Statistica Sinica **34** (2024), 999-1021
doi:https://doi.org/10.5705/ss.202021.0351

SHARP BOUNDS FOR VARIANCE OF TREATMENT EFFECT ESTIMATORS IN THE PRESENCE OF COVARIATES

Ruoyu Wang[1], Qihua Wang*[1,3], Wang Miao[2] and Xiaohua Zhou[2]

[1]*Chinese Academy of Sciences,* [2]*Peking University and* [3]*University of Chinese Academy of Sciences*

Abstract: In a completely randomized experiment, the variances of the treatment effect estimators in a finite population are usually not identifiable, and hence not estimable. Although some estimable bounds of such variances have been established in the literature, few are derived in the presence of covariates. We consider the difference-in-means estimator and the Wald estimator in completely randomized experiments with perfect compliance and noncompliance, respectively. We also establish sharp bounds for the variances of these two estimators when covariates are available. Furthermore, we obtain consistent estimators for such bounds that can be used to shorten the confidence intervals and improve the power of tests. The confidence intervals are constructed based on the consistent estimators of the upper bounds, and have coverage rates that are uniformly asymptotically guaranteed. We use analyses based on simulations and real data to evaluate and demonstrate the proposed methods.

Key words and phrases: Causal inference, partial identification, potential outcome, randomized experiment.

1. Introduction

Estimation and inference for the average treatment effect are extremely important in practice. Many studies assume that the observations are sampled from an infinite super-population (Hirano, Imbens and Ridder (2003); Imbens (2004); Belloni, Chernozhukov and Hansen (2014); Chan, Yam and Zhang (2016)). However, an infinite super-population seems contrived if we are interested in evaluating the treatment effect for a particular finite population (Li and Ding (2017)), for example, the patients enrolled in an experiment. In such cases, and the finite-population framework is more suitable. This framework views all potential outcomes as fixed, and the randomness of the data comes solely from the treatment assignment (Imbens and Rosenbaum (2005); Nolen and Hudgens (2011)). It also avoids assumptions about randomly sampling from some "vaguely defined super-population of study units" (Schochet (2013)), and the statistical analysis results under the framework are interpretable in the

*Corresponding author.

absence of a super-population. Theoretical guarantees in this framework rely on the treatment assignment, rather than unverifiable on sampling assumptions, such as the independent and identically distributed (i.i.d.) assumption. In randomized experiments, the finite-population framework has been widely used in data analysis since the work of Neyman (1990). A fundamental problem in completely randomized experiments under this framework is that the variance of the widely used difference-in-means estimator is unidentifiable. Thus, we cannot obtain a consistent variance estimator, and the standard inference based on a normal approximation fails. To mitigate this problem, Neyman (1990) adopted an estimable upper bound for the variance, which leads to a conservative inference. The precision of the bound is crucial for the power of a test and the width of the resulting confidence interval (CI). Thus, it is important to incorporate all available information to make the bound as precise as possible. The variance bound with binary outcomes is fairly well studied by Robins (1988), Ding and Dasgupta (2016), and Ding and Miratrix (2018), among others. For general outcomes, Aronow, Green and Lee (2014) improved the results of Neyman (1990) by deriving a sharp bound that cannot be improved without information other than the marginal distributions of potential outcomes.

In many randomized experiments, some covariates are observed in addition to the outcome. However, few approaches consider how to improve the variance bound using covariate information; an exception is the work of Ding, Feller and Miratrix (2019). Surprisingly, we observe that the upper bound for the variance of the treatment effect estimator given by Ding, Feller and Miratrix (2019) can be larger than that given by Aronow, Green and Lee (2014), in some situations. This is illustrated in Example 1 in Section 2.

The first main contribution of this study is to derive a sharp bound for the variance of the difference-in-means estimator in a finite population when covariates are available, and to obtain a consistent estimator of the bound. The proof of consistency is quite challenging. In our analysis of consistency, we allow the cardinality of the covariate support to diverge with the population size. This differs from the approach taken in many previous studies and increases the difficulty of the proof, owing to the lack of tools for analyzing the sample conditional quantile functions in such an estimator. Then, based on the consistent estimator of the variance bound, we obtain a shorter CI with a more accurate coverage rate. In addition, we show that the CI has an asymptotically guaranteed coverage rate, and the asymptotic result is uniform over a large class of finite populations. As discussed by Lehmann and Romano (2006), although this uniformity is crucial for inferences based on asymptotic results, it is omitted in many existing works.

The aforementioned results focus on completely randomized experiments in which units comply with the assigned treatments. However, noncompliance often occurs in randomized experiments. In such cases, the parameter of interest is

the local average treatment effect (LATE) (Angrist, Imbens and Rubin (1996); Abadie (2003)), and its inference is more complicated. For the LATE in a finite population, estimators include the Wald estimator and those proposed by Ding, Feller and Miratrix (2019) and Hong, Leung and Li (2020). The identification problem also exists for the variances of these estimators. However, to the best of our knowledge, no prior studies have derived a sharp bound for unidentifiable variances.

Another main contribution of this study is to extend the aforementioned results for the completely randomized experiment to the case with noncompliance. We establish a sharp bound for the variance of the Wald estimator, and propose a consistent estimator for the variance bound. The analysis of consistency is even more involved in this case, owing to the complexity of the estimator. Based on the consistent estimator of the upper bound, we construct a CI with a coverage rate that is uniformly asymptotically guaranteed. Note that the sharp bound without covariates can be derived as a special case of the resulting bound, which has not been investigated in prior studies. Simulations and an application to two real data sets from the randomized trial ACTG protocol 175 (Hammer et al. (1996)) and JOBS II (Vinokuir, Price and Schul (1995)) demonstrate the advantages of our methods.

The remainder of this paper is organized as follows. In Section 2, we establish the sharp variance bound in the presence of covariates for the difference-in-means estimator in a completely randomized experiment with perfect compliance. A consistent estimator is obtained for the bound. In Section 3, we consider the Wald estimator for the LATE in a completely randomized experiment in the presence of noncompliance; here, we establish a sharp variance bound for the Wald estimator in the presence of covariates, and obtain a consistent estimator for the bound. Simulation studies are conducted to evaluate the empirical performance of the proposed bound estimators in Section 4, followed by some applications to data from the randomized trial ACTG protocol 175 and JOBS II in Section 5. A discussion on possible extensions of our results is provided in Section 6. Proofs are relegated to the Supplementary Material.

2. Sharp variance bound for the difference-in-means estimator

2.1. Preliminaries

Suppose we are interested in the effect of a binary treatment on an outcome in a finite population consisting of N units. In a completely randomized experiment, n out of N units are sampled from the population, with n_1 assigned randomly to the treatment group and the other $n_0 = n - n_1$ assigned to the control group. Let $T_i = 1$ if unit i is assigned to the treatment group, and $T_i = 0$ if the unit is assigned to the control group; T_i is not defined if unit i is not enrolled in the experiment. For each unit i and $t = 0, 1$, let y_{ti} denote the potential outcome

that would be observed if unit i is assigned to treatment t. Let w_i denote a vector of covariates, with the constant 1 as its first component. The covariate vector w_i is observed if unit i is enrolled in the experiment (i.e., $T_i = 0$ or 1). Then, the characteristics of the population can be viewed as a matrix $\mathbf{U} = (y_1, y_0, w)$, where $y_1 = (y_{11}, y_{12}, \ldots, y_{1N})^T$, $y_0 = (y_{01}, y_{02}, \ldots, y_{0N})^T$, and $w = (w_1, \ldots, w_N)^{\mathrm{T}}$.

For any vector $a = (a_1, \ldots, a_N)^{\mathrm{T}}$, we let

$$\mu(a) = \frac{1}{N}\sum_{i=1}^{N} a_i, \quad \phi^2(a) = \frac{1}{N}\sum_{i=1}^{N}(a_i - \mu(a))^2.$$

Letting $\tau_i = y_{1i} - y_{0i}$ be the treatment effect for unit i and $\tau = (\tau_1, \ldots, \tau_N)^{\mathrm{T}}$, the parameter of interest is the average treatment effect,

$$\theta = \mu(\tau) = \frac{1}{N}\sum_{i=1}^{N} y_{1i} - \frac{1}{N}\sum_{i=1}^{N} y_{0i}.$$

Note that all parameters discussed in this paper depend on N, unless otherwise specified, and we omit the dependence in the notation for simplicity when there is no ambiguity. The treatment assignment is unrelated to the covariates in completely randomized experiments. Hence the average treatment effect can be estimated by the difference-in-means estimator

$$\hat{\theta} = \frac{1}{n_1}\sum_{T_i=1} y_{1i} - \frac{1}{n_0}\sum_{T_i=0} y_{0i}.$$

This estimator is widely used, owing to its simplicity and transparency, among other practical reasons (Shao, Yu and Zhong (2010); Lin (2013)). Moreover, it is the uniformly minimum variance unbiased estimator in the scenario presented in (Kallus (2018)). Following the literature (Imai (2008); Aronow and Middleton (2013); Shao, Yu and Zhong (2010); Kallus (2018); Ma, Tu and Liu (2020)), we consider the inference based on the difference-in-means estimator because of its popularity in practice and its theoretical importance.

