and $\pi_{c,X=0}$. $\hat{\alpha}_n^{t_E}$ is defined in equation (23).

Then, $\hat{\gamma}_c^{t_E}$ can be defined from equation (19) as

$$\hat{\gamma}_c^{t_E} = \hat{\mu}_{1c,X=0} - \hat{\alpha}_c^{t_E} \tag{25}$$

where $\hat{\mu}_{1c,X=0}$ is an ML estimate of $\mu_{1c,X=0}$, and $\hat{\alpha}_c^{t_E}$ is defined in equation (24).

The estimator of λ can be defined based on observable quantities from equations (17) and (18) (or from (19) and (20)) as

$$\hat{\lambda} = \hat{\mu}_{1n,X=1} - \hat{\mu}_{1n,X=0} \tag{26}$$

where $\hat{\mu}_{1n,X=1}$ and $\hat{\mu}_{1n,X=0}$ are ML estimates of $\mu_{1n,X=1}$ and $\mu_{1n,X=0}$.

Second, α_n, α_c and γ_c can be identified based on directly estimable quantities relying on the association between compliance and covariate, but ignoring the exclusion restriction. Let us call these tentative estimators $\hat{\alpha}_n^{t_X}$, $\hat{\alpha}_c^{t_X}$ and $\hat{\gamma}_c^{t_X}$. The key to the identification of $\hat{\alpha}_n^{t_X}$, $\hat{\alpha}_c^{t_X}$ and $\hat{\gamma}_c^{t_X}$ is the presence of covariate X, which has non-zero β_1 in equation (14) (that is, $\pi_{n,X=0} \neq \pi_{n,X=1}$ and $\pi_{c,X=0} \neq \pi_{c,X=1}$).

From equations (21) and (22), $\hat{\alpha}_n^{t_X}$ can be defined as

$$\hat{\alpha}_n^{t_X} = \frac{\hat{\pi}_{c,X=0}\hat{\mu}_{0,X=1} - \hat{\pi}_{c,X=1}\hat{\mu}_{0,X=0} - \hat{\pi}_{c,X=0}\hat{\lambda}}{\hat{\pi}_{c,X=0} - \hat{\pi}_{c,X=1}} \tag{27}$$

where $\hat{\pi}_{c,X=0}$, $\hat{\pi}_{c,X=1}$ and $\hat{\mu}_{0,X=1}$ are ML estimates of $\pi_{c,X=0}$, $\pi_{c,X=1}$ and $\mu_{0,X=1}$. $\hat{\lambda}$ is defined in equation (26).

From equation (21), $\hat{\alpha}_c^{t_X}$ can be defined as

$$\hat{\alpha}_c^{t_X} = \frac{\hat{\mu}_{0,X=0} - \hat{\pi}_{n,X=0}\hat{\alpha}_n^{t_X}}{\hat{\pi}_{c,X=0}} \tag{28}$$

where $\hat{\alpha}_n^{t_X}$ is defined in equation (27).

From equation (19), $\hat{\gamma}_c^{t_X}$ can then be defined as

$$\hat{\gamma}_c^{t_X} = \hat{\mu}_{1c,X=0} - \hat{\alpha}_c^{t_X} \tag{29}$$

where $\hat{\alpha}_c^{t_X}$ is defined in equation (28).

If there is no association between compliance and covariate (that is, $\beta_1 = 0$), $\pi_{n,X=0} = \pi_{n,X=1}$ and $\pi_{c,X=0} = \pi_{c,X=1}$ in equations (21) and (22). In other words, covariate X does not carry information to identify α_n and α_c. In this case, identification of CACE is completely dependent on the assumption of the exclusion restriction. Therefore, the CACE estimate is the same as $\hat{\gamma}_c^{t_E}$, which is an unbiased estimator of CACE if the exclusion restriction assumption holds. If the exclusion restriction is violated, $\hat{\gamma}_c^{t_E}$ is a biased estimate of CACE, where the degree of bias depends on the size of true γ_n and the compliance rate (see bias mechanism in Section 3.1).

If the covariate is a perfect predictor of compliance, the CACE estimate is the same as $\hat{\gamma}_c^{t_X}$, which is an unbiased estimator of CACE regardless of whether the exclusion restriction holds. That is, the estimator of α_c will not be biased due to the violation of the exclusion restriction, although the estimator of α_n can be still distorted due to the restriction in equations (17) and (18). Consequently, the CACE estimate will not be biased despite the model misspecification. However, covariates that are perfect predictors of compliance usually do not exist in practice, which excludes the possibility of complete elimination of bias based on covariate information.

