The Annals of Statistics
2014, Vol. 42, No. 3, 850–871
DOI: 10.1214/13-AOS1200
© Institute of Mathematical Statistics, 2014

SHARP BOUNDS ON THE VARIANCE IN RANDOMIZED EXPERIMENTS

BY PETER M. ARONOW, DONALD P. GREEN AND DONALD K. K. LEE

Yale University, Columbia University and Yale University

We propose a consistent estimator of sharp bounds on the variance of the difference-in-means estimator in completely randomized experiments. Generalizing Robins [*Stat. Med.* **7** (1988) 773–785], our results resolve a well-known identification problem in causal inference posed by Neyman [*Statist. Sci.* **5** (1990) 465–472. Reprint of the original 1923 paper]. A practical implication of our results is that the upper bound estimator facilitates the asymptotically narrowest conservative Wald-type confidence intervals, with applications in randomized controlled and clinical trials.

1. Introduction. We consider the long-standing problem of estimating the variance of the difference-in-means estimator as applied to a completely randomized experiment performed on a random sample of size n selected without replacement from a population of size N under a nonparametric model of deterministic potential outcomes. It has been known since Neyman [13] that neither unbiased nor consistent variance estimation is generally possible in this setting, due to the fact that the joint distribution of the potential outcomes can never be fully recovered from data.

In this paper, we propose an interval estimator that is consistent for sharp bounds, defined as the smallest interval containing all values of the variance that are compatible with the observable information. The upper bound is never larger than and often smaller than conventional approximations. Our estimator is also applicable to all possible cases of N and n ($n = N < \infty$, $n < N < \infty$, and $N = \infty$), thus providing a unified treatment of the problem. In the case where the outcomes are dichotomous and $n = N < \infty$, our estimator reproduces Robins [14] results. The case $n < N < \infty$ generalizes the settings considered by prior researchers. Unbiased variance estimation is not generally possible when $N < \infty$, but our estimator produces asymptotically sharp bounds. When the population size N is infinite, our estimator recovers the standard variance point estimator for mean differences between independent groups [13].

A practical implication of our work is that it facilitates confidence intervals that are often narrower than intervals produced by conventional methods: our upper

Received May 2013; revised December 2013.
MSC2010 subject classifications. Primary 62A01; secondary 62D99, 62G15.
Key words and phrases. Causal inference, finite populations, potential outcomes, randomized experiments, variance estimation.

bound variance estimator may be used to construct conservative Wald-type confidence intervals for the average treatment effect. Asymptotically, these intervals are the narrowest Wald-type intervals that are assured to have at least the nominal coverage. We illustrate empirical performance using data from an randomized controlled trial, discuss extensions and provide R code implementing our estimator. An implementation in Stata is also available from the authors.

2. Setting. Consider a population U_N consisting of $N \geq 4$ units. From U_N, n units are randomly sampled into the experimental sample, and the remaining $N - n$ units are left unsampled. Of the n sampled units, $m \geq 2$ units are randomly assigned to the treatment condition, and $n - m \geq 2$ units are randomly assigned to the control condition. Let the indicator variable X_i^T be one if unit i is assigned to the treatment condition, and let the indicator X_i^C be one if unit i is assigned to the control condition. If $X_i^T = X_i^C = 0$, then the unit is unsampled. Since units are sampled without replacement, $X_i^T + X_i^C \leq 1$. Without loss of generality, assume an index ordering $i = 1, \dots, N$ such that those assigned to treatment come first, $X_1^T, \dots, X_m^T = 1$, and those assigned to control come after, $X_{m+1}^C, \dots, X_n^C = 1$, and the remaining $N - n$ unsampled units, if any, come last.

Associated with each unit i are two potential outcomes [13, 15] under control and treatment, respectively: y_{0i} and y_{1i}. For each unit i, the analyst then observes y_{0i} when $X_i^C = 1$ and y_{1i} when $X_i^T = 1$. Given elements v_i, w_i for $i = 1, \dots, N$, we define the finite population mean $\mu_N(v)$, finite population variance $\sigma_N^2(v)$ and finite population covariance $\sigma_N(v, w)$, respectively, as

$$\mu_N(v) = \frac{1}{N}\sum_{i=1}^N v_i, \qquad \sigma_N^2(v) = \frac{1}{N}\sum_{i=1}^N \{v_i - \mu_N(v)\}^2,$$

$$\sigma_N(v, w) = \frac{1}{N}\sum_{i=1}^N \{v_i - \mu_N(v)\}\{w_i - \mu_N(w)\}.$$

The average treatment effect for the population U_N is $\tau_N = \mu_N(y_1) - \mu_N(y_0)$. The difference-in-means estimator of τ_N is

$$\hat{\tau}_N = \hat{\mu}_N(y_1) - \hat{\mu}_N(y_0) = \frac{1}{m}\sum_{i=1}^m y_{1i} - \frac{1}{n-m}\sum_{i=m+1}^n y_{0i}, \tag{1}$$

with $\mathrm{E}_X(\hat{\tau}_N) = \tau_N$, where the expectation operator E_X averages over all $\binom{N}{n}\binom{n}{m}$ possible treatment assignments.

Our inferential target is the variance of $\hat{\tau}_N$. Adapting Freedman [3], Proposition 1, the variance is

$$V_N = \frac{1}{N-1}\left\{\frac{N-m}{m}\sigma_N^2(y_1) + \frac{N-(n-m)}{n-m}\sigma_N^2(y_0) + 2\sigma_N(y_1, y_0)\right\}. \tag{2}$$

The unknown quantities in this expression are $\sigma_N^2(y_1)$, $\sigma_N^2(y_0)$ and $\sigma_N(y_1, y_0)$. By Cochran [1], Theorem 2.4, unbiased estimators of $\sigma_N^2(y_1)$ and $\sigma_N^2(y_0)$ are

$$\hat{\sigma}_N^2(y_1) = \frac{N-1}{N(m-1)} \sum_{i=1}^{m} \{y_{1i} - \hat{\mu}_N(y_1)\}^2,$$

$$\hat{\sigma}_N^2(y_0) = \frac{N-1}{N(n-m-1)} \sum_{i=m+1}^{n} \{y_{0i} - \hat{\mu}_N(y_0)\}^2.$$

Since both potential outcomes y_{0i} and y_{1i} for the same unit can never be observed simultaneously, consistent estimators do not generally exist for $\sigma_N(y_1, y_0)$ or for V_N when the population size N is finite. However, when the population being sampled from is infinite ($N = \infty$), Neyman [13] noted that the control and treatment units are effectively sampled independently from their respective distributions. Hence, the covariance term vanishes, and V_N is point identified. To see this, let $N \to \infty$ while holding m and n fixed so that $V_N \to \frac{1}{m}\sigma_N^2(y_1) + \frac{1}{n-m}\sigma_N^2(y_0)$, the sampling variance for the difference of independent means.

2.1. *Neyman* [13] *approximations when* $n = N$. When $n = N$, the sampling variance of the difference-in-means estimator reduces to

$$V_n = \frac{1}{n-1}\left\{\frac{n-m}{m}\sigma_n^2(y_1) + \frac{m}{n-m}\sigma_n^2(y_0) + 2\sigma_n(y_1, y_0)\right\}. \tag{3}$$

Neyman [13] proposed an estimator of V_n that uses the inequality $2\sigma_n(y_1, y_0) \le 2\{\sigma_n^2(y_1)\sigma_n^2(y_0)\}^{1/2} \le \sigma_n^2(y_1) + \sigma_n^2(y_0)$, by application of the Cauchy–Schwarz inequality and the inequality of arithmetic and geometric means. An upper bound estimate for V_n is obtained by setting $2\sigma_n(y_1, y_0) = \sigma_n^2(y_1) + \sigma_n^2(y_0)$ and substituting $\hat{\sigma}_n^2(y_1)$ and $\hat{\sigma}_n^2(y_0)$ for $\sigma_n^2(y_1)$ and $\sigma_n^2(y_0)$, respectively:

$$\hat{V}_n^a = \frac{n}{n-1}\left\{\frac{\hat{\sigma}_n^2(y_1)}{m} + \frac{\hat{\sigma}_n^2(y_0)}{n-m}\right\}. \tag{4}$$

Since $\mathrm{E}_X\{\hat{\sigma}_n^2(y_1)\} = \sigma_n^2(y_1)$ and $\mathrm{E}_X\{\hat{\sigma}_n^2(y_0)\} = \sigma_n^2(y_0)$, $\hat{V}_n^a$ is conservative as its bias is nonnegative:

$$\mathrm{E}_X(\hat{V}_n^a - V_n) = (n-1)^{-1}\{\sigma_n^2(y_1) + \sigma_n^2(y_0) - 2\sigma_n(y_1, y_0)\} \ge 0. \tag{5}$$

The estimate $\hat{V}_n^a$ is also produced by common estimators that presuppose sampling from an infinite superpopulation, including heteroskedasticity-robust variance estimators [12, 16] and the standard variance estimate for mean differences between independent groups [13]. Furthermore, $\hat{V}_n^a$ is known to be unbiased for V_n when effects are constant, as would hold when there exist no treatment effects whatsoever [5]. For these reasons, the estimate $\hat{V}_n^a$ is often recommended for the analysis of experimental data [4, 7].

