Journal of Econometrics 241 (2024) 105724

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

No star is good news: A unified look at rerandomization based on p-values from covariate balance tests

Anqi Zhao [a,*], Peng Ding [b,1]

[a] Fuqua School of Business, Duke University, United States of America
[b] Department of Statistics, University of California, Berkeley, United States of America

ARTICLE INFO

Keywords:
Constrained randomization
Design-based inference
Design of experiment
Gaussian correlation inequality
Potential outcomes
Regression adjustment

ABSTRACT

Randomized experiments balance all covariates on average and are considered the gold standard for estimating treatment effects. Chance imbalances are nonetheless common in realized treatment allocations. To inform readers of the comparability of treatment groups at baseline, contemporary scientific publications often report covariate balance tables with not only covariate means by treatment group but also the associated p-values from significance tests of their differences. The practical need to avoid small p-values as indicators of poor balance motivates balance check and rerandomization based on these p-values from covariate balance tests (ReP) as an attractive tool for improving covariate balance in designing randomized experiments. Despite the intuitiveness of such strategy and its possibly already widespread use in practice, the literature lacks results about its implications on subsequent inference, subjecting many effectively rerandomized experiments to possibly inefficient analyses. To fill this gap, we examine a variety of potentially useful schemes for ReP and quantify their impact on subsequent inference. Specifically, we focus on three estimators of the average treatment effect from the unadjusted, additive, and interacted linear regressions of the outcome on treatment, respectively, and derive their asymptotic sampling properties under ReP. The main findings are threefold. First, the estimator from the interacted regression is asymptotically the most efficient under all ReP schemes examined, and permits convenient regression-assisted inference identical to that under complete randomization. Second, ReP, in contrast to complete randomization, improves the asymptotic efficiency of the estimators from the unadjusted and additive regressions. Standard regression analyses are accordingly still valid but in general overconservative. Third, ReP reduces the asymptotic conditional biases of the three estimators and improves their coherence in terms of mean squared difference. These results establish ReP as a convenient tool for improving covariate balance in designing randomized experiments, and we recommend using the interacted regression for analyzing data from ReP designs.

1. Introduction

1.1. Rerandomization based on p-values

Covariate balance increases comparability of units under different treatment conditions, thereby strengthening the causal conclusions that can be drawn from data. Randomized experiments balance all observed and unobserved covariates on average,

* Corresponding author.
E-mail addresses: az171@duke.edu (A. Zhao), pengdingpku@berkeley.edu (P. Ding).

[1] Ding was partially funded by the U.S. National Science Foundation (grant 1945136). We thank David Broockman, Christopher Mann, Avi Feller, and Thomas Royen for inspiring discussions, and thank the editor and two referees for their most constructive suggestions that greatly helped to improve our paper.

https://doi.org/10.1016/j.jeconom.2024.105724
Received 18 April 2023; Received in revised form 26 February 2024; Accepted 26 February 2024

and provide the gold standard for estimating treatment effects. Chance imbalances are nonetheless common in realized allocations. Rerandomization, termed by Cox (1982) and Morgan and Rubin (2012), enforces covariate balance by discarding randomizations that do not satisfy a prespecified balance criterion. Bruhn and McKenzie (2009) conducted a survey of leading experimental researchers in development economics, and suggested that rerandomization is commonly used yet often poorly documented.

To inform readers of the comparability of treatment groups at baseline, contemporary scientific publications are often encouraged to report baseline covariate balance tables with not only covariate means by treatment group but also the associated p-values from significance tests of their differences. The practical need to avoid small p-values as indicators of poor balance motivates conducting rerandomization directly based on these p-values from balance tests (Bruhn and McKenzie, 2009; Ashraf et al., 2010). Formally, *rerandomization based on p-values* (ReP) runs one or more statistical tests to check the covariate balance of a realized randomization, and accepts the allocation if and only if the p-values of interest all exceed some prespecified thresholds. In their popular textbook on modern field experiments, Gerber and Green (2012, Chapter 4.5) made this recommendation as a way to "quickly approximate blocking".

