The Annals of Statistics
2022, Vol. 50, No. 6, 3439–3465
https://doi.org/10.1214/22-AOS2235

RERANDOMIZATION WITH DIMINISHING COVARIATE IMBALANCE AND DIVERGING NUMBER OF COVARIATES

BY YUHAO WANG[1,a] AND XINRAN LI[2,b]

[1]*Institute for Interdisciplinary Information Sciences, Tsinghua University,* [a]*yuhaow@tsinghua.edu.cn*
[2]*Department of Statistics, University of Illinois at Urbana-Champaign,* [b]*xinranli@illinois.edu*

Completely randomized experiments have been the gold standard for drawing causal inference because they can balance all potential confounding on average. However, they may suffer from unbalanced covariates for realized treatment assignments. Rerandomization, a design that rerandomizes the treatment assignment until a prespecified covariate balance criterion is met, has recently got attention due to its easy implementation, improved covariate balance and more efficient inference. Researchers have then suggested to use the treatment assignments that minimize the covariate imbalance, namely the optimally balanced design. This has caused again the long-time controversy between two philosophies for designing experiments: randomization versus optimal, and thus almost deterministic designs. Existing literature argued that rerandomization with overly balanced observed covariates can lead to highly imbalanced unobserved covariates, making it vulnerable to model misspecification. On the contrary, rerandomization with properly balanced covariates can provide robust inference for treatment effects while sacrificing some efficiency compared to the ideally optimal design. In this paper, we show it is possible that, by making the covariate imbalance diminishing at a proper rate as the sample size increases, rerandomization can achieve its ideally optimal precision that one can expect with perfectly balanced covariates, while still maintaining its robustness. We further investigate conditions on the number of covariates for achieving the desired optimality. Our results rely on a more delicate asymptotic analysis for rerandomization, allowing both diminishing covariate imbalance threshold (or equivalently the acceptance probability) and diverging number of covariates. The derived theory for rerandomization provides a deeper understanding of its large-sample property and can better guide its practical implementation. Furthermore, it also helps reconcile the controversy between randomized and optimal designs in an asymptotic sense.

1. Introduction. Since the seminal work of Fisher (1935), randomized experiments have become the gold standard for drawing causal inference, since they can balance all potential confounding factors, no matter observed or unobserved, on average. Moreover, they allow assumption-free inference for causal effects that uses only the randomization of the treatment assignment as the reasoned basis, without imposing any model or distributional assumption on the experimental units, such as independent and identically distributed (i.i.d.) sampling from some (often hypothetical) superpopulation or some model assumptions for the dependence of potential outcomes on covariates. This is often called randomization-based or design-based inference, as well as the finite population inference emphasizing its focus on the finite population of experimental units in hand; see Fisher (1935) and Splawa-Neyman (1990) for origins of this inferential framework.

Received April 2022; revised July 2022.
MSC2020 subject classifications. Primary 62K99; secondary 62K05.
Key words and phrases. Causal inference, optimal rerandomization, Mahalanobis distance, high-dimensional covariates, Berry–Esseen bound.

However, as pointed out by Morgan and Rubin (2012), the covariate distribution between two treatment groups are likely to be imbalanced for a realized treatment assignment, and a practical remedy hinted by Fisher is to simply rerandomize. The idea of rerandomization is intuitive and has a long history in the literature, traced back to Fisher (see Savage (1962), p. 88), Student (1938) and Cox (1982); see also Morgan and Rubin (2012) and references therein. Besides, it is often used implicitly in the design of experiments when the allocated treated and control groups exhibit undesired imbalances (see, e.g., Bruhn and McKenzie (2009), Heckman and Karapakula (2021)), although it is often not well documented. Nevertheless, the rerandomization design was formally proposed, analyzed and advocated until recently by Morgan and Rubin (2012). As noted by the authors, one main explanation for the less popularity of rerandomization is that it brings additional difficulty in analyzing the experiments, and in practice people often ignore rerandomization and analyze the experiments as if they were, say, completely randomized experiments. Morgan and Rubin (2012) then proposed randomization tests for sharp null hypotheses (e.g., the treatment has no effect for any unit) taking into account rerandomization. More recently, Li, Ding and Rubin (2018) studied the repeated sampling property of the difference-in-means estimator under rerandomization, which exhibits a non-Gaussian distribution, and further demonstrated that the estimator can be more precise under rerandomization than that under complete randomization. Importantly, rerandomization still allows assumption-free randomization-based inference as the complete randomization, and moreover, it provides more efficient difference-in-means estimator and shorter confidence intervals for the average treatment effect.

Researchers, for example, Kasy (2016) and Kallus (2018), have then suggested rerandomization with as small threshold as possible for the covariate imbalance, that is, an optimally balanced design, whose idea can be traced back to Student (1938), Kiefer (1959) and Taves (1974). With general continuous covariates, there is likely only one acceptable treatment assignment or two if the two treatment groups have equal sizes, resulting in an almost deterministic design. Obviously, due to the lack of randomization in the treatment assignment, randomization-based inference is no longer applicable or becomes powerless, since it is generally impossible to asymptotically approximate the randomization distribution of a certain estimator (which is a discrete distribution whose support consists of one or two points) and the minimum possible p-value from a randomization test is either 1 or 0.5 (Johansson, Rubin and Schultzberg (2021), Morgan and Rubin (2012)). Therefore, the statistical inference for an optimally balanced design is often driven by additional distributional assumptions on the experimental units, such as i.i.d. sampling of units from some superpopulation that is usually hypothetical (Johansson, Rubin and Schultzberg (2021)), and the criteria for choosing the optimal assignments are often based on some model assumptions for the dependence of potential outcomes on covariates.

Not surprisingly, there is a long-time debate between these two philosophies, randomized versus optimal (and thus almost deterministic) designs, for conducting experiments. Intuitively, it is similar in spirit to the classical trade-off between efficiency and robustness. Randomized design allows assumption-free and robust inference for treatment effects, while the optimal design tries to maximize the inference efficiency under some hypothesized data-generating models. Specifically, randomized design and its inference can use only the randomization of treatment assignments as the "reasoned basis" (Fisher (1935)), with all the potential outcomes being conditioned on or equivalently viewed as fixed constants. The optimal design often imposes some probabilistic models on the potential outcomes, and its efficiency and inference relies crucially on the randomness in the potential outcomes. Thus, these two designs use quite different sources of randomness. The randomized design has the advantage that the randomness in the treatment assignment is fully controlled by the experimenter and can be readily available for analysis. However, the optimal design relies on the

randomness of potential outcomes as well as their dependence on covariates, which may be misspecified in practice. For example, Kapelner et al. (2021) demonstrated that the "perfect" allocation with minimum covariate imbalance can endanger the estimation precision because unobserved covariates can be highly imbalanced, and Banerjee et al. (2020) suggested that targeting a fixed quantile of balance is safer than targeting an absolute balance objective from an ambiguity-averse decision-making perspective. Indeed, rerandomization can be viewed as a design standing between the completely randomized design and the optimally balanced design. More precisely, instead of ignoring covariate imbalance as in the completely randomized design or pursuing the minimum possible covariate imbalance as in the optimally balanced design, rerandomization repeatedly randomize treatment assignments until the induced covariate imbalance is below a certain threshold, which is chosen carefully to ensure there is sufficient randomness in the treatment assignment. As demonstrated in Li, Ding and Rubin (2018), under rerandomization with a fixed and positive covariate imbalance threshold, we can still conduct large-sample randomization-based inference; moreover, the difference-in-means estimator will be more precise (at least asymptotically) than that under complete randomization, which can further lead to shorter confidence intervals for the average treatment effect.

Nevertheless, there is still a gap in the current theory of rerandomization. Specifically, Li, Ding and Rubin (2018) showed that the smaller the covariate imbalance threshold or equivalently the *acceptance probability* (namely, the probability that the covariate imbalance for a completely randomized treatment assignment is below the threshold) is the more precise the difference-in-means estimator will be under rerandomization. However, this does not mean we should use as small threshold as possible, since it essentially leads to the optimal design for which randomization-based inference is not feasible or powerless. Technically, this is because the current asymptotic theory for rerandomization in Li, Ding and Rubin (2018) requires a fixed and positive covariate imbalance threshold that does not change with the sample size. This then raises the theoretical question that if we can conduct asymptotic analysis for rerandomization allowing a sample-size dependent covariate imbalance threshold, especially with the acceptance probability diminishing toward zero. Philosophically, we are interested in whether, by diminishing the acceptance probability to zero as sample size increases, we can asymptotically achieve the *ideally optimal precision* that one would expect with perfectly balanced covariates while still allowing robust randomization-based inference.

To answer the above questions, we will conduct more delicate finite population asymptotic analysis for rerandomization, allowing the covariate balance criterion including both the threshold and number of involved covariates to vary with the sample size. Our asymptotic analysis relies on a Berry–Esseen-type bound for the finite population central limit theorem. We derive the asymptotic distribution of the difference-in-means estimator under rerandomization and construct large-sample confidence intervals for the average treatment effect, which extends Li, Ding and Rubin (2018) that requires a fixed positive threshold and a fixed number of covariates for rerandomization. Moreover, we investigate whether rerandomization can achieve the ideally optimal precision. Specifically, we demonstrate that, when the number of covariates satisfies certain conditions (generally being a smaller order of the logarithm of the sample size), we can diminish the covariate imbalance threshold such that the corresponding acceptance probability converges to zero at a proper rate and the resulting difference-in-means estimator achieves the ideally optimal precision and becomes asymptotically Gaussian distributed, under which we can use the usual Wald-type confidence intervals. Note that this does not contradicts with the general asymptotic non-Gaussianity for rerandomization established in Li, Ding and Rubin (2018); it is because the non-Gaussian part in the limiting distribution can be asymptotically ignorable when we diminish the acceptance probability as the sample size increases.

The paper proceeds as follows. Section 2 introduces the framework and reviews existing results. Section 3 studies the multivariate Berry–Esseen-type bound for the finite population central limit theorem under complete randomization, which serves as the basis for studying the asymptotic properties of rerandomization in Section 4. Section 5 studies whether rerandomization can achieve the ideally optimal precision that we can expect with perfectly balanced covariates. Section 6 constructs large-sample confidence intervals for the average treatment effect under rerandomization. Section 7 investigates all the involved regularity conditions and discusses their practical implications, including both the covariate trimming and computational cost. Section 8 concludes with a short discussion.

2. Framework, notation and literature review.

2.1. *Potential outcomes and treatment assignment.* We consider an experiment with two treatment arms (labeled as treatment and control) and n units, among which n_1 units will be assigned to the treatment group and the remaining $n_0 = n - n_1$ units will be assigned to the control group, where n_1 and n_0 are predetermined fixed integers. We invoke the potential outcome framework (Rubin (1974), Splawa-Neyman (1990)) to define treatment effects, where each unit i has two potential outcomes $Y_i(1)$ and $Y_i(0)$ depending on its treatment assignment. The individual treatment effect for unit i is then $\tau_i = Y_i(1) - Y_i(0)$, and the corresponding average treatment effect for all units is $\tau = n^{-1}\sum_{i=1}^n \tau_i$, which is our estimand of interest. We use $\boldsymbol{X}_i \in \mathbb{R}^K$ to denote the available K-dimensional covariate vector for each unit i, and Z_i to denote the treatment assignment indicator, where $Z_i = 1$ if the unit receives treatment and $Z_i = 0$ otherwise. The observed outcome for unit i is then one of its two potential outcomes depending on its treatment assignment, that is, $Y_i = Z_i Y_i(1) + (1 - Z_i)Y_i(0)$.

Throughout the paper, we will conduct the finite population inference where all potential outcomes and covariates are viewed as fixed constants and the randomness in the observed data (e.g., Y_i's) comes solely from the random treatment assignments Z_i's. This is equivalent to conditional inference conditioning on all potential outcomes and covariates; see Li and Ding (2017) for a review of finite population inference with emphasis on applications to causal inference. The finite population inference has the advantage of imposing no model or distributional assumptions on the potential outcomes or covariates. Consequently, the distribution of the treatment assignments for all units, namely the treatment assignment mechanism, plays a crucial role for statistical inference. In a completely randomized experiment (CRE), the probability that the treatment assignment vector $\boldsymbol{Z} \equiv (Z_1, Z_2, \ldots, Z_n)^\top$ takes a particular value $\boldsymbol{z} \equiv (z_1, z_2, \ldots, z_n)^\top$ is $n_1!n_0!/n!$ if $z_i \in \{0, 1\}$ for all i and $\sum_{i=1}^n z_i = n_1$, and zero otherwise.

For descriptive convenience, we introduce several finite population quantities. For $z = 0, 1$, let $\bar{Y}(z)$ and $\bar{\boldsymbol{X}}$ be the finite population averages of potential outcome and covariates, $S_z^2 = (n-1)^{-1}\sum_{i=1}^n \{Y_i(z) - \bar{Y}(z)\}^2$, $\boldsymbol{S}_{\boldsymbol{X}}^2 = (n-1)^{-1}\sum_{i=1}^n (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^\top$ and $\boldsymbol{S}_{z,\boldsymbol{X}} = \boldsymbol{S}_{\boldsymbol{X},z}^\top = (n-1)^{-1}\sum_{i=1}^n \{Y_i(z) - \bar{Y}(z)\}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^\top$ be the finite population variance and covariances for potential outcomes and covariates. For the individual treatment effect, we define analogously its finite population variance $S_\tau^2 = (n-1)^{-1}\sum_{i=1}^n (\tau_i - \tau)^2$ and its finite population covariance with covariates $\boldsymbol{S}_{\tau,\boldsymbol{X}} = \boldsymbol{S}_{\boldsymbol{X},\tau}^\top = (n-1)^{-1}\sum_{i=1}^n (\tau_i - \tau)(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^\top$.