Neyman [13] also proposed a method for computing bounds on V_n. Given only knowledge of the second moments $\sigma_n^2(y_1)$ and $\sigma_n^2(y_0)$, the sharpest bound on $\sigma_n(y_1, y_0)$ is given by the Cauchy–Schwarz inequality: $-\{\sigma_n^2(y_1)\sigma_n^2(y_0)\}^{1/2} \le \sigma_n(y_1, y_0) \le \{\sigma_n^2(y_1)\sigma_n^2(y_0)\}^{1/2}$. By substituting $\hat{\sigma}_n^2(y_1)$ and $\hat{\sigma}_n^2(y_0)$ for $\sigma_n^2(y_1)$ and $\sigma_n^2(y_0)$, Neyman's bound estimator is

$$\hat{V}_n^{b\pm} = \frac{1}{n-1}\left[\frac{n-m}{m}\hat{\sigma}_n^2(y_1) + \frac{m}{n-m}\hat{\sigma}_n^2(y_0) \pm 2\{\hat{\sigma}_n^2(y_1)\hat{\sigma}_n^2(y_0)\}^{1/2}\right]. \tag{6}$$

The plus or minus sign is chosen depending on whether an upper or a lower bound estimate is desired. Neyman recommended choosing $\hat{V}_n^{b+}$ as a conservative approximation to the true variance, and suggested that it is "necessary" (page 471) to assume that the upper bound given by the Cauchy–Schwarz inequality holds.

3. Sharp bounds on V_N given marginal distributions of outcomes. Under the setting considered, estimates for the marginal distributions of y_1 and y_0 exist and can be used to obtain asymptotically sharp bounds on V_N given the information available. Let $G_N(y) = N^{-1}\sum_{i=1}^N I(y_{1i} \le y)$ and $F_N(y) = N^{-1}\sum_{i=1}^N I(y_{0i} \le y)$ be the marginal distribution functions of y_1 and y_0, respectively. Define their left-continuous inverses as $G_N^{-1}(u) = \inf\{y : G_N(y) \ge u\}$ and $F_N^{-1}(u) = \inf\{y : F_N(y) \ge u\}$. Define also

$$\begin{aligned} \sigma_N^H(y_1, y_0) &= \int_0^1 G_N^{-1}(u)F_N^{-1}(u)\,du - \mu_N(y_1)\mu_N(y_0), \\ \sigma_N^L(y_1, y_0) &= \int_0^1 G_N^{-1}(u)F_N^{-1}(1-u)\,du - \mu_N(y_1)\mu_N(y_0). \end{aligned} \tag{7}$$

LEMMA 1 (Hoeffding). *Given only G_N and F_N and no other information on the joint distribution of (y_1, y_0), the bound*

$$\sigma_N^L(y_1, y_0) \le \sigma_N(y_1, y_0) \le \sigma_N^H(y_1, y_0)$$

is sharp. The upper bound is attained if y_1 and y_0 are comonotonic, that is, $(y_1, y_0) \sim \{G_N^{-1}(U), F_N^{-1}(U)\}$ for a uniform random variable U on $[0, 1]$. The lower bound is attained if y_1 and y_0 are countermonotonic, that is, $(y_1, y_0) \sim \{G_N^{-1}(U), F_N^{-1}(1-U)\}$.

Lemma 1 implies that $[V_N^L, V_N^H]$ is the sharpest interval bound for V_N:

$$\begin{aligned} V_N^H &= \frac{1}{N-1}\left\{\frac{N-m}{m}\sigma_N^2(y_1) + \frac{N-(n-m)}{n-m}\sigma_N^2(y_0) + 2\sigma_N^H(y_1, y_0)\right\}, \\ V_N^L &= \frac{1}{N-1}\left\{\frac{N-m}{m}\sigma_N^2(y_1) + \frac{N-(n-m)}{n-m}\sigma_N^2(y_0) + 2\sigma_N^L(y_1, y_0)\right\}. \end{aligned}$$

In practice, we observe neither G_N nor F_N, but rather their estimates $\hat{G}_N(y) = m^{-1}\sum_{i=1}^N X_i^T I(y_{1i} \le y)$, $\hat{F}_N(y) = (n-m)^{-1}\sum_{i=1}^N X_i^C I(y_{0i} \le y)$ and left-continuous inverses

$$\hat{G}_N^{-1}(u) = \inf\{y : \hat{G}_N(y) \ge u\} = y_{1(\lceil mu \rceil)},$$
$$\hat{F}_N^{-1}(u) = \inf\{y : \hat{F}_N(y) \ge u\} = y_{0(m+\lceil (n-m)u \rceil)},$$

where $y_{1(1)} \le \cdots \le y_{1(m)}$ and $y_{0(m+1)} \le \cdots \le y_{0(n)}$ are the ordered observed outcomes, and $\lceil x \rceil$ denotes the smallest integer greater than or equal to x. Substituting $(\hat{G}_N, \hat{F}_N)$ for (G_N, F_N) in (7) yields an interval estimator $[\hat{\sigma}_N^L(y_1, y_0), \hat{\sigma}_N^H(y_1, y_0)]$ for $\sigma_N(y_1, y_0)$:

$$\hat{\sigma}_N^H(y_1, y_0) = \int_0^1 \hat{G}_N^{-1}(u)\hat{F}_N^{-1}(u)\,du - \hat{\mu}_N(y_1)\hat{\mu}_N(y_0),$$
$$\hat{\sigma}_N^L(y_1, y_0) = \int_0^1 \hat{G}_N^{-1}(u)\hat{F}_N^{-1}(1-u)\,du - \hat{\mu}_N(y_1)\hat{\mu}_N(y_0).$$

Let the $[0, 1]$-partition $\mathcal{P}_{m,n} = \{p_0, p_1, \ldots, p_P\}$ be the ordered distinct elements of $\{0, 1/m, 2/m, \ldots, 1\} \cup \{0, 1/(n-m), 2/(n-m), \ldots, 1\}$. Let $y_{1[i]} = y_{1(\lceil mp_i \rceil)}$ and $y_{0[i]} = y_{0\{m+\lceil (n-m)p_i \rceil\}}$. The inverses $\hat{G}_N^{-1}$ and $\hat{F}_N^{-1}$ are piecewise constant since $\hat{G}_N^{-1}(u) = y_{1[i]}$ and $\hat{F}_N^{-1}(u) = y_{0[i]}$ for $u \in (p_{i-1}, p_i]$. In addition, the symmetry $p_i = 1 - p_{P-i}$ implies that $p_i - p_{i-1} = p_{P+1-i} - p_{P-i}$. Thus, $[\hat{\sigma}_N^L(y_1, y_0), \hat{\sigma}_N^H(y_1, y_0)]$ reduces to

$$\hat{\sigma}_N^H(y_1, y_0) = \sum_{i=1}^P (p_i - p_{i-1})y_{1[i]}y_{0[i]} - \hat{\mu}_N(y_1)\hat{\mu}_N(y_0),$$
$$\hat{\sigma}_N^L(y_1, y_0) = \sum_{i=1}^P (p_i - p_{i-1})y_{1[i]}y_{0[P+1-i]} - \hat{\mu}_N(y_1)\hat{\mu}_N(y_0), \tag{8}$$

where $\hat{\mu}_N(y_1)$ and $\hat{\mu}_N(y_0)$ are as defined in (1).

Substituting $\hat{\sigma}_N^2(y_1)$, $\hat{\sigma}_N^2(y_0)$, and (8) for $\{\sigma_N^2(y_1), \sigma_N^2(y_0), \sigma_N(y_1, y_0)\}$ in the expressions for V_N^L and V_N^H, we obtain the interval estimator $[\hat{V}_N^L, \hat{V}_N^H]$ for V_N:

$$\hat{V}_N^H = \frac{1}{N-1}\left\{\frac{N-m}{m}\hat{\sigma}_N^2(y_1) + \frac{N-(n-m)}{n-m}\hat{\sigma}_N^2(y_0) + 2\hat{\sigma}_N^H(y_1, y_0)\right\},$$
$$\hat{V}_N^L = \frac{1}{N-1}\left\{\frac{N-m}{m}\hat{\sigma}_N^2(y_1) + \frac{N-(n-m)}{n-m}\hat{\sigma}_N^2(y_0) + 2\hat{\sigma}_N^L(y_1, y_0)\right\}. \tag{9}$$

Since Lemma 1 applies to the sample populations as well, it follows that $\hat{V}_N^H$ is never greater than $\hat{V}_N^{b+}$, and $\hat{V}_N^L$ is never smaller than $\hat{V}_N^{b-}$. R code to implement the estimators $\hat{V}_N^H$ and $\hat{V}_N^L$ is presented in Appendix B.

It is possible to demonstrate that, when outcomes are dichotomous and $n = N$, our estimator essentially reproduces the estimator proposed by Robins [14], equation (3), with a slight difference due to finite population corrections. See Copas [2], Gadbury, Iyer and Albert [6], Heckman, Smith and Clements [9] and Zhang et al. [19] for additional details on identification of the joint distribution of potential outcomes when outcomes are dichotomous.

4. Asymptotic sharpness of interval estimator. Let $\{U_N\}_N$ be a nested sequence of finite populations. The potential outcomes y_1 and y_0 of each unit are fixed, and hence the population grows deterministically. As in Isaki and Fuller [10], we do not assume that the sequences of treatment assignments are nested; instead, each U_N hosts its own random assignment. Let $H_N(\cdot,\cdot)$ be the joint distribution function of (y_1, y_0) for U_N. Under mild conditions on the scaling of U_N, the interval estimator $[\hat{V}_N^L, \hat{V}_N^H]$ converges to sharp bounds on V_N.