Despite its decade-long existence and close relevance to practice, the theory of ReP has not been addressed in the literature. Hansen and Bowers (2008) discussed two hypothesis testing-based techniques for balance check in randomized experiments but did not touch the issue of rerandomization and corresponding inference. Gerber and Green (2012) gave the practical recommendation yet did not discuss its theoretical implications. The existing discussion on rerandomization, on the other hand, focused mostly on balance criteria based on the Mahalanobis distance between covariate means by treatment group (Morgan and Rubin, 2012, 2015; Branson et al., 2016; Li et al., 2018, 2020; Li and Ding, 2020; Branson and Shao, 2021; Zhao and Ding, 2021, 2023; Johansson et al., 2021; Johansson and Schultzberg, 2022). The resulting procedure, also known as ReM, is convenient in theory but in general not a straightforward choice in practice. Another related literature is that on restricted randomization, also known as constrained randomization, which improves covariate balance by blocking, stratification, matched pairing, covariate-adaptive adjustment, etc. See, e.g., Bailey (1987), Imai et al. (2009), Bruhn and McKenzie (2009), Miratrix et al. (2013), Higgins et al. (2016), Bugni et al. (2018, 2019), Fogarty (2018), Liu and Yang (2020), Wang et al. (2021), Pashley and Miratrix (2021), Bai et al. (2022), and Ye et al. (2023). See also Johansson and Schultzberg (2022) for a discussion on the connection and comparison between stratification and rerandomization. The existing work in this literature however concerns restrictions distinct from those based on covariate balance tests. The gap between theory and practice causes many effectively ReP-based experiments to be analyzed as if they were completely randomized, risking overconservative inferences that hinder the detection of statistically significant findings (Bruhn and McKenzie, 2009). Glennerster and Takavarasha (2014) took an extreme stance and advised to "avoid using this technique (rerandomization), at least until there is more agreement in the literature about its pros and cons". This paper fills this gap and clarifies the theoretical implications of ReP.

1.2. Our contributions

First, we formalize ReP as a tool for improving covariate balance in randomized treatment-control experiments, and propose a variety of potentially useful schemes based on standard statistical tests. The proposed ReP schemes use p-values from two-sample t-tests, linear regression, and logistic regression to form the covariate balance criteria, allowing for easy implementation via standard software packages.

Next, we quantify for the first time the impact of the proposed ReP schemes on covariate balance and subsequent inference. Specifically, we focus on three estimators of the average treatment effect from the ordinary least squares (OLS) fits of the unadjusted, additive, and interacted linear regressions of the outcome on treatment, respectively, and evaluate their sampling properties under the proposed ReP schemes from the *design-based* perspective. In short, the design-based perspective assumes the finite-population framework and takes the physical act of randomization as the sole source of randomness in evaluating the sampling properties of quantities of interest (Neyman, 1923; Freedman, 2008b; Lin, 2013; Imbens and Rubin, 2015). The resulting inference is robust to model misspecification and hence also known as *model-assisted* inference; see Negi and Wooldridge (2021) and the references therein for the super-population counterpart. The main findings are threefold. First, ReP improves the covariate balance between treatment groups, which in turn (i) simplifies the interpretation of experimental results, (ii) reduces the conditional biases of the estimators (c.f. Ding, 2021b, Section 4.3), and (iii) promotes more coherent inferences between covariate-adjusted and unadjusted analyses (c.f. Zhao and Ding, 2023, Section S5.3). Second, the estimator from the interacted regression is asymptotically the most efficient under all ReP schemes considered, with the asymptotic sampling distribution unaffected by the rerandomization. It is thus our recommendation for subsequent analysis under the proposed ReP schemes, allowing for convenient regression-based inferences identical to that under complete randomization. Specifically, we can use the coefficient of the treatment in the OLS fit of the outcome on the treatment, covariates, and their interactions as the point estimator of the average treatment effect, and use the associated Eicker–Huber–White (EHW) standard error to estimate the true standard error. Third, ReP improves the asymptotic efficiency of the estimators from the unadjusted and additive regressions, rendering inference based on the usual normal approximation overconservative. This highlights the importance of rerandomization-specific inference under ReP when the unadjusted or additive regression is used and, by contrast, demonstrates the advantage of the interacted regression for efficient and straightforward inference by normal approximation. We thus recommend using the interacted regression for analyzing data from ReP designs.