According to Freedman (2008a), the variance of $\hat{\theta}$ is

$$\frac{1}{N-1}\left\{\frac{N}{n_1}\phi^2(y_1) + \frac{N}{n_0}\phi^2(y_0) - \phi^2(\tau)\right\},$$

and we denote this variance by $\sigma^2/(N-1)$. Under certain standard regularity conditions in a finite population, previous works (Freedman (2008a); Aronow, Green and Lee (2014); Li and Ding (2017)) have established that

$$\sqrt{N}\sigma^{-1}(\hat{\theta} - \theta) \xrightarrow{d} N(0, 1), \tag{2.1}$$

as n_1, n_0, and N go to infinity. We can perform a statistical inference based

on this asymptotic distribution. However, it is difficult to obtain a consistent estimator for σ^2. According to standard results in survey sampling (Cochran (1977)), $\phi^2(y_t)$ can be consistently estimated by

$$\hat{\phi}_t^2 = \frac{1}{n_t - 1} \sum_{T_i = t} \left(y_{ti} - \frac{1}{n_t} \sum_{T_j = t} y_{tj} \right)^2, \tag{2.2}$$

for $t = 0, 1$. However, $\phi^2(\tau)$, and hence σ^2, is not identifiable, because the potential outcomes y_1 and y_0 can never be observed simultaneously. To make an inference for θ based on (2.1), one can use an upper bound for σ^2 to construct a conservative CI. Alternatively, one may use an estimable lower bound for σ^2 to obtain a shorter confidence interval. However, the coverage rate of such a CI may not be guaranteed. To establish an estimable upper (lower) bound for σ^2, it suffices to establish an estimable lower (upper) bound for the unidentifiable term $\phi^2(\tau)$. We then derive the sharp bound for $\phi^2(\tau)$ and obtain its consistent estimator.

2.2. Sharp bound for $\phi^2(\tau)$

For any matrices $a = (a_1, \ldots, a_N)^{\mathrm{T}}$, $b = (b_1, \ldots, b_N)^{\mathrm{T}}$ and vectors $\bar{a}$, $\bar{b}$ that have dimensions equal to the number of columns of a and b, respectively, define

$$\begin{aligned}
P(a \le \bar{a}) &= \frac{1}{N} \sum_{i=1}^{N} 1\{a_i \le \bar{a}\}, \\
P(a = \bar{a}) &= \frac{1}{N} \sum_{i=1}^{N} 1\{a_i = \bar{a}\}, \\
P(a = \bar{a} \mid b = \bar{b}) &= \frac{\sum_{i=1}^{N} 1\{a_i = \bar{a}, b_i = \bar{b}\}}{\sum_{i=1}^{N} 1\{b_i = \bar{b}\}}, \\
P(a \le \bar{a} \mid b = \bar{b}) &= \frac{\sum_{i=1}^{N} 1\{a_i \le \bar{a}, b_i = \bar{b}\}}{\sum_{i=1}^{N} 1\{b_i = \bar{b}\}},
\end{aligned}$$

where $1\{\cdot\}$ is the indicator function and "$\le$" between two vectors corresponds to the component-wise inequality. Note that, in this paper, $P(\cdot)$ and $P(\cdot \mid \cdot)$ are some quantities that describe a vector, and we use $\mathbb{P}(\cdot)$ to denote the probability.

For any function H, we define $H^{-1}(u) = \inf\{s : H(s) \ge u\}$. In this paper, we adopt the convention $\inf \emptyset = \infty$. We let $\{\xi_1, \ldots, \xi_K\}$ be the set of all different values of w_i. Clearly, $K \le N$. We aim to derive bounds for $\phi^2(\tau)$ by using the covariate information efficiently. Define $\pi_k = P(w = \xi_k)$, for $k = 1, \ldots, K$. The quantities π_k $(k = 1, \ldots, K)$ and the functions $F_{t|k}$ $(t = 0, 1$ and $k = 1, \ldots, K)$ summarize the characteristics of the population and can be estimated using observed data. To obtain estimable bounds for $\phi^2(\tau)$, we focus on the bound that can be expressed as a functional of π_k and $F_{t|k}$ $(t = 0, 1$ and $k = 1, \ldots, K)$.

Define the set of lower bounds

$$\mathcal{B}_{\mathrm{L}} = \{b_{\mathrm{L}} : b_{\mathrm{L}} \text{ is a functional of } \pi_k \text{ and } F_{t|k} \text{ for } t = 0, 1 \text{ and } k = 1, \ldots, K; \\ b_{\mathrm{L}} \leq \phi^2(\tau)\}.$$

Define the set of upper bounds $\mathcal{B}_{\mathrm{H}}$ similarly. Then, the sharp bound is established in the following theorem.

Theorem 1. *A bound for $\phi^2(\tau)$ is $[\phi^2_{\mathrm{L}}, \phi^2_{\mathrm{H}}]$, where*

$$\phi^2_{\mathrm{L}} = \sum_{k=1}^{K} \pi_k \int_0^1 (F^{-1}_{1|k}(u) - F^{-1}_{0|k}(u))^2 du - \theta^2,$$
$$\phi^2_{\mathrm{H}} = \sum_{k=1}^{K} \pi_k \int_0^1 (F^{-1}_{1|k}(u) - F^{-1}_{0|k}(1-u))^2 du - \theta^2.$$

Moreover, the bound is sharp in the sense that ϕ^2_{L} is the largest lower bound in $\mathcal{B}_{\mathrm{L}}$, and ϕ^2_{H} is the smallest upper bound in $\mathcal{B}_{\mathrm{H}}$.

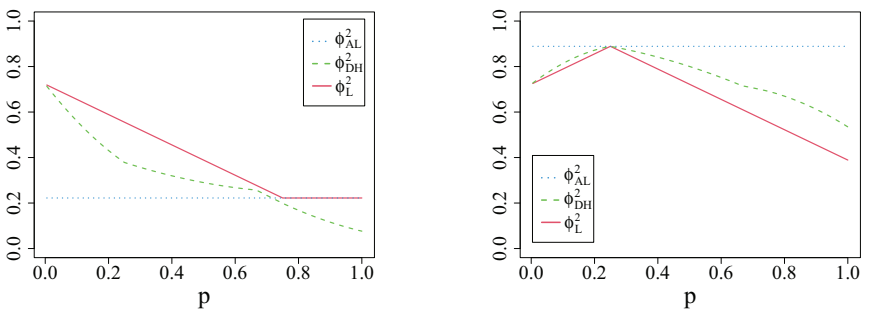
See the Supplementary Material for the proof of this theorem. Here, we compare this bound to previous bounds obtained by Aronow, Green and Lee (2014) and Ding, Feller and Miratrix (2019). By using the marginal distributions of the potential outcomes, Aronow, Green and Lee (2014) derive the following bound for $\phi^2(\tau)$:

$$\phi^2_{\mathrm{AL}} \coloneqq \int_0^1 (F_1^{-1}(u) - F_0^{-1}(u))^2 du - \theta^2 \leq \phi^2(\tau) \\ \leq \int_0^1 (F_1^{-1}(u) - F_0^{-1}(1-u))^2 du - \theta^2 \coloneqq \phi^2_{\mathrm{AH}},$$

where $F_t(y) = P(y_t \leq y)$, for $t = 0, 1$. The bound of Aronow, Green and Lee (2014) is sharp given the marginal distributions of the potential outcomes. In the presence of covariates, Ding, Feller and Miratrix (2019) proposed the following regression-based bound that may improve the bound of Aronow, Green and Lee (2014) in certain situations:

$$\phi^2_{\mathrm{DL}} \coloneqq \phi^2(\tau_w) + \int_0^1 (F^{-1}_{e_1}(u) - F^{-1}_{e_0}(u))^2 du \leq \phi^2(\tau) \\ \leq \phi^2(\tau_w) + \int_0^1 (F^{-1}_{e_1}(u) - F^{-1}_{e_0}(1-u))^2 du \coloneqq \phi^2_{\mathrm{DH}}.$$

where $\tau_w = (w_1^{\mathrm{T}}(\gamma_1 - \gamma_0), \ldots, w_N^{\mathrm{T}}(\gamma_1 - \gamma_0))^{\mathrm{T}}$, $F_{e_t}(s) = P(e_t \leq s)$, $e_t = (y_{t1} - w_1^{\mathrm{T}}\gamma_t, \ldots, y_{tN} - w_N^{\mathrm{T}}\gamma_t)^{\mathrm{T}}$, and γ_t is the least square regression coefficient of y_{ti} on w_i. The lower bound of Ding, Feller and Miratrix (2019) is not sharp, because it can be smaller even than that of Aronow, Green and Lee (2014), and thus may



(a) Relationship between ϕ^2_{AL}, ϕ^2_{DL}, and ϕ^2_{L}. (b) Relationship between ϕ^2_{AH}, ϕ^2_{DH}, and ϕ^2_{H}.