PROPOSITION 1. *Suppose the following conditions hold as $N \to \infty$:*

1. $(m/N, n/N) \to (\theta\rho, \theta)$ *for* $\theta \in (0, 1]$ *and* $\rho \in (0, 1)$;
2. H_N *converges weakly to a limit distribution* H *with marginals* $G(y) = H(y, \infty)$ *and* $F(y) = H(\infty, y)$;
3. $G_N(y) \to G(y)$ *at any discontinuity point of* G, *and* $F_N(y) \to F(y)$ *at any discontinuity point of* F;
4. *The sequences of distributions represented by* $\{G_N\}_N$ *and* $\{F_N\}_N$ *are uniformly square-integrable. That is, as* $\beta \to \infty$,

$$\sup_N \left\{ \frac{1}{N} \sum_{i: y_{1i}^2 \geq \beta}^{N} y_{1i}^2 \right\}, \qquad \sup_N \left\{ \frac{1}{N} \sum_{i: y_{0i}^2 \geq \beta}^{N} y_{0i}^2 \right\} \to 0.$$

Then for the collection $\mathcal{H}$ of all bivariate distributions with marginals G and F, the moments of each $h \in \mathcal{H}$ exist up to second order and

$$NV_N^H \to \frac{1-\theta\rho}{\theta\rho} \mathrm{Var}_H(y_1) + \frac{1-\theta(1-\rho)}{\theta(1-\rho)} \mathrm{Var}_H(y_0) + 2 \sup_{h \in \mathcal{H}} \mathrm{Cov}_h(y_1, y_0),$$

$$NV_N^L \to \frac{1-\theta\rho}{\theta\rho} \mathrm{Var}_H(y_1) + \frac{1-\theta(1-\rho)}{\theta(1-\rho)} \mathrm{Var}_H(y_0) + 2 \inf_{h \in \mathcal{H}} \mathrm{Cov}_h(y_1, y_0).$$

Moreover, $(\hat{V}_N^H - V_N^H, \hat{V}_N^L - V_N^L) = o_P(1/N)$.

REMARK 1. Condition 3 is used to establish the functional convergence of (G_N, F_N) to (G, F). When the units of U_N are independent and identically distributed samples from a superpopulation, the condition holds with probability one because of the strong law of large numbers. The condition is also satisfied if G and F are continuous, regardless of whether or not the units come from a superpopulation. We thank Professor A. W. van der Vaart for suggesting the latter as an alternate sufficient condition for convergence, which subsequently inspired condition 3.

REMARK 2. Given condition 2, any convergence of the marginal second moments of H_N to those of H (should they exist) necessarily implies condition 4. Thus, the condition is the weakest possible complement to conditions 1–3.

REMARK 3. If condition 4 of Proposition 1 is strengthened to require that y_1 and y_0 be bounded, then higher order rates of convergence can be obtained, namely that $\mathrm{P}(N|\hat{V}_N^H - V_N^H| > \varepsilon)$ and $\mathrm{P}(N|\hat{V}_N^L - V_N^L| > \varepsilon)$ are both of order $\mathcal{O}(1/N)$. Interested readers are referred to Proposition 2 in the Appendix.

Outline of proof. The random treatment assignment process can be expressed as a triangular array $\mathcal{X}$ where the Nth row $(\mathcal{X}_{N,1}, \dots, \mathcal{X}_{N,N}) = \{(X_1^T, X_1^C), \dots, (X_N^T, X_N^C)\}$ is the treatment/control assignment for population U_N. Since the treatment/control assignment for U_{N+1} is not related to that for U_N, each row of $\mathcal{X}$ is a random vector of a different probability space. As a result, the sequence of random distribution functions $(\hat{G}_N, \hat{F}_N)$ do not share a common probability space. However, by treating $(\hat{G}_N, \hat{F}_N)$ as random elements taking values in the product space of *càdlàg* functions $D([-\infty, \infty], \mathbb{R})^2$ endowed with the uniform metric, we show that $(\hat{G}_N, \hat{F}_N) \to (G, F)$ in probability. It then follows from the Skorohod representation that there exists a sequence of random elements $(\hat{G}'_N, \hat{F}'_N)$ defined on a common probability space that has the same law as $(\hat{G}_N, \hat{F}_N)$. Moreover, $(\hat{G}'_N, \hat{F}'_N)$ converges to (G, F) almost everywhere. Pathwise convergence of the moments of $(\hat{G}'_N, \hat{F}'_N)$ then implies probabilistic convergence of the moments of $(\hat{G}_N, \hat{F}_N)$ to the desired result. We refer the reader to the Appendix for details of the formal argument.

5. Confidence intervals for $\boldsymbol{\tau_N}$. The upper bound estimator $\hat{V}_N^H$ may be used as a basis for Wald-type confidence intervals for the average treatment effect. The proof of the following corollary follows directly from Freedman [3], Theorem 1, and associated remarks.

COROLLARY 1. *Suppose that the support of H is nonsingular and that conditions* 1–3 *of Proposition* 1 *hold. Suppose in addition that condition* 4 *is strengthened to require uniformly bounded third moments*:

$$\sup_N \left\{ \frac{1}{N} \sum_{i=1}^{N} |y_{1i}|^3 \right\}, \qquad \sup_N \left\{ \frac{1}{N} \sum_{i=1}^{N} |y_{0i}|^3 \right\} < \infty.$$

Then

$$\frac{\hat{\tau}_N - \tau_N}{(\gamma \hat{V}_N^H)^{1/2}}$$

converges weakly to the standard normal distribution where $\gamma = \lim_N (N V_N) / \lim_N (N V_N^H) \le 1$.

REMARK 4. As $\hat{V}_N^H$ is consistent for the sharp upper bound on V_N, then given large N, a confidence interval constructed as $\hat{\tau}_N \pm z_{1-\alpha/2}(\hat{V}_N^H)^{1/2}$ is asymptotically the narrowest Wald-type confidence interval assured to have at least the nominal coverage.

6. Application. We consider the randomized controlled trial reported by Harrison and Michelson [8], which assessed the intention-to-treat effects of an experimental phone call on donations to a nonprofit gay rights organization. The control phone call script contained a standard appeal. The experimental phone call script included an additional sentence that revealed the sexual orientation of the volunteer caller. The finite population U_N, which was not selected from any broader population, contains $N = n = 1561$ subjects, $m = 781$ of whom were randomly assigned to receive the experimental phone call. Outcomes were measured in terms of US dollars (USD) received per subject, ranging from \$0 to \$150. The mean donation given by subjects assigned to control was $\hat{\mu}_N(y_0) = \$1.397$, and the mean donation given by subjects assigned to treatment was $\hat{\mu}_N(y_1) = \$0.715$, yielding the difference-in-means estimate $\hat{\tau}_N = -\$0.682$.

In Table 1, we report the variance estimates and confidence intervals associated with Neyman's approximations and our proposed estimator. We find, as expected, that our estimates are sharper than Neyman's approximations. Compared to the conventional variance estimator $\hat{V}_N^a$, we find that our upper bound estimator yields a 7% reduction in the nominal variance. Importantly, if using $\hat{V}_N^a$ as a basis for conservative inference, one would need over 100 additional subjects in order to achieve the same nominal variance as that of our proposed upper bound estimate $\hat{V}_N^H$, all else equal. Similarly, if using $\hat{V}_N^{b+}$, one would need over 75 additional subjects to achieve the nominal variance of $\hat{V}_N^H$.

6.1. *Simulations.* We use the data from [8] to assess the operating characteristics of the upper bound estimators and associated Wald-type confidence intervals.

TABLE 1
Variance estimates and confidence intervals for Harrison and Michelson [8]

	Variance estimate (USD^2)	**95% confidence interval for τ_N**
Conventional ($\hat{V}_N^a$)	0.199	(−\$1.555, \$0.192)
Neyman upper bound ($\hat{V}_N^{b+}$)	0.196	(−\$1.548, \$0.185)
Neyman lower bound ($\hat{V}_N^{b-}$)	0.003	N/A
Sharp upper bound ($\widehat{V}_N^H$)	0.186	(−\$1.528, \$0.165)
Sharp lower bound ($\widehat{V}_N^L$)	0.098	N/A

These characteristics depend on the underlying joint distribution of potential outcomes, which cannot be directly observed and are instead hypothesized as part of these simulations. We thus impute the missing potential outcomes (potential control outcomes for treatment subjects, and potential treatment outcomes for control subjects) by asserting varying hypotheses about treatment effects. We simulate 25 million random assignments and, for each of these random assignments, compute the upper bound variance estimates $\hat{V}_N^a$, $\hat{V}_N^{b+}$ and $\hat{V}_N^H$, and associated confidence intervals that would have been obtained. For the collection of 25 million simulations, we calculate the mean variance estimate, the mean width of the associated 95% confidence intervals for τ_N and the fraction of simulated confidence intervals covering τ_N.

The first hypothesis that we evaluate is the sharp null hypothesis of no effect whatsoever. This hypothesis, denoted "Sharp Null," assumes that $y_{0i} = y_{1i}$ for all i. Under the Sharp Null, the treatment effect estimator variance is 0.199 USD^2. As can be seen in Table 2, Neyman's estimators predictably perform well since they implicitly assume that the outcomes are perfectly correlated: the bias (5) for $\hat{V}_N^a$ is zero because $\sigma_N^2(y_1) = \sigma_N^2(y_0) = \sigma_N(y_1, y_0)$. Due to the nonlinearity of the square root function, the Cauchy–Schwarz inequality implies that $\hat{V}_N^{b+}$ has nonpositive bias (-0.007 USD^2). The 95% confidence intervals associated with $\hat{V}_N^a$ and $\hat{V}_N^{b+}$ have coverage of 95.2% and 94.1%, respectively (the former is not exactly 95% because the sampling distribution of τ_N is not perfectly normal). Because $\hat{V}_N^H \leq \hat{V}_N^{b+}$, $\hat{V}_N^H$ is slightly more negatively biased (-0.010 USD^2) and has lower coverage (93.7%) than $\hat{V}_N^{b+}$.