Lastly, we extend the theory of ReP to experiments with more than two treatment arms and stratified experiments.

We make two novel technical contributions in the process. First, we clarify the value of the renowned Gaussian Correlation Inequality (Royen, 2014) for establishing the theoretical properties of rerandomization. Most existing proofs of the theory of ReM rely on the geometric properties of the Mahalanobis distance and do not generalize to rerandomization based on other balance criteria such as ReP. The Gaussian Correlation Inequality bridges this gap and provides a powerful tool for clarifying the theoretical guarantees of a wide class of rerandomization schemes including ReM and ReP. See Lemmas S3–S4 in the Supplementary Material. Second, we establish the asymptotic equivalence of the likelihood ratio test (LRT) and the Wald test for logistic and multinomial logistic regressions from the design-based perspective, complementing the recent discussions in Freedman (2008b), Hansen and Bowers (2008), and Guo and Basse (2023) on design-based inference from nonlinear regressions. See Theorem S1 in the Supplementary Material. Without invoking any assumption of the logistic or multinomial logistic model, we view logistic and multinomial logistic regressions as purely numeric procedures based on maximum likelihood estimation (MLE), and evaluate the sampling properties of the test statistics over the distribution of the treatment assignments.

1.3. Notation and definitions

For a set of tuples $\{(u_i, v_{i1}, \dots, v_{iL}) : u_i \in \mathbb{R},\ v_{il} \in \mathbb{R}^{K_l},\ i = 1, \dots, N,\ l = 1, \dots, L\}$, denote by $\texttt{lm}(u_i \sim v_{i1} + \cdots + v_{iL})$ the OLS fit of the linear regression of u_i on $(v_{i1}, \dots, v_{iL})$ with $(v_{i1}^{\mathrm{T}}, \dots, v_{iL}^{\mathrm{T}})^{\mathrm{T}}$ as the regressor vector, and by $\texttt{logit}(u_i \sim v_{i1} + \cdots + v_{iL})$ the MLE fit of the logistic or multinomial logistic regression of u_i on $(v_{i1}, \dots, v_{iL})$. We allow each v_{il} to be a scalar or a vector and use + to denote concatenation of regressors. Throughout, we focus on the numeric outputs of OLS and MLE without invoking any assumption of the corresponding linear or logistic model. Assume default tests and p-values from standard software packages throughout unless specified otherwise.

For two $K \times 1$ vectors $T = (t_1, \dots, t_K)^{\mathrm{T}}$ and $a = (a_1, \dots, a_K)^{\mathrm{T}}$, denote by $|T| \le a$ if $|t_k| \le a_k$ for all $k = 1, \dots, K$. Denote by $\mathrm{diag}(u_k)_{k=1}^K = \mathrm{diag}(u_1, \dots, u_K)$ the $K \times K$ diagonal matrix with u_k's on the diagonal. For a $K \times K$ symmetric matrix $V = (V_{kk'})_{k,k'=1,\dots,K}$ with $V_{kk} > 0$ for all $k = 1, \dots, K$, let $\sigma(V) = \mathrm{diag}(V_{kk}^{1/2})_{k=1}^K$ and $D(V) = \{\sigma(V)\}^{-1} V \{\sigma(V)\}^{-1}$. Intuitively, $D(V)$ gives the corresponding correlation matrix when V is a covariance matrix. Let $\|\epsilon\|_{\mathcal{M}} = \epsilon^{\mathrm{T}}\{\mathrm{cov}(\epsilon)\}^{-1}\epsilon$ denote the Mahalanobis distance of a random vector ϵ from the origin. Let $\rightsquigarrow$ denote convergence in distribution, and let $\mathbb{E}_{\mathrm{a}}$ denote the expectation of the asymptotic distribution.

Lastly, we use *peakedness* (Sherman, 1955) to quantify the relative efficiency between estimators.