2.2. *Covariate imbalance and rerandomization.* Under the CRE, the units are completely randomized into the two treatment arms, which guarantees that all pretreatment covariates, no matter observed or unobserved, are balanced *on average* between the two treatment groups. However, as pointed out by Morgan and Rubin (2012), the covariate imbalance is likely to occur for a realized treatment assignment. The classical literature in experimental design (see, e.g., Box, Hunter and Hunter (2005)) suggests blocking on pretreatment covariates which,

however, is not obvious to implement when the covariates are continuous. Recently, Morgan and Rubin (2012) formally proposed a design called rerandomization to actively avoid the unlucky covariate imbalance, by discarding those treatment assignments with unacceptable covariate imbalance. A general rerandomization design consists of the following steps:

1. Collect the covariate data for the experimental units, and specify a covariate balance criterion.
2. Randomly assign n_1 units to treatment group and the remaining n_0 units to control group.
3. Check the covariate balance for the treatment assignment from Step 2. If the balance criterion is satisfied, proceed to Step 4; otherwise, return to Step 2.
4. Conduct the experiment using the acceptable treatment assignment from Step 3.

The balance criterion in Step 1 is an accept–reject function of the treatment assignment vector $\boldsymbol{Z}$ and the pretreatment covariates $\boldsymbol{X}_i$'s. Morgan and Rubin (2012) suggested to use the Mahalanobis distance between covariate means in two treatment groups as the covariate balance criterion, which enjoys the affinely invariant property. Specifically, the difference-in-means of covariates between the two treatment groups is

$$\hat{\boldsymbol{\tau}}_{\boldsymbol{X}} \equiv \bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_0 = \frac{1}{n_1}\sum_{i=1}^{n} Z_i \boldsymbol{X}_i - \frac{1}{n_0}\sum_{i=1}^{n}(1-Z_i)\boldsymbol{X}_i, \tag{1}$$

where $\bar{\boldsymbol{X}}_z$ denotes the covariate mean for units under treatment arm z, and the corresponding Mahalanobis distance for measuring covariate imbalance is

$$\begin{aligned} M &= \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}^{\top}\mathrm{Cov}^{-1}(\hat{\boldsymbol{\tau}}_{\boldsymbol{X}})\hat{\boldsymbol{\tau}}_{\boldsymbol{X}} = \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}^{\top}\left(\frac{n}{n_1 n_0}\boldsymbol{S}_{\boldsymbol{X}}^2\right)^{-1}\hat{\boldsymbol{\tau}}_{\boldsymbol{X}} \\ &= \frac{n_1 n_0}{n}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_0)^{\top}(\boldsymbol{S}_{\boldsymbol{X}}^2)^{-1}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_0). \end{aligned}$$

Under rerandomization using the Mahalanobis distance (ReM) with a predetermined threshold a, a treatment assignment $\boldsymbol{Z}$ is acceptable if and only if the corresponding Mahalanobis distance is less than or equal to the threshold a, that is, $M \le a$. Throughout the paper, we will focus on ReM to illustrate our theory. Our results can be generalized to other covariate balance criteria as well.

2.3. *Recent results and challenges.* Rerandomization has a long history and has been utilized a lot in practice, although often implicitly. A formal proposition of rerandomization does not appear until Morgan and Rubin (2012), likely due to the critique that the classical Gaussian distribution theory is no longer valid for rerandomization; see Morgan and Rubin (2012) and references therein. Recently, Li, Ding and Rubin (2018) demonstrated that the usual difference-in-means estimator,

$$\hat{\tau} \equiv \bar{Y}_1 - \bar{Y}_0 = \frac{1}{n_1}\sum_{i=1}^{n} Z_i Y_i - \frac{1}{n_0}\sum_{i=1}^{n}(1-Z_i)Y_i, \tag{2}$$

is indeed asymptotically non-Gaussian distributed, where $\bar{Y}_z$ denotes the average observed outcome for units under treatment arm z. Specifically, they proved that, with a fixed positive threshold a and a fixed number of covariates K that do not vary with the sample size n, the asymptotic distribution for $\hat{\tau}$ under ReM has the following form:

$$\sqrt{n}(\hat{\tau} - \tau) \mid M \le a \dot{\sim} \sqrt{nV_{\tau\tau}}\left(\sqrt{1-R^2}\cdot\varepsilon + R\cdot L_{K,a}\right), \tag{3}$$

where $\dot{\sim}$ means that the distributions on both sides of (3) converge weakly to the same distribution. In (3), $\varepsilon \sim \mathcal{N}(0, 1)$ follows a standard Gaussian distribution, $L_{K,a} \sim D_1 \mid \boldsymbol{D}^\top \boldsymbol{D} \leq a$ follows a constrained Gaussian distribution with $\boldsymbol{D} = (D_1, D_2, \ldots, D_K)^\top \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_K)$, and ε and $L_{K,a}$ are mutually independent. Besides, $V_{\tau\tau}$ is the variance of $\hat{\tau}$ under the CRE, R^2 is the squared multiple correlation between $\hat{\tau}$ and $\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$ under the CRE, and we defer the explicit expressions for $V_{\tau\tau}$ and R^2 to Section 3.1. From (3), the difference-in-means estimator under ReM is asymptotically distributed as the convolution of a Gaussian and a constrained Gaussian random variables. Intuitively, the ε component represents the part of $\hat{\tau}$ that cannot be explained by $\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$, and the $L_{K,a}$ component represents the part that can be explained by $\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$, and thus it depends on both the threshold a for balance criterion and the number K of involved covariates.

The asymptotic derivation for (3) requires that both the threshold a and number of covariates K for the Mahalanobis distance criterion are fixed and do not change as the sample size n increases. However, both requirements are likely to be violated in practice. We first consider the choice of threshold for rerandomization. Generally, smaller threshold can provide us better covariate balance and more precise treatment effect estimator as indicated by (3); see Li, Ding and Rubin ((2018), Theorem 2). Therefore, researchers (Kasy (2016), Kallus (2018)) have suggested to use as small threshold as possible, say, the minimum Mahalanobis distance between covariate means in the two treatment groups. However, as argued by Morgan and Rubin (2012) and Li, Ding and Rubin (2018), too small threshold can lead to powerless randomization tests and inaccurate asymptotic approximations. For example, with general continuous covariates and using the minimum Mahalanobis distance as the covariate imbalance threshold, very likely there is only one (or two when $n_1 = n_0$) acceptable treatment assignment, and the corresponding minimum p-value that we can get from randomization tests is 1 (or 0.5 when $n_1 = n_0$), indicating no power to reject any hypothesis at a reasonable significance level. Besides, the resulting difference-in-means estimator is either deterministic or having only two possible values, under which it is impossible for the estimator to converge to any continuous distribution, and thus the asymptotic approximation derived in (3) no longer holds. Based on these observations, Li, Ding and Rubin (2018) suggested to use small, but not overly small threshold, for conducting rerandomization, which not only provides better covariate balance but also allows large-sample valid inference for the average treatment effect that bases only on the randomization of the treatment assignments.

Nevertheless, there is still a theoretical gap for the choice of the rerandomization threshold. The existing study assumes a fixed threshold a that does not vary with the sample size n. It is then natural to ask: can we decrease the threshold with the sample size such that the difference-in-means estimator under ReM converges weakly to a Gaussian distribution as the right-hand side of (3) with $a = 0$, the ideally optimal precision we expect when the covariates are perfectly balanced? This essentially requires a theoretical understanding of rerandomization when the threshold a (or the acceptance probability) varies and especially converges to zero as the sample size goes to infinity.

We then consider the number of covariates for rerandomization. With the rapidly growing ability for collecting data, it is common to have a large number of covariates for the experimental units. For example, Bloniarz et al. (2016), Wager et al. (2016) and Lei and Ding (2021) studied regression adjustment for the CRE in the analysis stage when the experiments were completed. However, only a few studies have considered a large number of covariates in the design stage of an experiment; two examples are Branson and Shao (2021) and Zhang, Yin and Rubin (2021) where the authors proposed ridge and PCA rerandomizations to deal with collinearity among covariates, an issue that becomes increasingly serious as the number of covariates increases with the sample size. There is even fewer studies on the theoretical property of rerandomization when the amount of covariate information increases as the sample

size grows. Note that practitioner often tends to balance as many covariates as possible with the hope to get more precise estimator. This is also hinted by the previous asymptotic result (3) in which the asymptotic distribution becomes more concentrated around zero as R^2 (a measure for the association between covariates and potential outcomes) increases. Therefore, it is important to establish a theory for rerandomization allowing diverging number of covariates, which can also provide guidelines on how to choose covariates for rerandomization in practice.

3. A multivariate Berry–Esseen-type bound for the finite population central limit theorem.

3.1. *Motivation and finite population central limit theorem for a fixed dimension.* The key for deriving the asymptotic property of ReM in (3) includes the following facts. First, the distribution of the difference-in-means estimator under ReM is essentially the same as its conditional distribution under the CRE given that the treatment assignment satisfies the Mahalanobis distance criterion, as indicated by the left-hand side of (3). This then motivates us to investigate the joint distribution of the difference-in-means of the outcome and covariates in (1) and (2) under the CRE. Second, by the finite population central limit theorem (Hájek (1960), Li and Ding (2017)), the joint distribution of $(\hat{\tau}, \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}^\top)^\top$ under the CRE is asymptotically Gaussian with mean and covariance matrix the same as its sampling mean and covariance matrix under the CRE: $\mathbb{E}(\hat{\tau}, \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}^\top)^\top = (\tau, \mathbf{0}_K^\top)^\top$ and

$$\begin{aligned}\operatorname{Cov}\begin{pmatrix}\hat{\tau}\\ \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}\end{pmatrix} &= \begin{pmatrix} n_1^{-1}S_1^2 + n_0^{-1}S_0^2 - n^{-1}S_\tau^2 & n_1^{-1}\boldsymbol{S}_{1,\boldsymbol{X}} + n_0^{-1}\boldsymbol{S}_{0,\boldsymbol{X}} \\ n_1^{-1}\boldsymbol{S}_{\boldsymbol{X},1} + n_0^{-1}\boldsymbol{S}_{\boldsymbol{X},0} & n/(n_1 n_0)\cdot \boldsymbol{S}_{\boldsymbol{X}}^2 \end{pmatrix} \\ &\equiv \boldsymbol{V} \equiv \begin{pmatrix} V_{\tau\tau} & \boldsymbol{V}_{\tau\boldsymbol{x}} \\ \boldsymbol{V}_{\boldsymbol{x}\tau} & \boldsymbol{V}_{\boldsymbol{x}\boldsymbol{x}} \end{pmatrix},\end{aligned} \tag{4}$$

where we introduce $\boldsymbol{V}$ to denote the covariance matrix of $(\hat{\tau}, \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}^\top)^\top$ under the CRE and $V_{\tau\tau}$ to denote the variance of $\hat{\tau}$ used in (3). Specifically, $\sqrt{n}(\hat{\tau} - \tau, \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}^\top)^\top \dot{\sim} \mathcal{N}(\mathbf{0}_{K+1}, n\boldsymbol{V})$, recalling that $\dot{\sim}$ means that the two distributions have the same weak limits. Based on these observations, Li, Ding and Rubin (2018) demonstrated that the asymptotic distribution of the difference-in-means estimator under ReM is essentially a conditional distribution from a multivariate Gaussian distribution, which simplifies to (3) and depends crucially on the squared multiple correlation between potential outcomes and covariates (or more precisely between $\hat{\tau}$ and $\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$ under the CRE):

$$R^2 = \operatorname{Corr}^2(\hat{\tau}, \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}) = \frac{\boldsymbol{V}_{\tau\boldsymbol{x}}\boldsymbol{V}_{\boldsymbol{x}\boldsymbol{x}}^{-1}\boldsymbol{V}_{\boldsymbol{x}\tau}}{V_{\tau\tau}} = \frac{n_1^{-1}S_{1|\boldsymbol{X}}^2 + n_0^{-1}S_{0|\boldsymbol{X}}^2 - n^{-1}S_{\tau|\boldsymbol{X}}^2}{n_1^{-1}S_1^2 + n_0^{-1}S_0^2 - n^{-1}S_\tau^2}, \tag{5}$$

where $S_{z|\boldsymbol{X}}^2 = \boldsymbol{S}_{z,\boldsymbol{X}}\boldsymbol{S}_{\boldsymbol{X}}^{-2}\boldsymbol{S}_{\boldsymbol{X},z}$ and $S_{\tau|\boldsymbol{X}}^2 = \boldsymbol{S}_{\tau,\boldsymbol{X}}\boldsymbol{S}_{\boldsymbol{X}}^{-2}\boldsymbol{S}_{\boldsymbol{X},\tau}$ are the finite population variances of the linear projections of potential outcomes $Y_i(z)$'s and individual effects τ_i's on the covariates $\boldsymbol{X}_i$'s.

Apparently, the above arguments require a fixed number of covariates K. Moreover, the weak convergence from the joint distribution to the conditional distribution requires that the probability of the conditioning event $\mathbb{P}(M \le a)$ has a positive limit, which implies a positive and nondiminishing threshold a for the Mahalanobis distance criterion. Otherwise, the original derivation in Li, Ding and Rubin (2018) will involve ratios between terms of order $o(1)$, for example, $\mathbb{P}\{\sqrt{n}(\hat{\tau} - \tau) \le c \mid M \le a\} = \mathbb{P}\{\sqrt{n}(\hat{\tau} - \tau) \le c, M \le a\}/\mathbb{P}(M \le a)$, of which the limits are unclear.

From the above discussion, it is obvious that the original form of finite population central limit theorem is not enough for studying the asymptotic property of rerandomization with a

diminishing threshold and a diverging number of covariates. Furthermore, it lefts the question that whether rerandomization with threshold or acceptance probability diminishing at a certain rate can lead to difference-in-means estimator with the ideally optimal precision. We will address these concerns in the remaining of the paper.

3.2. *Gaussian approximation under the completely randomized experiment.* We first study the convergence rate for the finite population central limit theorem under the CRE. More precisely, we will focus on the convergence rate for the Gaussian approximation of the joint distribution of the difference-in-means of the outcome and covariates under the CRE, and investigate explicitly how the convergence rate depends on the finite population including the dimension of the covariates.