We next consider two hypotheses that embed treatment effect heterogeneity, denoted "Heterogeneity A" and "Heterogeneity B." Under Heterogeneity A, we

TABLE 2
Simulated variance estimator properties under varying treatment effect hypotheses for Harrison and Michelson [8], *using* 25 *million simulated random assignments each*

Effect hypothesis	**Variance estimator**	**Mean var. estimate**	**Mean 95% CI width**	**Coverage for τ_N**
Sharp Null	Conventional ($\hat{V}_N^a$)	0.199	1.747	95.2%
(True Var.: 0.199)	Neyman upper bound ($\hat{V}_N^{b+}$)	0.193	1.724	94.1%
	Sharp upper bound ($\hat{V}_N^H$)	0.189	1.703	93.7%
Heterogeneity A	Conventional ($\hat{V}_N^a$)	0.279	2.067	96.7%
(True Var.: 0.238)	Neyman upper bound ($\hat{V}_N^{b+}$)	0.268	2.028	95.9%
	Sharp upper bound ($\hat{V}_N^H$)	0.258	1.987	95.4%
Heterogeneity B	Conventional ($\hat{V}_N^a$)	0.244	1.933	97.4%
(True Var.: 0.186)	Neyman upper bound ($\hat{V}_N^{b+}$)	0.226	1.860	96.5%
	Sharp upper bound ($\hat{V}_N^H$)	0.214	1.809	96.0%

assume that the sharp null hypothesis holds, with the exception of 10 subjects who had an observed $y_{0i} = 0$ USD under control. For these 10 subjects, we assume that $y_{1i} = 100$ USD. Under Heterogeneity A, the treatment effect estimator variance is 0.238 USD^2 and, as expected, all variance estimators are conservative (positively biased). However, the bias, confidence interval widths, and coverage for τ_N are all improved when $\hat{V}_N^H$ is used in place of either of Neyman's estimators. In formulating the Heterogeneity B hypothesis, we assume that Heterogeneity A holds, but, in addition, for all 6 subjects under treatment with an observed $y_{1i} \geq 50$ USD, we assume that $y_{0i} = 0$ USD. Under Heterogeneity B, the treatment effect estimator variance is 0.186 USD^2 and, again, while all estimators are conservative, $\hat{V}_N^H$ improves over Neyman's estimators.

In Appendix C, we further explore the relative performance of the upper bound estimates under varying assumptions about the distribution of potential outcomes. Using the Beta distribution family as an example to represent varying shapes of marginal treatment and control distributions, we show that it is possible for $\hat{V}_N^H$ to materially outperform $\hat{V}_N^a$ and $\hat{V}_N^{b+}$ as the two marginals diverge in shape. Our simulations therefore illustrate how $\hat{V}_N^H$ can improve upon Neyman's bounds under effect heterogeneity.

7. Discussion. The proposed variance estimator may also be extended to alternative designs. For block-randomized designs where the number of units per block grows asymptotically large, Proposition 1 holds within each block, and thus calculation of the overall variance is straightforward. In cluster-randomized designs with equally-sized clusters, the proposed estimator may be used with the unit of analysis being the cluster and the outcome being the cluster mean. It is also straightforward to adapt the estimator to completely randomized experiments with multiple treatments, which may be shown to be logically equivalent to sampling from a broader population. In addition, we note that our result can be generalized to characterize estimation error for arbitrary target populations within the sampling frame (e.g., unsampled units).

Finally, we remark on the scope of our findings, as our results presuppose deterministic potential outcomes. When the potential outcomes are stochastic, the total variance is greater than the conditional variance (2) because of the additional stochasticity. If one sought to estimate the total variance or bounds thereof, additional structure would need to be imposed on the stochastic process (e.g., independence across units and finite variances); otherwise it is possible for the identification set to be unbounded.

APPENDIX A: PROOFS

PROOF OF LEMMA 1. Let $H_N(y_1, y_0)$ be the joint distribution function of (y_1, y_0), and define two other distributions $H_N^H(y_1, y_0) = \min\{G_N(y_1), F_N(y_0)\}$

and $H_N^L(y_1, y_0) = \max\{0, G_N(y_1) + F_N(y_0) - 1\}$. All three distributions have the same marginals G_N and F_N. Defining E_Q as the expectation operator with respect to a distribution Q, a result by Hoeffding, recounted in Tchen [17], shows that

$$\mathrm{E}_{H_N^L}(y_1 y_0) \le \mathrm{E}_{H_N}(y_1 y_0) \le \mathrm{E}_{H_N^H}(y_1 y_0).$$

Since $\{G_N^{-1}(U), F_N^{-1}(U)\} \sim H_N^H$ and $\{G_N^{-1}(U), F_N^{-1}(1-U)\} \sim H_N^L$, the lower and upper bounds are equivalent to

$$\mathrm{E}_{H_N^H}(y_1 y_0) = \int_0^1 G_N^{-1}(u) F_N^{-1}(u)\, du,$$

$$\mathrm{E}_{H_N^L}(y_1 y_0) = \int_0^1 G_N^{-1}(u) F_N^{-1}(1-u)\, du.$$

The integrals exist because $|G_N^{-1}(u)|, |F_N^{-1}(u)| \le \max_{i=1}^N \max(|y_{1i}|, |y_{0i}|) < \infty$. □

Lemma 2 below will be required in the proofs of Propositions 1 and 2. In the special case where the units of U_N are independent and identically distributed samples from a superpopulation, the first part of the lemma reduces to the classical Glivenko–Cantelli theorem, and the convergence implied by the second part follows from the conditional bootstrap convergence results in van der Vaart and Wellner [18], Example 3.6.14. We thank an anonymous reviewer for suggesting a more elegant way for bounding (11) and (12) than our original approach.

LEMMA 2. *Suppose conditions* 1–3 *of Proposition* 1 *hold. Then*

$$\sup_y \bigl|G(y) - G_N(y)\bigr| \to 0 \quad \textit{and} \quad \sup_y \bigl|F(y) - F_N(y)\bigr| \to 0.$$

In addition, given $\eta_1, \eta_0 > 0$, *there exist two positive integers* $K_1(\eta_1)$ *and* $K_0(\eta_0)$ *such that*

$$\limsup_N \Bigl\{N\mathrm{P}\Bigl(\sup_y \bigl|G(y) - \hat{G}_N(y)\bigr| \ge \eta_1\Bigr)\Bigr\} \le \frac{(1-\theta\rho)K_1(\eta_1)}{\theta\rho\eta_1^2},$$

$$\limsup_N \Bigl\{N\mathrm{P}\Bigl(\sup_y \bigl|F(y) - \hat{F}_N(y)\bigr| \ge \eta_0\Bigr)\Bigr\} \le \frac{\{1-\theta(1-\rho)\}K_0(\eta_0)}{\theta(1-\rho)\eta_0^2}.$$

The integers are nonincreasing in η, *and depend also on the limiting distribution* H *of* (y_1, y_0).

PROOF. For the first part of the lemma, we follow the argument used in the Glivenko–Cantelli theorem. Given $\eta_1 > 0$, there exists a partition $-\infty = s_0 < s_1 < \cdots < s_{K_1(\eta_1)} = \infty$ such that $G(s_i-) < G(s_{i-1}) + \eta_1/2$. For any $1 \le i \le K_1(\eta_1)$ and $s_{i-1} \le s < s_i$,

$$G(s_i-) - G_N(s_i-) - \eta_1/2 < G(s) - G_N(s) < G(s_{i-1}) - G_N(s_{i-1}) + \eta_1/2,$$

hence $\sup_y |G(y) - G_N(y)| < \eta_1$ if $|G(s_{i-1}) - G_N(s_{i-1})| < \eta_1/2$ and $|G(s_i-) - G_N(s_i-)| < \eta_1/2$ for all i. By conditions 2 and 3, this is satisfied for all N sufficiently large. The uniform convergence of F_N follows in the same way.

To establish the second part of the lemma, note that $\sup_y |G(y) - \hat{G}_N(y)| < \eta_1$ on the set

$$\bigcap_{i=1}^{K_1(\eta_1)} \{|G(s_{i-1}) - \hat{G}_N(s_{i-1})|, |G(s_i-) - \hat{G}_N(s_i-)| < \eta_1/2\}.$$

Since $\mathrm{P}\{(\bigcap_i A_i)^c\} = \mathrm{P}(\bigcup_i A_i^c) \leq \sum_i \mathrm{P}(A_i^c)$, we have

$$\begin{aligned}
&\mathrm{P}\Big\{\sup_y |G(y) - \hat{G}_N(y)| \geq \eta_1\Big\} \\
&\quad \leq \sum_{i=1}^{K_1(\eta_1)} \mathrm{P}\{|G(s_{i-1}) - \hat{G}_N(s_{i-1})| \geq \eta_1/2\} \\
&\qquad + \sum_{i=1}^{K_1(\eta_1)} \mathrm{P}\{|G(s_i-) - \hat{G}_N(s_i-)| \geq \eta_1/2\} \\
&\quad \leq \sum_{i=1}^{K_1(\eta_1)} \mathrm{P}\{|\hat{G}_N(s_{i-1}) - G_N(s_{i-1})| \geq \eta_1/2 - o(1)\} \qquad (10)\\
&\qquad + \sum_{i=1}^{K_1(\eta_1)} \mathrm{P}\{|\hat{G}_N(s_i-) - G_N(s_i-)| \geq \eta_1/2 - o(1)\} \\
&\quad \leq \sum_{i=1}^{K_1(\eta_1)} \frac{\mathrm{Var}_X\{\hat{G}_N(s_{i-1})\} + \mathrm{Var}_X\{\hat{G}_N(s_i-)\}}{\{\eta_1/2 - o(1)\}^2},
\end{aligned}$$

where the second inequality follows from $|G(y) - G_N(y)| = o(1)$, and the last inequality from Chebyshev's inequality and the fact that $\mathrm{E}_X \hat{G}_N(y) = G_N(y)$.

The argument used to derive (2) can also be used to bound the variances in (10). Noting that $\sigma_N^2(I\{y_1 \leq y\}) = G_N(y)\{1 - G_N(y)\} \leq 1/4$ and similarly $\sigma_N^2(I\{y_0 \leq y\}) \leq 1/4$, we have for all y,

$$\mathrm{Var}_X\, \hat{G}_N(y) = \frac{N-m}{(N-1)m}\sigma_N^2(I\{y_1 \leq y\}) = \frac{N-m}{4(N-1)m}, \tag{11}$$

$$\mathrm{Var}_X\, \hat{F}_N(y) = \frac{N-(n-m)}{(N-1)(n-m)}\sigma_N^2(I\{y_0 \leq y\}) \leq \frac{N-(n-m)}{4(N-1)(n-m)}. \tag{12}$$

Plugging (11) into (10) and taking limits yields the desired result for $\hat{G}_N$, after absorbing the factor of 2 into $K_1(\eta_1)$. The result for $\hat{F}_N$ can be obtained in the same manner. □

PROOF OF PROPOSITION 1. As indicated in the proof outline, we proceed in several stages.