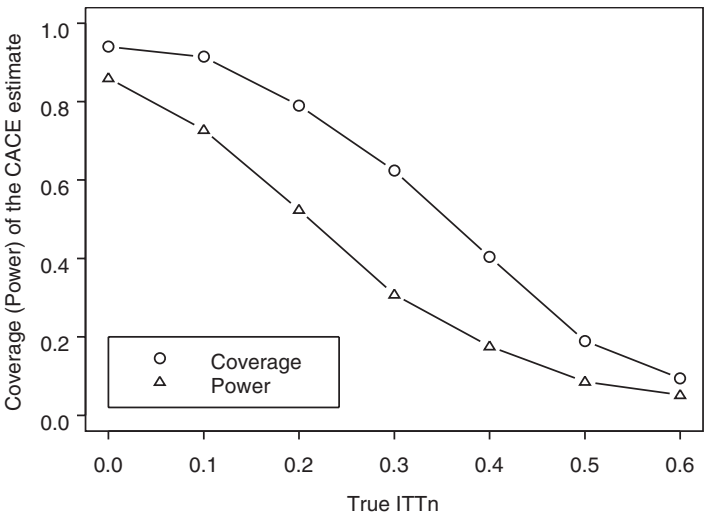

Figure 1. Simulation: sensitivity of the CACE estimate to violation of the exclusion restriction assumption when there is no covariate.

In most situations, where the association between compliance and covariates is neither zero nor perfect, both the exclusion restriction assumption and covariate information affect the identification of CACE. When the assumption of the exclusion restriction is violated, covariate information plays a critical role in reducing the bias in the CACE estimate. As the level of association between compliance and the covariate increases, the bias due to the violation of the exclusion restriction decreases, because parameter estimates are adjusted more towards $\hat{\alpha}_{\mathrm{n}}^{tx}$, $\hat{\alpha}_{\mathrm{c}}^{tx}$ and $\hat{\gamma}_{\mathrm{c}}^{tx}$, which are not affected by the exclusion restriction. Even when the assumption of the exclusion restriction holds, having covariates that are good predictors of compliance is important, because this increases the precision of the CACE estimate [25].

4. SIMULATION STUDIES

4.1. CACE estimation with no covariate

This section demonstrates the quality of the CACE estimate and statistical power when the exclusion restriction assumption is violated at different levels. The simulation results presented in Figure 1 and Table I are based on 500 replications with a sample size of 500. The true values of key parameters were chosen based on the JOBS II Study example. Equal probability of treatment/control assignment and 50 per cent compliance rate were used as true values for all simulation settings in this study.

Without covariates, the continuous outcome Y generated for CACE estimation can be described as

$$Y_i = \alpha_{\mathrm{n}} n_i + \alpha_{\mathrm{c}} c_i + \gamma_{\mathrm{n}} n_i Z_i + \gamma_{\mathrm{c}} c_i Z_i + \varepsilon_i \tag{30}$$

Table I. Simulation: estimation of CACE with no covariate (true CACE $=-0.6$, true $\alpha_n = 1.0$, true $\alpha_c = 1.5$).

ITT_n	CACE		α_n		α_c	
True	Estimate	SE	Estimate	SE	Estimate	SE
0.0	−0.583	0.180	1.008	0.087	1.484	0.155
0.1	−0.492	0.193	1.104	0.094	1.393	0.170
0.2	−0.387	0.197	1.206	0.096	1.288	0.175
0.3	−0.279	0.196	1.310	0.095	1.180	0.173
0.4	−0.178	0.187	1.410	0.091	1.079	0.164
0.5	−0.093	0.176	1.503	0.085	0.994	0.151
0.6	−0.025	0.165	1.589	0.081	0.926	0.139

Under the assumption of the exclusion restriction, ITT_n (γ_n) is fixed at zero in the estimation of CACE. How seriously the exclusion restriction assumption is violated depends on the true value of ITT_n, which was set at various values in the data generation. Opposite directions are chosen for the true values of CACE and ITT_n in simulation settings based on JOBS II. The true treatment effect for compliers (CACE) is -0.6, which is approximately 0.6 in terms of effect size, assuming that decrease in the outcome is desirable (for example, depression in JOBS II). The true treatment effect for never-takers (ITT_n) ranges from 0.0 to 0.6, which is approximately 0.0 to -0.6 in terms of effect size, assuming that the positive value of ITT_n represents a negative effect of the intervention on never-takers. The true residual variance (σ^2) is 1.0, the true control group complier mean (α_c) is 1.5, and the true control group never-taker mean (α_n) is 1.0.

