Asymptotic theory of rerandomization in treatment–control experiments

Xinran Li[a], Peng Ding[b], and Donald B. Rubin[a,1]

[a]Department of Statistics, Harvard University, Cambridge, MA 02138; and [b]Department of Statistics, University of California, Berkeley, CA 94720

Contributed by Donald B. Rubin, June 29, 2018 (sent for review May 17, 2018; reviewed by Robert J. Tibshirani and C. F. Jeff Wu)

Although complete randomization ensures covariate balance on average, the chance of observing significant differences between treatment and control covariate distributions increases with many covariates. Rerandomization discards randomizations that do not satisfy a predetermined covariate balance criterion, generally resulting in better covariate balance and more precise estimates of causal effects. Previous theory has derived finite sample theory for rerandomization under the assumptions of equal treatment group sizes, Gaussian covariate and outcome distributions, or additive causal effects, but not for the general sampling distribution of the difference-in-means estimator for the average causal effect. We develop asymptotic theory for rerandomization without these assumptions, which reveals a non-Gaussian asymptotic distribution for this estimator, specifically a linear combination of a Gaussian random variable and truncated Gaussian random variables. This distribution follows because rerandomization affects only the projection of potential outcomes onto the covariate space but does not affect the corresponding orthogonal residuals. We demonstrate that, compared with complete randomization, rerandomization reduces the asymptotic quantile ranges of the difference-in-means estimator. Moreover, our work constructs accurate large-sample confidence intervals for the average causal effect.

causal inference | covariate balance | geometry of rerandomization | Mahalanobis distance | quantile range

Ever since Fisher's (1–3) seminal work, randomized experiments have become the "gold standard" for drawing causal inferences. Complete randomization balances the covariate distributions between treatment groups in expectation, thereby ensuring the existence of unbiased estimators of average causal effects. Covariate imbalance, however, often occurs in specific randomized experiments, as recognized by Fisher (2) and later researchers (e.g., refs. 4–9). The standard approach advocated by Fisher (3), stratification or blocking, ensures balance with a few discrete covariates (e.g., refs. 10–12).

When a randomized allocation is unbalanced, it is reasonable to discard that allocation and redraw another one until a certain predetermined covariate balance criterion is satisfied. This is rerandomization, an experimental design hinted at by Fisher (cf. ref. 13, p. 88) and Cox (14, 15) and formally proposed by Rubin (16) and Morgan and Rubin (17). Morgan and Rubin (17) showed that the difference-in-means estimator is generally unbiased for the average causal effect under rerandomization with equal-sized treatment groups and obtained the sampling variance of this estimator under additional assumptions of Gaussian covariate and outcome distributions and additive causal effects. When rerandomization is applied but these assumptions do not hold, statistical inference becomes more challenging, because the theory that is justified by the central limit theorem under complete randomization (18, 19) no longer generally holds. Some applied researchers believe that "the only analysis that we can be completely confident in is a permutation test or rerandomization test" (ref. 7, p. 219). However, permutation tests based on randomization require sharp null hypotheses that imply all missing potential outcomes are known.

Analogous to the repeated sampling properties for complete randomization (11, 20), we evaluate the repeated sampling properties of the difference-in-means estimator when rerandomization is used, where all potential outcomes and covariates are regarded as fixed quantities and all randomness arises solely from the random treatment assignments. The geometry of rerandomization reveals non-Gaussian asymptotic distributions, which serve as the foundation for constructing large-sample confidence intervals for average causal effects. Furthermore, we compare the lengths of quantile ranges of the asymptotic distributions of the difference-in-means estimator under rerandomization and complete randomization, extending Morgan and Rubin's (17, 21) comparison of their sampling variances.

Framework, Notation, and Basic Results

Covariate Imbalance and Rerandomization. Inferring the causal effect of some binary treatment on an outcome Y is of central interest in many studies. We consider an experiment with n units, with n_1 assigned to treatment and n_0 assigned to control, $n = n_1 + n_0$. Before conducting the experiment, we collect K covariates with values $\boldsymbol{X}_i = (X_{1i}, X_{2i}, \ldots, X_{Ki})'$ for the ith unit, which can possibly include transformations of basic covariates and their interactions. Let Z_i be the treatment indicator for unit i ($Z_i = 1$ if the active treatment level; $Z_i = 0$ if the control level) and $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_n)'$ be the treatment assignment vector with $n_1 \equiv \sum_{i=1}^n Z_i$ and $n_0 \equiv \sum_{i=1}^n (1 - Z_i)$. In a completely randomized experiment (CRE), n_1 and n_0 are fixed, and the distribution of $\boldsymbol{Z}$ is such that each possible value, $\boldsymbol{z} = (z_1, \ldots, z_n)'$, of $\boldsymbol{Z}$ has probability $n_1!n_0!/n!$. The difference-in-means vector of the covariates between treatment and control groups is $\hat{\boldsymbol{\tau}}_{\boldsymbol{X}} = n_1^{-1} \sum_{i=1}^n Z_i \boldsymbol{X}_i - n_0^{-1} \sum_{i=1}^n (1 - Z_i) \boldsymbol{X}_i$. Although on average $\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$ has mean zero over all $n!/(n_1!n_0!)$ randomizations, for any realized value of $\boldsymbol{Z}$, imbalancedness in covariate distributions

Significance

Rerandomization refers to experimental designs that enforce covariate balance. This paper studies the asymptotic properties of the difference-in-means estimator under rerandomization, based on the randomness of the treatment assignment without imposing any parametric modeling assumptions on the covariates or outcome. The non-Gaussian asymptotic distribution allows for constructing large-sample confidence intervals for the average treatment effect and demonstrates the advantages of rerandomization over complete randomization.

Author contributions: X.L., P.D., and D.B.R. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

Reviewers: R.J.T., Stanford University; and C.F.J.W., Georgia Institute of Technology.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. Email: dbrubin@mac.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1808191115/-/DCSupplemental.

Published online August 27, 2018.

between treatment groups often occurs. For 10 uncorrelated covariates, with probability 40%, at least one of the absolute values of the t statistics comparing covariate means will be larger than 1.96, the 0.975 quantile of the standard Gaussian distribution (17).

When covariate imbalance arises in a drawn allocation, it is reasonable to discard that allocation and draw another until some a priori covariate balance criterion is satisfied. This is rerandomization, an intuitive experimental design tool apparently personally advocated by R. A. Fisher (discussion in ref. 16) and formally discussed by Morgan and Rubin (17).