(i) *Functional convergence of random distribution functions.* Let $D([-\infty,\infty],\mathbb{R})^2$ be the Cartesian product of the space of *càdlàg* functions with itself, endowed with the uniform metric induced by the norm $\|(v,u)\| = \max\{\sup_y |v(y)|, \sup_y |u(y)|\}$. Thus, $D([-\infty,\infty],\mathbb{R})^2$ is a nonseparable metric space. Lemma 2 shows that the distribution functions $(\hat{G}_N, \hat{F}_N)$ converge in probability to (G, F) in $D([-\infty,\infty],\mathbb{R})^2$. That is, $\mathrm{P}(\|(\hat{G}_N - G, \hat{F}_N - F)\| \geq \varepsilon) \to 0$ for every $\varepsilon > 0$. As is the case with the lemma, the statement does not require the use of outer measures because for each N, $(\hat{G}_N, \hat{F}_N)$ can take on at most $\binom{N}{n}\binom{n}{m}$ distinct values in $D([-\infty,\infty],\mathbb{R})^2$; therefore, $\|(\hat{G}_N - G, \hat{F}_N - F)\|$ is finite discrete valued.

(ii) *Existence of random distributions $(\hat{G}'_N, \hat{F}'_N)$ defined on a common probability space.* Since the limit (G, F) is deterministic, the support of the limiting probability measure on $D([-\infty,\infty],\mathbb{R})^2$ is a singleton. Applying the Skorohod representation [18], Theorem 1.10.3, to $(\hat{G}_N, \hat{F}_N)$ yields new random elements $(\hat{G}'_N, \hat{F}'_N)$ on $D([-\infty,\infty],\mathbb{R})^2$ that have the same law as $(\hat{G}_N, \hat{F}_N)$. Furthermore, $(\hat{G}'_N, \hat{F}'_N)$ converges to (G, F) almost everywhere, in the sense that along each sample path ω' (in a set of measure one), the distribution functions converge uniformly:

$$\sup_y \left|\hat{G}'_N(y;\omega') - G(y)\right| \to 0 \quad \text{and} \quad \sup_y \left|\hat{F}'_N(y;\omega') - F(y)\right| \to 0.$$

(iii) *Convergence of* $\mathrm{E}_{\hat{G}_N}(y_1^p)$ *and* $\mathrm{E}_{\hat{F}_N}(y_0^p)$ *for p = 1, 2.* Define E_Q as the expectation operator with respect to a distribution Q. Under condition 1, there exists N_0 such that $1/m \leq 2/(\theta\rho N)$ and $1/(n-m) \leq 2/\{\theta(1-\rho)N\}$ for $N \geq N_0$. Then for each $N \geq N_0$ and every realization of $(\hat{G}_N, \hat{F}_N)$, condition 4 implies that as $\beta \to \infty$,

$$\mathrm{E}_{\hat{G}_N}\left(y_1^2 I\{y_1^2 \geq \beta\}\right) = \sum_{i:y_{1i}^2 \geq \beta}^{N} \frac{X_i^T y_{1i}^2}{m} \leq \frac{2}{\theta\rho} \sup_{N \geq N_0} \left\{\sum_{i:y_{1i}^2 \geq \beta}^{N} \frac{y_{1i}^2}{N}\right\} \to 0,$$

$$\mathrm{E}_{\hat{F}_N}\left(y_0^2 I\{y_0^2 \geq \beta\}\right) = \sum_{i:y_{0i}^2 \geq \beta}^{N} \frac{X_i^C y_{0i}^2}{n-m} \leq \frac{2}{\theta(1-\rho)} \sup_{N \geq N_0} \left\{\sum_{i:y_{0i}^2 \geq \beta}^{N} \frac{y_{0i}^2}{N}\right\} \to 0.$$

Recall that both $(\hat{G}'_N, \hat{F}'_N)$ and $(\hat{G}_N, \hat{F}_N)$ share the same finite discrete distribution. Thus, for almost all sample paths ω' in the probability space of $(\hat{G}'_N, \hat{F}'_N)$, the sequences of distributions represented by $\{\hat{G}'_N(\cdot;\omega')\}_N$ and $\{\hat{F}'_N(\cdot;\omega')\}_N$ are uniformly square-integrable. Moreover, since $\hat{G}'_N(\cdot;\omega') \to G(\cdot)$, the random moments $\{\mathrm{E}_{\hat{G}'_N}(y_1), \mathrm{E}_{\hat{G}'_N}(y_1^2)\}$ converge to $\{\mathrm{E}_H(y_1), \mathrm{E}_H(y_1^2)\}$ almost everywhere, with the limits being finite. Similarly, $\{\mathrm{E}_{\hat{F}'_N}(y_0), \mathrm{E}_{\hat{F}'_N}(y_0^2)\} \to \{\mathrm{E}_H(y_0), \mathrm{E}_H(y_0^2)\}$

almost everywhere as well. Translating this back into convergence in probability for the first two random moments of $\hat{G}_N$ and $\hat{F}_N$, we have

$$\hat{\sigma}_N^2(y_1) = \frac{N-1}{N}\frac{m}{m-1}\bigl[\mathrm{E}_{\hat{G}_N}(y_1^2) - \{\mathrm{E}_{\hat{G}_N}(y_1)\}^2\bigr] \to \mathrm{Var}_H(y_1), \tag{13}$$

$$\hat{\sigma}_N^2(y_0) = \frac{N-1}{N}\frac{n-m}{n-m-1}\bigl[\mathrm{E}_{\hat{F}_N}(y_0^2) - \{\mathrm{E}_{\hat{F}_N}(y_0)\}^2\bigr] \to \mathrm{Var}_H(y_0) \tag{14}$$

in probability.

(iv) *Convergence of* $\hat{\sigma}_N^H(y_1, y_0)$ *and* $\hat{\sigma}_N^L(y_1, y_0)$. Define the distributions $H^H(y_1, y_0) = \min\{G(y_1), F(y_0)\}$ and $H^L(y_1, y_0) = \max\{0, G(y_1) + F(y_0) - 1\}$, both of which have marginals G and F. Using Hoeffding's result from the proof of Lemma 1, we have that

$$\mathrm{E}_{H^H}(y_1 y_0) = \sup_{h\in\mathcal{H}} \mathrm{E}_h(y_1 y_0),$$
$$\mathrm{E}_{H^L}(y_1 y_0) = \inf_{h\in\mathcal{H}} \mathrm{E}_h(y_1 y_0).$$

Now fix a sample path and define two sequences of distributions $\hat{H}_N^{H'}(y_1, y_0; \omega') = \min\{\hat{G}'_N(y_1; \omega'), \hat{F}'_N(y_0; \omega')\}$ and $\hat{H}_N^{L'}(y_1, y_0; \omega') = \max\{0, \hat{G}'_N(y_1; \omega') + \hat{F}'_N(y_0; \omega') - 1\}$. It is clear that $\hat{H}_N^{H'}(\cdot,\cdot;\omega')$ converges to $H^H(\cdot,\cdot)$ and $\hat{H}_N^{L'}(\cdot,\cdot;\omega')$ converges to $H^L(\cdot,\cdot)$ pointwise. Given that the product $y_1 y_0$ is also uniformly integrable with respect to almost all sequences $\{\hat{H}_N^{H'}(\cdot,\cdot;\omega')\}_N$ and $\{\hat{H}_N^{L'}(\cdot,\cdot;\omega')\}_N$ because $\{|XY| \geq \beta^2\} \subset \{|X| \geq \beta\} \cup \{|Y| \geq \beta\}$, it follows that $\mathrm{E}_{\hat{H}_N^{H'}}(y_1 y_0) \to \sup_{h\in\mathcal{H}} \mathrm{E}_h(y_1 y_0)$ and $\mathrm{E}_{\hat{H}_N^{L'}}(y_1 y_0) \to \inf_{h\in\mathcal{H}} \mathrm{E}_h(y_1 y_0)$ almost everywhere. Thus,

$$\hat{\sigma}_N^H(y_1, y_0) = \mathrm{E}_{\hat{H}_N^H}(y_1 y_0) - \mathrm{E}_{\hat{G}_N}(y_1)\mathrm{E}_{\hat{F}_N}(y_0) \to \sup_{h\in\mathcal{H}} \mathrm{Cov}_h(y_1, y_0), \tag{15}$$

$$\hat{\sigma}_N^L(y_1, y_0) = \mathrm{E}_{\hat{H}_N^L}(y_1 y_0) - \mathrm{E}_{\hat{G}_N}(y_1)\mathrm{E}_{\hat{F}_N}(y_0) \to \inf_{h\in\mathcal{H}} \mathrm{Cov}_h(y_1, y_0) \tag{16}$$

in probability. Plugging (13)–(16) into (9) then yields the proposition. □

PROPOSITION 2. *Suppose conditions* 1–3 *of Proposition* 1 *hold, and that* y_1 *and* y_0 *are bounded*: $|y_{1i}|, |y_{0i}| \leq C < \infty$ *for all* i. *Given* $\varepsilon > 0$, *for any* $\varepsilon_1, \varepsilon_2, \varepsilon_3 > 0$ *such that* $\sum_i \varepsilon_i = \varepsilon$,

$$\mathrm{P}(N|\hat{V}_N^H - V_N^H| \geq \varepsilon) \leq \mathcal{O}\biggl(\frac{C^4}{N}\kappa_1(\varepsilon_1)\biggr),$$

$$\mathrm{P}(N|\hat{V}_N^L - V_N^L| \geq \varepsilon) \leq \mathcal{O}\biggl(\frac{C^4}{N}\{1/\varepsilon_1^2 + \kappa_2(\varepsilon_2) + \kappa_3(\varepsilon_3)\}\biggr),$$

where $\kappa_1(\varepsilon_1)$, $\kappa_2(\varepsilon_2)$ *and* $\kappa_3(\varepsilon_3)$ *depend on the limiting distribution* H.