Figure 1 shows the sensitivity of the CACE estimate to violation of the exclusion restriction. In this study, coverage is defined as the proportion of replications out of 500 replications that are covered by the 95 per cent confidence interval. Power is defined as the proportion of replications out of 500 replications where the CACE estimate is significantly different from zero ($\alpha = 0.05$). Coverage and power at $ITT_n = 0$ show the quality of the CACE estimate when the exclusion restriction is not violated (coverage $= 0.940$, power $= 0.860$). Coverage and power of the CACE estimate are still acceptable when the exclusion restriction is minimally violated (that is, $ITT_n = 0.1$). However, the violation of the exclusion restriction starts to affect the quality of the CACE estimate substantially as the size of ITT_n increases. When the absolute magnitude of ITT_n is one-half that of CACE (that is, $ITT_n = 0.3$), coverage is 0.624 and power is 0.308. When the absolute magnitude of ITT_n is the same as that of CACE (that is, $ITT_n = 0.6$), coverage decreases to 0.094. When a large effect size is expected for compliers in interventions, it will be realistic to assume that the size of ITT_n will be relatively small and the impact of misspecification also will be small. However, if a small CACE is expected, and if it is questionable whether the exclusion restriction will hold, one needs to be more careful in interpreting the CACE estimate. As shown above, violation of the exclusion restriction may eliminate or double the CACE estimate.

Table I shows the estimates of CACE, α_n and α_c that deviate from true values as the exclusion restriction assumption is more severely violated. In Section 3.1, it was demonstrated that the estimate of α_n will be biased as much as γ_n (ITT_n), the estimate of α_c will be biased as much as $-\pi_n\gamma_n/\pi_c$, and the estimate of CACE will be biased as much as $\pi_n\gamma_n/\pi_c$ by

misspecifying ITT_n as zero. Since true π_c is 0.5, the size of bias in the estimate of α_c will be approximately $-\gamma_n$, and the size of bias in the CACE estimate will be approximately γ_n in this simulation setting. It is shown in Table I that the bias in the estimates of CACE, α_n and α_c increases proportionally according to the size of ITT_n.

The quality of the CACE estimate reported in Figure 1 and Table I will vary in different settings. That is, compliance rate, α_c and α_n values, and other available auxiliary information such as from covariates affect the sensitivity of the CACE estimate to violation of the exclusion restriction assumption. Compliance rate has a direct influence on the quality of the CACE estimate; however, it is often very difficult to control compliance behaviour of human participants. α_c and α_n values also affect the quality of the CACE estimate. If outcome distributions of compliers and never-takers are distant from each other, distinguishing two distributions will be relatively easier than when the distributions are close to each other. If the assumption of the exclusion restriction holds, precision in the CACE estimate will increase as the distance between α_c and α_n increases [25]. However, if the assumption does not hold, systematic influence of α_c and α_n values cannot be expected, since the estimates of α_c and α_n will also be distorted, as shown in Table I. In addition, it is not practical to depend on information from unknown outcome distributions in the control condition. In contrast, the association between compliance and covariates can be estimated based on information from the treatment condition without relying on the exclusion restriction. As a way to improve the quality of the CACE estimate, this study focuses on the use of covariate information, which is also a relatively controllable factor in randomized trial practice.

4.2. CACE estimation with a covariate

This section demonstrates how covariate information affects the sensitivity of CACE estimation due to violation of the exclusion restriction. The simulation setting is the same as in the previous section except that a covariate is added in the model. The model used for data generation and CACE estimation is described in equations (13) and (14). For simplicity, one continuous covariate X is used in the simulation study, where $X_i \sim N(0, 1)$. Both the outcome and compliance are regressed on the same covariate X. The true direct effect of X on the outcome (λ) is -0.3. Two different levels of association between compliance and the covariate (β_1) are considered in the simulation. The two association levels studied are 0.5 and 0.3 in terms of the odds ratio (OR) in the logistic regression of c on X. As described in equation (13), constant effect of CACE that does not vary across different values of X_i is assumed.

Figure 2 shows the sensitivity of the CACE estimate to the violation of the exclusion restriction when there is a covariate. The association between compliance and the covariate is 0.5 in terms of the odds ratio. Coverage and power at $ITT_n = 0$ show the quality of CACE estimate when the exclusion restriction is not violated (coverage$=0.964$, power$=0.948$). Coverage and power remain more stable than in the model without covariates (see Figure 1) as the size of ITT_n increases. When the absolute magnitude of ITT_n is one-half that of CACE (that is, $ITT_n = 0.3$), coverage is 0.742 and power is 0.470. When the absolute magnitude of ITT_n is the same as that of CACE (that is, $ITT_n = 0.6$), coverage is 0.192 and power is 0.106.