In general, rerandomization entails the following steps: (*i*) Collect covariate data; (*ii*) specify a balance criterion to determine whether a randomization is acceptable or not; (*iii*) randomize the units to treatment and control groups; (*iv*) if the balance criterion is satisfied, proceed to step *v*, and otherwise, return to step *iii*; (*v*) conduct the experiment using the final randomization obtained in step *iv*; and (*vi*) analyze the data, taking into account the rerandomization used in steps *ii*–*iv*.

Although the balance criterion in step *ii* can be general, Morgan and Rubin (17) suggested using the Mahalanobis distance between covariate means in treatment and control groups, and they (21) suggested considering tiers of covariates according to their presumed importance in predicting the outcomes in this experiment. We discuss these two types of rerandomization in detail and apposite statistical inference after these rerandomizations as implied by step *vi*. We then extend the theory to general rerandomizations in *SI Appendix*, section A1.

Potential Outcomes and Definitions of Finite Population Quantities. We use the potential outcomes framework to define causal effects and let $Y_i(1)$ and $Y_i(0)$ denote the potential outcomes of unit i under active treatment and control, respectively. On the difference scale, the individual causal effect for unit i is $\tau_i = Y_i(1) - Y_i(0)$, and the average causal effect for the finite population of n units is $\tau = \sum_{i=1}^n \tau_i/n$. Let $\bar{Y}(z) = \sum_{i=1}^n Y_i(z)/n$ be the finite population average of potential outcomes under treatment arm z and $\bar{\boldsymbol{X}} = \sum_{i=1}^n \boldsymbol{X}_i/n$ be the finite population average of covariates. Let $S^2_{Y(z)} = \sum_{i=1}^n \{Y_i(z) - \bar{Y}(z)\}^2/(n-1)$ be the finite population variance (with divisor $n-1$) of the potential outcomes under treatment arm z, $S^2_\tau = \sum_{i=1}^n (\tau_i - \tau)^2/(n-1)$ be the finite population variance of the individual causal effects, $S_{Y(z),\boldsymbol{X}} = S'_{\boldsymbol{X},Y(z)} = \sum_{i=1}^n \{Y_i(z) - \bar{Y}(z)\}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})'/(n-1)$ be the finite population covariance between potential outcomes and covariates, and $\boldsymbol{S}^2_{\boldsymbol{X}} = \sum_{i=1}^n (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})'/(n-1)$ be the finite population covariance matrix of covariates. These fixed quantities depend on n implicitly, but do not depend on the randomization or rerandomization scheme.

Repeated Sampling Inference in a CRE. The observed outcome for unit i is $Y_i = Z_i Y_i(1) + (1 - Z_i) Y_i(0)$, a function of treatment assignment and potential outcomes. In a CRE, Neyman (20) showed that, for estimating τ, the difference-in-means estimator $\hat{\tau} = n_1^{-1} \sum_{i=1}^n Z_i Y_i - n_0^{-1} \sum_{i=1}^n (1 - Z_i) Y_i$ is unbiased (the expectation of $\hat{\tau}$ over all possible randomizations is τ) and obtained its sampling variance over all randomizations for constructing a large-sample confidence interval for τ. However, Neyman's (20) interval is not accurate if rerandomization is used.

Let $r_1 = n_1/n$ and $r_0 = n_0/n$ be the proportions of units receiving treatment and control. According to the finite population central limit theorem (19), under some regularity conditions, the large-n sampling distribution, over all randomizations, of $\sqrt{n}(\hat{\tau} - \tau, \hat{\boldsymbol{\tau}}'_{\boldsymbol{X}})$ is Gaussian with mean zero and covariance matrix

$$\boldsymbol{V} = \begin{pmatrix} V_{\tau\tau} & \boldsymbol{V}_{\tau\boldsymbol{x}} \\ \boldsymbol{V}_{\boldsymbol{x}\tau} & \boldsymbol{V}_{\boldsymbol{xx}} \end{pmatrix} = \begin{pmatrix} r_1^{-1} S^2_{Y(1)} + r_0^{-1} S^2_{Y(0)} - S^2_\tau & r_1^{-1} \boldsymbol{S}_{Y(1),\boldsymbol{X}} + r_0^{-1} \boldsymbol{S}_{Y(0),\boldsymbol{X}} \\ r_1^{-1} \boldsymbol{S}_{\boldsymbol{X},Y(1)} + r_0^{-1} \boldsymbol{S}_{\boldsymbol{X},Y(0)} & (r_1 r_0)^{-1} \boldsymbol{S}^2_{\boldsymbol{X}} \end{pmatrix}.$$

We are conducting randomization-based inference, where all of the covariates and potential outcomes are fixed numbers, and randomness comes solely from the treatment assignment. We embed n units into an infinite sequence of finite populations with increasing sizes, and a sufficient condition for the asymptotic Gaussianity of $\sqrt{n}(\hat{\tau} - \tau, \hat{\boldsymbol{\tau}}'_{\boldsymbol{X}})$ is as follows (19).

Condition 1: As $n \to \infty$, for $z = 0, 1$, (*i*) r_z, the proportion of units under treatment arm z, has positive limits; (*ii*) the finite population variances and covariances $S^2_{Y(z)}, S^2_\tau, \boldsymbol{S}^2_{\boldsymbol{X}}$ and $S_{\boldsymbol{X},Y(z)}$ have finite limiting values, and the limit of $\boldsymbol{S}^2_{\boldsymbol{X}}$ is nonsingular; and (*iii*) $\max_{1 \le i \le n} |Y_i(z) - \bar{Y}(z)|^2/n \to 0$ and $\max_{1 \le i \le n} \|\boldsymbol{X}_i - \bar{\boldsymbol{X}}\|_2^2/n \to 0$.

We introduce the notation $\dot{\sim}$ for two sequences of random vectors converging weakly to the same distribution. Therefore, under the CRE and *Condition 1*, $\sqrt{n}(\hat{\tau} - \tau, \hat{\boldsymbol{\tau}}'_{\boldsymbol{X}}) \dot{\sim} (A, \boldsymbol{B}')$, where $(A, \boldsymbol{B}')$ is a random vector from $\mathcal{N}(\boldsymbol{0}, \boldsymbol{V})$ (19).