Let $r_1 = n_1/n$ and $r_0 = n_0/n$ be the proportions of treated and control units, and for each unit $1 \le i \le n$, let $\boldsymbol{u}_i = (r_0 Y_i(1) + r_1 Y_i(0), \boldsymbol{X}_i^\top)^\top \in \mathbb{R}^{K+1}$ be a vector consisting of a weighted average of the two potential outcomes and the covariates. By the definitions in (1) and (2), we can verify that the difference-in-means vector $(\hat{\tau}, \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}^\top)^\top$ has the following equivalent form:

$$\begin{pmatrix} \hat{\tau} \\ \hat{\boldsymbol{\tau}}_{\boldsymbol{X}} \end{pmatrix} = \frac{n}{n_1 n_0} \sum_{i=1}^{n} Z_i \boldsymbol{u}_i - \frac{n}{n_0} \begin{pmatrix} \bar{Y}(0) \\ \bar{\boldsymbol{X}} \end{pmatrix}, \tag{6}$$

which, up to a linear transformation, is essentially the summation of a simple random sample of size n_1 from the finite population $\mathcal{U}_n \equiv \{\boldsymbol{u}_i : i = 1, 2, \ldots, n\}$. Thus, the sampling property of the difference-in-means $(\hat{\tau}, \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}^\top)^\top$ can be fully characterized by the population $\mathcal{U}_n$. Let $\bar{\boldsymbol{u}} = n^{-1} \sum_{i=1}^n \boldsymbol{u}_i$ and $\boldsymbol{S}_{\boldsymbol{u}}^2 = (n-1)^{-1} \sum_{i=1}^n (\boldsymbol{u}_i - \bar{\boldsymbol{u}})(\boldsymbol{u}_i - \bar{\boldsymbol{u}})^\top$ be the finite population mean and covariance matrix for $\mathcal{U}_n$, and let $\boldsymbol{S}_{\boldsymbol{u}}^{-1}$ denote the inverse of the positive semidefinite square root of $\boldsymbol{S}_{\boldsymbol{u}}^2$. Define

$$\gamma_n \equiv \frac{(K+1)^{1/4}}{\sqrt{n r_1 r_0}} \frac{1}{n} \sum_{i=1}^{n} \left\| \boldsymbol{S}_{\boldsymbol{u}}^{-1} (\boldsymbol{u}_i - \bar{\boldsymbol{u}}) \right\|_2^3, \tag{7}$$

which is the third moment of the standardized finite population $\{\boldsymbol{S}_{\boldsymbol{u}}^{-1}(\boldsymbol{u}_i - \bar{\boldsymbol{u}}) : i = 1, 2, \ldots, n\}$ up to a certain scale. For descriptive convenience, we define γ_n to be infinity when r_1 or r_0 equals zero or $\boldsymbol{S}_{\boldsymbol{u}}^2$ is singular. In (7), we use the subscript n to emphasize the dependence of γ_n on the finite population $\mathcal{U}_n$ of size n. Note that γ_n is uniquely determined by r_1, r_0 and the potential outcomes and covariates of the n experimental units.

Below we consider the Berry–Esseen-type bound for the Gaussian approximation of the difference-in-means vector in (6) under the CRE. Note that, under the CRE, $(\hat{\tau}, \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}^\top)^\top$ has mean $(\tau, \boldsymbol{0}_{1\times K})^\top$ and covariance matrix $\boldsymbol{V}$ as in (4). Let $\mathcal{C}_{K+1}$ denote the collection of all measurable convex sets in $\mathbb{R}^{K+1}$. We then focus on bounding the supremum of the absolute difference between the probabilities of being in any measurable convex set for the standardized difference-in-means vector and the standard Gaussian random vector:

$$\Delta_n \equiv \sup_{\mathcal{Q} \in \mathcal{C}_{K+1}} \left| \mathbb{P}\left\{ \boldsymbol{V}^{-1/2} \begin{pmatrix} \hat{\tau} - \tau \\ \hat{\boldsymbol{\tau}}_{\boldsymbol{X}} \end{pmatrix} \in \mathcal{Q} \right\} - \mathbb{P}(\boldsymbol{\varepsilon} \in \mathcal{Q}) \right|. \tag{8}$$

By some algebra, $\boldsymbol{V} = (n r_1 r_0)^{-1} \boldsymbol{S}_{\boldsymbol{u}}^2$, and thus Δ_n is well defined as long as $\gamma_n < \infty$. For descriptive convenience, we define Δ_n to be 1 when r_1 or r_0 equals zero or $\boldsymbol{S}_{\boldsymbol{u}}^2$ is singular. The bound for (8) is a natural multivariate extension of the classical univariate Berry–Esseen bound for the absolute difference between two distribution functions. More importantly, it suffices for our asymptotic analysis of rerandomization, noticing that the acceptance region for $\hat{\boldsymbol{\tau}}_{\boldsymbol{X}}$ under the Mahalanobis distance criterion is indeed a convex set in $\mathbb{R}^K$. From (6),

we essentially need to understand the Berry–Esseen-type bound for the central limit theorem under simple random sampling, which itself is also a special case of the combinatorial central limit theorem. Below we give a brief literature review.

Berry (1941) and Esseen (1942) independently discovered the original Berry–Esseen theorem when studying the convergence rate for Gaussian approximation of summations of independent univariate random variables. Bentkus (2003, 2004) and Raič (2019) then extended it to the multivariate case, considering the Gaussian approximation for probabilities of being in any measurable convex sets. Recently, Chernozhukov, Chetverikov and Kato (2017), Chernozhuokov, Chetverikov and Koike (2020) and Fang and Koike (2021) achieved tighter bounds by focusing only on Gaussian approximation for probabilities of being in hyperrectangles (or more generally sparsely convex sets), where the bounds can vanish even when the dimension of random vectors is much larger than the sample size in the summation. Note that all of these results are for independent summands.

In the context of combinatorial central limit theorem (including the central limit theorem for simple random sampling as a special case), the summands become weakly dependent. Bikelis (1969) and Höglund (1978) studied the corresponding Berry–Esseen-type bound in the univariate case. However, there has been much less study for the multivariate case, in contrast to the rich literature for independent summands. One exception is Bolthausen and Götze (1993), who established the Berry–Esseen-type bound for the multivariate combinatorial central limit theorem. Based on their results, we can show that there exists an absolute constant C_K that depends only on the dimension K such that $\Delta_n \le C_K \gamma_n$, with γ_n and Δ_n defined in (7) and (8). However, the authors did not characterize how the constant C_K may increase with the dimension K, and thus the bound is not sufficient for studying rerandomization with diverging number of covariates. To the best of our knowledge, there has not been any formal result of the Berry–Esseen-type bound for the combinatorial central limit theorem with explicit dependence on the dimension, except for an informal result presented by Raič (2015) at a workshop. Based on the result in Raič (2015), we can show that there exists an absolute constant C such that $\Delta_n \le C\gamma_n$, noting that the definition of γ_n in (7) involves a term of $(K+1)^{1/4}$ that depends explicitly on the dimension of the difference-in-means vector in (6).

Since the result in Raič (2015) has not been proved yet, we also derive a Berry–Esseen-type bound for the central limit theorem under simple random sampling ourselves. Our proof makes use of the multivariate Berry–Esseen-type bound for sum of independent random vectors (see, e.g., Raič (2019)) and the coupling between simple random sampling and Bernoulli independent sampling utilized by Hájek (1960). Based on our proof, we can derive that $\Delta_n \le 174\gamma_n + 7\gamma_n^{1/3}$, where the first term of γ_n is from the Bernoulli independent sampling or more generally the Berry–Esseen-type bound for sum of independent random vectors, and the additional term of $\gamma_n^{1/3}$ comes from the coupling between simple random sampling and Bernoulli independent sampling. There is actually a tighter bound than $\gamma_n^{1/3}$ for the coupling, but we present the bound $\gamma_n^{1/3}$ for the ease of understanding; see the Supplementary Material (Wang and Li (2022)) for more details. Obviously, our rate of convergence is slower than that conjectured in Raič (2015). Nevertheless, it is still able to reveal the interesting property of rerandomization with diminishing threshold for covariate imbalance and diverging number of covariates, as studied in detail shortly. We summarize these results for bounding Δ_n in the following theorem.

THEOREM 1. *For any $n \ge 2$, $K \ge 0$, $r_1, r_0 \in (0,1)$, and any finite population $\Pi_n \equiv \{(Y_i(1), Y_i(0), \boldsymbol{X}_i) : i = 1, 2, \ldots, n\}$ with nonsingular $\boldsymbol{V}$ defined as in (4), define γ_n and Δ_n as in (7) and (8). Then*:

(i) *there exists an absolute constant* C_K *that depends only on* K *such that* $\Delta_n \le C_K \gamma_n$;
(ii) *if the conjecture in Raič (2015) hold, then there exists a universal constant* C *such that* $\Delta_n \le C\gamma_n$;
(iii) $\Delta_n \le 174\gamma_n + 7\gamma_n^{1/3}$.

3.3. *Gaussian approximation with stronger moment conditions*. Theorem 1 provides Berry–Esseen-type bounds for the finite population central limit theorem under complete randomization, with (ii) and (iii) characterizing explicit dependence on the dimension of covariates K, and thus crucial for studying rerandomization with diverging number of covariates. Compared to that conjectured by Raič (2015), our derived bound in Theorem 1(iii) has an additional term of order $\gamma_n^{1/3}$. As discussed before, this additional term is due to the coupling between simple random sampling and Bernoulli independent sampling. Intuitively, under simple random sampling (or equivalently complete randomization), the treatment indicators for all units are dependent, because the total number of units assigned to the active treatment is constrained to be n_1. Such a dependence among these indicators makes it more challenging to bound the error for Gaussian approximation. Ignoring the dependence on K, γ_n is of order $n^{-1/2}$, making $\gamma_n^{1/3}$ of order $n^{-1/6}$ and thus the bound in Theorem 1(iii) larger than usual Berry–Essen-type bounds.

Below we also provide more accurate bounds for the coupling between simple random sampling and Bernoulli independent sampling and consequently improve the Berry–Esseen-type bound for the Gaussian approximation in Theorem 1(iii), at least in terms of the explicit dependence on the sample size n, with stronger moment conditions on the centered finite population $\{\boldsymbol{S}_{\boldsymbol{u}}^{-1}(\boldsymbol{u}_i - \bar{\boldsymbol{u}}) : 1 \le i \le n\}$. We summarize the results in the following theorem.

THEOREM 2. *Under the same setting as Theorem 1,*

$$\Delta_n \le 180\gamma_n + \frac{3(\log n)^{3/4}(K+1)^{3/4}}{n^{1/4}\sqrt{r_1 r_0}} \cdot \max_{1\le i\le n} \left\|\boldsymbol{S}_{\boldsymbol{u}}^{-1}(\boldsymbol{u}_i - \bar{\boldsymbol{u}})\right\|_\infty,$$

and, for any $\iota \ge 2$, *there exists a universal constant* C_ι *depending only on* ι *such that*

$$\Delta_n \le 174\gamma_n + \frac{C_\iota (K+1)^{3\iota/\{4(\iota+1)\}}}{n^{\iota/\{4(\iota+1)\}}\{r_1 r_0\}^{\iota/2}} \cdot \frac{1}{n}\sum_{i=1}^n \left\|\boldsymbol{S}_{\boldsymbol{u}}^{-1}(\boldsymbol{u}_i - \bar{\boldsymbol{u}})\right\|_\iota^\iota.$$

From Theorem 2 and ignoring the dependence on K, if all coordinates of the centered finite population have bounded ιth moments for some $\iota \ge 2$, then the additional term in the Berry–Esseen-type bound is of order $n^{-\iota/\{4(\iota+1)\}}$; if further they are bounded, then the additional term becomes of order $n^{-1/4}$ except for a $\log n$ term. These bounds are more accurate than that of order $n^{-1/6}$ in Theorem 1(iii), but there are still gaps between them and that of order $n^{-1/2}$ conjectured by Raič (2015), which requires future work. Nevertheless, the derived bounds are already sufficient to discover interesting properties of rerandomization and provide almost the same quantitative message for the design of rerandomization (see, e.g., Table 1 and its discussion).

It is worth mentioning that there are other approaches for deriving Berry–Esseen-type bound, such as Stein's method that was actually used by Bolthausen and Götze (1993) and helps justify Theorem 1(i); see also the recent work by Shi and Ding (2022). Here, we use the coupling approach, mainly because we can utilize the recent results on Berry–Esseen-type bounds for Gaussian approximations of summations of independent random vectors (Bentkus (2004), Raič (2019)). It will be interesting to investigate whether other approaches can give tighter Berry–Esseen-type bounds or even prove the conjectured rate in Raič (2015). In addition, the coupling approach also justifies central limit theorems for stratified randomized

experiments (Bickel and Freedman (1984), Liu, Ren and Yang (2022)), and our results can be useful for deriving the corresponding Berry–Esseen-type bounds. We leave these for future work.