Figure 3 shows the sensitivity of the CACE estimate to the violation of the exclusion restriction when there is a covariate that is highly associated with compliance (OR $=0.3$). Coverage and power at $ITT_n = 0$ show the quality of the CACE estimate when the exclusion restriction is not violated (coverage$=0.964$, power$=0.970$). Coverage and power show

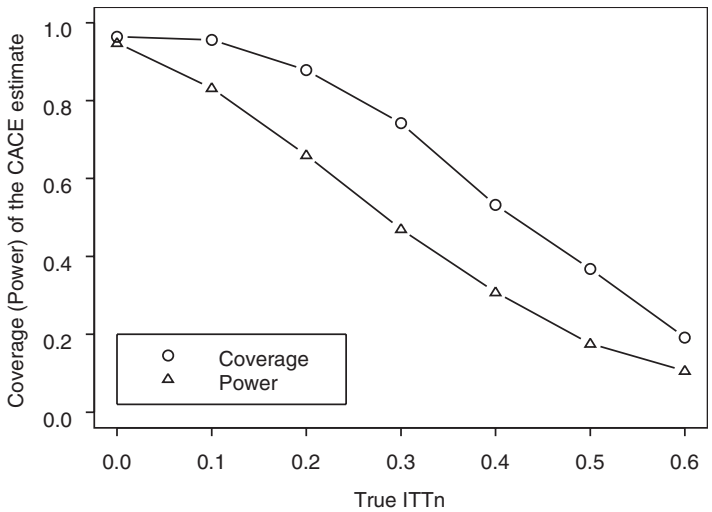

Figure 2. Simulation: sensitivity of the CACE estimate to violation of the exclusion restriction assumption when there is a covariate with OR = 0.5.

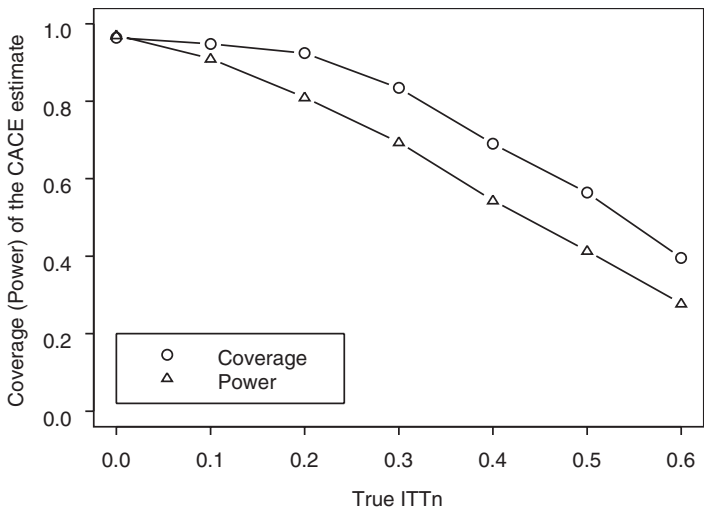

Figure 3. Simulation: sensitivity of the CACE estimate to violation of the exclusion restriction assumption when there is a covariate with OR = 0.3.

noticeable improvement compared to Figure 2 (that is, OR = 0.5). The improvement is even more substantial compared to the model without covariates (see Figure 1). When the absolute magnitude of ITT_n is one-half of that of CACE (that is, $ITT_n = 0.3$), coverage is 0.834 and power is 0.694. When the absolute magnitude of ITT_n is the same as that of CACE (that is, $ITT_n = 0.6$), coverage is 0.396 and power is 0.278.

Table II. Simulation: estimation of CACE with a covariate (OR = 0.3, true CACE = −0.6, true $\alpha_n = 1.0$, true $\alpha_c = 1.5$).

ITT_n	CACE		α_n		α_c	
True	Estimate	SE	Estimate	SE	Estimate	SE
0.0	−0.607	0.162	0.996	0.091	1.503	0.146
0.1	−0.552	0.167	1.082	0.095	1.440	0.154
0.2	−0.490	0.171	1.172	0.098	1.370	0.159
0.3	−0.424	0.173	1.265	0.099	1.294	0.162
0.4	−0.357	0.175	1.358	0.101	1.218	0.163
0.5	−0.291	0.172	1.451	0.099	1.144	0.160
0.6	−0.232	0.167	1.541	0.096	1.078	0.150

Table II shows the estimates of CACE, α_n and α_c when there is a covariate that is highly associated with compliance (OR = 0.3). It is shown in Table II that the bias in the estimates of α_n and α_c increases at a slower rate than in the model without covariates (see Table I) as the size of ITT_n increases. Consequently, the CACE estimate is much less biased in the model with a covariate (Table II) compared to the CACE estimate without covariates (Table I). For example, when the absolute magnitude of ITT_n is one-half that of CACE (that is, $ITT_n = 0.3$), the size of the average CACE estimate in the model without covariates is about two-thirds that in the model with a covariate.