Rerandomization Using the Mahalanobis Distance

Mahalanobis Distance. The Mahalanobis distance between the covariate means in treatment and control groups is

$$M = \hat{\boldsymbol{\tau}}'_{\boldsymbol{X}} \{\mathrm{Var}(\hat{\boldsymbol{\tau}}_{\boldsymbol{X}})\}^{-1} \hat{\boldsymbol{\tau}}_{\boldsymbol{X}} = \left(\sqrt{n}\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}\right)' \boldsymbol{V}_{\boldsymbol{xx}}^{-1} \left(\sqrt{n}\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}\right),$$

where $\boldsymbol{V}_{\boldsymbol{xx}} = (r_1 r_0)^{-1} \boldsymbol{S}^2_{\boldsymbol{X}}$ is a fixed and known $K \times K$ matrix in our setting. A rerandomization scheme proposed by Morgan and Rubin (17) accepts only those randomizations with the Mahalanobis distance less than or equal to a, a prespecified threshold. Let $\mathcal{M}$ denote the event that a treatment assignment $\boldsymbol{Z}$ is accepted; that is, $M \le a$. Below we use rerandomization using the Mahalanobis distance (ReM) to denote rerandomization using this criterion, which, as a design, depends on both the covariates and the threshold a.

When we allow transformations and interactions of $\boldsymbol{X}$, ReM can incorporate a wide class of rerandomization schemes. For small sample sizes, there may not exist any randomization satisfying some balance criterion. However, according to the finite population central limit theorem (19), the acceptance probability of a randomization is asymptotically $p_a = P(\chi^2_K \le a)$. Therefore, for relatively large sample sizes, there usually exist many randomizations satisfying the balance criterion. In practice, we want to choose the asymptotic acceptance probability to be small, e.g., $p_a = 0.001$. We comment on this issue in *Discussion*.

Multiple Correlation Between $\hat{\tau}$ and $\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$. The sampling distribution of $\hat{\tau}$ under ReM depends on the squared multiple correlation between $\hat{\tau}$ and $\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$ under the CRE, which is also the proportion of the variance of $\hat{\tau}$ explained by $\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$ in linear projection: $R^2 = \mathrm{Cov}(\hat{\tau}, \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}) \mathrm{Var}(\hat{\boldsymbol{\tau}}_{\boldsymbol{X}})^{-1} \mathrm{Cov}(\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}, \hat{\tau})/\mathrm{Var}(\hat{\tau}) = \boldsymbol{V}_{\tau\boldsymbol{x}} \boldsymbol{V}_{\boldsymbol{xx}}^{-1} \boldsymbol{V}_{\boldsymbol{x}\tau}/V_{\tau\tau}$. Define the variance of the linear projection of $Y(z)$ on $\boldsymbol{X}$ as $S^2_{Y(z)|\boldsymbol{X}} = \boldsymbol{S}_{Y(z),\boldsymbol{X}} \left(\boldsymbol{S}^2_{\boldsymbol{X}}\right)^{-1} \boldsymbol{S}_{\boldsymbol{X},Y(z)}$ for $z = 0, 1$. We similarly define $S^2_{\tau|\boldsymbol{X}}$, the variance of the linear projection of τ on $\boldsymbol{X}$.

Proposition 1. *R^2 can be expressed in terms of the variances of the potential outcomes and of their projections on $\boldsymbol{X}$:*

$$R^2 = \frac{r_1^{-1} S^2_{Y(1)|\boldsymbol{X}} + r_0^{-1} S^2_{Y(0)|\boldsymbol{X}} - S^2_{\tau|\boldsymbol{X}}}{r_1^{-1} S^2_{Y(1)} + r_0^{-1} S^2_{Y(0)} - S^2_\tau}.$$

When the causal effect is additive, $S^2_\tau = 0$, $\boldsymbol{S}_{\tau,\boldsymbol{X}} = 0$, and $\boldsymbol{S}_{Y(1),\boldsymbol{X}} = \boldsymbol{S}_{Y(0),\boldsymbol{X}}$, and then $R^2 = S^2_{Y(0)|\boldsymbol{X}}/S^2_{Y(0)}$ is the squared multiple correlation between $\boldsymbol{X}$ and $Y(0)$.

Asymptotic Sampling Distribution of $\hat{\tau}$ Under ReM. Simply stated, $\sqrt{n}(\hat{\tau}-\tau)$ has two parts: the part unrelated to the covariates, which we call ε_0, and is thus unaffected by rerandomization, and the other part related to the covariates, which we call $L_{K,a}$, and is thus affected by rerandomization. Therefore, the asymptotic distribution of $\hat{\tau}$ is a linear combination of two independent random variables: $\varepsilon_0 \sim \mathcal{N}(0,1)$ is a standard Gaussian random variable, and $L_{K,a} \sim D_1 \,|\, \boldsymbol{D}'\boldsymbol{D} \le a$, where $\boldsymbol{D} = (D_1, \ldots, D_K)' \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_K)$.

Theorem 1. *Under ReM and Condition 1,*

$$\sqrt{n}(\hat{\tau}-\tau) \,|\, \mathcal{M} \dot{\sim} \sqrt{V_{\tau\tau}}\left(\sqrt{1-R^2}\cdot\varepsilon_0 + \sqrt{R^2}\cdot L_{K,a}\right), \qquad [1]$$

where ε_0 is independent of $L_{K,a}$.

The coefficients of the linear combination are functions of R^2, which measures the association between the potential outcomes and the covariates. When $R^2 = 0$, the right-hand side of **[1]** becomes a Gaussian random variable, the same as the asymptotic distribution of $\sqrt{n}(\hat{\tau}-\tau)$ in *Repeated Sampling Inference in a CRE*. Importantly, the definition of R^2 is based on linear projections instead of linear models of the potential outcomes. Our asymptotic theory is based on the distribution of the randomization without imposing any modeling assumptions on the potential outcomes.

Representation of the Asymptotic Distribution Under ReM. The asymptotic distribution in **[1]** involves a random variable $L_{K,a}$ that does not appear in standard statistical problems. The spherical symmetry of the standard Gaussian vector allows us to represent $L_{K,a}$ using some known distributions, which allows for easy simulation of $L_{K,a}$.