Figure 1. Comparison of three bounds under different values of p.

lead to more conservative CIs in spite of the available covariate information. This situation does not occur with our bound. It can be verified that the bounds ϕ^2_{AL}, ϕ^2_{AH}, ϕ^2_{DL}, and ϕ^2_{DH} are all functionals of π_k and $F_{t|k}$ ($t = 0, 1$ and $k = 1, \ldots, K$). Thus, we have

$$\phi^2_{\mathrm{L}} \geq \max\{\phi^2_{\mathrm{AL}}, \phi^2_{\mathrm{DL}}\}, \ \phi^2_{\mathrm{H}} \leq \min\{\phi^2_{\mathrm{AH}}, \phi^2_{\mathrm{DH}}\}, \tag{2.3}$$

according to Theorem 1. When there is no covariate, our bound reduces to $[\phi^2_{\mathrm{AL}}, \phi^2_{\mathrm{AH}}]$ by letting $K = 1$, $\xi_1 = 1$, and $w_i = 1$, for $i = 1, \ldots, N$. The following example illustrates the improvement of our bound as the association between the covariates and the potential outcomes varies.

Example 1. Consider a population with $N = 600$ units. Suppose the potential outcomes and the covariate are binary, with $P(w = 1) = 1/3$, $P(y_1 = 1) = 2/3$, $P(y_0 = 1) = 1/3$, $P(y_0 = 1 \mid w = 1) = 3/4$, and $P(y_0 = 1 \mid w = 0) = 1/8$. Let $p = P(y_1 = 1 \mid w = 1)$ ($p \in \{1/200, \ldots, 1\}$); then, $P(y_1 = 1 \mid w = 0) = 1 - p/2$. Figure 1 presents the three bounds under different values of p.

Figure 1 shows that $\phi^2_{\mathrm{L}} \geq \max\{\phi^2_{\mathrm{AL}}, \phi^2_{\mathrm{DL}}\}$ and $\phi^2_{\mathrm{H}} \leq \min\{\phi^2_{\mathrm{AH}}, \phi^2_{\mathrm{DH}}\}$ under all settings of p, and in many situations, the inequalities are strict. For $p \leq 143/200$, the bound of Ding, Feller and Miratrix (2019) is tighter than that of Aronow, Green and Lee (2014). However, for $p > 143/200$, $\phi^2_{\mathrm{DL}} < \phi^2_{\mathrm{AL}}$, even though covariate information is used in the approach of Ding, Feller and Miratrix (2019).

2.3. Estimation of the sharp bound and the CI

To estimate ϕ^2_{L} and ϕ^2_{H} and study the asymptotic properties of their proposed estimators, we adopt the following standard framework (Li and Ding (2017)) for our theoretical development. Suppose there is a sequence of finite populations $\mathbf{U}_N$ of size N. For each N, n_1 units are assigned randomly to the treatment group, and n_0 units are assigned to the control group. As the population size

$N \to \infty$, the sizes of the treatment and the control groups satisfy $n_1/N \to \rho_1$, $n_0/N \to \rho_0$, with $\rho_1, \rho_0 \in (0,1)$ and $\rho_1+\rho_0 \leq 1$. Here, we assume that the number of covariate values K is known and is allowed to grow at a certain rate with the population size N. To accommodate continuous covariates, we can stratify them and increase the number of strata with the sample size. We estimate π_k and $F_{t|k}(y)$ using the empirical probabilities $\hat{\pi}_k = 1/n \sum_{T_i \in \{0,1\}} 1\{w_i = \xi_k\}$ and $\hat{F}_{t|k}(y) = \sum_{T_i=t} 1\{y_{ti} \leq y, w = \xi_k\} / \sum_{T_i=t} 1\{w_i = \xi_k\}$, respectively. By plugging in these estimators, we obtain the following estimators for ϕ^2_{L} and ϕ^2_{H}:

$$\hat{\phi}^2_{\mathrm{L}} = \sum_{k=1}^{K} \hat{\pi}_k \int_0^1 (\hat{F}^{-1}_{1|k}(u) - \hat{F}^{-1}_{0|k}(u))^2 du - \hat{\theta}^2,$$
$$\hat{\phi}^2_{\mathrm{H}} = \sum_{k=1}^{K} \hat{\pi}_k \int_0^1 (\hat{F}^{-1}_{1|k}(1-u) - \hat{F}^{-1}_{0|k}(u))^2 du - \hat{\theta}^2. \tag{2.4}$$

These estimators involve sample conditional quantile functions $\hat{F}^{-1}_{t|k}(u)$, which have complicated statistical properties in the finite-population framework. Many of the theoretical results for such functions in the super-population framework cannot be applied to the scenario we consider. Moreover, the number of covariate values K is allowed to diverge as the population size increases, which further complicates the problem. Thus, it is not trivial to analyze the asymptotic properties of these estimators. However, we observe that the first term of these estimators is actually the weighted sum of the Wasserstein distances between some distributions. By invoking a representation theorem of Wasserstein distances, we prove the consistency results for the estimators with a careful analysis of the error terms. See the Supplementary Material for more details. We assume that the population $\mathbf{U}_N$ satisfies the following two conditions.

Condition 1. *There is some constant C_M that does not depend on N such that $1/N \sum_{i=1}^{N} y^4_{ti} \leq C_M$, for $t = 0, 1$.*

Condition 2. *There is some constant C_π that does not depend on N such that $\pi_k \geq C_\pi/K$, for $k = 1, \ldots, K$.*

Furthermore, we assume K satisfies the following condition.

Condition 3. $K^2 \log K/N \to 0$ *as* $N \to \infty$.

Condition 1 requires that the potential outcomes have uniformly bounded fourth moments. Condition 2 requires that the proportion of units with each covariate value not be too small. Condition 3 imposes some upper bound on the number of values the covariate may take. If the covariate w is some subgroup indicator, then Condition 3 can be satisfied easily if the number of subgroups is not too large. Continuous components in w can be stratified to meet Condition 3. If w contains many components, then the number of covariate values may

be too large, even after stratifying the continuous components. In this case, one can partition units with similar covariate values into subgroups, and use the subgroup label as the new covariate to apply the proposed method. Alternatively, one can first employ dimension reduction or variable selection methods to obtain a low-dimensional covariate, and then apply the proposed method using the obtained covariate. The following theorem establishes the consistency of the bound estimators.

Theorem 2. *Under Conditions* 1, 2, *and* 3, *we have*

$$(\hat{\phi}^2_{\mathrm{L}}, \hat{\phi}^2_{\mathrm{H}}) - (\phi^2_{\mathrm{L}}, \phi^2_{\mathrm{H}}) \xrightarrow{P} 0$$

as $N \to \infty$.

The proof of this theorem is relegated to the Supplementary Material. The lower bound ϕ^2_{L} for $\phi^2(\tau)$ implies an upper bound for σ^2. The consistency result of $\hat{\phi}^2_{\mathrm{L}}$ is sufficient for constructing a conservative CI. A conservative $1-\alpha$ CI for θ is given by

$$I_N = \left[\hat{\theta} - q_{\alpha/2}\hat{\sigma}N^{-1/2}, \hat{\theta} + q_{\alpha/2}\hat{\sigma}N^{-1/2}\right], \tag{2.5}$$

where

$$\hat{\sigma}^2 = \left(\frac{N}{n_1}\hat{\phi}^2_1 + \frac{N}{n_0}\hat{\phi}^2_0 - \hat{\phi}^2_L\right)$$

and $q_{\alpha/2}$ is the upper $\alpha/2$ quantile of a standard normal distribution.