5. APPLICATION TO JOBS II

This section demonstrates the estimation of CACE in practice when the assumption of the exclusion restriction is possibly violated. In CACE analysis examples using JOBS II, the level of depression six months after the intervention is used as the outcome (depression at T6). Depression was measured with a subscale of 11 items based on the Hopkins symptom checklist [26]. The present study focused on the high-risk status group based on previous studies [10, 27], which indicated that the job search intervention had its primary impact on high-risk respondents. Risk score was computed based on risk variables (depression, financial strain and assertiveness) in the screening data [27] predicting depressive symptoms at follow-up. A total sample size of 486 was analysed in this study after deleting cases that had missingness in covariates and outcome variables. Among 486 individuals, 328 are in the treatment condition and 158 are in the control condition. The response rate at follow-up six months after the intervention was 87 per cent. The variables used in the current study are shown in Table III.

In addition to demographic information, all participants in JOBS II were asked before randomization how highly they expected they would comply with intervention activities, which was intended to better predict actual compliance. Table IV shows the responses from treatment group individuals to the question ‘How likely or unlikely is it that you would participate in the one-week job seminar if you were offered the opportunity during the next three weeks?’ Since compliance behaviour is actually observed in the treatment condition, expected compliance can be compared to actual compliance (that is, showing up). It is shown that the actual compliance rate is sharply discriminated between the individuals who answered that it is

Table III. JOBS II: sample statistics.

Variable	Control group		No-shows ($Z=1$)		Shows ($Z=1$)	
	Mean	SD	Mean	SD	Mean	SD
Depression at T6	2.090	0.759	2.043	0.704	1.902	0.708
Depression at T0	2.491	0.280	2.441	0.315	2.420	0.297
Expected compliance (0/1)	0.329	0.471	0.218	0.414	0.442	0.498
Age in years	36.127	9.854	33.312	9.678	39.717	9.586
School grade completed	13.340	1.940	12.880	1.920	13.800	2.070
Assertiveness at T0	3.029	0.879	3.218	0.925	2.973	0.924
Not married (0/1)	0.570	0.497	0.626	0.486	0.646	0.479
Economic hardship at T0	3.456	0.953	3.771	0.822	3.600	0.804
Non-White (0/1)	0.171	0.378	0.238	0.427	0.155	0.363
Male (0/1)	0.386	0.488	0.388	0.489	0.475	0.501

Table IV. JOBS II: expected versus actual compliance ($Z=1$).

Expected compliance (n)	Actual compliance	
	No	Yes
Extremely unlikely (1)	1	0
Very unlikely (1)	0	1
Unlikely (2)	2	0
Neither nor (26)	15	11
Likely (93)	51	42
Very likely (93)	46	47
Extremely likely (112)	32	80

extremely likely that they would attend the intervention seminars and those who answered differently. A dichotomous variable (expected compliance: 0 = low, 1 = high) is created based on these two categories of individuals.

Table V shows the results from CACE estimation with various pretreatment covariates including expected compliance. A slightly different analysis using JOBS II has been previously presented by Little and Yau [5]. The effect size of the differential treatment effect estimate is calculated by dividing the outcome difference in treatment and control condition means after treatment by the square root of the variance pooled across the control and treatment groups. The CACE estimate is significant and has a meaningful level of effect size (CACE = -0.338, effect size = 0.466). The level of depression was lower for compliers in the intervention condition compared to that for control condition individuals who could have complied if they have had been assigned to the intervention condition. λ represents the direct effect of covariates on the outcome, where $\lambda=(\lambda_1,\lambda_2,\lambda_3,\lambda_4,\lambda_5,\lambda_6,\lambda_7,\lambda_8,\lambda_9)'$. Economic hardship before the intervention was found to be a significant predictor of the outcome. Depression level decreased more among individuals who did not have economic hardship.

The same covariates that are used as predictors of the outcome are also used as predictors of compliance. For nine covariates, logit coefficients $\beta_1=(\beta_{11},\beta_{12},\beta_{13},\beta_{14},\beta_{15},\beta_{16},\beta_{17},\beta_{18},\beta_{19})'$. Table V shows that expected compliance has high association with compliance ($\beta_{12}=1.264$). The estimated odds of being a complier are 3.54 times higher for individuals who expected

Table V. JOBS II: CACE estimation with various covariates.