4. Asymptotic properties of rerandomization with sample-size dependent Mahalanobis distance criterion. Throughout the paper, we conduct finite population asymptotic analysis for rerandomization. Specifically, we embed the finite population of size n into a sequence of finite populations with increasing sizes. Importantly, we allow both the threshold a and dimension of covariates K for the Mahalanobis distance criterion to depend on the sample size n, and will write them explicitly as a_n and K_n, using the subscript n to emphasize such dependence. We further define $p_n \equiv \mathbb{P}(\chi^2_{K_n} \leq a_n)$, where $\chi^2_{K_n}$ denotes a random variable following the chi-square distribution with degrees of freedom K_n. By the definition of Δ_n in (8), we can derive that the acceptance probability of a completely randomized treatment assignment under ReM, $\mathbb{P}(M \leq a_n)$, is bounded between $p_n - \Delta_n$ and $p_n + \Delta_n$. Thus, we can intuitively understand p_n as the approximate acceptance probability for rerandomization; specifically, $\mathbb{P}(M \leq a_n)/p_n = 1 + o(1)$ when $p_n \gg \Delta_n$. In practice, given the number of covariates K_n, the choice of the threshold a_n is often based on the approximate acceptance probability p_n, that is, a_n is the p_nth quantile of the chi-square distribution with degrees of freedom K_n. For example, Morgan and Rubin (2012) and Li, Ding and Rubin (2018) suggested to choose small but not overly small approximate acceptance probablity, for example, $p_n = 0.001$. Therefore, in the remaining discussion, we will mainly focus on the approximate acceptance probability p_n and number of covariates K_n, since they are more relevant for the practical implementation of ReM, and view the threshold a_n as a deterministic function of p_n and K_n. For descriptive convenience, we sometimes call p_n simply as the acceptance probability, while emphasizing $\mathbb{P}(M \leq a_n)$ as the actual acceptance probability.

4.1. *Asymptotic distribution under ReM.* We first invoke the following regularity condition on the sequence of finite populations, which, by Theorem 1, implies the Gaussian approximation for the difference-in-means of the outcome and covariates under the CRE.

CONDITION 1. As $n \to \infty$, the sequence of finite populations satisfies that $\gamma_n \to 0$.

Recall the definition of γ_n in (7). Condition 1 requires that, for sufficiently large sample size n, there are positive proportions of units in both treatment and control groups (i.e., $r_1 > 0$ and $r_0 > 0$), and the covariance matrix $\boldsymbol{V}$ in (4) for the difference-in-means vector $(\hat{\tau}, \hat{\boldsymbol{\tau}}_{\boldsymbol{X}}^\top)^\top$ is nonsingular. The latter essentially requires that the covariates are not collinear, which can be guaranteed by our design, and that the potential outcomes cannot be perfectly explained by covariates, in the sense that the corresponding R^2 in (5) is strictly less than 1, which is likely to hold in most applications. Besides, as demonstrated in the Supplementary Material (Wang and Li (2022)),

$$\gamma_n \geq 2^{-3/2}(nr_1r_0)^{-1/2}(K_n + 1)^{7/4}. \tag{9}$$

Thus, a necessary condition for Condition 1 is $K_n = o((nr_1r_0)^{2/7}) = o(n^{2/7})$, that is, the number of covariates increases at most a polynomial rate of the sample size as n goes to infinity. If the standardized finite population $\{\boldsymbol{S}_{\boldsymbol{u}}^{-1}(\boldsymbol{u}_i - \bar{\boldsymbol{u}}) : 1 \leq i \leq n\}$ is coordinatewise bounded, and the proportions of treated and control units r_1 and r_0 are bounded away from zero, then $K_n = o(n^{2/7})$ is also sufficient for Condition 1; see the Supplementary Material (Wang and Li (2022)) for details. We defer more detailed discussions about Condition 1 to Section 7.

We then invoke the following regularity condition on the choice of the acceptable probability, which is coherent with our intuition that too small threshold can prevent asymptotic approximation for ReM based on Gaussian and constrained Gaussian distributions, as discussed in Section 2.3.

CONDITION 2. As $n \to \infty$, $p_n/\Delta_n \to \infty$.

We are now ready to present our formal result for the asymptotic distribution of the difference-in-means estimator $\hat{\tau}$ under ReM. Recall that $V_{\tau\tau}$ in (4) is the variance of $\hat{\tau}$ under the CRE, R^2 in (5) is the squared multiple correlation between the difference-in-means of the outcome and covariates under the CRE, and ε_0 and L_{K_n,a_n} are independent standard Gaussian and constrained Gaussian random variables defined as in Section 2.3. To emphasize its dependence on the sample size n, we will write R^2 explicitly as R_n^2.

THEOREM 3. *Under ReM and Conditions 1 and 2, as* $n \to \infty$,

$$\sup_{c\in\mathbb{R}} \left|\mathbb{P}\left\{V_{\tau\tau}^{-1/2}(\hat{\tau}-\tau) \le c \mid M \le a_n\right\} - \mathbb{P}\left(\sqrt{1-R_n^2}\varepsilon_0 + \sqrt{R_n^2}L_{K_n,a_n} \le c\right)\right| \to 0. \tag{10}$$

From Theorem 3, under ReM, the difference between the difference-in-means estimator and the true average treatment effect, $\hat{\tau}-\tau$, follows asymptotically the distribution of $V_{\tau\tau}^{1/2}(\sqrt{1-R_n^2}\varepsilon_0 + \sqrt{R_n^2}L_{K_n,a_n})$, a convolution of standard Gaussian and constrained Gaussian random variables, with coefficients depending on $V_{\tau\tau}$ and R_n^2. Compared to (3), the asymptotic distribution in Theorem 3 has the same form as that in Li, Ding and Rubin ((2018), Theorem 1), which is not surprising given that both theorems focus on the same estimator $\hat{\tau}$ and the same design ReM. However, Theorem 3 is more general and it covers Li, Ding and Rubin ((2018), Theorem 1) as a special case. Specifically, when both K_n and $a_n > 0$ are fixed and do not change with n, p_n is a fixed positive constant and Condition 2 holds immediately from Condition 1 and Theorem 1. Besides, Condition 1 is almost implied by the regularity condition involved in Li, Ding and Rubin (2018); see the Supplementary Material (Wang and Li (2022)) for more details.

More importantly, Theorem 3 allows the dimension of covariates and the acceptance probability, as well as the rerandomization threshold, to vary with the sample size. Note that from Theorem 1, $\gamma_n \to 0$ implies $\Delta_n \to 0$. Thus, Condition 2 holds naturally if we choose p_n to be any fixed positive number. Moreover, we can also let p_n decrease with n and eventually converge to zero as $n \to \infty$, while maintaining that $p_n \gg \Delta_n$, under which the actual acceptance probability $\mathbb{P}(M \le a_n) = p_n(1+o(1))$ also converges to zero as $n \to \infty$. In simple words, Theorem 3 allows the acceptance probability to converge to zero as the sample size goes to infinity.

Note that all potential confounding factors, no matter observed or unobserved, can always be viewed as potential outcomes unaffected by the treatment. Therefore, Theorem 3 also implies that any potential confounding factor is asymptotically balanced between the two treatment groups.

4.2. *Asymptotic improvement from ReM.* We now investigate the improvement from ReM compared to the CRE, and in particular how such improvement depends on the acceptance probability and the covariate information. Note that the CRE can be viewed as a special case of ReM with $\boldsymbol{X} = \varnothing$ and $a_n = \infty$. By the same logic as Theorem 3, we can derive that

$$\sup_{c\in\mathbb{R}} \left|\mathbb{P}\left\{V_{\tau\tau}^{-1/2}(\hat{\tau}-\tau) \le c\right\} - \mathbb{P}(\varepsilon_0 \le c)\right| \to 0, \tag{11}$$

that is, $\hat{\tau} - \tau$ is asymptotically Gaussian distributed with mean zero and variance $V_{\tau\tau}$ under the CRE. Obviously, the asymptotic distribution in (11) is a special form of that in (10) with $R_n^2 = 0$, which is intuitive in the sense that ReM with irrelevant covariates is asymptotically equivalent to the CRE. However, when $R_n^2 > 0$, which is likely to hold in practice, we expect ReM to provide more precise difference-in-means estimator than the CRE, as discussed in detail below.

From Theorem 3 and by the same logic as Li, Ding and Rubin ((2018), Corollaries 1–3), we can derive the following asymptotic properties of ReM, demonstrating its advantage over the CRE. For $\alpha \in (0, 1)$, let $v_{\alpha,K,a}(R^2)$ denote the αth quantile of the distribution of $\sqrt{1 - R^2}\varepsilon_0 + \sqrt{R^2}L_{K,a}$, and z_α denote the αth quantile of the standard Gaussian distribution. We further introduce $v_{K,a} = \mathrm{Var}(L_{K,a})$ to denote the variance of the constrained Gaussian random variable. From Morgan and Rubin (2012), $v_{K,a} = \mathbb{P}(\chi^2_{K+2} \le a)/\mathbb{P}(\chi^2_K \le a)$, where χ^2_{K+2} and χ^2_K follow chi-square distributions with degrees of freedom $K + 2$ and K, respectively.

COROLLARY 1. *Under ReM and Conditions 1 and 2, the asymptotic distribution of* $V_{\tau\tau}^{-1/2}(\hat{\tau} - \tau)$, $\sqrt{1 - R_n^2}\varepsilon_0 + \sqrt{R_n^2}L_{K_n,a_n}$, *as shown in* (10) *is symmetric and unimodal around zero. Compared to the asymptotic distribution in* (11) *under the CRE, the percentage reductions in asymptotic variance and length of asymptotic* $1 - \alpha$ *symmetric quantile range for* $\alpha \in (0, 1)$ *are, respectively,*

$$(1 - v_{K_n,a_n})R_n^2 \quad \text{and} \quad 1 - \frac{v_{1-\alpha/2,K_n,a_n}(R_n^2)}{z_{1-\alpha/2}}. \tag{12}$$

Both percentage reductions in (12) *are nonnegative and are uniquely determined by* (R_n^2, p_n, K_n), *and they are nondecreasing in* R_n^2 *and nonincreasing in* p_n *and* K_n.

First, from Corollary 1, the difference-in-means estimator $\hat{\tau}$ is asymptotically unbiased for the average treatment effect τ. As pointed out by Morgan and Rubin (2012), when the treated and control groups have different sizes (i.e., $n_1 \neq n_0$), the difference-in-means estimator is generally biased. Corollary 1 shows that the bias goes away as the sample size goes to infinity when Condition 1 holds, which requires that r_1 and r_0 are not too close to zero as implied by the definition in (7). Second, the improvement from ReM on estimation precision is nondecreasing in the strength of the association between potential outcomes and covariates measured by R_n^2. Generally, the more covariates involved in rerandomization, the larger the R_n^2 will be. However, this does not mean that we should use as many covariates as possible. If the additional covariates provide little increment for R_n^2, the gain from ReM will deteriorate since the percentage reductions in (12) are nonincreasing in K_n with fixed R_n^2 and p_n. Third, both percentage reductions in (12) are nonincreasing in the acceptance probability p_n, and approach their maximum values R_n^2 and $1 - \sqrt{1 - R_n^2}$ when p_n equals 0. Again, this does not mean that we should use as small threshold as possible, since the asymptotic derivation in Corollary 1 requires that $p_n/\Delta_n \to \infty$ as $n \to \infty$. This then raises the question that if we can choose p_n such that both percentage reductions in (12) achieve their maximum values as $n \to \infty$ or the asymptotic distribution of $V_{\tau\tau}^{-1/2}(\hat{\tau} - \tau)$ under ReM becomes essentially Gaussian with mean zero and variance $1 - R_n^2$ (i.e., the asymptotic distribution in (10) with L_{K_n,a_n} replaced by zero), while still maintaining $p_n/\Delta_n \to \infty$. In other words, can we choose the acceptable probability such that rerandomization achieves its ideally optimal precision? We will answer this question in the next section.

5. Optimal rerandomization with diminishing acceptance probability. In this section, we investigate whether rerandomization can achieve its ideally optimal precision by diminishing the acceptance probability to zero at a proper rate as the sample size increases. Specifically, we wonder if the asymptotic approximation in (10) can hold with L_{K_n,a_n} replaced by zero, and what conditions we need to impose on the sequence of finite populations as well as the choice of acceptance probability. These questions rely crucially on the asymptotic behavior of the constrained Gaussian random variable L_{K_n,a_n}, and in particular its dependence on the acceptance probability p_n. Below we will first study the asymptotic property of L_{K_n,a_n}, and then investigate whether we are able to achieve the ideally optimal rerandomization.

5.1. *Asymptotic properties of the constrained Gaussian random variable.* In this subsection, we will mainly focus on the asymptotic behavior of the variance of L_{K_n,a_n}, which as mentioned earlier has the following equivalent form: $v_{K_n,a_n} \equiv \mathrm{Var}(L_{K_n,a_n}) = \mathbb{P}(\chi^2_{K_n+2} \le a_n)/\mathbb{P}(\chi^2_{K_n} \le a_n)$, due to the following two reasons. First, as $n \to \infty$, $L_{K_n,a_n} \xrightarrow{\mathbb{P}} 0$ if and only if $v_{K_n,a_n} \to 0$. This is because the random variables $L^2_{K_n,a_n}$ for all n are always uniformly integrable, regardless of how K_n and p_n vary with n, as demonstrated in the Supplementary Material (Wang and Li (2022), Proposition A2). Second, as shown in Corollary 1, the percentage reduction in asymptotic variance under ReM is $(1 - v_{K_n,a_n})R^2_n$, and its relative difference from the ideally optimal percentage reduction is $1 - (1 - v_{K_n,a_n})R^2_n/R^2_n = v_{K_n,a_n}$. Thus, the variance of L_{K_n,a_n} characterizes how different ReM is from the ideally optimal one in terms of the improvement on estimation precision, and such a difference will become asymptotically negligible if and only if $v_{K_n,a_n} \to 0$ as $n \to \infty$.

The following theorem shows the asymptotic behavior of v_{K_n,a_n}, which depends crucially on the asymptotic behavior of the ratio between $\log(p_n^{-1})$ and K_n.

THEOREM 4. *As $n \to \infty$:*

(i) *if* $\log(p_n^{-1})/K_n \to \infty$, *then* $v_{K_n,a_n} \to 0$;
(ii) *if* $\limsup_{n\to\infty} \log(p_n^{-1})/K_n < \infty$, *then* $\liminf_{n\to\infty} v_{K_n,a_n} > 0$;
(iii) *if* $\liminf_{n\to\infty} \log(p_n^{-1})/K_n > 0$, *then* $\limsup_{n\to\infty} v_{K_n,a_n} < 1$;
(iv) *if* $\log(p_n^{-1})/K_n \to 0$, *then* $v_{K_n,a_n} \to 1$.