PROOF. Define the bivariate distribution functions $H_N^H(y_1, y_0) = \min\{G_N(y_1), F_N(y_0)\}$, $H_N^L(y_1, y_0) = \max(0, G_N(y_1) + F_N(y_0) - 1)$, $\hat{H}_N^H(y_1,$

$y_0) = \min\{\hat{G}_N(y_1), \hat{F}_N(y_0)\}$, and $\hat{H}_N^L(y_1, y_0) = \max(0, \hat{G}_N(y_1) + \hat{F}_N(y_0) - 1)$. Let E_Q be the expectation operator with respect to a distribution Q. Using another result by Hoeffding as recounted in Lehmann [11], Lemma 2, the following covariances can be expressed as

$$\begin{aligned}
\hat{\sigma}_N^H(y_1, y_0) &= \int_{-C}^{C}\int_{-C}^{C} \hat{H}_N^H(y_1, y_0)\,dy_1\,dy_0 \\
&\quad - \int_{-C}^{C} \hat{G}_N(y_1)\,dy_1 \int_{-C}^{C} \hat{F}_N(y_0)\,dy_0 \\
&= \int_{-C}^{C}\int_{-C}^{C} \hat{H}_N^H(y_1, y_0)\,dy_1\,dy_0 - C^2 + C\mathrm{E}_{\hat{G}_N}(y_1) \\
&\quad + C\mathrm{E}_{\hat{F}_N}(y_0) - \mathrm{E}_{\hat{G}_N}(y_1)\mathrm{E}_{\hat{F}_N}(y_0), \\
\sigma_N^H(y_1, y_0) &= \int_{-C}^{C}\int_{-C}^{C} H_N^H(y_1, y_0)\,dy_1\,dy_0 \\
&\quad - \int_{-C}^{C} G_N(y_1)\,dy_1 \int_{-C}^{C} F_N(y_0)\,dy_0 \\
&= \int_{-C}^{C}\int_{-C}^{C} H_N^H(y_1, y_0)\,dy_1\,dy_0 - C^2 + C\mathrm{E}_{G_N}(y_1) \\
&\quad + C\mathrm{E}_{F_N}(y_0) - \mathrm{E}_{G_N}(y_1)\mathrm{E}_{F_N}(y_0),
\end{aligned}$$

where the second equality follows from the identity $\mathrm{E}(W) = C - \int_{-C}^{C} \mathrm{P}(W \leq w)\,dw$ for any random variable W bounded by C. Then

$$\begin{aligned}
N(\hat{V}_N^H - V_N^H) &= \frac{N-m}{m-1}\left\{\mathrm{E}_{\hat{G}_N}(y_1^2) - \frac{N(m-1)}{m(N-1)}\mathrm{E}_{G_N}(y_1^2)\right\} \\
&\quad - \frac{N-m}{m-1}\left[\{\mathrm{E}_{\hat{G}_N}(y_1)\}^2 - \frac{N(m-1)}{m(N-1)}\{\mathrm{E}_{G_N}(y_1)\}^2\right] \\
&\quad + \frac{N-(n-m)}{n-m-1}\left\{\mathrm{E}_{\hat{F}_N}(y_0^2) - \frac{N(n-m-1)}{(N-1)(n-m)}\mathrm{E}_{F_N}(y_0^2)\right\} \\
&\quad - \frac{N-(n-m)}{n-m-1}\left[\{\mathrm{E}_{\hat{F}_N}(y_0)\}^2 - \frac{N(n-m-1)}{(N-1)(n-m)}\{\mathrm{E}_{F_N}(y_0)\}^2\right] \\
&\quad + \frac{2N}{N-1}C\{\mathrm{E}_{\hat{G}_N}(y_1) - \mathrm{E}_{G_N}(y_1) + \mathrm{E}_{\hat{F}_N}(y_0) - \mathrm{E}_{F_N}(y_0)\} \\
&\quad - \frac{2N}{N-1}\{\mathrm{E}_{\hat{G}_N}(y_1)\mathrm{E}_{\hat{F}_N}(y_0) - \mathrm{E}_{G_N}(y_1)\mathrm{E}_{F_N}(y_0)\} \\
&\quad + \frac{2N}{N-1}\int_{[-C,C]^2}\{\hat{H}_N^H(y_1, y_0) - H_N^H(y_1, y_0)\}\,dy_1\,dy_0.
\end{aligned}$$

To obtain the desired result, we proceed by bounding the probability that each of the seven terms are large. Let $\nu_1, \ldots, \nu_8 > 0$ be a tuple whose sum is ε_1. For the first term,

$$\begin{aligned}
&\mathrm{P}\left\{\frac{N-m}{m-1}\left|\mathrm{E}_{\hat{G}_N}(y_1^2) - \frac{N(m-1)}{m(N-1)}\mathrm{E}_{G_N}(y_1^2)\right| \geq \nu_1\right\} \\
&\quad \leq \mathrm{P}\left\{|\mathrm{E}_{\hat{G}_N}(y_1^2) - \mathrm{E}_{G_N}(y_1^2)| \geq \frac{(m-1)\nu_1}{N-m} - \frac{(N-m)C^2}{m(N-1)}\right\} \\
&\quad \leq \mathrm{Var}_X\{\mathrm{E}_{\hat{G}_N}(y_1^2)\} \Big/ \left\{\frac{(m-1)\nu_1}{N-m} - o(1)\right\}^2 \\
&\quad \leq \frac{(N-m)C^4}{(N-1)m} \Big/ \left\{\frac{(m-1)\nu_1}{N-m} - o(1)\right\}^2,
\end{aligned}$$

where the first inequality follows from $|\mathrm{E}_{\hat{G}_N}(y_1^2) - \beta_N \mathrm{E}_{G_N}(y_1^2)| \leq |\mathrm{E}_{\hat{G}_N}(y_1^2) - \mathrm{E}_{G_N}(y_1^2)| + |(1-\beta_N)|\mathrm{E}_{G_N}(y_1^2)$, and the second inequality from Chebyshev's inequality and the fact that $\mathrm{E}_X \mathrm{E}_{\hat{G}_N}(y_1^p) = \mathrm{E}_{G_N}(y^p)$. The bound on the variance is obtained in the same way as (11). Thus,

$$\begin{aligned}
&\limsup_N N\mathrm{P}\left\{\frac{N-m}{m-1}\left|\mathrm{E}_{\hat{G}_N}(y_1^2) - \frac{N(m-1)}{m(N-1)}\mathrm{E}_{G_N}(y_1^2)\right| \geq \nu_1\right\} \\
&\quad \leq \frac{(1-\theta\rho)^3 C^4}{\theta^3\rho^3\nu_1^2}.
\end{aligned} \tag{17}$$

For the second term,

$$\begin{aligned}
&\mathrm{P}\left\{\frac{N-m}{m-1}\left|\{\mathrm{E}_{\hat{G}_N}(y_1)\}^2 - \frac{N(m-1)}{m(N-1)}\{\mathrm{E}_{G_N}(y_1)\}^2\right| \geq \nu_2\right\} \\
&\quad \leq \mathrm{P}\left\{\left|\mathrm{E}_{\hat{G}_N}(y_1) - \left\{\frac{N(m-1)}{m(N-1)}\right\}^{1/2}\mathrm{E}_{G_N}(y_1)\right| \right. \\
&\qquad\qquad \left. \geq \frac{(m-1)\nu_2/\{C(N-m)\}}{1+\{N(m-1)/(m(N-1))\}^{1/2}}\right\} \\
&\quad \leq \mathrm{P}\left\{|\mathrm{E}_{\hat{G}_N}(y_1) - \mathrm{E}_{G_N}(y_1)| \geq \frac{(m-1)\nu_2}{\{2+o(1)\}(N-m)C} - o(1)\right\} \\
&\quad \leq \mathrm{Var}_X\{\mathrm{E}_{\hat{G}_N}(y_1)\} \Big/ \left[\frac{(m-1)\nu_2}{\{2+o(1)\}(N-m)C} - o(1)\right]^2 \\
&\quad \leq \frac{(N-m)C^2}{(N-1)m} \Big/ \left[\frac{(m-1)\nu_2}{\{2+o(1)\}(N-m)C} - o(1)\right]^2,
\end{aligned}$$

where the first inequality follows from the identity $u^2 - v^2 = (u+v)(u-v)$. Hence,

$$\limsup_N N\mathrm{P}\left\{\frac{N-m}{m-1}\left|\{\mathrm{E}_{\hat{G}_N}(y_1)\}^2 - \frac{N(m-1)}{m(N-1)}\{\mathrm{E}_{G_N}(y_1)\}^2\right| \geq \nu_2\right\} \leq \frac{4(1-\theta\rho)^3C^4}{\theta^3\rho^3\nu_2^2}. \tag{18}$$

The same arguments apply to the third and fourth terms:

$$\limsup_N N\mathrm{P}\left\{\frac{N-(n-m)}{n-m-1}\left|\mathrm{E}_{\hat{F}_N}(y_0^2) - \frac{N(n-m-1)}{(N-1)(n-m)}\mathrm{E}_{F_N}(y_0^2)\right| \geq \nu_3\right\} \leq \frac{\{1-\theta(1-\rho)\}^3C^4}{\theta^3(1-\rho)^3\nu_3^2}, \tag{19}$$

$$\limsup_N N\mathrm{P}\left\{\frac{N-(n-m)}{n-m-1}\left|\{\mathrm{E}_{\hat{F}_N}(y_0)\}^2 - \frac{N(n-m-1)}{(N-1)(n-m)}\{\mathrm{E}_{F_N}(y_0)\}^2\right| \geq \nu_4\right\} \leq \frac{4\{1-\theta(1-\rho)\}^3C^4}{\theta^3(1-\rho)^3\nu_4^2}. \tag{20}$$