Definition 1. For two symmetric random vectors A and B in $\mathbb{R}^K$, we say A is *more peaked* than B if $\mathbb{P}(A \in C) \ge \mathbb{P}(B \in C)$ for all symmetric convex sets C in $\mathbb{R}^K$, denoted by $A \succeq B$.

For $K = 1$, a more peaked random variable has narrower central quantile ranges. For A and B with finite second moments, $A \succeq B$ implies $\mathrm{cov}(A) - \mathrm{cov}(B)$ is negative semidefinite (Li et al., 2020, Proposition 4). For A and B that are both normal with zero means, $A \succeq B$ is equivalent to $\mathrm{cov}(A) - \mathrm{cov}(B)$ being negative semidefinite. This suggests peakedness as a more refined measure for comparing relative efficiency of estimators than covariance. We formalize the intuition in Definition 2 below.

Definition 2. Assume that $\hat{\theta}_1$ and $\hat{\theta}_2$ are two consistent estimators for parameter $\theta \in \mathbb{R}^K$ as the sample size N tends to infinity, with $\sqrt{N}(\hat{\theta}_1 - \theta) \rightsquigarrow A_1$ and $\sqrt{N}(\hat{\theta}_2 - \theta) \rightsquigarrow A_2$ for some symmetric random vectors A_1 and A_2. We say

(i) $\hat{\theta}_1$ and $\hat{\theta}_2$ are *asymptotically equally efficient* if A_1 and A_2 have the same distribution, denoted by $\hat{\theta}_1 \overset{\cdot}{\sim} \hat{\theta}_2$;
(ii) $\hat{\theta}_1$ is *asymptotically more efficient* than $\hat{\theta}_2$ if $A_1 \succeq A_2$, denoted by $\hat{\theta}_1 \succeq_\infty \hat{\theta}_2$.

By Definition 2, an asymptotically more efficient scalar estimator has not only a smaller asymptotic variance but also narrower central quantile ranges.

2. Basic setting of the treatment-control experiment

2.1. Regression-based inference under complete randomization

Consider an intervention of two levels, indexed by $q = 0, 1$, and a finite population of N units, indexed by $i = 1, \dots, N$. Let $Y_i(q) \in \mathbb{R}$ be the potential outcome of unit i under treatment level $q \in \{0, 1\}$ (Neyman, 1923; Imbens and Rubin, 2015). The individual treatment effect is $\tau_i = Y_i(1) - Y_i(0)$ for unit i, and the finite-population average treatment effect is $\tau = N^{-1}\sum_{i=1}^N \tau_i = \bar{Y}(1) - \bar{Y}(0)$, where $\bar{Y}(q) = N^{-1}\sum_{i=1}^N Y_i(q)$.

For some prespecified, fixed integer $N_1 > 0$, complete randomization draws a random sample of N_1 units to receive level 1 of the intervention and then assigns the remaining $N_0 = N - N_1 > 0$ units to level 0. Let $e_q = N_q/N$ denote the proportion of units under treatment level $q \in \{0, 1\}$.

Let $Z_i \in \{0, 1\}$ denote the treatment level received by unit i. The observed outcome equals $Y_i = Z_i Y_i(1) + (1 - Z_i) Y_i(0)$. Let $\hat{Y}(q) = N_q^{-1}\sum_{i:Z_i=q} Y_i$ denote the average observed outcome under treatment level $q \in \{0, 1\}$. The difference in means $\hat{\tau}_{\mathrm{N}} = \hat{Y}(1) - \hat{Y}(0)$ is unbiased for τ under complete randomization (Neyman, 1923), and can be computed as the coefficient of Z_i from the simple, unadjusted linear regression $\texttt{lm}(Y_i \sim 1 + Z_i)$ over $i = 1, \dots, N$.