Parameter	Estimate	SE
CACE	−0.338	0.142
α_c	2.210	0.435
α_n	1.911	0.391
σ^2	0.492	0.036
Y on x		
Depression at T0 (λ_1)	0.070	0.107
Expected compliance (λ_2)	−0.027	0.073
Age in years (λ_3)	−0.001	0.004
School grade comp (λ_4)	−0.030	0.017
Assertiveness at T0 (λ_5)	−0.038	0.038
Not married (λ_6)	−0.123	0.076
Economic hardship (λ_7)	0.150	0.041
Non-White (λ_8)	0.066	0.092
Male (λ_9)	−0.102	0.069
c on x		
Intercept (β_0)	−4.841	1.631
Depression at T0 (β_{11})	−0.420	0.423
Expected compliance (β_{12})	1.264	0.279
Age in years (β_{13})	0.078	0.016
School grade comp (β_{14})	0.310	0.071
Assertiveness at T0 (β_{15})	−0.362	0.149
Not married (β_{16})	0.659	0.290
Economic hardship (β_{17})	−0.191	0.156
Non-White (β_{18})	−0.460	0.331
Male (β_{19})	0.369	0.261

Table VI. JOBS II: sensitivity of the CACE estimate to covariate information.

Present covariates	CACE estimate	SE	Effect size
No covariates	−0.237	0.152	0.327
Depression T0	−0.237	0.155	0.327
Depression T0, expected compliance	−0.283	0.152	0.390
Depression T0, expected compliance, age	−0.298	0.150	0.411
Depression T0, expected compliance, age, grade	−0.313	0.145	0.432
All nine covariates	−0.338	0.142	0.466

that it was extremely likely for them to attend the job seminar than for individuals who expected a lower possibility of attending the seminar. In addition to expected compliance, several covariates were found to be significant predictors of compliance behaviour. Individuals complied more if they were older, more educated less assertive and not married.

The summary results from six CACE analyses are shown in Table VI. The same CACE model used in Table V is used, but different numbers of covariates are present in six analyses. In each analysis, the same covariates that are used as predictors of the outcome are also used as predictors of compliance. The results demonstrate the sensitivity of the CACE estimate in JOBS II to covariate information.

When there are no covariates present in the model, the magnitude of the CACE estimate is about 70 per cent of that when all nine covariates are present in the model (see Table V for full results). Adding baseline depression (depression T0), which is the weakest predictor of compliance, has little impact on the CACE estimate. However, adding covariates that are strong predictors of compliance results in a noticeable change in the CACE estimate. Given randomization of intervention assignment, the fluctuation shown in Table VI can be considered as substantial, implying possible violation of the exclusion restriction. As demonstrated in previous sections, the magnitude of bias in the CACE estimate due to violation of the exclusion restriction may be reduced if covariates are good predictors of compliance. However, covariate information would not completely eliminate bias due to violation of the exclusion restriction. Therefore, the CACE estimate adjusted for covariates should be considered as less biased, but not completely unbiased.

6. DISCUSSION

This study examined the bias mechanism in the estimation of CACE when the assumption of the exclusion restriction is violated. The magnitude of bias is affected mainly by compliance rate and the size of treatment assignment effect on non-compliers (ITT_n). The magnitude of bias in the CACE estimate increases as the size of ITT_n increases and compliance rate decreases. If the compliance rate is very low, violation of the exclusion restriction assumption can cause substantial bias in the CACE estimate even when the effect of treatment assignment on non-compliers is trivial. When the exclusion restriction is violated, the CACE estimate not only can be understated, but also can be exaggerated, depending on how the assignment of treatment affects non-compliers.

Compliance rate and the effect of treatment assignment on non-compliers directly affect bias in the CACE estimate; however, these factors are not easily controllable in practice. The current study focused on the use of covariate information to test and to reduce bias due to violation of the exclusion restriction. It was demonstrated that bias in the CACE estimate due to model misspecification can be reduced substantially by including covariates that are good predictors of compliance. In JOBS II, one kind of covariate information collected before the intervention assignment was about study participants' expected level of compliance, which turned out to be a good predictor of actual compliance. To increase the precision in classifying individuals based on compliance types, one can use more aggressive ways of monitoring compliance (for example, by blood test in pharmaceutical trials), or have a run-in period before the main trial to select potential compliers [28]. However, these methods are not always applicable, especially in social-behavioural field experiments. Given that, readily observable covariates such as expected compliance, motivation, baseline outcome measures and background variables are valuable sources of information that can be acquired at a relatively low cost.

The limitation of using covariate information without relaxing the exclusion restriction is that the bias due to violation of the assumption cannot be completely eliminated unless covariates are perfect predictors of compliance. To test the ultimate magnitude of the bias and to estimate the assignment effect of treatment on non-compliers, the assumption of the exclusion restriction needs to be relaxed. Without assuming the exclusion restriction, the identifiability of CACE models relies on auxiliary information such as from covariates [7]

and proper priors [4, 6]. Practicality of the CACE estimation methods relying on auxiliary information is still under investigation.