Next, we study the property of the CI I_N in (2.5). As discussed in (Lehmann and Romano (2006)), inferences based on asymptotic results are not reassuring unless some uniform convergence results can be established. We next show that I_N is uniformly asymptotically level $1-\alpha$ over a class of finite populations. For some constants $L_1, L_2, L_3 > 0$, we introduce the following class of finite populations:

$$\begin{aligned}\mathcal{P}_N = \Bigg\{ & \mathbf{U}^*_N = (y^*_1, y^*_0, w^*) : \ \mathbf{U}^*_N \text{ is of size } N, \text{ and} \\ & \text{(a) } \frac{1}{N}\sum_{i=1}^{N} y^{*4}_{ti} \le L_1; \quad \text{(b) } \phi^2(y^*_t) \ge L_2; \quad \text{(c) } P(w^* = \xi_k) \ge \frac{L_3}{K} \\ & \text{for } t = 0, 1 \text{ and } k = 1, \ldots, K \Bigg\}. \end{aligned} \tag{2.6}$$

Under Conditions 1 and 2, if the variances of the potential outcomes are bounded away from zero, then $\mathbf{U}_N$ belongs to $\mathcal{P}_N$, for some L_1, L_2, and L_3. Constraint (a) in the definition of $\mathcal{P}_N$ requires that the fourth moments of the potential outcomes be uniformly bounded. Bounded fourth moments are required for the theoretical development in many existing works (Freedman (2008a,b)). Constraint (b)

requires that the variance of the potential outcomes be bounded away from zero. According to the Cauchy–Schwartz inequality and some straightforward calculations, Constraint (b) implies that the variance of $\sqrt{N}(\hat{\theta} - \theta)$ is bounded away from zero, at least for sufficiently large N, if $\mathbf{U}_N \in \mathcal{P}_N$. Constraint (c) requires that the units with each covariate value not be too rare. Constraints (b) and (c) ensure that the denominators of some quantities in the theoretical analysis are not too small, which is important to establish the desired properties. The set $\mathcal{P}_N$ contains a large class of finite populations. For an illustration, suppose (y_{1i}, y_{0i}, w_i), for $i = 1, \ldots, n$, are i.i.d. observations of some random variables (Y_1, Y_0, W). Then, according to the strong law of large numbers, $\mathbf{U}_N$ belongs to $\mathcal{P}_N$ for sufficiently large N with probability one, as long as (Y_1, Y_0, W) satisfies $E[Y_t^4] < L_1$, $\text{var}[Y_t] > L_2$, and $\mathbb{P}(W = \xi_k) > L_3/K$, for $t = 0, 1$ and $k = 1, \ldots, K$.

Next, we show that I_N has a uniformly asymptotically guaranteed coverage rate over $\mathcal{P}_N$ for any $L_1, L_2, L_3 > 0$.

Theorem 3. *Under Condition* 3*, the CI* I_N *in* (2.5) *is uniformly asymptotically level* $1 - \alpha$ *over* $\mathcal{P}_N$ *for any* $L_1, L_2, L_3 > 0$*; that is,*

$$\liminf_{N \to \infty} \inf_{\mathbf{U}_N \in \mathcal{P}_N} \mathbb{P}(\theta \in I_N) \geq 1 - \alpha.$$

The proof of this theorem is given in the Supplementary Material.

3. Sharp Variance Bound for the Wald Estimator

3.1. Sharp bound for the unidentifiable term in the variance

In the previous section, we discussed the variance bound in completely randomized experiments with perfect compliance, where each unit takes the treatment assigned by the randomization procedure. However, noncompliance often arises in randomized experiments, resulting in some units taking a treatment different to the assigned treatment, following their own will or for other reasons. For each unit i and $t = 0, 1$, we let $d_{ti} \in \{0, 1\}$ denote the treatment that unit i actually takes if assigned to treatment t. In this case, the units can be classified into four groups according to the value of (d_{1i}, d_{0i}) (Angrist, Imbens and Rubin (1996); Frangakis and Rubin (2002)):

$$g_i = \begin{cases} \text{Always Taker (a)} & \text{if } d_{1i} = 1 \text{ and } d_{0i} = 1, \\ \text{Complier (c)} & \text{if } d_{1i} = 1 \text{ and } d_{0i} = 0, \\ \text{Never Taker (n)} & \text{if } d_{1i} = 0 \text{ and } d_{0i} = 0, \\ \text{Defier (d)} & \text{if } d_{1i} = 0 \text{ and } d_{0i} = 1. \end{cases}$$

Let $g = (g_1, \ldots, g_N)$. Then, the characteristics of the population can be viewed as a matrix $\mathbf{U}_\text{c} = (y_1, y_0, w, g)$, where y_1, y_0, and w are defined in Section 2. For

$t = 0, 1$, $k = 1, \ldots, K$ and $h = \mathrm{a}, \mathrm{c}, \mathrm{n}, \mathrm{d}$, let $F_{t|(k,h)}(y) = P(y_t \leq y \mid w = \xi_k, g = h)$, $\pi_{k|h} = P(w = \xi_k \mid g = h)$ and, $\pi_h = P(g = h)$.

In this section, we maintain the following standard assumptions when analyzing the randomized experiment with noncompliance.

Assumption 1. *(i) Monotonicity: $d_{1i} \geq d_{0i}$; (ii) exclusion restriction: $y_{1i} = y_{0i}$ if $d_{1i} = d_{0i}$; and (iii) strong instrument: $\pi_{\mathrm{c}} \geq C_0$, where C_0 is a positive constant.*

Assumption 1 (i) rules out the existence of defiers, and is usually easy to assess. For example, it holds automatically if units in the control group do not have access to the treatment. Assumption 1 (ii) means that the treatment assignment affects the potential outcome only by affecting the treatment that a unit actually receives. Assumption 1 (iii) ensures the existence of compliers. Assumption 1 is commonly adopted to identify the causal effect in the presence of noncompliance; see Angrist, Imbens and Rubin (1996) and Abadie (2003) for detailed discussions of Assumption 1. In a randomized experiment with noncompliance, the parameter of interest is the LATE (Angrist, Imbens and Rubin (1996); Abadie (2003)),

$$\theta_{\mathrm{c}} = \frac{\sum_{i=1}^{N} \tau_i 1\{g_i = \mathrm{c}\}}{\sum_{i=1}^{N} 1\{g_i = \mathrm{c}\}},$$

which is the average effect of the treatment for the compliers.

Under the monotonicity and the exclusion restriction, we have $1\{g_i = \mathrm{c}\} = d_{1i} - d_{0i}$ and $(d_{1i} - d_{0i})\tau_i = \tau_i = y_{1i} - y_{0i}$. Thus,

$$\pi_{\mathrm{c}} = \frac{1}{N}\sum_{i=1}^{N} 1\{g_i = \mathrm{c}\} = \frac{1}{N}\sum_{i=1}^{N}(d_{1i} - d_{0i}) = \mu(d_1) - \mu(d_0),$$

$$\frac{1}{N}\sum_{i=1}^{N}(d_{1i} - d_{0i})\tau_i = \frac{1}{N}\sum_{i=1}^{N}\tau_i = \mu(\tau),$$

and $\theta_{\mathrm{c}} = \pi_{\mathrm{c}}^{-1}\theta$. Hence, θ_{c} can be estimated by the "Wald estimator"

$$\hat{\theta}_{\mathrm{c}} = \hat{\pi}_{\mathrm{c}}^{-1}\hat{\theta},$$

where $\hat{\theta} = \sum_{T_i=1} y_{1i}/n_1 - \sum_{T_i=0} y_{0i}/n_0$ and $\hat{\pi}_{\mathrm{c}} = \sum_{T_i=1} d_{1i}/n_1 - \sum_{T_i=0} d_{0i}/n_0$. Let $z_i = (y_{1i}, y_{0i}, d_{1i}, d_{0i})^{\mathrm{T}}$, $\bar{z} = \sum_{i=1}^{N} z_i/N$, and

$$V_N = \frac{1}{N}\sum_{i=1}^{N}(z_i - \bar{z})(z_i - \bar{z})^{\mathrm{T}}.$$

The asymptotic normality is established under the following regularity condition.

Condition 4. *There is some constant C_λ that does not dependent on N such that the eigenvalues of V_N are not smaller than C_λ.*

Now, we are ready to state the asymptotic normality result.