The presence of covariates promises the opportunity to improve estimation efficiency. Let $x_i = (x_{i1}, \dots, x_{iJ})^{\mathrm{T}}$ denote the J pretreatment covariates for unit i, centered at $\bar{x} = N^{-1}\sum_{i=1}^N x_i = 0_J$. Fisher (1935) suggested a covariate-adjusted estimator $\hat{\tau}_{\mathrm{F}}$ for τ, as the coefficient of Z_i from the additive linear regression $\texttt{lm}(Y_i \sim 1 + Z_i + x_i)$ over $i = 1, \dots, N$. Freedman (2008a) criticized the possible efficiency loss by $\hat{\tau}_{\mathrm{F}}$ compared to $\hat{\tau}_{\mathrm{N}}$. Lin (2013) recommended an improved estimator, denoted by $\hat{\tau}_{\mathrm{L}}$, as the coefficient of Z_i from the interacted linear regression $\texttt{lm}(Y_i \sim 1 + Z_i + x_i + Z_i x_i)$ over $i = 1, \dots, N$, and showed its asymptotic efficiency over

$\hat{\tau}_{\text{N}}$ and $\hat{\tau}_{\text{F}}$. In addition, Lin (2013) also showed that the corresponding EHW standard errors are asymptotically conservative for estimating the true standard errors of $\hat{\tau}_{\text{F}}$ and $\hat{\tau}_{\text{L}}$. This justifies large-sample Wald-type inference of τ based on OLS.

We adopt the *design-based* perspective for all theoretical statements in this article, which views the physical act of randomization, as represented by $(Z_i)_{i=1}^N$, as the sole source of randomness in evaluating the sampling properties of quantities of interest. Accordingly, despite the estimators $\hat{\tau}_{\text{N}}$, $\hat{\tau}_{\text{F}}$, and $\hat{\tau}_{\text{L}}$ are all outputs from linear regressions, we invoke no assumption of the corresponding linear models but evaluate the sampling properties of $\hat{\tau}_*$ ($* = \text{N}, \text{F}, \text{L}$) over the distribution of $(Z_i)_{i=1}^N$ conditioning on the potential outcomes and covariates. All theoretical guarantees of the $\hat{\tau}_*$'s are therefore design-based and hold even when the linear models are misspecified; see Negi and Wooldridge (2021) and the references therein for the super-population counterpart.

2.2. Covariate balance and rerandomization

The regression adjustment by Fisher (1935) and Lin (2013) can be viewed as adjusting for imbalances in covariate means. Let $\hat{\tau}_x = \hat{x}(1) - \hat{x}(0)$ denote the difference in covariate means between treatment groups, with $\hat{x}(q) = N_q^{-1}\sum_{i:Z_i=q} x_i$ for $q = 0, 1$. Let $\hat{\gamma}_{\text{F}}$ be the coefficient vector of x_i from the additive regression $\texttt{lm}(Y_i \sim 1 + Z_i + x_i)$ over $i = 1, \dots, N$, and let $\hat{\gamma}_{\text{L}} = e_0\hat{\gamma}_{\text{L},1} + e_1\hat{\gamma}_{\text{L},0}$, where $\hat{\gamma}_{\text{L},q}$ denotes the coefficient vector of x_i from the treatment group-specific regression $\texttt{lm}(Y_i \sim 1 + x_i)$ over $\{i : Z_i = q\}$. Zhao and Ding (2021, Proposition 1) showed that

$$\hat{\tau}_* = \hat{\tau}_{\text{N}} - \hat{\tau}_x^{\text{T}}\hat{\gamma}_* \quad \text{for} \quad * = \text{F}, \text{L}.$$

This expresses $\hat{\tau}_{\text{F}}$ and $\hat{\tau}_{\text{L}}$ as variants of $\hat{\tau}_{\text{N}}$ after adjusting for a linear function of the difference in covariate means.

Rerandomization, on the other hand, enforces covariate balance in the design stage, and accepts an allocation if and only if it satisfies some prespecified covariate balance criterion (Cox, 1982; Morgan and Rubin, 2012). Assume complete randomization for the initial allocation. Morgan and Rubin (2012) and Li et al. (2018) studied a special type of rerandomization, known as ReM, that uses the Mahalanobis distance of $\hat{\tau}_x$ as the balance criterion, and accepts a randomization if and only if $\|\hat{\tau}_x\|_{\mathcal{M}} \le a_0$ for some prespecified threshold a_0. The practical need to avoid small p-values in baseline covariate balance tables instead motivates ReP that accepts a randomization if and only if the p-values from relevant balance tests all exceed some prespecified thresholds. To fill the gap in the literature regarding the theoretical properties of ReP, we examine nine hypothesis testing-based covariate balance criteria for conducting ReP under the completely randomized treatment-control experiment, and quantify their respective impact on subsequent inference from the design-based perspective. We start with three two-sample t-test-based criteria in Section 3 given their direct connections with the balance tables in practice, and extend the discussion to six regression-based alternatives in Section 4. The results provide the basis for generalizations to experiments with more than two treatment arms and stratified experiments, which we formalize in Sections 5 and 6.