Let $\chi^2_{K,a} \sim \chi^2_K \,|\, \chi^2_K \le a$ be a truncated χ^2 random variable, U_K be the first coordinate of the uniform random vector over the $(K-1)$-dimensional unit sphere, S be a random sign taking ± 1 with probability $1/2$, and $\beta_K \sim \text{Beta}\,(1/2, (K-1)/2)$ be a Beta random variable degenerating to point mass at 1 when $K = 1$.

Proposition 2. *$L_{K,a}$ can be represented as*

$$L_{K,a} \sim D_1 \,|\, \boldsymbol{D}'\boldsymbol{D} \le a \sim \chi_{K,a}\, U_K \sim \chi_{K,a}\, S\sqrt{\beta_K}, \qquad [2]$$

where $(\chi_{K,a}, U_K)$ are mutually independent, and $(\chi_{K,a}, S, \beta_K)$ are jointly independent. $L_{K,a}$ is symmetric and unimodal around zero, with $\text{Var}(L_{K,a}) = v_{K,a} = P(\chi^2_{K+2} \le a)/P(\chi^2_K \le a) < 1$.

Because both ε_0 and $L_{K,a}$ are symmetric and both are unimodal at zero, their linear combination is also symmetric and unimodal at zero according to Wintner's (22) theorem. The same is true for the asymptotic distribution of $\sqrt{n}(\hat{\tau}-\tau)$ in **[1]**. Moreover, the unimodal property plays an important role in the conservativeness of confidence intervals, discussed shortly for ReM. The representation in **[2]** allows for easy simulation of $L_{K,a}$, as well as the asymptotic distribution of $\sqrt{n}(\hat{\tau}-\tau)$ in **[1]**, which is relevant for statistical inference discussed later.

In *SI Appendix*, section A2, we give more detailed explanations regarding the geometry and the shape of the asymptotic distribution in **[1]**.

Asymptotic Unbiasedness, Sampling Variance, and Quantile Ranges. *Theorem 1* characterizes the asymptotic behavior of $\hat{\tau}$ over ReM, which immediately implies the following conclusions.

First, the asymptotic distribution in **[1]** is symmetric around 0, implying that $\hat{\tau}$ is asymptotically unbiased for τ. Let $\mathbb{E}_a(\cdot)$ denote the expectation of the asymptotic sampling distribution of a sequence of random vectors.

Corollary 1. *Under ReM and Condition 1,* $\mathbb{E}_a\{\sqrt{n}(\hat{\tau}-\tau)|\mathcal{M}\} = 0$.

Morgan and Rubin (17) gave a counterexample showing that, in an experiment with unequal treatment group sizes, $\hat{\tau}$ can be biased for τ over ReM. Our result confirms a conjecture in ref. 21 that the bias is often small with large samples. *Corollary 1* extends their theorem 2.1 (17) and ensures the asymptotic unbiasedness of $\hat{\tau}$ for experiments with any ratio of n_1/n_0.

Covariates, whether observed or unobserved, are variables unaffected by the treatments. Therefore, the average causal effect on any covariate is 0, and *Corollary 1* implies that any covariate asymptotically has the same means under treatment and control.

Furthermore, from *Proposition 2* and *Theorem 1*, we can calculate the asymptotic sampling variances of $\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$ and $\hat{\tau}$ and the percentage reductions in asymptotic sampling variances (PRIASV) under ReM compared with the CRE. Recalling that $v_{K,a} = P(\chi^2_{K+2} \le a)/P(\chi^2_K \le a)$, we summarize the results in *Corollary 2*.

Corollary 2. *Under ReM and Condition 1, the asymptotic sampling covariance of $\sqrt{n}\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$ is $v_{K,a}\boldsymbol{V}_{\boldsymbol{xx}}$, and the PRIASV of any component of $\sqrt{n}\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$ is $1 - v_{K,a}$. The asymptotic sampling variance of $\sqrt{n}(\hat{\tau}-\tau)$ is $V_{\tau\tau}\left\{1-(1-v_{K,a})R^2\right\}$, and the PRIASV of $\sqrt{n}(\hat{\tau}-\tau)$ is $(1-v_{K,a})R^2$.*

Rigorously, the asymptotic sampling covariance and variance of $\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$ and $\hat{\tau}$ should be the limits of $v_{K,a}\boldsymbol{V}_{\boldsymbol{xx}}$ and $V_{\tau\tau}\{1-(1-v_{K,a})R^2\}$ in the sequence of finite populations. However, for descriptive convenience, we omit these limit signs when discussing the expectation and covariance of asymptotic sampling distributions. When a is close to 0, that is, when the asymptotic acceptance probability is small, the asymptotic sampling variance $V_{\tau\tau}\left\{1-(1-v_{K,a})R^2\right\}$ reduces to $V_{\tau\tau}(1-R^2)$, which is identical to the asymptotic sampling variance of the regression-adjusted estimator under the CRE (18). Therefore, rerandomization accomplishes covariate adjustment in the design stage, whereas regression accomplishes covariate adjustment in the analysis stage.

When the causal effect is additive, R^2 equals the finite population squared multiple correlation between $\boldsymbol{X}$ and $Y(0)$. Therefore, *Corollary 2* is an asymptotic extension of theorem 3.2 in Morgan and Rubin (17).

Under ReM, in addition to the sampling variance reduction result concerning $\hat{\tau}$ in *Corollary 2*, we consider the reduction in the length of the $(1-\alpha)$ quantile range of $\hat{\tau}$ compared with that under the CRE. We choose the length of the $(1-\alpha)$ quantile range, because of its connection to constructing confidence intervals as discussed shortly.