In examining the bias mechanism in CACE estimation, this study assumed an additive effect of treatment assignment (that is, no interaction effects) for both compliers and non-compliers. However, in practice the effect of treatment assignment may vary depending on covariate values. The presence of interaction effects further complicates the bias mechanism. Also, covariate information may have weaker and inconsistent effects on treatment effect estimates. More investigation is needed to identify the bias mechanism with interaction effects. This study examined the bias mechanism in CACE estimation focusing on intervention settings where the intensity of compliance is the same among compliers. However, in practice, compliance often has varying intensity [29, 30]. Further study is needed to clarify the relationship between causal effect estimates and model misspecification in more general situations.

APPENDIX: BIAS MECHANISM IN A MISSPECIFIED MODEL WITH NEVER-TAKERS AND ALWAYS-TAKERS

If there are both never-takers and always-takers, compliance type can be observed only for always-takers among individuals assigned to the control condition, and only for never-takers among individuals assigned to the treatment condition. Given that, the observed-data likelihood function is

$$\begin{aligned} L(\theta \mid \text{data}) \propto & \prod_{i \in \{Z_i=1, D_i=0\}} \pi_{\text{n}} f(y_i \mid \mu_{1\text{n}}, \sigma^2) \\ & \times \prod_{i \in \{Z_i=0, D_i=1\}} \pi_{\text{a}} f(y_i \mid \mu_{0\text{a}}, \sigma^2) \\ & \times \prod_{i \in \{Z_i=1, D_i=1\}} [\pi_{\text{c}} f(y_i \mid \mu_{1\text{c}}, \sigma^2) + \pi_{\text{a}} f(y_i \mid \mu_{1\text{a}}, \sigma^2)] \\ & \times \prod_{i \in \{Z_i=0, D_i=0\}} [\pi_{\text{n}} f(y_i \mid \mu_{0\text{n}}, \sigma^2) + \pi_{\text{c}} f(y_i \mid \mu_{0\text{c}}, \sigma^2)] \end{aligned} \tag{A1}$$

where π_{a} is the proportion of always-takers in the population, $\mu_{1\text{a}}$ denotes population mean potential outcome for always-takers if $Z=1$, and $\mu_{0\text{a}}$ denotes population mean potential outcome for always-takers if $Z=0$.

Based on equation (A1), four observable population means can be expressed in terms of model parameters as

$$\mu_{1\text{n}} = \alpha_{\text{n}} + \gamma_{\text{n}} \tag{A2}$$

$$\mu_{0\text{a}} = \alpha_{\text{a}} \tag{A3}$$

$$\mu_1 = \pi_{\text{n}}(\alpha_{\text{n}} + \gamma_{\text{n}}) + \pi_{\text{c}}(\alpha_{\text{c}} + \gamma_{\text{c}}) + \pi_{\text{a}}(\alpha_{\text{a}} + \gamma_{\text{a}}) \tag{A4}$$

$$\mu_0 = \pi_{\text{n}}\alpha_{\text{n}} + \pi_{\text{c}}\alpha_{\text{c}} + \pi_{\text{a}}\alpha_{\text{a}} \tag{A5}$$

where γ_{a} is the average causal effect of treatment assignment for always-takers.

If the exclusion restriction assumption holds, $\gamma_n = 0$ and $\gamma_a = 0$. Therefore, α_c can be identified from equations (A2), (A3) and (A5) as

$$\alpha_c = \frac{\mu_0 - \pi_n \mu_{1n} - \pi_a \mu_{0a}}{\pi_c} \tag{A6}$$

Then, γ_c (CACE) can be identified from equations (A2), (A3), (A4) and (A6) as

$$\gamma_c = \frac{\mu_1 - \mu_0}{\pi_c} \tag{A7}$$

If the assumption of the exclusion restriction is violated for both never-takers and always-takers, $\gamma_n \neq 0$ and $\gamma_a \neq 0$. In this case, the correct specification of α_c from equations (A2), (A3) and (A5) is

$$\alpha_c = \frac{\mu_0 - \pi_n \mu_{1n} - \pi_a \mu_{0a}}{\pi_c} + \frac{\pi_n \gamma_n}{\pi_c} \tag{A8}$$

Then, the correct specification of γ_c (CACE) from equations (A2), (A3), (A4) and (A8) is

$$\gamma_c = \frac{\mu_1 - \mu_0}{\pi_c} - \frac{\pi_n \gamma_n}{\pi_c} - \frac{\pi_a \gamma_a}{\pi_c} \tag{A9}$$

Equation (A9) shows that the estimator of CACE (γ_c) will be biased as much as $\pi_n \gamma_n / \pi_c$, if γ_n is misspecified as zero, and as much as $\pi_a \gamma_a / \pi_c$, if γ_a is misspecified as zero. If the exclusion restriction is violated for both never-takers and always-takers, CACE (γ_c) will be biased as much as $\pi_n \gamma_n / \pi_c + \pi_a \gamma_a / \pi_c$.