Theorem 4. *Under Assumption* 1*and Conditions* 1 *and* 4*, we have*

$$\sqrt{N}\sigma_{\mathrm{c}}^{-1}(\hat{\theta}_{\mathrm{c}} - \theta_{\mathrm{c}}) \xrightarrow{d} N(0,1),$$

provided that σ_{c}^2 *is bounded away from zero, where*

$$\sigma_{\mathrm{c}}^2 = \frac{1}{\pi_{\mathrm{c}}^2}\left(\frac{N}{n_1}\phi^2(\tilde{y}_1) + \frac{N}{n_0}\phi^2(\tilde{y}_0) - \phi^2(\tilde{\tau})\right),$$

$\tilde{y}_t = (y_{t1} - \theta_{\mathrm{c}} d_{t1}, \ldots, y_{tN} - \theta_{\mathrm{c}} d_{tN})^{\mathrm{T}}$*, for* $t = 0, 1$*, and* $\tilde{\tau} = \tilde{y}_1 - \tilde{y}_0$*.*

The proof of this theorem is provided in the Supplementary Material. Let $\hat{y}_{ti} = y_{ti} - \hat{\theta}_{\mathrm{c}} d_{ti}$; then, under the conditions of Theorem 4, $\phi^2(\tilde{y}_t)$ can be consistently estimated using

$$\check{\phi}_t^2 = \frac{1}{n_t - 1}\sum_{T_i = t}\left(\hat{y}_{ti} - \frac{1}{n_t}\sum_{T_j = t}\hat{y}_{tj}\right)^2, \tag{3.1}$$

and π_{c} can be consistently estimated using $\hat{\pi}_{\mathrm{c}}$. However, analogously to $\phi^2(\tau)$, $\phi^2(\tilde{\tau})$ is unidentifiable. Here, we construct a sharp bound for $\phi^2(\tilde{\tau})$ using covariate information. Note that the sharp bound has not been obtained, even in the absence of covariates. Define $\tilde{F}_{t|k}(y) = P(\tilde{y}_t \le y \mid w = \xi_k, g = \mathrm{c})$ as a set of lower bounds for $\phi^2(\tilde{\tau})$. Let $\tilde{\mathcal{B}}_{\mathrm{L}} = \{\tilde{b}_{\mathrm{L}} : \tilde{b}_{\mathrm{L}}$ is a functional of π_{c}, $\pi_{k|\mathrm{c}}$, and $\tilde{F}_{t|k}$, for $t = 0, 1$ and $k = 1, \ldots, K$; $\tilde{b}_{\mathrm{L}} \le \phi^2(\tilde{\tau})\}$. Define the set of upper bounds $\tilde{\mathcal{B}}_{\mathrm{H}}$ similarly. Then, we can establish the following sharp bound for $\phi^2(\tilde{\tau})$.

Theorem 5. *A bound for* $\phi^2(\tilde{\tau})$ *is* $[\tilde{\phi}_{\mathrm{L}}^2, \tilde{\phi}_{\mathrm{H}}^2]$*, where*

$$\tilde{\phi}_{\mathrm{L}}^2 = \sum_{k=1}^{K}\pi_{\mathrm{c}}\pi_{k|\mathrm{c}}\int_0^1 (\tilde{F}_{1|k}^{-1}(u) - \tilde{F}_{0|k}^{-1}(u))^2 du,$$

$$\tilde{\phi}_{\mathrm{H}}^2 = \sum_{k=1}^{K}\pi_{\mathrm{c}}\pi_{k|\mathrm{c}}\int_0^1 (\tilde{F}_{1|k}^{-1}(u) - \tilde{F}_{0|k}^{-1}(1-u))^2 du.$$

Moreover, the bound is sharp in the sense that $\tilde{\phi}_{\mathrm{L}}^2$ *is the largest lower bound in* $\tilde{\mathcal{B}}_{\mathrm{L}}$*, and* $\tilde{\phi}_{\mathrm{H}}^2$ *is the smallest upper bound in* $\tilde{\mathcal{B}}_{\mathrm{H}}$*.*

See the Supplementary Material for the proof of this theorem. If there is no covariate, we let $K = 1$ and $\xi_1 = 1$, and define $w_i = 1$, for $i = 1, \ldots, N$. Then, we can obtain a bound without covariates, which has not previously been considered in the literature.

3.2. Estimation of the sharp bound and the CI

To estimate the bounds, we need to estimate π_{c}, $\pi_{k|\mathrm{c}}$, and $\tilde{F}_{t|k}(y)$. Here, π_{c} can be estimated using $\hat{\pi}_{\mathrm{c}}$. We let

$$\hat{\lambda}_{1k} = \frac{1}{n_1}\sum_{T_i=1} d_{1i}1\{w_i = \xi_k\} - \frac{1}{n_0}\sum_{T_i=0} d_{0i}1\{w_i = \xi_k\}$$
$$\hat{\lambda}_{0k} = \frac{1}{n_0}\sum_{T_i=0} (1-d_{0i})1\{w_i = \xi_k\} - \frac{1}{n_1}\sum_{T_i=1} (1-d_{1i})1\{w_i = \xi_k\}.$$

Under Assumption 1, we estimate $\pi_{k|\mathrm{c}}$ and $\tilde{F}_{t|k}(y)$ as

$$\hat{\pi}_{k|\mathrm{c}} = \hat{\pi}_{\mathrm{c}}^{-1}\hat{\lambda}_{1|k},$$
$$\check{F}_{1|k}(y) = \hat{\lambda}_{1|k}^{-1}\left(\frac{1}{n_1}\sum_{T_i=1} d_{1i}1\{\hat{y}_{1i} \le y\}1\{w_i = \xi_k\} - \frac{1}{n_0}\sum_{T_i=0} d_{0i}1\{\hat{y}_{0i} \le y\}1\{w_i = \xi_k\}\right),$$

and

$$\check{F}_{0|k}(y) = \hat{\lambda}_{0|k}^{-1}\left(\frac{1}{n_0}\sum_{T_i=0} (1-d_{0i})1\{\hat{y}_{0i} \le y\}1\{w_i = \xi_k\} - \frac{1}{n_1}\sum_{T_i=1} (1-d_{1i})1\{\hat{y}_{1i} \le y\}1\{w_i = \xi_k\}\right),$$

where $\hat{y}_{ti} = y_{ti} - \hat{\theta}_{\mathrm{c}} d_{ti}$, for $t = 0, 1$ and $i = 1, \ldots, N$.

We then obtain estimators for $\tilde{\phi}_{\mathrm{L}}^2$ and ϕ_{H}^2 by plugging these estimators into the expressions of Theorem 5:

$$\begin{aligned}\check{\phi}_{\mathrm{L}}^2 &= \sum_{k=1}^{K} \hat{\pi}_{\mathrm{c}}\hat{\pi}_{k|\mathrm{c}} \int_0^1 (\check{F}_{1|k}^{-1}(u) - \check{F}_{0|k}^{-1}(u))^2 du,\\ \check{\phi}_{\mathrm{H}}^2 &= \sum_{k=1}^{K} \hat{\pi}_{\mathrm{c}}\hat{\pi}_{k|\mathrm{c}} \int_0^1 (\check{F}_{1|k}^{-1}(u) - \check{F}_{0|k}^{-1}(1-u))^2 du.\end{aligned} \tag{3.2}$$

The estimator $\check{F}_{t|k}(y)$ may not be monotonic with respect to y, resulting in difficulties in a theoretical analysis. However, $\check{F}_{t|k}^{-1}(u)$ is still well defined, and in the next theorem, we show that the non-monotonicity of $\check{F}_{t|k}(y)$ does not diminish the consistency of $\check{\phi}_{\mathrm{L}}^2$ and $\check{\phi}_{\mathrm{H}}^2$.

In the following asymptotic analysis, we denote $\mathbf{U}_{\mathrm{c}}$ by $\mathbf{U}_{\mathrm{c},N}$. We assume that $\mathbf{U}_{\mathrm{c},N}$ satisfies Assumption 1, Condition 1, and the following two conditions. The following conditions are modified versions of Conditions 2 and 3 in the presence

of noncompliance.

Condition 5. *There is some constant $C_{\pi,\mathrm{c}}$ that does not depend on N such that $\pi_{k|\mathrm{c}}\pi_{\mathrm{c}} \geq C_{\pi,\mathrm{c}}/K$, for $k = 1, \ldots, K$.*

Condition 6. $K^2 \log K \max\{C_N^4, 1\}/N \to 0$ *as* $N \to \infty$, *where* $C_N = \max_{t,i} |y_{ti}|$.

Then, we are ready to establish the consistency of the estimators proposed in (3.2).

Theorem 6. *Under Assumption* 1 *and Conditions* 1, 5, *and* 6, *we have*

$$(\check{\phi}_{\mathrm{L}}^2, \check{\phi}_{\mathrm{H}}^2) - (\tilde{\phi}_{\mathrm{L}}^2, \tilde{\phi}_{\mathrm{H}}^2) \xrightarrow{P} 0.$$