Let z_ξ be the ξth quantile of a standard Gaussian distribution. Let $\nu_\xi(R^2, p_a, K)$ be the ξth quantile of $\sqrt{1-R^2}\cdot\varepsilon_0 + \sqrt{R^2}\cdot L_{K,a}$, with $\nu_\xi(0, p_a, K) = z_\xi$. Because p_a and K are usually known by design, we write $\nu_\xi(R^2, p_a, K)$ as $\nu_\xi(R^2)$ for notational simplicity. Under ReM, the $(1-\alpha)$ quantile range of the asymptotic distribution of $\sqrt{n}(\hat{\tau}-\tau)$ is

$$\text{QR}_\alpha(V_{\tau\tau}, R^2) = \left[\nu_{\alpha/2}(R^2)\sqrt{V_{\tau\tau}},\ \ \nu_{1-\alpha/2}(R^2)\sqrt{V_{\tau\tau}}\right], \qquad [3]$$

and the corresponding quantile range under the CRE is

$$\text{QR}_\alpha(V_{\tau\tau}, 0) = \left[z_{\alpha/2}\sqrt{V_{\tau\tau}},\ \ z_{1-\alpha/2}\sqrt{V_{\tau\tau}}\right]. \qquad [4]$$

Theorem 2. *Under Condition 1, the length of the $(1-\alpha)$ quantile range of the asymptotic sampling distribution of $\sqrt{n}(\hat{\tau}-\tau)$ under ReM is less than or equal to that under the CRE, with the difference nondecreasing in R^2 and nonincreasing in p_a and K.*

Sampling Variance Estimation and Confidence Intervals. Asymptotic sampling variance and quantile ranges for $\hat{\tau}$ depend on $V_{\tau\tau}$ and

R^2, which are determined by the covariances among potential outcomes and covariates. To obtain a sampling variance estimator and to construct an asymptotic confidence interval for τ, we need to estimate these variances and covariances. Let $s^2_{Y(z)}$, $s^2_{Y(z)|\boldsymbol{X}}$, and $\boldsymbol{s}_{Y(z),\boldsymbol{X}}$ be the sample variance of Y, sample variance of linear projection of Y on $\boldsymbol{X}$, and sample covariance between Y and $\boldsymbol{X}$ in treatment arm z. We show in *SI Appendix*, section A4 that under ReM they are consistent for their population analogues. Therefore, we can then estimate $S^2_{\tau|\boldsymbol{X}}$ by

$$s^2_{\tau|\boldsymbol{X}} = (\boldsymbol{s}_{Y(1),\boldsymbol{X}} - \boldsymbol{s}_{Y(0),\boldsymbol{X}})(\boldsymbol{S}^2_{\boldsymbol{X}})^{-1}(\boldsymbol{s}_{\boldsymbol{X},Y(1)} - \boldsymbol{s}_{\boldsymbol{X},Y(0)})$$

and $V_{\tau\tau}$ by (23)

$$\hat{V}_{\tau\tau} = r_1^{-1}s^2_{Y(1)} + r_0^{-1}s^2_{Y(0)} - s^2_{\tau|\boldsymbol{X}}.$$

We can then estimate R^2 by

$$\hat{R}^2 = \hat{V}^{-1}_{\tau\tau}\left\{r_1^{-1}s^2_{Y(1)|\boldsymbol{X}} + r_0^{-1}s^2_{Y(0)|\boldsymbol{X}} - s^2_{\tau|\boldsymbol{X}}\right\}. \qquad [5]$$

We set $\hat{R}^2$ to be 0 if the estimator in **[5]** is negative.

According to *Corollary 2*, we can estimate the asymptotic sampling variance of $\hat{\tau}$ by $\hat{V}_{\tau\tau}\{1-(1-v_{K,a})\hat{R}^2\}/n$, and according to **[3]**, we can construct a large sample $(1-\alpha)$ confidence interval for τ using $\hat{\tau} - \mathrm{QR}_\alpha(\hat{V}_{\tau\tau}, \hat{R}^2)/\sqrt{n}$. Not surprisingly, similar to Neyman's (20) analysis of the CRE, unless the residual from the linear projection of the individual causal effect on the covariates is constant, the above sampling variance estimator and the associated confidence interval are both asymptotically conservative, in the sense that the probability limit of the variance estimator is larger than or equal to the actual sampling variance, and the limit of coverage probability of the confidence interval is larger than or equal to $(1-\alpha)$.

Moreover, the sampling variance estimator is smaller than Neyman's (20) sampling variance estimator for the CRE, and the confidence interval is shorter than Neyman's (20) confidence interval for the CRE. Therefore, if we conduct ReM in the design stage but analyze data as in the CRE, the consequential sampling variance estimator and confidence intervals will be overly conservative.

The above results are all intuitive, and we present the algebraic details for the proofs of these results in *SI Appendix*, section A4. Interestingly, as shown in *SI Appendix*, section A4, we do not need conditions beyond *Condition 1* to ensure the asymptotic properties of the sampling variance estimator and the confidence intervals.

Rerandomization with Tiers of Covariates

Mahalanobis Distance with Tiers of Covariates. When covariates are thought to have different levels of importance for the outcomes, Morgan and Rubin (21) proposed rerandomization using the Mahalanobis distance with differing criteria for different tiers of covariates. We partition the covariates into T tiers indexed by $t = 1, \ldots, T$ with decreasing importance, with k_t covariates in tier t. Let $\boldsymbol{X}_i = (\boldsymbol{X}_i[1], \ldots, \boldsymbol{X}_i[T])$, where $\boldsymbol{X}_i[t]$ denotes the covariates in tier t. Define $\boldsymbol{X}_i[\overline{t}] = (\boldsymbol{X}_i[1], \ldots, \boldsymbol{X}_i[t])$, the covariates in the first t tiers. Following Morgan and Rubin (21), let $\boldsymbol{S}^2_{\boldsymbol{X}[\overline{t-1}]}$ be the finite population covariance matrix of the covariates in the first $t-1$ tiers and $\boldsymbol{S}_{\boldsymbol{X}[t],\boldsymbol{X}[\overline{t-1}]}$ be the finite population covariance between $\boldsymbol{X}[t]$ and $\boldsymbol{X}[\overline{t-1}]$. We first apply a block-wise Gram–Schmidt orthogonalization to the covariates to create the orthogonalized covariates. Let $\boldsymbol{E}_i[1] = \boldsymbol{X}_i[1]$, and for $2 \le t \le T$, let

$$\boldsymbol{E}_i[t] = \boldsymbol{X}_i[t] - \boldsymbol{S}_{\boldsymbol{X}[t],\boldsymbol{X}[\overline{t-1}]}\left(\boldsymbol{S}^2_{\boldsymbol{X}[\overline{t-1}]}\right)^{-1}\boldsymbol{X}_i[\overline{t-1}],$$