For the fifth term,

$$\begin{aligned}
&\mathrm{P}\left[\frac{2NC}{N-1}\left|\mathrm{E}_{\hat{G}_N}(y_1) - \mathrm{E}_{G_N}(y_1) + \mathrm{E}_{\hat{F}_N}(y_0) - \mathrm{E}_{F_N}(y_0)\right| < \nu_5 + \nu_6\right] \\
&\quad \geq \mathrm{P}\left\{\left|\mathrm{E}_{\hat{G}_N}(y_1) - \mathrm{E}_{G_N}(y_1)\right| < \frac{(N-1)\nu_5}{2NC}, \right. \\
&\qquad\qquad \left. \left|\mathrm{E}_{\hat{F}_N}(y_0) - \mathrm{E}_{F_N}(y_0)\right| < \frac{(N-1)\nu_6}{2NC}\right\} \\
&\quad \geq 1 - \mathrm{P}\left\{\left|\mathrm{E}_{\hat{G}_N}(y_1) - \mathrm{E}_{G_N}(y_1)\right| \geq \frac{(N-1)\nu_5}{2NC}\right\} \\
&\qquad - \mathrm{P}\left\{\left|\mathrm{E}_{\hat{F}_N}(y_0) - \mathrm{E}_{F_N}(y_0)\right| \geq \frac{(N-1)\nu_6}{2NC}\right\} \\
&\quad \geq 1 - \mathrm{Var}_X\{\mathrm{E}_{\hat{G}_N}(y_1)\}\Big/\left\{\frac{(N-1)\nu_5}{2NC}\right\}^2 - \mathrm{Var}_X\{\mathrm{E}_{\hat{F}_N}(y_0)\}\Big/\left\{\frac{(N-1)\nu_6}{2NC}\right\}^2,
\end{aligned}$$

so we have

$$\limsup_N N\mathrm{P}\left[\frac{2NC}{N-1}\left|\mathrm{E}_{\hat{G}_N}(y_1) - \mathrm{E}_{G_N}(y_1) + \mathrm{E}_{\hat{F}_N}(y_0) - \mathrm{E}_{F_N}(y_0)\right| \geq \nu_5 + \nu_6\right] \leq \frac{4(1-\theta\rho)C^4}{\theta\rho\nu_5^2} + \frac{4\{1-\theta(1-\rho)\}C^4}{\theta(1-\rho)\nu_6^2}. \tag{21}$$

For the sixth term, we use the fact that $|uv - u'v'| \leq |uv - u'v| + |u'v - u'v'|$ to obtain

$$\begin{aligned}
&\mathrm{P}\biggl\{\frac{2N}{N-1}\bigl|\mathrm{E}_{\hat{G}_N}(y_1)\mathrm{E}_{\hat{F}_N}(y_0) - \mathrm{E}_{G_N}(y_1)\mathrm{E}_{F_N}(y_0)\bigr| < \nu_7 + \nu_8\biggr\} \\
&\quad \geq 1 - \mathrm{P}\biggl\{\bigl|\mathrm{E}_{\hat{G}_N}(y_1) - \mathrm{E}_{G_N}(y_1)\bigr| \geq \frac{(N-1)\nu_7}{2NC}\biggr\} \\
&\quad\quad - \mathrm{P}\biggl\{\bigl|\mathrm{E}_{\hat{F}_N}(y_0) - \mathrm{E}_{F_N}(y_0)\bigr| \geq \frac{(N-1)\nu_8}{2NC}\biggr\}.
\end{aligned}$$

Following the rest of the derivation of (21) gives

$$\begin{aligned}
&\limsup_N N\mathrm{P}\biggl\{\frac{2N}{N-1}\bigl|\mathrm{E}_{\hat{G}_N}(y_1)\mathrm{E}_{\hat{F}_N}(y_0) - \mathrm{E}_{G_N}(y_1)\mathrm{E}_{F_N}(y_0)\bigr| \geq \nu_7 + \nu_8\biggr\} \\
&\quad \leq \frac{4(1-\theta\rho)C^4}{\theta\rho\nu_7^2} + \frac{4\{1-\theta(1-\rho)\}C^4}{\theta(1-\rho)\nu_8^2}.
\end{aligned} \tag{22}$$

To bound the probability that the last term exceeds $1-\varepsilon_1$, first note that $|u - u'| < \eta$ and $|v - v'| < \eta$ implies $|\min(u, v) - \min(u', v')| < \eta$. This gives the third inequality below:

$$\begin{aligned}
&\mathrm{P}\biggl[\frac{2N}{N-1}\biggl|\int_{[-C,C]^2}\{\hat{H}_N^H(y_1, y_0) - H_N^H(y_1, y_0)\}\,dy_1\,dy_0\biggr| < 1 - \varepsilon_1\biggr] \\
&\quad \geq \mathrm{P}\biggl\{\sup_{y_1,y_0}\bigl|\hat{H}_N^H(y_1, y_0) - H_N^H(y_1, y_0)\bigr| < \frac{(N-1)(1-\varepsilon_1)}{8NC^2}\biggr\} \\
&\quad \geq \mathrm{P}\biggl[\sup_{y_1,y_0}\bigl|\min\{\hat{G}_N(y_1), \hat{F}_N(y_0)\} - \min\{G_N(y_1), F_N(y_0)\}\bigr| \\
&\qquad\qquad < \frac{(N-1)(1-\varepsilon_1)}{8NC^2}\biggr] \\
&\quad \geq \mathrm{P}\biggl\{\sup_y\bigl|G_N(y) - \hat{G}_N(y)\bigr|, \sup_y\bigl|F_N(y) - \hat{F}_N(y)\bigr| < \frac{(N-1)(1-\varepsilon_1)}{8NC^2}\biggr\} \\
&\quad \geq \mathrm{P}\biggl\{\sup_y\bigl|G(y) - \hat{G}_N(y)\bigr|, \sup_y\bigl|F(y) - \hat{F}_N(y)\bigr| < \frac{(N-1)(1-\varepsilon_1)}{8NC^2} - o(1)\biggr\} \\
&\quad \geq 1 - \mathrm{P}\biggl\{\sup_y\bigl|G(y) - \hat{G}_N(y)\bigr| \geq \frac{(N-1)(1-\varepsilon_1)}{8NC^2} - o(1)\biggr\} \\
&\quad\quad - \mathrm{P}\biggl\{\sup_y\bigl|F(y) - \hat{F}_N(y)\bigr| \geq \frac{(N-1)(1-\varepsilon_1)}{8NC^2} - o(1)\biggr\}.
\end{aligned}$$

The fourth inequality follows from Lemma 2 which shows that $\sup_y |G(y) - G_N(y)| = o(1)$ and $\sup_y |F(y) - F_N(y)| = o(1)$. We can now apply the second part of Lemma 2 to bound the probability above. Given $\xi > 0$,

$$\begin{aligned}
&\limsup_N N\mathrm{P}\left[\frac{2N}{N-1}\left|\int_{[-C,C]^2}\{\hat{H}_N^H(y_1, y_0) - H_N^H(y_1, y_0)\}\,dy_1\,dy_0\right| \geq 1-\varepsilon_1\right] \\
&\quad\leq \limsup_N N\mathrm{P}\left\{\sup_y |G(y) - \hat{G}_N(y)| \geq \frac{1-\varepsilon_1}{8C^2} - \xi\right\} \\
&\qquad + \limsup_N N\mathrm{P}\left\{\sup_y |F(y) - \hat{F}_N(y)| \geq \frac{1-\varepsilon_1}{8C^2} - \xi\right\} \\
&\quad\leq \frac{(1-\theta\rho)K_1(((1-\varepsilon_1)/(8C^2)) - \xi)}{\theta\rho\{((1-\varepsilon_1)/(8C^2)) - \xi\}^2} \\
&\qquad + \frac{\{1-\theta(1-\rho)\}K_0(((1-\varepsilon_1)/(8C^2)) - \xi)}{\theta(1-\rho)\{((1-\varepsilon_1)/(8C^2)) - \xi\}^2}.
\end{aligned}$$

Since ξ is arbitrary and both $K_1(\cdot)$ and $K_0(\cdot)$ are nonincreasing, there exists $\kappa_1(\varepsilon_1)$ such that

$$\begin{aligned}
&\limsup_N N\mathrm{P}\left[\frac{2N}{N-1}\left|\int_{[-C,C]^2}\{\hat{H}_N^H(y_1, y_0) - H_N^H(y_1, y_0)\}\,dy_1\,dy_0\right| \geq 1-\varepsilon_1\right] \\
&\quad\leq C^4\kappa_1(\varepsilon_1).
\end{aligned} \tag{23}$$

The bounds (17)–(23) imply that

$$\limsup_N N\mathrm{P}(N|\hat{V}_N^H - V_N| \geq \varepsilon) \leq C^4\left(\sum_{i=1}^{8}\frac{c_i}{\nu_i^2} + \kappa_1(\varepsilon_1)\right).$$