From Theorem 4, the smaller the p_n and K_n, the larger the ratio $\log(p_n^{-1})/K_n$, and the smaller v_{K_n,a_n} tends to be. This is intuitive noting that Corollary 1 implicitly implies that v_{K_n,a_n}, viewed as a function of (p_n, K_n), is nondecreasing in p_n and K_n. Theorem 4 has the following implications. First, v_{K_n,a_n} or equivalently L_{K_n,a_n} becomes asymptotically negligible if and only if $\log(p_n^{-1})/K_n \to \infty$ as $n \to \infty$. Intuitively, this means that the constrained Gaussian term in the asymptotic distribution in (10) becomes asymptotically negligible if and only if the acceptance probability p_n decreases superexponentially with respect to the dimension of covariates, that is, $p_n = \exp(-c_n K_n)$ with $c_n \to \infty$ as $n \to \infty$. Second, if $\log(p_n^{-1})/K_n \to 0$, then the variance of the constrained Gaussian variable L_{K_n,a_n} becomes asymptotically equivalent to that of the unconstrained standard Gaussian random variable, and from Corollary 1, ReM asymptotically provides no gain compared to the CRE in the sense that the percentage reduction in asymptotic variance converges to zero as $n \to \infty$. Thus, when the acceptance probability is too large and in particular decreases subexponentially with respect to the dimension of the covariates, that is, $p_n = \exp(-o(1) \cdot K_n)$, then ReM is essentially equivalent to the CRE in large samples. Third, when the acceptance probability decreases exponentially with respect to the dimension of covariates, that is, $p_n = \exp(-c_n K_n)$ with c_n being of constant order, and assume that the squared multiple correlation R^2_n is of

constant order, and thus does not diminish to zero, asymptotically, ReM provides strictly more precise difference-in-means estimator than the CRE, although there is still a gap from the ideally optimal one.

From the above, the performance of ReM, in particular its asymptotic improvement over the CRE, depends crucially on the ratio between $\log(p_n^{-1})$ and K_n, as well as R_n^2 measuring the association between potential outcomes and covariates. To minimize v_{K_n,a_n}, Corollary 1 and Theorem 4 suggest to use as small acceptance probability and number of covariates as possible. However, there is trade-off for the choice of both of them. First, although smaller K_n implies smaller v_{K_n,a_n}, it at the same time reduces the outcome-covariates association R_n^2. Second, although smaller p_n decreases v_{K_n,a_n}, it at the same time renders the asymptotic approximation inaccurate and may eventually invalidate the asymptotic approximation in (10); see Condition 2. While the former trade-off for K_n involves more subjective judgments concerning the unknown outcome-covariates dependence structure, the latter trade-off for p_n lies more on the technical side and will be studied in more detail in the next subsection.

5.2. *Optimal rerandomization and its implication.* From Theorem 4, to achieve the ideally optimal rerandomization with given number of covariates, we want to choose the acceptance probability such that the following condition holds.

CONDITION 3. As $n \to \infty$, $\log(p_n^{-1})/K_n \to \infty$.

We further assume that the outcome-covariates association R_n^2 is bounded away from 1, under which the first term $\sqrt{1-R_n^2}$ in the asymptotic approximation in (10) is not negligible as $n \to \infty$. This condition on R_n^2 is likely to hold in practice, since we generally do not expect that the covariates can perfectly explain the potential outcomes, which is too ideal for most applications. Moreover, if R_n^2 indeed converges to 1 as $n \to \infty$, then the asymptotic approximation in (10) can be of $o_{\mathbb{P}}(1)$ itself, implying that $\hat{\tau} - \tau$ can converge to zero faster than the usual $n^{-1/2}$ rate, under which the causal effect estimation becomes much simpler.

CONDITION 4. As $n \to \infty$, $\limsup_{n\to\infty} R_n^2 < 1$.

The following theorem shows that, under certain regularity conditions, ReM can achieve its ideally optimal precision.

THEOREM 5. *Under ReM and Conditions 1–4,*

$$\sup_{c\in\mathbb{R}} \left|\mathbb{P}\{V_{\tau\tau}^{-1/2}(\hat{\tau}-\tau) \le c \mid M \le a_n\} - \mathbb{P}\left(\sqrt{1-R_n^2}\varepsilon_0 \le c\right)\right| \to 0. \tag{13}$$

From Theorem 5, $\hat{\tau}$ can be asymptotically Gaussian distributed under ReM. Moreover, its asymptotic distribution is the same as that of the regression-adjusted estimator under the CRE (Lin (2013), Lei and Ding (2021)). Therefore, rerandomization and regression adjustment are essentially dual of each other (Li and Ding (2020)), and Theorem 5 closes the previous gap between them that is due to the constrained Gaussian random variable L_{K_n,a_n}. This new insight is important for practitioners who may worry about efficiency loss of rerandomization compared to regression adjustment. Moreover, compared to regression adjustment in the analysis stage, rerandomization in the design stage is blind of outcomes and has the advantage of avoiding data snooping (Lin (2013)). Besides, the difference-in-means estimator is simpler and provides more transparent analysis for treatment effects (Cox (2007), Freedman (2008), Rosenbaum (2010)).

Theorem 5 has important implications. First, it shows that, by diminishing the imbalance threshold as the sample size grows, rerandomization can achieve its ideally optimal precision that we can expect from an optimal design with even perfectly balanced covariates. Second, we should not choose too small rerandomization threshold (or acceptance probability), as implied by Condition 2, so that it is still possible to conduct robust randomization-based inference as completely randomized experiments, without imposing any distributional assumptions on potential outcomes and covariates. These imply that rerandomization with properly diminishing covariate imbalance threshold can achieve optimal efficiency as an ideal optimal design while maintaining robustness as a randomized design, that is, such an optimal rerandomization can enjoy advantages from both optimal and randomized designs. Therefore, Theorem 5 helps reconcile the long-time controversies between the two philosophies (randomized versus optimal) for conducting experiments in an asymptotic sense.

Theorem 5 also provides important insights for practitioners that are more used to optimal designs. First, the usual optimal designs and the corresponding inference are often sensitive to their model assumptions. Our theory shows that the optimal rerandomization can always achieve the ideally optimal precision, while still being robust to model misspecification. Second, it helps mitigate the computation burden for conducting optimal designs. In particular, Theorem 5 suggests that we should not pursue the best allocation minimizing the covariate imbalance between the two treatment groups, which is generally NP-hard. Instead, we only need to randomly choose one from the best approximately p_n proportion of all assignments, with p_n satisfying both Conditions 2 and 3. As discussed later in Section 7, p_n can often decrease at a polynomial rate of the sample size n. This indicates that, in expectation, the computational complexity for getting one acceptable assignment is often of polynomial order of the sample size; see Section 7.3 for the more explicit rate. In sum, the optimal rerandomization can maintain the efficiency gain as an ideal optimal design, while being more robust and requiring less computation. In the Supplementary Material (Wang and Li (2022)), we provide more discussions on its connection with optimal designs under certain model assumptions.

The validity of the asymptotic Gaussian approximation for rerandomization depends crucially on Conditions 1–4, among which Conditions 2 and 3 involves the choice of the acceptance probability. Below we assume that the number of covariates K_n involved in rerandomization has been given and that Conditions 1 and 4 hold, and focus on investigating the existence of the choice of acceptance probability p_n such that both Conditions 2 and 3 hold, or equivalently such that rerandomization can achieve its ideally optimal precision as in (13). It turns out the existence of such a choice of p_n relies crucially on the ratio between K_n and $\log(\Delta_n^{-1})$, recalling the definition of Δ_n in (8).

THEOREM 6. *Under ReM and Conditions 1 and 4*:

(i) *if and only if* $\log(\Delta_n^{-1})/K_n \to \infty$, *there exists a sequence* $\{p_n\}$ *such that both Conditions 2 and 3 hold, under which ReM achieves its ideally optimal precision and the asymptotic Gaussian approximation in* (13) *holds*;

(iii) *if* $\limsup_{n\to\infty} \log(\Delta_n^{-1})/K_n < \infty$, *then for any sequence* $\{p_n\}$ *satisfying Condition 2 such that the asymptotic approximation in* (10) *holds*, $\liminf_{n\to\infty} v_{K_n,a_n} > 0$;

(iii) *if* $\liminf_{n\to\infty} \log(\Delta_n^{-1})/K_n > 0$, *then there exists a sequence* $\{p_n\}$ *satisfying Condition 2 such that* (10) *holds and* $\limsup_{n\to\infty} v_{K_n,a_n} < 1$;

(iv) *if* $\log(\Delta_n^{-1})/K_n \to 0$, *then for any sequence* $\{p_n\}$ *satisfying Condition 2 such that* (10) *holds, the corresponding* $v_{K_n,a_n} \to 1$ *as* $n \to \infty$, *under which ReM asymptotically provides no gain on estimation precision compared to the CRE.*

From Theorem 6, the optimal precision that ReM can achieve depends crucially on the asymptotic behavior of the ratio $\log(\Delta_n^{-1})/K_n$. For any fixed n, in general, as K_n increases,

Δ_n will increase, and thus the ratio $\log(\Delta_n^{-1})/K_n$ will decrease. Therefore, intuitively, the smaller the number of involved covariates for rerandomization, the more likely we are able to satisfy the condition in Theorem 6(i), and consequently, to achieve the ideally optimal precision. Moreover, when ReM involves more and more covariates, it will eventually lose its advantage over the CRE. Therefore, Theorem 6 suggests that we should not use too many covariates when performing rerandomization. In practice, we should try to use a moderate number of covariates that are most relevant for the potential outcomes of interest, as measured by the corresponding R_n^2. For example, when γ_n has the same order as its lower bound in (9), and r_1 and r_0 are bounded away from zero, then we can choose $K_n = O(\log n)$ number of covariates, under which $\liminf_{n\to\infty} \log(\Delta_n^{-1})/K_n$ must be positive and ReM can provide nonnegligible gain over the CRE as implied by Theorem 6(iii). As γ_n becomes further from its lower bound, we generally want to use fewer covariates. We defer more detailed discussion on the rate of $\log(\Delta_n^{-1})/K_n$ to Section 7.

6. Large-sample inference under rerandomization. Theorems 3 and 5 provide asymptotic approximations for the distribution of the difference-in-means estimator $\hat{\tau}$ under ReM, based on which we can construct large-sample confidence intervals for the average treatment effect τ. From the asymptotic approximations in (10) and (13), the asymptotic distribution of $\hat{\tau}$ under ReM depends on the variance $V_{\tau\tau}$ for the CRE and the squared multiple correlation R_n^2, both of which are determined by the finite population variances of the potential outcomes and individual effects as well as their linear projections on covariates. For each treatment group $z \in \{0, 1\}$, recall that $\bar{Y}_z$ and $\bar{\boldsymbol{X}}_z$ are the average observed outcome and covariates, and define $s_z^2 = (n_z - 1)^{-1} \sum_{i:Z_i=z} (Y_i - \bar{Y}_z)^2$ and $\boldsymbol{s}_{z,\boldsymbol{X}} = \boldsymbol{s}_{\boldsymbol{X},z}^\top = (n_z - 1)^{-1} \sum_{i:Z_i=z} (Y_i - \bar{Y}_z)(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_z)^\top$ as the sample variance and covariance for the observed outcome and covariates. We further introduce $s_{z\backslash\boldsymbol{X}}^2 = s_z^2 - \boldsymbol{s}_{z,\boldsymbol{X}} \boldsymbol{S}_{\boldsymbol{X}}^{-2} \boldsymbol{s}_{\boldsymbol{X},z}$ and $s_{\tau|\boldsymbol{X}}^2 = (\boldsymbol{s}_{1,\boldsymbol{X}} - \boldsymbol{s}_{0,\boldsymbol{X}}) \boldsymbol{S}_{\boldsymbol{X}}^{-2} (\boldsymbol{s}_{\boldsymbol{X},1} - \boldsymbol{s}_{\boldsymbol{X},0})$. We can then estimate $V_{\tau\tau}$ and R_n^2 in (4) and (5) through replacing the finite population variances in their definitions by the corresponding sample analogues:

$$\hat{V}_{\tau\tau} = n_1^{-1} s_1^2 + n_0^{-1} s_0^2 - n^{-1} s_{\tau|\boldsymbol{X}}^2, \qquad \hat{R}_n^2 = 1 - \hat{V}_{\tau\tau}^{-1} (n_1^{-1} s_{1\backslash\boldsymbol{X}}^2 + n_0^{-1} s_{0\backslash\boldsymbol{X}}^2). \tag{14}$$

Note that the finite population variance of individual effects S_τ^2 is generally not identifiable, since we can never observe the individual effect for any unit. Thus, we do not expect any consistent estimator for it as well as $V_{\tau\tau}$ (Splawa-Neyman (1990)). However, because S_τ^2 is bounded below by $S_{\tau|\boldsymbol{X}}^2$ that has sample analogue $s_{\tau|\boldsymbol{X}}^2$, we can still estimate $V_{\tau\tau}$ conservatively as in (14).

Based on the asymptotic approximation established in Theorem 3 and the estimators in (14), for any $\alpha \in (0, 1)$, we can then construct the following $1 - \alpha$ confidence interval for τ:

$$\hat{\mathcal{C}}_\alpha = [\hat{\tau} - \hat{V}_{\tau\tau}^{1/2} \cdot \nu_{1-\alpha/2,K_n,a_n}(\hat{R}_n^2), \quad \hat{\tau} + \hat{V}_{\tau\tau}^{1/2} \cdot \nu_{1-\alpha/2,K_n,a_n}(\hat{R}_n^2)], \tag{15}$$

recalling that $\nu_{1-\alpha/2,K,a}(R^2)$ is the $(1 - \alpha/2)$th quantile of the distribution of $\sqrt{1 - R^2}\varepsilon_0 + \sqrt{R^2} L_{K,a}$.