where $\boldsymbol{E}_i[t]$ is the residual of the projection of the covariates $\boldsymbol{X}_i[t]$ in tier t onto the space spanned by the covariates in previous tiers. Let $\boldsymbol{E}_i = (\boldsymbol{E}_i[1], \ldots, \boldsymbol{E}_i[T])$. Let $\hat{\boldsymbol{\tau}}_{\boldsymbol{E}[t]}$ be the difference-in-means vector of $\boldsymbol{E}_i[t]$ between treatment and control groups and $\boldsymbol{S}^2_{\boldsymbol{E}[t]}$ be the finite population covariance matrix of $\boldsymbol{E}_i[t]$. The Mahalanobis distance in tier t is $M_t = (n_1 n_0)/n \cdot \hat{\boldsymbol{\tau}}'_{\boldsymbol{E}[t]}(\boldsymbol{S}^2_{\boldsymbol{E}[t]})^{-1}\hat{\boldsymbol{\tau}}_{\boldsymbol{E}[t]}$, and rerandomization using the Mahalanobis distance with tiers of covariates (ReMT) accepts those treatment assignments with $M_t \le a_t$, where the a_ts are predetermined constants $(1 \le t \le T)$. If $T = 1$, then ReMT is simply ReM. Let $\mathcal{T}$ denote the event that a treatment assignment $\boldsymbol{Z}$ is accepted under ReMT. The theory below extends Morgan and Rubin (21), using the concepts from our *Rerandomization Using the Mahalanobis Distance* section.

Multiple Correlation Between $\hat{\tau}$ and $\hat{\boldsymbol{\tau}}_{\boldsymbol{E}[t]}$. Let ρ^2_t be the squared multiple correlation between $\hat{\tau}$ and the difference-in-means vector of the orthogonalized covariates in tier t $\hat{\boldsymbol{\tau}}_{\boldsymbol{E}[t]}$: $\rho^2_t = \mathrm{Cov}\left(\hat{\tau}, \hat{\boldsymbol{\tau}}_{\boldsymbol{E}[t]}\right)\mathrm{Var}\left(\hat{\boldsymbol{\tau}}_{\boldsymbol{E}[t]}\right)^{-1}\mathrm{Cov}\left(\hat{\boldsymbol{\tau}}_{\boldsymbol{E}[t]}, \hat{\tau}\right)/\mathrm{Var}(\hat{\tau})$. Define the finite population variance of the projection of $Y(z)$ on $\boldsymbol{E}[t]$ as $S^2_{Y(z)|\boldsymbol{E}[t]} = \boldsymbol{S}_{Y(z),\boldsymbol{E}[t]}\left(\boldsymbol{S}^2_{\boldsymbol{E}[t]}\right)^{-1}\boldsymbol{S}_{\boldsymbol{E}[t],Y(z)}$ for $z = 0, 1$, where $\boldsymbol{S}_{Y(z),\boldsymbol{E}[t]}$ is the finite population covariance between potential outcomes and orthogonalized covariates in tier t. We can similarly define $S^2_{\tau|\boldsymbol{E}[t]}$. According to *Proposition 1*,

$$\rho^2_t = \frac{r_1^{-1}S^2_{Y(1)|\boldsymbol{E}[t]} + r_0^{-1}S^2_{Y(0)|\boldsymbol{E}[t]} - S^2_{\tau|\boldsymbol{E}[t]}}{r_1^{-1}S^2_{Y(1)} + r_0^{-1}S^2_{Y(0)} - S^2_\tau}, \quad (1 \le t \le T).$$

When the causal effect is additive, $\rho^2_t = S^2_{Y(0)|\boldsymbol{E}[t]}/S^2_{Y(0)}$ reduces to the squared multiple correlation between $Y(0)$ and $\boldsymbol{E}[t]$. For descriptive simplicity, we introduce $\rho^2_{T+1} = 1 - \sum_{t=1}^T \rho^2_t = 1 - R^2$ for later discussion.

Asymptotic Sampling Distribution of $\hat{\tau}$ Under ReMT. Intuitively, $\sqrt{n}(\hat{\tau} - \tau)$ can be decomposed into $(T+1)$ parts: the part unrelated to covariates and the T projections onto the spaces spanned by the orthogonalized covariates in T tiers. Due to their construction, these $(T+1)$ parts are orthogonal to each other, and the constraint for balance in tier t affects only the tth projection.

As earlier, let $\varepsilon_0 \sim \mathcal{N}(0,1)$, and extending earlier notation using the subscript t, let $L_{k_t,a_t} \sim D_{t1} \,|\, \boldsymbol{D}'_t\boldsymbol{D}_t \le a_t$, where $\boldsymbol{D}_t = (D_{t1}, \ldots, D_{tk_t}) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{k_t})$ for $1 \le t \le T$.

Theorem 3. *Under ReMT and Condition 1,*

$$\sqrt{n}(\hat{\tau} - \tau) \,|\, \mathcal{T} \dot{\sim} \sqrt{V_{\tau\tau}}\left(\rho_{T+1}\cdot\varepsilon_0 + \sum_{t=1}^T \rho_t \cdot L_{k_t,a_t}\right), \qquad [6]$$

where $(\varepsilon_0, L_{k_1,a_1}, \ldots, L_{k_T,a_T})$ *are jointly independent.*

In **[6]**, ε_0 is the part of $\sqrt{n}(\hat{\tau} - \tau)$ that is unrelated to the covariates, and L_{k_t,a_t} is the part related to the orthogonalized covariates $\boldsymbol{E}_i[t]$ in tier t. According to *Proposition 2*, the distribution in *Theorem 3* is easy to simulate.

Asymptotic Unbiasedness, Sampling Variance, and Quantile Ranges. *Theorem 3* characterizes the asymptotic behavior of $\sqrt{n}(\hat{\tau} - \tau)$ under ReMT, which extends Morgan and Rubin (21) as follows.

First, the asymptotic distribution in **[6]** is symmetric around 0, implying that $\hat{\tau}$ is asymptotically unbiased for τ. Therefore, all observed or unobserved covariates have asymptotically zero difference in means.

Corollary 3. *Under ReMT and Condition 1,* $\mathbb{E}_a\{\sqrt{n}(\hat{\tau} - \tau) \,|\, \mathcal{T}\} = 0$.

The asymptotic sampling covariance of $\hat{\tau}_X$ under ReMT has a complicated but conceptually obvious form, and we give it in *SI Appendix*, section A3. Below we present only the PRIASV of $\hat{\tau}$; the PRIASVs for covariates are special cases of the same corollary because covariates are formally "outcomes" unaffected by the treatment. Recall that $v_{k_t,a_t} = P(\chi^2_{k_t+2} \le a_t)/P(\chi^2_{k_t} \le a_t)$.