By minimizing the right-hand side over $\nu_1, \dots, \nu_8 > 0$ subject to the constraint $\nu_1 + \cdots + \nu_8 = \varepsilon_1$, the sum in the parenthesis can be absorbed into $\kappa_1(\varepsilon_1)$, yielding the desired convergence rate for $N\hat{V}_N^H$. To get the rate for $N\hat{V}_N^L$, we repeat the argument used to derive (23). First note that $|u - u'| < \eta$ and $|v - v'| < \zeta$ implies $|\max(0, u+v-1) - \max(0, u'+v'-1)| < \eta + \zeta$. This gives the second inequality below:

$$\begin{aligned}
&\mathrm{P}\left[\frac{2N}{N-1}\left|\int_{[-C,C]^2}\{\hat{H}_N^L(y_1, y_0) - H_N^L(y_1, y_0)\}\,dy_1\,dy_0\right| < \varepsilon_2 + \varepsilon_3\right] \\
&\quad\geq \mathrm{P}\Bigl[\sup_{y_1, y_0}\bigl|\max\{0, \hat{G}_N(y_1) + \hat{F}_N(y_0) - 1\} \\
&\qquad\qquad - \max\{0, G_N(y_1) + F_N(y_0) - 1\}\bigr| \\
&\qquad\qquad\qquad < \frac{(N-1)(\varepsilon_2 + \varepsilon_3)}{8NC^2}\Bigr] \\
&\quad\geq \mathrm{P}\left\{\sup_y |G(y) - \hat{G}_N(y)| < \frac{(N-1)\varepsilon_2}{8NC^2} - o(1),\right.
\end{aligned}$$

$$\sup_y\left|F(y)-\hat{F}_N(y)\right| < \frac{(N-1)\varepsilon_3}{8NC^2} - o(1)\Bigg\}$$

$$\geq 1 - \mathrm{P}\left\{\sup_y\left|G(y)-\hat{G}_N(y)\right| \geq \frac{(N-1)\varepsilon_2}{8NC^2} - o(1)\right\}$$

$$- \mathrm{P}\left\{\sup_y\left|F(y)-\hat{F}_N(y)\right| \geq \frac{(N-1)\varepsilon_3}{8NC^2} - o(1)\right\}.$$

Thus there exist $\kappa_2(\varepsilon_2)$ and $\kappa_3(\varepsilon_3)$ such that

$$\begin{aligned}&\limsup_N N\mathrm{P}\left[\frac{2N}{N-1}\left|\int_{[-C,C]^2}\left\{\hat{H}_N^L(y_1,y_0)-H_N^L(y_1,y_0)\right\}dy_1\,dy_0\right| \geq \varepsilon_2+\varepsilon_3\right]\\ &\quad\leq C^4\left\{\kappa_2(\varepsilon_2)+\kappa_3(\varepsilon_3)\right\}.\end{aligned} \tag{24}$$

□

APPENDIX B: R CODE FOR IMPLEMENTING ESTIMATOR

Here, we present R code for the function `sharp.var`, which outputs the bound estimates $\hat{V}_N^H$ (given input `upper=TRUE`) and $\hat{V}_N^L$ (given input `upper=FALSE`). The other inputs are `yt` (the observed outcomes under treatment), `yc` (the observed outcomes under control) and `N` (the total number of units in the population).

```
sharp.var <- function(yt,yc,N=length(c(yt,yc)),upper=TRUE) {
 m <- length(yt)
 n <- m + length(yc)
 FPvar <- function(x,N) (N-1)/(N*(length(x)-1))
   * sum((x - mean(x))^2)
 yt <- sort(yt)
 if(upper == TRUE) yc <- sort(yc) else
   yc <- sort(yc,decreasing=TRUE)
 p_i <- unique(sort(c(seq(0,n-m,1)/(n-m),seq(0,m,1)/m))) -
 .Machine$double.eps^.5
 p_i[1] <- .Machine$double.eps^.5
 yti <- yt[ceiling(p_i*m)]
 yci <- yc[ceiling(p_i*(n-m))]
 p_i_minus <- c(NA,p_i[1: (length(p_i)-1)])
 return(((N-m)/m * FPvar(yt,N) + (N-(n-m))/(n-m) * FPvar(yc,N)
   + 2*sum(((p_i-p_i_minus)*yti*yci)[2:length(p_i)])
   - 2*mean(yt)*mean(yc))/(N-1))
}
```

APPENDIX C: ILLUSTRATIVE UPPER BOUND IMPROVEMENTS

In Table 3, we present illustrations of the improvements in the variance upper bounds by varying the marginal distributions of potential outcomes over the Beta distribution family: the control potential outcomes are assumed to be distributed according to Beta(α_0,β_0), and the treatment potential outcomes according to Beta(α_1,β_1). Strictly speaking, since finite populations cannot have continuous marginals, the Beta distributions represent approximations to plausible marginals

TABLE 3
Illustrative upper bound ratios given Beta distributed potential outcomes

	α_0	β_0	α_1	β_1	V_N^H/V_N^a	V_N^H/V_N^{b+}
1	0.1	0.1	0.1	0.1	1.00	1.00
2	0.1	0.1	0.1	1	0.68	0.79
3	0.1	0.1	0.1	2	0.61	0.81
4	0.1	0.1	1	1	0.92	0.97
5	0.1	0.1	1	2	0.86	0.95
6	0.1	0.1	2	2	0.86	0.96
7	1	1	0.1	0.1	0.92	0.97
8	1	1	0.1	1	0.81	0.84
9	1	1	0.1	2	0.71	0.83
10	1	1	1	1	1.00	1.00
11	1	1	1	2	0.98	0.99
12	1	1	2	2	0.98	1.00
13	2	2	0.1	0.1	0.86	0.96
14	2	2	0.1	1	0.85	0.85
15	2	2	0.1	2	0.76	0.83
16	2	2	1	1	0.98	1.00
17	2	2	1	2	0.99	0.99
18	2	2	2	2	1.00	1.00

when N is large. We report the ratios V_N^H/V_N^a and V_N^H/V_N^{b+} (the limits of $\hat{V}_N^H/\hat{V}_N^a$ and $\hat{V}_N^H/\hat{V}_N^{b+}$) under different values of $(\alpha_0, \beta_0, \alpha_1, \beta_1)$ while holding $m = n/2$ and $n = N$ fixed.

Table 3 presents 18 scenarios, wherein $(\alpha_0, \beta_0) \in \{(0.1, 0.1), (1, 1), (2, 2)\}$, and $\alpha_1, \beta_1 \in \{0.1, 1, 2\}$. The results are identical for Beta(α_1, β_1) and Beta(β_1, α_1); thus, we omit redundant results. The ratios were computed via numerical quadrature using the `NIntegrate` command in Mathematica 7.0.1.0 under the default settings.

Our results illustrate that when the marginal distributions are identical (i.e., cases 1, 10 and 18), all upper bounds are identical, since the Cauchy–Schwarz and AM-GM inequalities hold exactly. However, as the marginal distributions diverge in shape (e.g., cases 3, 9 and 15), our proposed upper bound V_N^H materially outperforms Neyman's bounds V_N^a and V_N^{b+}.

Acknowledgements. The authors thank Allison Carnegie, Ed Kaplan, Winston Lin, Cyrus Samii, Aad van der Vaart and the review team for helpful comments.

REFERENCES

[1] COCHRAN, W. G. (1977). *Sampling Techniques*, 3rd ed. Wiley, New York. MR0474575

[2] COPAS, J. B. (1973). Randomization models for the matched and unmatched 2 × 2 tables. *Biometrika* **60** 467–476. MR0448746

[3] FREEDMAN, D. A. (2008). On regression adjustments in experiments with several treatments. *Ann. Appl. Stat.* **2** 176–196. MR2415599

[4] FREEDMAN, D. A., PISANI, R. and PURVES, R. A. (2007). *Statistics*, 4th ed. Norton, New York.

[5] GADBURY, G. L. (2001). Randomization inference and bias of standard errors. *Amer. Statist.* **55** 310–313. MR1939365

[6] GADBURY, G. L., IYER, H. K. and ALBERT, J. M. (2004). Individual treatment effects in randomized trials with binary outcomes. *J. Statist. Plann. Inference* **121** 163–174. MR2038814

[7] GERBER, A. S. and GREEN, D. P. (2012). *Field Experiments: Design, Analysis, and Interpretation*. Norton, New York.

[8] HARRISON, B. F. and MICHELSON, M. R. (2012). Not that there's anything wrong with that: The effect of personalized appeals on marriage equality campaigns. *Political Behavior* **34** 325–344.

[9] HECKMAN, J. J., SMITH, J. and CLEMENTS, N. (1997). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *Rev. Econom. Stud.* **64** 487–535. MR1485826

[10] ISAKI, C. T. and FULLER, W. A. (1982). Survey design under the regression superpopulation model. *J. Amer. Statist. Assoc.* **77** 89–96. MR0648029

[11] LEHMANN, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37** 1137–1153. MR0202228

[12] LIN, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *Ann. Appl. Stat.* **7** 295–318. MR3086420

[13] NEYMAN, J. S. (1990). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statist. Sci.* **5** 465–472. Reprint of the original 1923 paper. MR1092986

[14] ROBINS, J. M. (1988). Confidence intervals for causal parameters. *Stat. Med.* **7** 773–785.

[15] RUBIN, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Ann. Statist.* **6** 34–58. MR0472152

[16] SAMII, C. and ARONOW, P. M. (2012). On equivalencies between design-based and regression-based variance estimators for randomized experiments. *Statist. Probab. Lett.* **82** 365–370. MR2875224

[17] TCHEN, A. H. (1980). Inequalities for distributions with given marginals. *Ann. Probab.* **8** 814–827. MR0577318

[18] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York. MR1385671

[19] ZHANG, Z., WANG, C., NIE, L. and SOON, G. (2013). Assessing the heterogeneity of treatment effects via potential outcomes of individual patients. *J. Roy. Statist. Soc. Ser. C* **62** 687–704.

P. M. ARONOW
DEPARTMENT OF POLITICAL SCIENCE
YALE UNIVERSITY
NEW HAVEN, CONNECTICUT 06520
USA
E-MAIL: peter.aronow@yale.edu

D. P. GREEN
DEPARTMENT OF POLITICAL SCIENCE
COLUMBIA UNIVERSITY
NEW YORK, NEW YORK 10027
USA
E-MAIL: dpg2110@columbia.edu

D. K. K. LEE
SCHOOL OF MANAGEMENT (O.R.)
YALE UNIVERSITY
NEW HAVEN, CONNECTICUT 06520
USA
E-MAIL: donald.lee@yale.edu