The proof of this theorem is relegated to the Supplementary Material. The lower bound $\tilde{\phi}_{\mathrm{L}}^2$ for $\phi^2(\tilde{\tau})$ implies an upper bound for σ_{c}^2. By Theorems 4 and 6, we can construct a conservative $1 - \alpha$ CI for θ_{c},

$$I_{\mathrm{c},N} = \left[\hat{\theta}_{\mathrm{c}} - q_{\alpha/2}\hat{\sigma}_{\mathrm{c}} N^{-1/2}, \hat{\theta}_{\mathrm{c}} + q_{\alpha/2}\hat{\sigma}_{\mathrm{c}} N^{-1/2}\right], \tag{3.3}$$

where

$$\hat{\sigma}_{\mathrm{c}}^2 = \frac{1}{\hat{\pi}_{\mathrm{c}}^2}\left(\frac{N}{n_1}\check{\phi}_1^2 + \frac{N}{n_0}\check{\phi}_0^2 - \check{\phi}_L^2\right) \tag{3.4}$$

and $q_{\alpha/2}$ is the upper $\alpha/2$ quantile of a standard normal distribution. We then show that $I_{\mathrm{c},N}$ is uniformly asymptotically level $1 - \alpha$ over a class of finite populations. In the following asymptotic analysis, we denote $\mathbf{U}_{\mathrm{c}}$ by $\mathbf{U}_{\mathrm{c},N}$. For any finite population $\mathbf{U}_{\mathrm{c},N}^* = (y_1^*, y_0^*, w^*, g^*)$, we define $\tilde{y}_t^*$ similarly to $\tilde{y}_t$, for $t = 0, 1$. Let Λ_N^* be the smallest eigenvalue of

$$\frac{1}{N}\sum_{i=1}^{N}(z_i^* - \bar{z}^*)(z_i^* - \bar{z}^*)^{\mathrm{T}},$$

where $z_i^* = (y_{1i}^*, y_{0i}^*, d_{1i}^*, d_{0i}^*)^{\mathrm{T}}$ and $\bar{z}^* = \sum_{i=1}^N z_i^*/N$. For some constants $L_0, L_1, L_2, L_3 > 0$, we introduce the following class of finite populations:

$$\begin{aligned}\mathcal{P}_{\mathrm{c},N} = \Bigg\{ & \mathbf{U}_{\mathrm{c},N}^* = (y_1^*, y_0^*, w^*, g^*) : \ \mathbf{U}_{\mathrm{c},N}^* \text{ is of size } N \text{ and satisfies} \\ & \text{(a) Assumption 1; (b) } \Lambda_N^* \geq L_0; \text{ (c) } \frac{1}{N}\sum_{i=1}^{N} y_{ti}^{*4} \leq L_1; \\ & \text{(d) } \phi^2(\tilde{y}_t^*) \geq L_2; \quad \text{and (e) } P(w^* = \xi_k \mid g^* = \mathrm{c})P(g^* = \mathrm{c}) \geq \frac{L_3}{K} \\ & \text{for } t = 0, 1 \text{ and } k = 1, \ldots, K \Bigg\}.\end{aligned}$$

Table 1. Bounds ϕ^2_{AL}, ϕ^2_{AH}, ϕ^2_{DL}, ϕ^2_{DH}, ϕ^2_{L}, and ϕ^2_{H} and the true value of $\phi^2(\tau)$ under different population sizes.

	ϕ^2_{AL}	ϕ^2_{AH}	ϕ^2_{DL}	ϕ^2_{DH}	ϕ^2_{L}	ϕ^2_{H}	$\phi^2(\tau)$
$N = 400$	0.98	58.14	1.02	58.04	9.04	42.92	16.72
$N = 800$	0.78	54.79	0.74	54.76	8.83	40.53	17.70
$N = 2{,}000$	0.88	58.23	0.86	58.22	9.14	42.25	18.56

Constraint (a) is required for the identification of the LATE. Constraint (b) is a regularity condition that ensures the asymptotic normality of $\hat{\theta}_{\mathrm{c}}$. Constraints (c), (d), and (e) are similar to those in the definition of $\mathcal{P}_N$ in Section 2.3. Here, $\mathcal{P}_{\mathrm{c},N}$ can be a large class of finite populations if L_0, L_2, and L_3 are small and L_1 is large. The class of finite populations $\mathcal{P}_{\mathrm{c},N}$ can be related to some class of generic distributions in the same way as discussed before Theorem 3. The details are omitted here. Similar arguments to those in the proof of Theorem 3 show the following result.

Theorem 7. *Under Condition 6, the CI* $I_{\mathrm{c},N}$ *in* (3.3) *is uniformly asymptotically level* $1-\alpha$ *over* $\mathcal{P}_{\mathrm{c},N}$*; that is,*

$$\liminf_{N\to\infty} \inf_{\mathbf{U}_{\mathrm{c},N}\in\mathcal{P}_{\mathrm{c},N}} \mathbb{P}(\theta \in I_{\mathrm{c},N}) \geq 1-\alpha.$$

4. Simulations

4.1. Completely randomized experiments with perfect compliance

In this subsection, we evaluate the performance of the bounds and the estimators $\hat{\phi}_{\mathrm{L}}$, $\hat{\phi}_{\mathrm{H}}$ proposed in Section 2 using simulations. We first generate a finite population of size N by drawing i.i.d. samples from the following data-generation process:

(i) W takes a value in $\{1,2,3,4\}$ with equal probability;

(ii) for $w = 1,2,3,4$, $Y_1 \mid W = w \sim N(\mu_w, \phi^2_w)$, $V \mid W = w \sim N(0, 6-\phi^2_w)$, $Y_1 \perp\!\!\!\perp V \mid W$, and $Y_0 = 0.3Y_1 + V$, where $(\mu_1,\mu_2,\mu_3,\mu_4) = (3,0,-2,4)$ and $(\phi^2_1,\phi^2_2,\phi^2_3,\phi^2_4) = (2,1.5,5,4)$.

The generated values are viewed as the fixed finite population. We take $N = 400, 800$ and 2000, to demonstrate the performance of the proposed method under different population sizes. The following table shows the $\phi^2(\tau)$ and the true value of different bounds under different population sizes.

It can be seen that the intervals $[\phi^2_{\mathrm{AL}}, \phi^2_{\mathrm{AH}}]$, $[\phi^2_{\mathrm{DL}}, \phi^2_{\mathrm{DH}}]$, and $[\phi^2_{\mathrm{L}}, \phi^2_{\mathrm{H}}]$ all contain $\phi^2(\tau)$, and hence the bounds are all valid. Under all population sizes, the lower bound ϕ^2_{L} is much larger than the other two lower bounds, and the upper bound ϕ^2_{H} is much smaller than the other two upper bounds.

Table 2. RMSE of the estimators for ϕ^2_{L} and ϕ^2_{H} under different population sizes ($n_1 = n_0 = N/2$).

	$N = 400$	$N = 800$	$N = 2{,}000$
ϕ^2_{L}	1.71	0.87	0.66
ϕ^2_{H}	2.31	1.36	1.00

Table 3. Average widths (AWs) and coverage rates (CRs) of 95% CIs based on the naive bound, ϕ^2_{AL}, ϕ^2_{DL}, and ϕ^2_{L} under different population sizes ($n_1 = n_0 = N/2$).

Method	naive		ϕ^2_{AL}		ϕ^2_{DL}		ϕ^2_{L}	
	AW	CR	AW	CR	AW	CR	AW	CR
$N = 400$	1.511	98.0%	1.495	97.8%	1.493	97.8%	1.383	96.6%
$N = 800$	1.033	97.8%	1.025	97.7%	1.025	97.7%	0.945	96.7%
$N = 2{,}000$	0.674	98.4%	0.668	98.3%	0.668	98.3%	0.619	97.3%

The above results show the effectiveness of our bounds at the population level. Next, we consider the performance of the proposed bound estimators in completely randomized experiments. In the simulation, half of the units are assigned randomly to the treatment group, and the other half are assigned to the control group. Then, we estimate the proposed bounds using the estimators defined in (2.4). The randomized assignment is repeated 1,000 times. The root mean squared error (RMSE) of the bound estimator under different N is summarized in Table 2, showing that the RMSE decreases as the sample size increases, which confirms the consistency result in Theorem 2.

Next, we evaluate the performance of the bound estimators in terms of constructing CIs. Different CIs can be constructed based on the asymptotic normality in (2.1) and different lower bounds for $\phi^2(\tau)$. To obtain the CIs, we use $\hat{\phi}^2_t$ defined in (2.2) to estimate $\phi^2(y_t)$, for $t = 0, 1$, and replace $\phi^2(\tau)$ in σ^2 with the estimators of the different lower bounds. We also estimate the bounds of Aronow, Green and Lee (2014) and Ding, Feller and Miratrix (2019) using plug-in estimators, as suggested in these works, and estimate the proposed lower bound using the estimators defined in (2.4). The following table shows the average width (AW) and coverage rate (CR) of the 95% CIs based on the naive lower bound zero (Neyman (1990)), the estimator of ϕ^2_{AL} (Aronow, Green and Lee (2014)), the estimator of ϕ^2_{DL} (Ding, Feller and Miratrix (2019)), and the estimator of ϕ^2_{L}.

The CIs based on the estimator of ϕ^2_{L} are the narrowest, and the corresponding coverage rate is the closest to 95% of the four CIs under all population sizes. See the Supplementary Material for additional simulation results on the performance of the proposed CI.

Table 4. Bounds constructed with and without covariates, and the RMSE of their estimators under different population sizes. LNC: lower bound without covariate; HNC: upper bound without covariate; LC: lower bound with covariate; HC: upper bound with covariate ($n_1 = n_0 = N/2$).