Corollary 4. *Under ReMT and Condition 1, the asymptotic sampling variance of* $\sqrt{n}(\hat{\tau}-\tau)$ *is* $V_{\tau\tau}\{1-\sum_{t=1}^{T}(1-v_{k_t,a_t})\rho_t^2\}$, *and the PRIASV of* $\sqrt{n}(\hat{\tau}-\tau)$ *is* $\sum_{t=1}^{T}(1-v_{k_t,a_t})\rho_t^2$.

When the causal effect is additive, ρ_t^2 becomes the squared multiple correlation between $\boldsymbol{E}[t]$ and $Y(0)$. Therefore, *Corollary 4* is an asymptotic extension of Morgan and Rubin's theorem 4.2 (21). When the thresholds a_ts are close to zero, the asymptotic sampling variance $V_{\tau\tau}\left\{1-\sum_{t=1}^{T}(1-v_{k_t,a_t})\rho_t^2\right\}$ reduces to $V_{\tau\tau}(1-\sum_{t=1}^{T}\rho_t^2) = V_{\tau\tau}(1-R^2)$, which is identical to that of the regression-adjusted estimator under the CRE (18).

We now compare the quantile range under ReMT to that under the CRE. Let $\nu_\xi(\rho_1^2,\rho_2^2,\ldots,\rho_T^2)$ be the ξth quantile of $\rho_{T+1}\varepsilon_0+\sum_{t=1}^{T}\rho_t L_{k_t,a_t}$. Although $\nu_\xi(\rho_1^2,\rho_2^2,\ldots,\rho_T^2)$ depends also on p_{a_t} and k_t $(1\le t\le K)$, we suppress the dependence to avoid notational clutter. The $(1-\alpha)$ quantile range of the asymptotic distribution of $\sqrt{n}(\hat{\tau}-\tau)$ under ReMT is

$$\mathrm{QR}_\alpha(V_{\tau\tau},\rho_1^2,\ldots,\rho_T^2) = \left[\nu_{\alpha/2}(\rho_1^2,\ldots,\rho_T^2)\sqrt{V_{\tau\tau}},\ \nu_{1-\alpha/2}(\rho_1^2,\ldots,\rho_T^2)\sqrt{V_{\tau\tau}}\right]. \quad [7]$$

The stronger the squared correlation is between the outcome and the orthogonalized covariates in tier t, the more reduction in quantile range when using ReMT rather than the CRE. The following *Theorem 4* is intuitive.

Theorem 4. *Under Condition 1, the* $(1-\alpha)$ *quantile range of the asymptotic distribution of* $\sqrt{n}(\hat{\tau}-\tau)$ *under ReMT is less than, or equal to, the range under the CRE, and the reduction in length is nondecreasing in* ρ_t^2 *and nonincreasing in* p_{a_t} *and* k_t, *for all* $1\le t\le T$.

Sampling Variance Estimation and Confidence Interval. We can estimate $V_{\tau\tau}$ and ρ_t^2 $(1\le t\le T)$ in the same way as in ReM, and we estimate ρ_{T+1}^2 by $1-\hat{R}^2$. In practice, we set $\hat{\rho}_t^2$ $(1\le t\le T)$ to 0 when it is negative due to sampling variability and standardize their sum to $\hat{R}^2$. According to *Corollary 4* and **[7]**, we can estimate the sampling variance of $\hat{\tau}$ and construct confidence intervals for τ by replacing the unknown quantities with their point estimates. The sampling variance estimator is smaller than Neyman's (20) sampling variance estimator for the CRE, and the confidence interval is shorter than Neyman's (20) confidence interval for the CRE; both are asymptotically conservative in general, and only when the residual from the linear projection of the individual causal effect on the covariates is constant, are they asymptotically exact. Therefore, analyzing data from ReMT as if they arose from a CRE, the resulting sampling variance estimator and confidence intervals are overly conservative. These intuitive statements appear to require lengthy proofs, which are relegated to *SI Appendix*, section A4.

Table 1. Three tiers of covariates

Tier	Covariates
Tier 1	High-school GPA
Tier 2	Whether lives at home, gender, age
	Whether rarely puts off studying for tests
Tier 3	Whether mother/father is a college graduate
	Whether mother/father is a high-school graduate
	Whether never puts off studying for tests
	Whether wants more than a bachelor degree
	Whether intends to finish in 4 y
	Whether plans to work while in school
	Whether at the first choice school, mother tongue

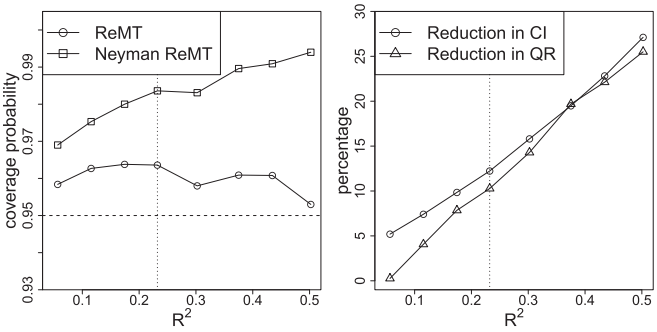

Fig. 1. Eight datasets simulated based on the Student Achievement and Retention Project. *Left* shows the empirical coverage probabilities of our and Neyman's (20) 95% confidence intervals under ReMT, and *Right* shows the percentage reductions of average lengths of confidence intervals and quantile ranges comparing ReMT with a CRE.

An Education Example with Tiers of Covariates

We illustrate our theory using the data from the Student Achievement and Retention Project (24), a randomized evaluation of academic services and incentives at one of the satellite campuses of a large Canadian university. A treatment group of 150 students was offered an array of support services and substantial cash awards for meeting a target first-year grade-point average (GPA), and a control group of many more (1,006) students received only standard university support services.

To illustrate the benefit of rerandomization, we use the 15 covariates listed in Table 1 and exclude students with missing values, resulting in $n_1 = 118$ and $n_0 = 856$. To make the simulation relevant to the real data, we fix unknown parameters based on some simple model fitting: We fit a linear regression of the observed first-year GPA on the treatment indicator, all covariates and their interactions, and use the fitted model to generate all potential outcomes eschewing additivity. Note that the generating models for the potential outcomes are nonlinear in the basic covariates. To make the data-generating process realistic, we simulate eight pseudosets of potential outcomes, using the fitted model with different choices for the variance of the residuals. The error terms for $Y(1)$ and $Y(0)$ are independent, and therefore conditional on the covariates, and the potential outcomes are simulated as uncorrelated, but they have a positive correlation marginally. The final potential outcomes are all truncated to lie on $[0,4]$, mimicking the value of the GPA. We choose different variances of residuals such that the values of R^2 for the eight simulated datasets are located approximately evenly within interval $[0,0.5]$. One choice for the variance of residuals is the one estimated from the fitted linear model, and the corresponding R^2 is about 0.23.