To ensure the asymptotic validity of the confidence interval in (15), we invoke the following regularity condition. We defer more detailed discussion of the condition to Section 7. For $z = 0, 1$, let $S_{z\backslash\boldsymbol{X}}^2 = S_z^2 - S_{z|\boldsymbol{X}}^2$ denote the finite population variance of the residuals from the linear projection of potential outcomes $Y_i(z)$'s on covariates $\boldsymbol{X}_i$'s.

CONDITION 5. As $n \to \infty$,

$$\frac{\max_{z\in\{0,1\}} \max_{1\le i\le n} \{Y_i(z) - \bar{Y}(z)\}^2}{r_0 S_{1\backslash\boldsymbol{X}}^2 + r_1 S_{0\backslash\boldsymbol{X}}^2} \cdot \frac{\max\{K_n, 1\}}{r_1 r_0} \cdot \sqrt{\frac{\max\{1, \log K_n, -\log p_n\}}{n}} \to 0. \tag{16}$$

The following theorem shows that the confidence interval in (15) is asymptotically conservative, and becomes asymptotically exact when $S^2_{\tau\setminus \boldsymbol{X}} \equiv S^2_\tau - S^2_{\tau|\boldsymbol{X}}$ is asymptotically negligible.

THEOREM 7. *Under ReM and Conditions 1, 2 and 5, as* $n \to \infty$:

(i) *the estimators in (14) are asymptotically conservative in the sense that*

$$\max\{|\hat{V}_{\tau\tau}(1-\hat{R}^2_n) - V_{\tau\tau}(1-R^2_n) - S^2_{\tau\setminus \boldsymbol{X}}/n|, |\hat{V}_{\tau\tau}\hat{R}^2_n - V_{\tau\tau}R^2_n|\}$$
$$= o_{\mathbb{P}}(V_{\tau\tau}(1-R^2_n) + S^2_{\tau\setminus \boldsymbol{X}}/n);$$

(ii) *for any* $\alpha \in (0,1)$, *the resulting* $1-\alpha$ *confidence interval in (15) is asymptotically conservative, in the sense that* $\liminf_{n\to\infty} \mathbb{P}(\tau \in \hat{\mathcal{C}}_\alpha \mid M \le a_n) \ge 1-\alpha$;

(iii) *if further* $S^2_{\tau\setminus \boldsymbol{X}} = nV_{\tau\tau}(1-R^2_n)\cdot o(1)$, *the* $1-\alpha$ *confidence interval in (15) becomes asymptotically exact, in the sense that* $\lim_{n\to\infty} \mathbb{P}(\tau \in \hat{\mathcal{C}}_\alpha \mid M \le a_n) = 1-\alpha$.

Note that in Theorem 7, we do not make the typical assumption that R^2_n and $V_{\tau\tau}$ have limiting values as the sample size goes to infinity (see, e.g., Li, Ding and Rubin (2018)), since such an assumption may not be very satisfying given that we allow the number of covariates K_n to vary with the sample size (which consequently affects R^2_n). This brings additional challenge to the proof of Theorem 7. Theorem 7 also highlights the conservativeness in the finite population inference that dates back to Splawa-Neyman's (1990) analysis for the CRE. The conservativeness of the confidence interval comes mainly from $S^2_{\tau\setminus \boldsymbol{X}}$ as indicated in Theorem 7(i), which characterizes the individual effect heterogeneity after taking into account the covariates. The intervals become asymptotically exact when the individual effect heterogeneity is asymptotically linearly explained by the covariates, as shown in Theorem 7(iii).

If further Conditions 3 and 4 hold as in Theorem 5, not surprisingly, we can then use the usual Wald-type confidence intervals based on Gaussian quantiles, ignoring the term involving the constrained Gaussian random variable. Specifically, for any $\alpha \in (0,1)$, let

$$\tilde{\mathcal{C}}_\alpha = \big[\hat{\tau} - \sqrt{\hat{V}_{\tau\tau}(1-\hat{R}^2_n)}\cdot z_{1-\alpha/2}, \quad \hat{\tau} + \sqrt{\hat{V}_{\tau\tau}(1-\hat{R}^2_n)}\cdot z_{1-\alpha/2}\big]. \tag{17}$$

The following theorem shows the asymptotic validity of the above Wald-type confidence intervals.

THEOREM 8. *Under ReM and Conditions 1–5, Theorem 7(i)–(iii) still hold with* $\hat{\mathcal{C}}_\alpha$ *in (15) replaced by* $\tilde{\mathcal{C}}_\alpha$ *in (17).*

Theorem 8 shows that the usual Wald-type confidence intervals can be asymptotically valid for inferring the average treatment effect under ReM. However, we want to emphasize that it does involve additional regularity conditions, in particular Condition 3 on the choice of acceptance probability. While Theorem 8 provides confidence intervals of more convenient forms, we still recommend constructing confidence intervals based on Theorem 7, which takes into account explicitly the constrained Gaussian random variable that is determined by the acceptance probability and the number of covariates for rerandomization. The intervals from Theorem 7 not only requires fewer regularity conditions, but also becomes asymptotically equivalent to the ones from Theorem 8 when the additional Conditions 3 and 4 hold, under which the constrained Gaussian random variable L_{K_n,a_n} is of order $o_{\mathbb{P}}(1)$ and is thus negligible asymptotically.

Below we give a brief remark on improving the finite-sample performance of the confidence intervals. Note that $s^2_{1\setminus \boldsymbol{X}}$ and $s^2_{0\setminus \boldsymbol{X}}$ are almost equivalent to the sample variances

of the residuals from the linear projection of observed outcome on covariates in treated and control groups, respectively. Inspired by the regression analysis (see, e.g., MacKinnon (2013), Lei and Ding (2021)), we can consider rescaling the residuals to improve their finite-sample performance. For example, letting $\hat{e}_i$ denote the residual from the linear projection of observed outcome on covariates for unit i, we can rescale $\hat{e}_i$ to be $\kappa_i \hat{e}_i$, with $\kappa_i = 1$ for HC0, $\kappa_i = \sqrt{(n_{z_i} - 1)/(n_{z_i} - K_n - 1)}$ for HC1, $\kappa_i = 1/\sqrt{1 - H_{z_i,ii}}$ for HC2 and $\kappa_i = 1/(1 - H_{z_i,ii})$ for HC3, where $H_{z,ii}$ is the leverage of unit i for the covariate matrix consisting of the intercept and K_n covariates under treatment arm z.

Finally, we remark that throughout Sections 4 to 6 we assume Conditions 1 and 2, which are stronger than that required by completely randomized experiments. When these conditions fail, the asymptotic properties of rerandomization may fail, making it challenging to conduct robust randomization-based inference for rerandomization; see Appendix A1 in the Supplementary Material (Wang and Li (2022)) for a finite-sample worst-case analysis and its application for practical diagnosis of rerandomization.

7. Regularity conditions and practical implications. From the previous discussion, the asymptotic approximation of the difference-in-mean estimator $\hat{\tau}$ under ReM relies on Conditions 1 and 2. Once further Conditions 3 and 4 hold, we can achieve the "optimal" rerandomization, under which $\hat{\tau}$ becomes asymptotically Gaussian distributed. The large-sample inference for the average treatment effect τ under ReM relies additionally on Condition 5. These regularity conditions depend crucially on the convergence rate of Δ_n or its upper bound γ_n. In the following, we first investigate the convergence rate of γ_n when units are i.i.d. samples from some superpopulation, and discuss its implication on the validity of these regularity conditions. We then consider practical strategies that can make these regularity conditions more likely to hold, utilizing the advantage of finite population inference that requires no model or distributional assumptions on the potential outcomes and covariates. Finally, we discuss the computational cost of (optimal) rerandomization.

7.1. *Regularity conditions under i.i.d. sampling and their implications.* Throughout this subsection, we assume that the potential outcomes and covariates $(Y_i(1), Y_i(0), \boldsymbol{X}_i)$, for $i = 1, 2, \ldots, n$, are i.i.d. from a superpopulation that depends implicitly on the sample size n, noting that the dimension of covariates is allowed to vary with n. Recall that $\boldsymbol{u}_i = (r_0 Y_i(1) + r_1 Y_i(0), \boldsymbol{X}_i^\top)^\top \in \mathbb{R}^{K_n+1}$, and γ_n in (7) is uniquely determined by the treatment group proportions r_1, r_0, the dimension of covariates K_n, the sample size n and the finite population $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$, which is further assumed to consist of i.i.d. draws from a superpopulation. Below we impose some moment conditions on the sequence of superpopulations as $n \to \infty$.

CONDITION 6. For each sample size n, $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ are i.i.d. random vectors in $\mathbb{R}^{K_n+1}$ with finite and nonsingular covariance matrix. The standardized random vector $\boldsymbol{\xi}_i = \mathrm{Cov}(\boldsymbol{u}_i)^{-1/2}(\boldsymbol{u}_i - \mathbb{E}\boldsymbol{u}_i) \in \mathbb{R}^{K_n+1}$ satisfies that $\sup_{\boldsymbol{v} \in \mathbb{R}^{K_n+1}: \boldsymbol{v}^\top \boldsymbol{v} = 1} \mathbb{E}|\boldsymbol{v}^\top \boldsymbol{\xi}_i|^\delta = O(1)$ for some $\delta > 2$, that is, there exists an absolute constant $C < \infty$ such that $\mathbb{E}|\boldsymbol{v}^\top \boldsymbol{\xi}_i|^\delta \le C$ for all n and all unit vector $\boldsymbol{v}$ in $\mathbb{R}^{K_n+1}$.

From Lei and Ding ((2021), Proposition F.1), a sufficient condition for the uniform boundedness of $\sup_{\boldsymbol{v} \in \mathbb{R}^{K_n+1}: \boldsymbol{v}^\top \boldsymbol{v} = 1} \mathbb{E}|\boldsymbol{v}^\top \boldsymbol{\xi}_i|^\delta$ in Condition 6 is that $\boldsymbol{\xi}_i$ has independent coordinates whose absolute δ-th moment is uniformly bounded by a certain finite constant. The following proposition, which is a direct corollary of Lei and Ding ((2021), Lemma H.1), gives a stochastic upper bound of γ_n under Condition 6.

PROPOSITION 1. *If Condition 6 holds and the dimension of covariates $K_n = O(n^{\omega})$ for some $\omega \in (0, 1)$, then*

$$\gamma_n = O_{\mathbb{P}}\Big(\frac{1}{\sqrt{r_1 r_0}} \frac{(K_n+1)^{7/4}}{n^{1/2-1/\delta}}\Big).$$

Note that $\gamma_n \geq 2^{-3/2}(nr_1r_0)^{-1/2}(K_n+1)^{7/4}$ as shown in (9). Therefore, the rate in Proposition 1 is precise up to an $n^{1/\delta}$ factor; when $\delta = \infty$, the rate matches the lower bound rate in (9). From Proposition 1, we can immediately derive the following implications. Recall that r_1 and r_0 are the proportions of treated and control units. In practice, both treatment groups are likely to have nonnegligible proportions of units, and thus it is reasonable to assume that both r_1^{-1} and r_0^{-1} are of order $O(1)$. Recall that R_n^2 in (5) denotes the finite population squared multiple correlation between potential outcomes and covariates. We introduce $R^2_{\text{sup},n} = \text{Corr}^2(r_1 Y(0) + r_0 Y(1), \boldsymbol{X})$ to denote the superpopulation analogue with $(Y(1), Y(0), \boldsymbol{X})$ following the superpopulation distribution at sample size n.

COROLLARY 2. *If Condition 6 holds, $r_z^{-1} = O(1)$ for $z = 0, 1$ and $K_n = o(n^{2/7-4/(7\delta)})$, then*:

(i) *$\gamma_n = o_{\mathbb{P}}(1)$, and thus Δ_n in (8) is of order $o_{\mathbb{P}}(1)$;*

(ii) *the finite population and superpopulation squared multiple correlations R_n^2 and $R^2_{\text{sup},n}$ are asymptotically equivalent, in the sense that $R_n^2 - R^2_{\text{sup},n} = o_{\mathbb{P}}(1)$;*

(iii) *if further the standardized potential outcomes have bounded bth moments for some $b > 4$, both $\text{Var}(Y(1))$ and $\text{Var}(Y(0))$ are of the same order as $\text{Var}(r_0 Y(1) + r_1 Y(0))$, $\limsup_{n\to\infty} R^2_{\text{sup},n} < 1$ and $K_n = O(n^c)$ and $-\log p_n = o(n^{1-4/b-2c})$ for some $c < 1/2 - 2/b$, then the quantity on the left-hand side of (16) is of order $o_{\mathbb{P}}(1)$.*

Corollary 2(i) implies that Condition 1 holds with high probability, which, based on Theorem 3, further implies that the asymptotic approximation for the difference-in-means estimator under ReM using Gaussian and constrained Gaussian distributions is valid with high probability, given that the acceptance probability is chosen to satisfy Condition 2 with high probability (i.e., $\Delta_n / p_n = o_{\mathbb{P}}(1)$). With additional regularity conditions on the superpopulation variance and squared multiple correlations, Corollary 2(iii) implies that Condition 5 holds with high probability, which further implies that the large-sample inference for the average treatment effect discussed in Section 6 is valid with high probability. From the above, when the experimental units are i.i.d. from some superpopulation satisfying the moment conditions in Condition 6 and Corollary 2(iii), the number of covariates is not too large, and the acceptance probability is not too small, with high probability, we are able to asymptotically approximate the distribution of $\hat{\tau}$ under ReM, as well as constructing asymptotically valid confidence intervals for the average treatment effect τ.

Furthermore, Corollary 2(ii) implies that Condition 4 holds with high probability when $R^2_{\text{sup},n} \leq 1 - c$ for some absolute constant $c > 0$. Thus, based on Theorem 5, as long as the choice of acceptance probability satisfies Condition 3, with high probability, we can approximate the distribution of $\hat{\tau}$ under ReM by a Gaussian distribution, under which ReM achieves its ideally optimal precision. Based on Theorem 5, we can then derive the following corollary. Recall that $\mathcal{U}_n = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$. In the corollary below, we will write the conditioning on $\mathcal{U}_n$ explicitly to emphasize that we are considering the randomization distribution of $\hat{\tau}$ under ReM.