	LNC		HNC		LC		HC	
	Value	RMSE	Value	RMSE	Value	RMSE	Value	RMSE
$N = 400$	0.78	0.75	40.05	5.41	9.82	4.62	29.42	5.94
$N = 800$	0.73	0.45	39.61	3.49	10.42	3.92	28.79	5.35
$N = 2{,}000$	0.72	0.27	37.19	2.10	9.52	2.45	28.45	4.55

4.2. Completely randomized experiments with noncompliance

Here, we show the simulation performance of the bounds and the estimators $\check{\phi}_{\mathrm{L}}$, $\check{\phi}_{\mathrm{H}}$ proposed in Section 3. First, we generate finite populations of size $N = 400, 800$, and 2000 by drawing i.i.d. samples from the following data-generation process:

(i) generate the compliance type $G \in \{\mathrm{a}, \mathrm{c}, \mathrm{n}\}$, with the probability that $G = \mathrm{a}$, c, and n being 0.2, 0.7, and 0.1, respectively;

(ii) for $h = \mathrm{a}, \mathrm{c}$, and n, the conditional distribution $W \mid G = h$ has probability mass p_{1h}, p_{2h}, p_{3h}, and p_{4h} at $1, 2, 3$, and 4, respectively, where $(p_{1\mathrm{a}}, p_{2\mathrm{a}}, p_{3\mathrm{a}}, p_{4\mathrm{a}}) = (0.15, 0.2, 0.3, 0.35)$, $(p_{1\mathrm{c}}, p_{2\mathrm{c}}, p_{3\mathrm{c}}, p_{4\mathrm{c}}) = (0.25, 0.25, 0.25, 0.25)$, and $(p_{1\mathrm{n}}, p_{2\mathrm{n}}, p_{3\mathrm{n}}, p_{4\mathrm{n}}) = (0.35, 0.3, 0.2, 0.15)$;

(iii) for $w = 1, 2, 3, 4$, $Y_1 \mid W = w \sim N(\mu_w, \phi_w^2)$, where $(\mu_1, \mu_2, \mu_3, \mu_4) = (3, 0, -2, 4)$ and $(\phi_1^2, \phi_2^2, \phi_3^2, \phi_4^2) = (2, 1.5, 5, 4)$;

(iv) $Y_0 \mid G = \mathrm{c}, W = w \sim N(0.3w, 6 - \phi_w^2)$, and $Y_0 = Y_1$ if $G = \mathrm{a}$ or n.

Theorem 5 can also provide a bound without using covariates. So to illustrate the usefulness of the covariates, we compare the bounds constructed with and without them. In the following, the lower bounds constructed with and without covariates are denoted by "LC" and "LNC", respectively, and the upper bounds constructed with and without covariates are denoted by "HC" and "HNC", respectively. As in Section 4.1, half of the units are assigned randomly to the treatment group, and the other half are assigned to the control group. Then, we estimate the bounds for each of these random assignments using the estimators proposed in (3.2). The randomized assignment is repeated 1,000 times. In the simulation $\phi^2(\tilde{\tau})$ is equal to $9.97, 10.41$, and 9.56 when $N = 400, 600$, and 2000, respectively. The following table shows the true values of various bounds and the RMSE of their estimators under different population sizes.

Table 4 shows that LC is much larger than LNC, and HC is much smaller than HNC. Therefore, covariates are useful in terms of sharpening the bounds.

Table 5. Average widths (AWs) and coverage rates (CRs) of 95% CIs based on the naive bound, HNL and HL under different population sizes. LNC: lower bound without covariate; LC: lower bound with covariate ($n_1 = n_0 = N/2$)

Method	naive		LNC		LC	
	AW	CR	AW	CR	AW	CR
$N = 400$	2.188	98.5%	2.168	98.4%	1.990	97.3%
$N = 800$	1.553	97.9%	1.542	97.8%	1.399	96.4%
$N = 2{,}000$	0.980	98.2%	0.973	98.2%	0.893	96.6%

In general, the RMSE of the bound estimators generally decreases as the sample size increases, validating the consistency result in Theorem 6.

Next, we construct CIs based on the asymptotic normality in Theorem 4 and different lower bounds for $\phi^2(\tilde{\tau})$. To obtain the CIs, we use $\check{\phi}_t^2$ defined in (3.1) to estimate $\phi^2(\tilde{y}_t)$, for $t = 0, 1$, and replace $\phi^2(\tilde{\tau})$ in σ_c^2 with the estimators of various lower bounds. The following table shows the average width (AW) and coverage rate (CR) of the 95% CIs based on the naive lower bound zero and the estimator of the lower bound in Theorem 5, construct with and without covariates.

The results show that the CI based on the estimator of LC is the narrowest, and that the corresponding coverage rate is the closest to 95% of the three CIs under all population sizes. This demonstrates the usefulness of covariates in terms of constructing CIs.

5. Real-Data Applications

5.1. Application to ACTG protocol 175

In this section, we apply our approach proposed in Section 2 to a data set from the randomized trial ACTG protocol 175 (Hammer et al. (1996)). The data used in this subsection are available from the R package "speff2trial" (`https://cran.r-project.org/web/packages/speff2trial/index.html`). The ACTG 175 study evaluated four therapies for subjects infected with the human immunodeficiency virus whose CD4 cell counts (a measure of immunologic status) were between 200 and 500 mm^{-3}. Here, we regard the 2,139 enrolled subjects as a finite population, and consider two treatment arms: the standard zidovudine monotherapy (denoted by "arm 0"), and the combination therapy with zidovudine and didanosine (denoted by "arm 1"). The parameter of interest is the average treatment effect of the combination therapy on the CD8 cell count, measured at 20 ± 5 weeks post baseline, compared with that of the monotherapy. In the randomized trial, 532 subjects are assigned randomly to arm 0, and 522 subjects are assigned randomly to arm 1. The available covariate is the age of each subject. In order to meet Condition 3, we divide the subjects into $\lfloor N^{1/4} \rfloor = 6$ groups according to age: less than 20 years old, between 21 and 30 years old, between

31 and 40 years old, between 41 and 50 years old, between 51 and 60 years old, and older than 60 years old. We then use age group, gender, and antiretroviral history as covariates, and apply our proposed method. The proposed bounds ϕ^2_{L} and ϕ^2_{H} are estimated using the estimators defined in (2.4). We also estimate the bounds proposed in the literature (Aronow and Middleton (2013); Ding, Feller and Miratrix (2019)). Bounds ϕ^2_{AL}, ϕ^2_{AH}, ϕ^2_{DL}, and ϕ^2_{DH} are estimated using plug-in estimators, as suggested in Aronow, Green and Lee (2014) and Ding, Feller and Miratrix (2019). The estimates of the lower bounds ϕ^2_{AL}, ϕ^2_{DL}, and ϕ^2_{L} are 0.12, 0.27, and 4.75, respectively; the estimates of the upper bounds ϕ^2_{AH}, ϕ^2_{DH}, and ϕ^2_{H} are 70.79, 69.87, and 65.67, respectively (values are divided by 10,000). The estimate of the proposed lower bound is the largest among the three lower bounds, and the estimate of the proposed upper bound is the smallest among the three upper bounds. The 95% CI constructed using zero as a lower bound for $\phi^2(\tau)$ is $[-13.27, 92.79]$. Using the lower bound estimate of Aronow, Green and Lee (2014) leads to the CI $[-13.25, 92.77]$, and using that of Ding, Feller and Miratrix (2019) leads to the CI $[-13.23, 92.75]$. The CI $[-12.46, 91.98]$ is obtained by using $\hat{\phi}^2_{\mathrm{L}}$ given in (2.4). The widths of the four CIs are 106.07, 106.02, 105.98, and 104.44, respectively. Comparing the CI width of the naive method with the widths based on the lower bound of Aronow and Middleton (2013) and Ding, Feller and Miratrix (2019), and the proposed lower bound for $\phi^2(\tau)$, the CI width reductions are 0.04, 0.09, and 1.62, respectively. The reductions for the three methods are not very large compared with that of the naive method. The reason may be that $N\hat{\phi}^2_1/n_1 + N\hat{\phi}^2_0/n_0$ is too large relative to the estimator of the lower bound for $\phi^2(\tau)$ in this specific problem, where $\hat{\phi}^2_1$ and $\hat{\phi}^2_0$ are the estimators for $\phi^2(y_1)$ and $\phi^2(y_0)$, respectively. Note that the CI width is proportional to $\sqrt{N\hat{\phi}^2_1/n_1 + N\hat{\phi}^2_0/n_0 - \hat{\phi}^2_{\mathrm{B}}}$, where $\hat{\phi}^2_{\mathrm{B}}$ is the estimator of the lower bound for $\phi^2(\tau)$. This implies that the lower bound estimator $\hat{\phi}^2_{\mathrm{B}}$ does not play an important role in the CI width if $N\hat{\phi}^2_1/n_1 + N\hat{\phi}^2_0/n_0$ is much larger than $\hat{\phi}^2_{\mathrm{B}}$.