Table 1 partitions the 15 covariates into three tiers. We choose a_t such that $P(\chi^2_{k_t} \le a_t) = (0.001)^{1/3} = 0.1$ for $t = 1,2,3$. We simulate data under ReMT and obtain the confidence intervals based on our asymptotic theory for ReMT and Neyman's (20) results for the CRE. Fig. 1, *Left* shows the empirical coverage probabilities of our and Neyman's (20) confidence intervals, showing that Neyman's (20) CRE confidence intervals are very conservative.

To evaluate the performance of ReMT compared with a CRE, we compare the average length of Neyman's (20) confidence

STATISTICS

interval under a CRE with the confidence interval under ReMT. From Fig. 1, *Right*, the percentage reduction in average lengths of the 95% confidence intervals under ReMT compared with Neyman's (20) under a CRE is nondecreasing in R^2. We also compare the empirical 95% quantile range of $\hat{\tau}$ under ReMT and a CRE: The percentage reductions in the lengths of quantile ranges are close to the percentage reductions for average lengths of confidence intervals. When R^2 is close to that of the real dataset (i.e., 0.23), the percentage increase in the effective sample size, that is, the sample size needed with a CRE for $\hat{\tau}$ to achieve the same 95% quantile range under ReMT, is about 24%. When R^2 is about twice as large as with the real data (i.e., 0.5), the percentage increase in the effective sample size increases to 80%.

Discussion

Our theory suggests that choosing a small p_a will lead to more precise difference in means in general. However, we do not suggest choosing p_a to be too small, such as accepting only those assignments with the smallest Mahalanobis distance. The extreme rerandomization choosing an allocation that balances observed covariates as well as possible has an undesirable consequence. Because it is deterministic, randomization distributions are degenerate, rendering randomization inference impossible. Even if we randomize over all allocations satisfying the best allocation, randomization inference has little power due to very few possible allocations. How to choose p_a remains an open problem.

Materials and Methods

We did not conduct the experiment, and we are analyzing secondary data without any personal identifying information. As such, this study is exempt from human subjects review. The original experiments underwent human subjects review in Canada (24).

ACKNOWLEDGMENTS. P.D. acknowledges support from the National Science Foundation (DMS 1713152). D.B.R. acknowledges support from the National Institute of Allergy and Infectious Diseases/NIH (R01AI102710), National Science Foundation (IIS-1409177), Office of Naval Research (N00014-17-1-2131), and a Google Faculty Fellowship.

1. Fisher RA (1925) *Statistical Methods for Research Workers* (Oliver and Boyd, Edinburgh), 1st Ed.
2. Fisher RA (1926) The arrangement of field experiments. *J Minist Agric G B* 33:503–513.
3. Fisher RA (1935) *The Design of Experiments* (Oliver and Boyd, Edinburgh), 1st Ed.
4. Student (1938) Comparison between balanced and random arrangements of field plots. *Biometrika* 29:363–378.
5. Greevy R, Lu B, Silber JH, Rosenbaum P (2004) Optimal multivariate matching before randomization. *Biostatistics* 5:263–275.
6. Hansen BB, Bowers J (2008) Covariate balance in simple, stratified and clustered comparative studies. *Stat Sci* 23:219–236.
7. Bruhn M, McKenzie D (2009) In pursuit of balance: Randomization in practice in development field experiments. *Am Econ J Appl Econ* 1:200–232.
8. Krieger AM, Azriel D, Kapelner A (2016) Nearly random designs with greatly improved balance. arXiv:1612.02315.
9. Athey S, Imbens GW (2017) The econometrics of randomized experiments. *Handbook of Economic Field Experiments*, eds Banerjee A, Esther D (North-Holland, Amsterdam), Vol 1, pp 73–140.
10. Cochran WG, Cox GM (1992) *Experimental Designs* (Wiley, New York), 2nd Ed.
11. Imbens GW, Rubin DB (2015) *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction* (Cambridge Univ Press, Cambridge, UK).
12. Higgins MJ, Sävje F, Sekhon JS (2016) Improving massive experiments with threshold blocking. *Proc Natl Acad Sci USA* 113:7369–7376.
13. Savage LJ (1962) *The Foundations of Statistical Inference* (Methuen and Co Ltd, London).
14. Cox DR (1982) Randomization and concomitant variables in the design of experiments. *Statistics and Probability: Essays in Honor of C. R. Rao,* eds Krishnaiah PR, Kallianpur G, Ghosh JK (North-Holland, Amsterdam), pp 197–202.
15. Cox DR (2009) Randomization in the design of experiments. *Int Stat Rev* 77:415–429.
16. Rubin DB (2008) Comment to W. R. Shadish, M. H. Clark and P. M Steiner. *J Am Stat Assoc* 103:1350–1353.
17. Morgan KL, Rubin DB (2012) Rerandomization to improve covariate balance in experiments. *Ann Stat* 40:1263–1282.
18. Lin W (2013) Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *Ann Appl Stat* 7:295–318.
19. Li X, Ding P (2017) General forms of finite population central limit theorems with applications to causal inference. *J Am Stat Assoc* 112:1759–1769.
20. Neyman J (1923) On the application of probability theory to agricultural experiments. Essay on principles (with discussion). Section 9 (translated). Reprinted. *Stat Sci* 5:465–472.
21. Morgan KL, Rubin DB (2015) Rerandomization to balance tiers of covariates. *J Am Stat Assoc* 110:1412–1421.
22. Wintner A (1936) On a class of Fourier transforms. *Am J Math* 58:45–90.
23. Ding P, Feller A, Miratrix L (2018) Decomposing treatment effect variation. *J Am Stat Assoc*, in press.
24. Angrist J, Lang D, Oreopoulos P (2009) Incentives and services for college achievement: Evidence from a randomized trial. *Am Econ J Appl Econ* 1:136–163.