COROLLARY 3. *If Condition 6 holds, $r_z^{-1} = O(1)$ for $z = 0, 1$, $R^2_{\sup,n} \le 1 - c$ for some absolute constant $c > 0$, and $K_n = o(\log n)$, then there exists a sequence of acceptance probabilities p_n (or equivalently a sequence of thresholds a_n) such that, with probability converging to 1, the distribution of the difference-in-means estimator under ReM can be asymptotically approximated by a Gaussian distribution with mean zero and variance $V_{\tau\tau}(1 - R_n^2)$, that is,*

$$\sup_{c\in\mathbb{R}} \left| \mathbb{P}\{V_{\tau\tau}^{-1/2}(\hat{\tau} - \tau) \le c \mid M \le a_n, \mathcal{U}_n\} - \mathbb{P}\left(\sqrt{1 - R_n^2}\varepsilon_0 \le c \mid \mathcal{U}_n\right) \right| = o_{\mathbb{P}}(1).$$

Below we further consider the choice of acceptance probability under various cases for the number of covariates involved in rerandomization. We assume that Condition 6 holds and both treatment groups have nonnegligible proportions of units, that is, $r_z^{-1} = O(1)$ for $z = 0, 1$, and consider the cases where the number of covariates K_n increases sublogarithmically, logarithmically and polynomially with the sample size n, as shown in the first column of Table 1. Proposition 1 then gives an upper bound of the stochastic rate of γ_n, as shown in the second column of Table 1, all of which are of order $o_{\mathbb{P}}(1)$. Note that in the polynomial increase case, the rate of K_n is restricted to $K_n \asymp n^{\zeta}$ with $\zeta \in (0, \frac{2}{7} - \frac{4}{7\delta})$; otherwise, γ_n may not converge to zero in probability, under which Condition 1 may fail and the asymptotic inference will become difficult. From Theorem 1, under all the three cases, we can choose the acceptance probability p_n to decrease polynomially with the sample size such that Condition 2 holds with high probability, although there are various constraints on the exact polynomial decay rate. This is shown in the third column of Table 1, where κ can at least take value $1/3$ based on our result in Theorem 1(iii) and it can take value 1 if Raič's (2015) conjecture as in Theorem 1(ii) holds. The last column of Table 1 then shows the corresponding asymptotic behavior of the variance v_{K_n,a_n} of the constrained Gaussian random variable. Recall that the gain from ReM on estimation precision depends crucially on v_{K_n,a_n} as shown in Corollary 1. From Theorem 4, (i) when K_n increases sublogarithmically with n, v_{K_n,a_n} converges to zero and rerandomization achieves its ideally optimal precision; (ii) when K_n increases logarithmically with n, v_{K_n,a_n} is strictly between 0 and 1 when n is sufficiently large, and thus rerandomization provides gain in estimation precision compared to the CRE, although there is still a gap from the ideally optimal gain; (iii) when K_n increases polynomially with n, v_{K_n,a_n} converges to 1 as $n \to \infty$ and rerandomization provides no gain over the CRE. Therefore, in practice, we do not recommend using too many covariates, under which we will essentially lose the advantage from rerandomization. As a side note, these observations hold no matter which Berry–Esseen bound we use for the asymptotic Gaussian approximation (either the conjectured (ii), our derived (iii) in Theorem 1, or the ones in Theorem 2), except for the explicit rate of the polynomially decaying acceptance probability.

TABLE 1

Asymptotic properties for rerandomization under various rates for the number of involved covariates. Column 1 shows the asymptotic rate of the number of covariates K_n as n increases. Column 2 shows the corresponding stochastic rate of γ_n based on Proposition 1. Column 3 shows the choice of acceptance probability that is sufficient for Condition 2 with high probability. Column 4 shows the asymptotic rate for the variance v_{K_n,a_n} of the corresponding constrained Gaussian random variable

K_n	γ_n	p_n	v_{K_n,a_n}
$K_n = o(\log n)$	$o_{\mathbb{P}}(\frac{(\log n)^{7/4}}{n^{1/2-1/\delta}})$	$p_n \asymp n^{-\beta}$, $\beta \in (0, \frac{\kappa}{2} - \frac{\kappa}{\delta})$	$o(1)$
$K_n \asymp \log n$	$O_{\mathbb{P}}(\frac{(\log n)^{7/4}}{n^{1/2-1/\delta}})$	$p_n \asymp n^{-\beta}$, $\beta \in (0, \frac{\kappa}{2} - \frac{\kappa}{\delta})$	$(0, 1)$
$K_n \asymp n^{\zeta}$, $\zeta \in (0, \frac{2}{7} - \frac{4}{7\delta})$	$O_{\mathbb{P}}(n^{-(\frac{1}{2}-\frac{1}{\delta}-\frac{7\zeta}{4})})$	$p_n \asymp n^{-\beta}$, $\beta \in (0, \frac{\kappa}{2} - \frac{\kappa}{\delta} - \frac{7\zeta\kappa}{4})$	$1 - o(1)$

7.2. *Practical implication*. From the previous discussion, the asymptotic approximation of rerandomization depends crucially on γ_n (which involves K_n) and p_n. Below we provide suggestions and guidance on how to leverage these factors to improve the performance of rerandomization in practice.

We first consider K_n. From Theorem 6 and the discussion in Section 7.1, we suggest to choose at most $O(\log n)$ covariates, otherwise we may lose the advantage of rerandomization. Note that our finite population inference does not impose any model or distributional assumptions on the potential outcomes and covariates as well as their dependence structure. Thus, we have the flexibility to preprocess the covariates in an arbitrary way. Note that the covariates, although do not affect our asymptotic inference as long as the corresponding regularity conditions hold, do affect the improvement from rerandomization as indicated by the outcome-covariates association R_n^2. Therefore, we suggest to choose a moderate subset of covariates or a moderate dimensional transformation of original covariates to conduct rerandomization. For example, we can choose a subset of covariates of size $O(\log n)$ based on our subjective knowledge about the importance of these covariates for predicting/explaining the potential outcomes or based on some pilot studies. Recently, in the presence of high-dimensional covariates, Zhang, Yin and Rubin (2021) proposed to use principle component analysis to find some proper subspace of covariates to conduct rerandomization.

We then consider γ_n. Below we give an equivalent form of γ_n that can be more informative for its practical implications. For $z = 0, 1$ and $1 \le i \le n$, let $e_i(z) = Y_i(z) - \bar{Y}(z) - \boldsymbol{S}_{z,\boldsymbol{X}}(\boldsymbol{S}_{\boldsymbol{X}}^2)^{-1}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})$ be the residual from the linear projection of potential outcome $Y_i(z)$ on covariates $\boldsymbol{X}_i$. Then $r_0 e_i(1) + r_1 e_i(0)$ is the residual from the projection of $r_0 Y_i(1) + r_1 Y_i(0)$ on $\boldsymbol{X}_i$. We further introduce e_i to denote the corresponding standardized residual for unit i, that is, $r_0 e_i(1) + r_1 e_i(0)$ standardized by its finite population mean and standard deviation. Let $\tilde{\boldsymbol{X}} = (\boldsymbol{X}_1, \boldsymbol{X}_2 \ldots, \boldsymbol{X}_n)^\top \in \mathbb{R}^{n \times K_n}$ be the matrix consisting of the covariates for all n units, $\boldsymbol{H} = \tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}^\top \tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}^\top \in \mathbb{R}^{n \times n}$ be the corresponding projection or hat matrix, and H_{ii} be the ith diagonal element of $\boldsymbol{H}$, which is usually called the leverage score for unit i in regression analysis. As demonstrated in the Supplementary Material (Wang and Li (2022)), γ_n in (7) can be bounded by

$$\gamma_n = \frac{(K_n+1)^{1/4}}{\sqrt{nr_1r_0}} \frac{1}{n} \sum_{i=1}^{n} \big(e_i^2 + (n-1)H_{ii}\big)^{3/2} \in \bigg[\frac{1}{4\sqrt{2}}\tilde{\gamma}_n, \sqrt{2}\tilde{\gamma}_n\bigg],$$

with

$$\tilde{\gamma}_n = \frac{(K_n+1)^{1/4}}{\sqrt{r_1r_0n}} \frac{1}{n} \sum_{i=1}^{n} |e_i|^3 + \frac{(K_n+1)^{1/4}}{\sqrt{r_1r_0}} \sum_{i=1}^{n} H_{ii}^{3/2}. \tag{18}$$

Obviously, γ_n and $\tilde{\gamma}_n$ are of the same order, and it is thus equivalent to consider $\tilde{\gamma}_n$. From (18), $\tilde{\gamma}_n$ depends crucially on the absolute third moment of the standardized residuals $n^{-1}\sum_{i=1}^n |e_i|^3$ and the summation of the leverage scores to the power of $3/2$ over all n units. Note that e_i's depend on the true potential outcomes and are generally unknown in the design stage of an experiment. When the potential outcomes $Y_i(1)$ and $Y_i(0)$, or more precisely their residuals $e_i(1)$ and $e_i(0)$ are not too heavy tailed, we expect $n^{-1}\sum_{i=1}^n |e_i|^3$ to be of constant order, under which the first term in (18) is likely to be well controlled. The second term in (18) seems to be more complicated, but fortunately it is known in the design stage since it depends only on the pretreatment covariates. Thus, we can and should check the leverage scores for all units before conducting rerandomization. If $\sum_{i=1}^n H_{ii}^{3/2}$ is small, then we expect the asymptotic approximation for rerandomization to work well; otherwise, we need to be careful. As discussed before, under the finite population inference framework, we have the flexibility to preprocess the covariates in an arbitrary way. In particular, we can

try to shrink the leverage scores via trimming, a practical strategy that is also recommended when conducting regression adjustment for completely randomized experiments (Lei and Ding (2021)). Note that too much trimming may reduce the outcome-covariates association R_n^2, and thus deteriorate the improvement from rerandomization. Besides, as demonstrated in the Supplementary Material (Wang and Li (2022)), $\sum_{i=1}^n H_{ii}^{3/2}$ is always lower bounded by $K_n^{3/2}/\sqrt{n}$. In practice, we may consider performing the minimum possible trimming such that the resulting $\sum_{i=1}^n H_{ii}^{3/2}$ is close to its minimum value $K_n^{3/2}/\sqrt{n}$. As a side note, both ridge and PCA rerandomizations recently proposed by Branson, Dasgupta and Rubin (2016) and Zhang, Yin and Rubin (2021) can also help reduce leverage scores. Thus, our theory also provides some justification for these two designs. Moreover, compared to their preprocessing on the covariates, trimming can be more robust to outliers. Besides, it does not change the original covariates (as well as their meaning) much and may be more helpful in preserving the explainability of original covariates (more precisely, their squared multiple correlation with the potential outcomes).

Finally, we consider p_n. The discussion in Section 7.1 suggests choosing the acceptance probability p_n such that it decays polynomially with the sample size n. However, such results may not be helpful for the choice of p_n under a finite sample size. Below we give some practical guideline on the choice of p_n with a moderate number of covariates K_n. First, from Corollary 1, the gap between rerandomization with a certain p_n and the ideally optimal one is characterized by v_{K_n,a_n}, the variance of the constrained Gaussian random variable. Therefore, we suggest choosing p_n such that the corresponding v_{K_n,a_n} is small, say, 0.01. Second, with a given p_n, we can check the asymptotic approximation by using some pseudo potential outcomes as proxies, for example, some linear/nonlinear combinations of covariates based on some prior knowledge. Third, note that, with acceptance probability p_n, the number of randomizations needed for getting an acceptable treatment assignment is about $1/p_n$. In practice, our choice of p_n can also take into account this computation cost; see also Section 7.3.

7.3. *Computational cost of rerandomization*. Below we briefly discuss the computational cost for getting one acceptable treatment assignment from ReM with n units, K-dimensional covariates ($K \le n$) and acceptance probability $p > 0$. To facilitate the computation, we can first standardize the covariates, that is, getting $\boldsymbol{S}_{\boldsymbol{X}}^{-1}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})$ for each i, which has a complexity of $O(nK^2)$. Then the Mahalanobis distance M is equivalently the Euclidean norm of the difference-in-means of standardized covariates up to some scale, whose computation has complexity of $O(nK)$ and in expectation needs to be done approximately p^{-1} times. Besides, in each iteration, one has to do a complete randomization of experimental units into treatment and control groups, which has complexity of $O(n)$ (Fan, Muller and Rezucha (1962)); see also Meng (2013). Consequently, in expectation, the computational complexity for getting one acceptable assignment from ReM is approximately of $O(nK(K + p^{-1}))$. From the discussion in Section 7.1 and in particular Table 1, to achieve the ideally optimal rerandomization, we will choose $K = o(\log n)$ and can choose $p^{-1} = n^{\beta}$ for sufficiently small β, under which the computational complexity for getting an acceptable assignment from ReM is in expectation approximately of order $o(n^{1+\beta}\log n)$ (i.e., polynomial in n with exponent slightly greater than 1). This implies that the optimal rerandomization is computationally feasible even for relatively large sample size.

8. Conclusion and discussion. There is a long-time controversy between the two philosophies for designing an experiment: randomization versus optimal, and thus often deterministic assignment. In the context of balancing covariates in randomized experiments, the optimal design tries to find the treatment assignment minimizing the covariate imbalance

(say the Mahalanobis distance), while rerandomization tries to restrict the covariate imbalance and at the same time maintain sufficient randomness of the design, which is necessary for robust inference of treatment effects. In this paper, we demonstrated that, by letting the acceptance probability diminish to zero at a property rate (e.g., a polynomial rate), rerandomization can still have sufficient randomness for robust randomization-based inference of treatment effects, and more importantly, it can achieve the ideally optimal precision that one can expect from the optimally balanced design. Note that our theory also helps mitigate the computation burden for usual optimal designs. In particular, to achieve the ideally optimal precision, we only need to randomly select assignments from a small proportion (which generally decreases with the sample size polynomially) of all assignments with the best covariate balance, whose computational complexity is generally of a polynomial order of the sample size with exponent slightly greater than 1 as discussed in Section 7.3.