5.2. Application to JOBS II

In this section, we apply our approach proposed in Section 3 to a data set from the randomized trial JOBS II (Vinokuir, Price and Schul (1995)). The data used in this subsection are available from `https://www.icpsr.umich.edu/web/ICPSR/studies/2739`. The JOBS II intervention trial studied the efficacy of a job training intervention in preventing depression caused by job loss, and in prompting high-quality re-employment. The treatment consisted of five half-day training seminars designed to enhance the participants' job search strategies. The control group receives a booklet with some brief tips. After some screening procedures, 1,801 respondents were enrolled in this study, with 552 and 1,249 respondents in the control and treatment groups, respectively. Of

the respondents assigned to the treatment group, only 54% participated in the treatment. Thus, there is a large proportion of noncompliance in this study. The parameter of interest is the LATE of the treatment on the depression score (a larger score indicates more severer depression). We use gender, initial risk status and economic hardship as the covariates, and apply our proposed method. The estimates for $\tilde{\phi}_{\mathrm{L}}^2$ and $\tilde{\phi}_{\mathrm{H}}^2$ are 0.23 and 0.81, respectively. The 95% CIs constructed using the naive bound zero and $\tilde{\phi}_{\mathrm{L}}$ are $[-0.2428, 0.0271]$ and $[-0.2360, 0.0202]$, respectively. Our method shortens the CI by 0.014. When testing the null hypothesis that $\theta_{\mathrm{c}} = 0$ against the alternative hypothesis that $\theta_{\mathrm{c}} < 0$, the naive method gives a p-value of 0.059, whereas our method gives a p-value of 0.049. Thus, our method can detect the treatment effect at the 0.05 significance level, whereas the naive method cannot.

6. Discussion

In this paper, we establish sharp variance bounds for the widely used difference-in-means estimator and Wald estimator in the presence of covariates in completely randomized experiments. These bounds can help to improve the performance of inference procedures based on a normal approximation. We do not impose any assumption on the support of the outcomes; hence, our results are general and apply to both binary and continuous outcomes. The variances of the difference-in-means estimator in matched pair randomized experiments (Imai (2008)) and those of the Horvitz–Thompson estimator in stratified randomized experiments and clustered randomized experiments (Miratrix, Sekhon and Yu (2013); Mukerjee, Dasgupta and Rubin (2018); Middleton and Aronow (2015)) share a similar unidentifiable term to that considered in this study. Moreover, the unidentifiable phenomenon appears in the asymptotic variance of the regression adjustment estimators; see Lin (2013), Freedman (2008a), and Bloniarz et al. (2016). The insights provided in this paper also apply in these settings, although we omit the details here. It would be of great interest to extend our work to randomized experiments with other randomization schemes, such as a 2^2 factorial design (Lu (2017)) or some other complex assignment mechanism (Mukerjee, Dasgupta and Rubin (2018)).

Supplementary Material

The online Supplementary Material contains further simulation results and proofs for all theoretical results presented in this paper.

Acknowkedgments

Wang's research was supported by the National Natural Science Foundation of China (General program 12271510, General program 11871460, and program

for Innovative Research Group 61621003), and a grant from the Key Lab of Random Complex Structure and Data Science, CAS. Miao's research was supported by the National Key R&D Program (2022YFA1008100) and the National Natural Science Foundation of China (Genral program 12071015).

References

Abadie, A. (2003). Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics* **113**, 231–263.

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

Aronow, P. M., Green, D. P. and Lee, D. K. K. (2014). Sharp bounds on the variance in randomized experiments. *The Annals of Statistics* **42**, 850–871.

Aronow, P. M. and Middleton, J. A. (2013). A class of unbiased estimators of the average treatment effect in randomized experiments. *Journal of Causal Inference* **1**, 135–154.

Belloni, A., Chernozhukov, V. and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies* **81**, 608–650.

Bloniarz, A., Liu, H., Zhang, C.-H., Sekhon, J. S. and Yu, B. (2016). Lasso adjustments of treatment effect estimates in randomized experiments. *Proceedings of the National Academy of Sciences* **113**, 7383–7390.

Chan, K. C. G., Yam, S. C. P. and Zhang, Z. (2016). Globally efficient non-parametric inference of average treatment effects by empirical balancing calibration weighting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **78**, 673–700.

Cochran, W. G. (1977). *Sampling Techniques*. 3rd Edition. Wiley, New York.

Ding, P. and Dasgupta, T. (2016). A potential tale of two-by-two tables from completely randomized experiments. *Journal of the American Statistical Association* **111**, 157–168.

Ding, P., Feller, A. and Miratrix, L. (2019). Decomposing treatment effect variation. *Journal of the American Statistical Association* **114**, 304–317.

Ding, P. and Miratrix, L. W. (2018). Model-free causal inference of binary experimental data. *Scandinavian Journal of Statistics* **46**, 200–214.

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

Freedman, D. A. (2008a). On regression adjustments in experiments with several treatments. *The Annals of Applied Statistics* **2**, 176–196.

Freedman, D. A. (2008b). On regression adjustments to experimental data. *Advances in Applied Mathematics* **40**, 180–193.

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundacker, H., Schooley, R. T., Haubrich, R. H. et al. (1996). A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *The New England Journal of Medicine* **335**, 1081–1090.

Hirano, K., Imbens, G. W. and Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* **71**, 1161–1189.

Hong, H., Leung, M. P. and Li, J. (2020). Inference on finite-population treatment effects under limited overlap. *The Econometrics Journal* **23**, 32–47.

Imai, K. (2008). Variance identification and efficiency analysis in randomized experiments under the matched-pair design. *Statistics in Medicine* **27**, 4857–4873.

Imbens, G. W. (2004). Nonparametric estimation of average effects under exogeneity: A review. *The Review of Economics and Statistics* **86**, 4–29.

Imbens, G. W. and Rosenbaum, P. R. (2005). Robust accurate confidence intervals with a weak instrument. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **25**, 305–327.

Kallus, N. (2018). Optimal a priori balance in the design of controlled experiments. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80**, 85–112.

Lehmann, E. L. and Romano, J. P. (2006). *Testing Statistical Hypotheses*. Springer Science & Business Media.

Li, X. and Ding, P. (2017). General forms of finite population central limit theorems with applications to causal inference. *Journal of the American Statistical Association* **112**, 1759–1769.

Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining freedman's critique. *The Annals of Applied Statistics* **7**, 295–318.

Lu, J. (2017). Sharpening randomization-based causal inference for 22 factorial designs with binary outcomes. *Statistical Methods in Medical Research* **28**, 1064–1078.

Ma, W., Tu, F. and Liu, H. (2020). Regression analysis for covariate-adaptive randomization: A robust and efficient inference perspective. *arXiv:2009.02287*.

Middleton, J. A. and Aronow, P. M. (2015). Unbiased estimation of the average treatment effect in cluster-randomized experiments. *Statistics, Politics and Policy* **6**, 39–75.

Miratrix, L. W., Sekhon, J. S. and Yu, B. (2013). Adjusting treatment effect estimates by post-stratification in randomized experiments. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **75**, 369–396.

Mukerjee, R., Dasgupta, T. and Rubin, D. B. (2018). Using standard tools from finite population sampling to improve causal inference for complex experiments. *Journal of the American Statistical Association* **113**, 868–881.

Neyman, J. (1990). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science* **5**, 465–472.

Nolen, T. L. and Hudgens, M. G. (2011). Randomization-based inference within principal strata. *Journal of the American Statistical Association* **106**, 581–593.

Robins, J. M. (1988). Confidence intervals for causal parameters. *Statistics in Medicine* **7**, 773–785.

Schochet, P. Z. (2013). Estimators for clustered education RCTs using the neyman model for causal inference. *Journal of Educational and Behavioral Statistics* **38**, 219–238.

Shao, J., Yu, X. and Zhong, B. (2010). A theory for testing hypotheses under covariate-adaptive randomization. *Biometrika* **97**, 347–360.

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

Ruoyu Wang

Academy of Mathematics and Systems Science, Chinese Academy of Sciences, 55 Zhongguancun East Road, Haidian District, Beijing 100190, China.

E-mail: wangruoyu17@mails.ucas.edu.cn

Qihua Wang

Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China.

University of Chinese Academy of Sciences, Beijing 101408, China.

E-mail: qhwang@amss.ac.cn

Wang Miao

School of Mathematical Sciences, Peking University, 5 Summer Palace Road, Haidian District, Beijing 100871, China.

E-mail: mwfy@pku.edu.cn

Xiaohua Zhou

Beijing International Center for Mathematical Research and Department of Biostatistics, Peking University, 5 Summer Palace Road, Haidian District, Beijing 100871, China.

E-mail: azhou@math.pku.edu.cn

(Received October 2021; accepted September 2022)