ACKNOWLEDGEMENTS

This study was supported by NIMH grant P30 MH38330, Michigan Prevention Research Center at the Institute for Social Research, Richard H. Price, P.I., University of Michigan. This study was also supported by grant KO2 AA 00230 from NIAAA and by NIMH grant P50 MH38725, Epidemiologic Prevention Center for Early Risk Behaviors, Philip Leaf, P.I., and grant RO1 DA11796 from NIDA, Follow-up of Two Preventive Intervention Trials, Nicholas Ialongo, P.I., Department of Mental Hygiene, Bloomberg School of Public Health, Johns Hopkins University. I would like to thank Amiram Vinokur for providing data and stimulating input. I also thank Guido Imbens, Keisuke Hirano and Bengt Muthén for helpful advice.

REFERENCES

1. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 1996; **91**:444–455.
2. Bloom HS. Accounting for no-shows in experimental evaluation designs. *Evaluation Review* 1984; **8**:225–246.
3. Imbens GW, Angrist J. Identification and estimation of local average treatment effects. *Econometrica* 1994; **62**:467–476.
4. Imbens GW, Rubin DB. Bayesian inference for causal effects in randomized experiments with non-compliance. *Annals of Statistics* 1997; **25**:305–327.
5. Little RJA, Yau LHY. Statistical techniques for analyzing data from prevention trials: treatment of no-shows using Rubin's causal model. *Psychological Methods* 1998; **3**:147–159.
6. Hirano K, Imbens GW, Rubin DB, Zhou XH. Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics* 2000; **1**:69–88.
7. Jo B. Estimation of intervention effects with non-compliance: alternative model specifications. *Journal of Educational and Behavioral Statistics* (in press).

8. West SG, Sagarin BJ. Participant selection and loss in randomized experiments. In *Research Design*, Bickman L (ed.). Sage: Thousand Oaks, CA, 2000; 117–154.
9. McDonald C, Hiu S, Tierney W. Effects of computer reminders for influenza vaccination on morbidity during influenza epidemics. *MD Computing* 1992; **9**:304–312.
10. Vinokur AD, Price RH, Schul Y. Impact of the JOBS intervention on unemployed workers varying in risk for depression. *American Journal of Community Psychology* 1995; **23**:39–74.
11. Vinokur AD, Schul Y. Mastery and inoculation against setbacks as active ingredients in intervention for the unemployed. *Journal of Consulting and Clinical Psychology* 1997; **65**:867–877.
12. Barling J. *Employment, Stress and Family Functioning*. Wiley: New York, 1990.
13. Fryer D, Payne R. Being unemployed: a review of the literature on the psychological experience of unemployment. *International Review of Industrial and Organizational Psychology* 1986; 235–278.
14. Vinokur AD, Price RH, Caplan RD. From field experiments to program implementation: assessing the potential outcomes an experimental intervention program for unemployed persons. *American Journal of Community Psychology* 1991; **19**:543–562.
15. Holland PW. Statistics and causal inference. *Journal of the American Statistical Association* 1986; **81**: 945–970.
16. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 1974; **66**:688–701
17. Rubin DB. Bayesian inference for causal effects: the role of randomization. *Annals of Statistics* 1978; **6**:34–58.
18. 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.
19. Rubin DB. Comment on 'Neyman (1923) and causal inference in experiments and observational studies'. *Statistical Science* 1990; **5**:472–480.
20. Dempster A, Laird N, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* 1977; **39**:1–38.
21. Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. Wiley: New York, 1987.
22. McLachlan GJ, Krishnan T. *The EM Algorithm and Extensions*. Wiley: New York, 1997.
23. Tanner M. *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. Springer: New York, 1996.
24. Muthén LK, Muthén BO. M*plus User's Guide*. Muthén & Muthén: Los Angeles, CA, 1998–2001.
25. Jo B. Statistical power in randomized intervention studies with non-compliance. *Psychological Methods* 2002; **7**:178–193.
26. Derogatis LR, Lipman RS, Rickles K, Uhlenuth EH, Covi L. The Hopkins Symptom Checklist (HSCL). In *Psychological Measurements in Psychopharmacology: Modern Problems in Pharmacopsychiatry*, Pichot P (ed.). Karger, Basel: New York, 1974; 79–110.
27. Price RH, van Ryn M, Vinokur AD. Impact of a preventive job search intervention on the likelihood of depression among the unemployed. *Journal of Health and Social Behavior* 1992; **33**:158–167.
28. Sheiner LB, Rubin, DB. Intention to treat analysis and the goals of clinical trials. *Clinical Pharmacology and Therapeutics* 1995; **57**:6–15.
29. Angrist JD, Imbens GW. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association* 1995; **90**:431–442.
30. Goetghebeur 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.