The derived theory for rerandomization also allows for a diverging number of covariates. In particular, we found that, when the number of covariates is too large, not only will the asymptotic approximation for rerandomization become inaccurate, but also rerandomization will lose its gain on efficiency. Therefore, we suggest practitioners to use a moderate number of covariates, especially those important and useful for explaining the potential outcomes, and to also perform trimming as suggested in Section 7. Importantly, because our finite population inference imposes no distributional assumptions on potential outcomes and covariates, we are free to adjust the covariates in an arbitrary way, such as trimming or transforming using, for example, principle components. In the Supplementary Material (Wang and Li (2022)), we also provide additional finite-sample diagnosis tools, discuss the choice of covariates and threshold for rerandomization, and conduct a simulation study.

In this paper, we mainly focused on the asymptotic properties of rerandomized treatment-control experiments using the Mahalanobis distance criterion; see the Supplementary Material (Wang and Li (2022)) for extension to regression adjustment under rerandomization. Beyond that, the derived theory, including both the finite population central limit theorem and the asymptotic behavior of the constrained Gaussian random variable, can also be useful for analyzing other covariate balance criteria, such as the Mahalanobis distance criterion with tiers of covariates (Li, Ding and Rubin (2018), Morgan and Rubin (2015)). It will also be interesting to extend the theory to rerandomization in more complex experiments, such as blocked experiments (Johansson and Schultzberg (2022), Wang, Wang and Liu (2021)), factorial experiments (Branson, Dasgupta and Rubin (2016), Li, Ding and Rubin (2020)), survey experiments (Yang, Qu and Li (2021)) and sequential experiments (Zhou et al. (2018)). Besides, we mainly considered finite population inference focusing on the average treatment effect of the experimental units in hand. It will be interesting to also consider superpopulation inference of some population average treatment effect when the units are randomly sampled from some superpopulation (Schultzberg and Johansson (2020)).

Acknowledgments. We thank the Editor, the Associate Editor and three reviewers for constructive comments.

Yuhao Wang is also affiliated with Shanghai Qi Zhi Institute.

Funding. The work of Yuhao Wang was supported by Tsinghua New Faculty Start-up Fund and the 2030 Innovation Megaprojects of China (Programme on New Generation Artificial Intelligence) Grant No. 2021AAA0150000.

SUPPLEMENTARY MATERIAL

Supplement to "Rerandomization with diminishing covariate imbalance and diverging number of covariates" (DOI: 10.1214/22-AOS2235SUPP; .pdf). First, we provide additional finite-sample diagnosis tools for rerandomization, and conduct a simulation study.

Second, we extend the asymptotic theory to regression adjustment under rerandomization. Third, we study the Berry–Esseen-type bound for finite population central limit theorem under simple random sampling. Fourth, we prove all the theorems, corollaries and propositions. Fifth, we connect rerandomization with usual optimal designs.

REFERENCES

BANERJEE, A. V., CHASSANG, S., MONTERO, S. and SNOWBERG, E. (2020). A theory of experimenters: Robustness, randomization, and balance. *Am. Econ. Rev.* **110** 1206–1230.

BENTKUS, V. (2003). On the dependence of the Berry–Esseen bound on dimension. *J. Statist. Plann. Inference* **113** 385–402. MR1965117 https://doi.org/10.1016/S0378-3758(02)00094-0

BENTKUS, V. (2004). A Lyapunov type bound in $\mathbf{R}^d$. *Teor. Veroyatn. Primen.* **49** 400–410. MR2144310 https://doi.org/10.1137/S0040585X97981123

BERRY, A. C. (1941). The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc.* **49** 122–136. MR0003498 https://doi.org/10.2307/1990053

BICKEL, P. J. and FREEDMAN, D. A. (1984). Asymptotic normality and the bootstrap in stratified sampling. *Ann. Statist.* **12** 470–482. MR0740906 https://doi.org/10.1214/aos/1176346500

BIKELIS, A. (1969). The estimation of the remainder term in the central limit theorem for samples taken from finite sets. *Studia Sci. Math. Hungar.* **4** 345–354. MR0254902

BLONIARZ, A., LIU, H., ZHANG, C.-H., SEKHON, J. S. and YU, B. (2016). Lasso adjustments of treatment effect estimates in randomized experiments. *Proc. Natl. Acad. Sci. USA* **113** 7383–7390. MR3531136 https://doi.org/10.1073/pnas.1510506113

BOLTHAUSEN, E. and GÖTZE, F. (1993). The rate of convergence for multivariate sampling statistics. *Ann. Statist.* **21** 1692–1710. MR1245764 https://doi.org/10.1214/aos/1176349393

BOX, G. E. P., HUNTER, J. S. and HUNTER, W. G. (2005). *Statistics for Experimenters*: *Design*, *Innovation*, *and Discovery*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley-Interscience, Hoboken, NJ. MR2140250

BRANSON, Z., DASGUPTA, T. and RUBIN, D. B. (2016). Improving covariate balance in 2^K factorial designs via rerandomization with an application to a New York City Department of Education high school study. *Ann. Appl. Stat.* **10** 1958–1976. MR3592044 https://doi.org/10.1214/16-AOAS959

BRANSON, Z. and SHAO, S. (2021). Ridge rerandomization: An experimental design strategy in the presence of covariate collinearity. *J. Statist. Plann. Inference* **211** 287–314. MR4125927 https://doi.org/10.1016/j.jspi.2020.07.002

BRUHN, M. and MCKENZIE, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *Am. Econ. J. Appl. Econ.* **1** 200–232.

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45** 2309–2352. MR3693963 https://doi.org/10.1214/16-AOP1113

CHERNOZHUKOV, V., CHETVERIKOV, D. and KOIKE, Y. (2020). Nearly optimal central limit theorem and bootstrap approximations in high dimensions. Preprint. Available at arXiv:2012.09513.

COX, D. R. (1982). Randomization and concomitant variables in the design of experiments. In *Statistics and Probability*: *Essays in Honor of C. R. Rao* 197–202. North-Holland, Amsterdam. MR0659470

COX, D. R. (2007). Applied statistics: A review. *Ann. Appl. Stat.* **1** 1–16. MR2393838 https://doi.org/10.1214/07-AOAS113

ESSEEN, C.-G. (1942). On the Liapounoff limit of error in the theory of probability. *Ark. Mat. Astron. Fys.* **28A** 19. MR0011909

FAN, C. T., MULLER, M. E. and REZUCHA, I. (1962). Development of sampling plans by using sequential (item by item) selection techniques and digital computers. *J. Amer. Statist. Assoc.* **57** 387–402. MR0139243

FANG, X. and KOIKE, Y. (2021). High-dimensional central limit theorems by Stein's method. *Ann. Appl. Probab.* **31** 1660–1686. MR4312842 https://doi.org/10.1214/20-aap1629

FISHER, R. A. (1935). *The Design of Experiments*, 1st ed. Oliver and Boyd, Edinburgh, London.

FREEDMAN, D. A. (2008). Editorial: Oasis or mirage? *Chance* **21** 59–61. MR2422783 https://doi.org/10.1007/s00144-008-0047-x

HÁJEK, J. (1960). Limiting distributions in simple random sampling from a finite population. *Magy. Tud. Akad. Mat. Kut. Intéz. Közl.* **5** 361–374. MR0125612

HECKMAN, J. J. and KARAPAKULA, G. (2021). Using a satisficing model of experimenter decision-making to guide finite-sample inference for compromised experiments. *Econom. J.* **24** C1–C39. MR4281220 https://doi.org/10.1093/ectj/utab009

HÖGLUND, T. (1978). Sampling from a finite population: A remainder term estimate. *Scand. J. Stat.* **5** 69–71. MR0471130

JOHANSSON, P., RUBIN, D. B. and SCHULTZBERG, M. (2021). On optimal rerandomization designs. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 395–403. MR4250281 https://doi.org/10.1111/rssb.12417

JOHANSSON, P. and SCHULTZBERG, M. (2022). Rerandomization: A complement or substitute for stratification in randomized experiments? *J. Statist. Plann. Inference* **218** 43–58. MR4327401 https://doi.org/10.1016/j.jspi.2021.09.002

KALLUS, N. (2018). Optimal *a priori* balance in the design of controlled experiments. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 85–112. MR3744713 https://doi.org/10.1111/rssb.12240

KAPELNER, A., KRIEGER, A. M., SKLAR, M., SHALIT, U. and AZRIEL, D. (2021). Harmonizing optimized designs with classic randomization in experiments. *Amer. Statist.* **75** 195–206. MR4256124 https://doi.org/10.1080/00031305.2020.1717619

KASY, M. (2016). Why experimenters might not always want to randomize, and what they could do instead. *Polit. Anal.* **24** 324–338.

KIEFER, J. (1959). Optimum experimental designs. *J. Roy. Statist. Soc. Ser. B* **21** 272–319. MR0113263

LEI, L. and DING, P. (2021). Regression adjustment in completely randomized experiments with a diverging number of covariates. *Biometrika* **108** 815–828. MR4341353 https://doi.org/10.1093/biomet/asaa103

LI, X. and DING, P. (2017). General forms of finite population central limit theorems with applications to causal inference. *J. Amer. Statist. Assoc.* **112** 1759–1769. MR3750897 https://doi.org/10.1080/01621459.2017.1295865

LI, X. and DING, P. (2020). Rerandomization and regression adjustment. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 241–268. MR4060984

LI, X., DING, P. and RUBIN, D. B. (2018). Asymptotic theory of rerandomization in treatment-control experiments. *Proc. Natl. Acad. Sci. USA* **115** 9157–9162. MR3856113 https://doi.org/10.1073/pnas.1808191115

LI, X., DING, P. and RUBIN, D. B. (2020). Rerandomization in 2^K factorial experiments. *Ann. Statist.* **48** 43–63. MR4065152 https://doi.org/10.1214/18-AOS1790

LIN, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *Ann. Appl. Stat.* **7** 295–318. MR3086420 https://doi.org/10.1214/12-AOAS583

LIU, H., REN, J. and YANG, Y. (2022). Randomization-based Joint Central Limit Theorem and Efficient Covariate Adjustment in Randomized Block 2^K Factorial Experiments. *J. Amer. Statist. Assoc.* In press. https://doi.org/10.1080/01621459.2022.2102985

MACKINNON, J. G. (2013). Thirty years of heteroskedasticity-robust inference. In *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis* 437–461. Springer, Berlin.

MENG, X. (2013). Scalable simple random sampling and stratified sampling. In *International Conference on Machine Learning. PMLR* 531–539.

MORGAN, K. L. and RUBIN, D. B. (2012). Rerandomization to improve covariate balance in experiments. *Ann. Statist.* **40** 1263–1282. MR2985950 https://doi.org/10.1214/12-AOS1008

MORGAN, K. L. and RUBIN, D. B. (2015). Rerandomization to balance tiers of covariates. *J. Amer. Statist. Assoc.* **110** 1412–1421. MR3449036 https://doi.org/10.1080/01621459.2015.1079528

RAIČ, M. (2015). Multivariate normal approximation: Permutation statistics, local dependence and beyond.

RAIČ, M. (2019). A multivariate Berry–Esseen theorem with explicit constants. *Bernoulli* **25** 2824–2853. MR4003566 https://doi.org/10.3150/18-BEJ1072

ROSENBAUM, P. R. (2010). *Design of Observational Studies. Springer Series in Statistics*. Springer, New York. MR2561612 https://doi.org/10.1007/978-1-4419-1213-8

RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **66** 688–701.

SAVAGE, L. J. (1962). *The Foundations of Statistical Inference*. Methuen, London.

SCHULTZBERG, M. and JOHANSSON, P. (2020). Asymptotic inference for optimal rerandomization designs. *Open Stat.* **1** 49–58. https://doi.org/10.1515/stat-2020-0102

SHI, L. and DING, P. (2022). Berry–Esseen bounds for design-based causal inference with possibly diverging treatment levels and varying group sizes. Preprint. Available at arXiv:2209.12345.

SPLAWA-NEYMAN, J. (1990). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statist. Sci.* **5** 465–472. MR1092986

STUDENT (1938). Comparison between balanced and random arrangements of field plots. *Biometrika* **29** 363–378.

TAVES, D. R. (1974). Minimization: A new method of assigning patients to treatment and control groups. *Clin. Pharmacol. Ther.* **15** 443–453.

WAGER, S., DU, W., TAYLOR, J. and TIBSHIRANI, R. J. (2016). High-dimensional regression adjustments in randomized experiments. *Proc. Natl. Acad. Sci. USA* **113** 12673–12678. MR3576188 https://doi.org/10.1073/pnas.1614732113

WANG, Y. and LI, X. (2022). Supplement to "Rerandomization with diminishing covariate imbalance and diverging number of covariates." https://doi.org/10.1214/22-AOS2235SUPP

WANG, X., WANG, T. and LIU, H. (2021). Rerandomization in Stratified Randomized Experiments. *J. Amer. Statist. Assoc.* In press. https://doi.org/10.1080/01621459.2021.1990767

YANG, Z., QU, T. and LI, X. (2021). Rejective Sampling, Rerandomization, and Regression Adjustment in Survey Experiments. *J. Amer. Statist. Assoc.* In press. https://doi.org/10.1080/01621459.2021.1984926

ZHANG, H., YIN, G. and RUBIN, D. B. (2021). PCA Rerandomization. Preprint. Available at arXiv:2102.12262.

ZHOU, Q., ERNST, P. A., MORGAN, K. L., RUBIN, D. B. and ZHANG, A. (2018). Sequential rerandomization. *Biometrika* **105** 745–752. MR3842897 https://doi.org/10.1093/biomet/asy031