3. ReP based on two-sample t-tests

3.1. Marginal, joint, and consensus rules

The difference in covariate means provides an intuitive measure of covariate balance under the treatment-control experiment. Depending on whether we examine the J covariates separately or together, this motivates three two-sample t-test-based criteria for ReP.

To begin with, recall x_{ij} as the jth covariate of unit i. A common approach to balance check is to run one two-sample t-test for each covariate $j \in \{1, \dots, J\}$ based on $(x_{ij}, Z_i)_{i=1}^N$, and use the resulting two-sided p-value, denoted by $p_{j,\text{t}}$, to measure the balance of $(x_{ij})_{i=1}^N$ between treatment groups. This yields J marginal p-values, $\{p_{j,\text{t}} : j = 1, \dots, J\}$, that occupy the last column of the covariate balance tables. An intuitive, and possibly already widely used, criterion for ReP is then to accept a randomization if and only if $p_{j,\text{t}} \ge \alpha_j$ for all $j = 1, \dots, J$ for some prespecified thresholds $\alpha_j \in (0, 1)$ (Bruhn and McKenzie, 2009). We call this the *marginal rule* based on J marginal tests of individual covariates. This generalizes the "big stick" method discussed by Bruhn and McKenzie (2009).

Alternatively, we can test the difference in means of all J covariates together by a multivariate analog of the two-sample t-test, and accept a randomization if and only if the p-value from this joint test exceeds some prespecified threshold. Let $\hat{\Omega}$ be the pooled estimated covariance of $\hat{\tau}_x$. The two-sample Hotelling's T^2 test takes $W_{\text{t}} = \hat{\tau}_x^{\text{T}}\hat{\Omega}^{-1}\hat{\tau}_x$ as the test statistic, and computes a one-sided p-value, denoted by $p_{0,\text{t}}$, by comparing W_{t} against the Hotelling's T^2 distribution. Alternatively, we can replace the Hotelling's T^2 distribution with the asymptotically equivalent χ^2_J distribution and compute the p-value based on the joint Wald test. A *joint rule* then accepts a randomization if and only if $p_{0,\text{t}} \ge \alpha_0$ for some prespecified threshold $\alpha_0 \in (0, 1)$.

In situations where both marginal and joint balances are desired, we can adopt a *consensus rule* that accepts a randomization if and only if it is acceptable under both the marginal and joint rules with $p_{j,\text{t}} \ge \alpha_j$ for all $j = 0, 1, \dots, J$.

Index by "mg", "jt", and "cs" the marginal, joint, and consensus rules, respectively. This defines three ReP schemes by two-sample t-tests, summarized in Definition 3 below. Of interest are their implications on the subsequent inference based on $\hat{\tau}_*$ ($* = \text{N}, \text{F}, \text{L}$). We address this question in Sections 3.2 and 3.3 below.

Definition 3. Assume ReP by two-sample t-tests. Let $\mathcal{A}_{\text{t,mg}} = \{p_{j,\text{t}} \ge \alpha_j \text{ for all } j = 1, \dots, J\}$, $\mathcal{A}_{\text{t,jt}} = \{p_{0,\text{t}} \ge \alpha_0\}$, and $\mathcal{A}_{\text{t,cs}} = \mathcal{A}_{\text{t,mg}} \cap \mathcal{A}_{\text{t,jt}} = \{p_{j,\text{t}} \ge \alpha_j \text{ for all } j = 0, 1, \dots, J\}$ denote the acceptance criteria under the marginal, joint, and consensus rules, respectively.

3.2. Asymptotic theory

We derive in this subsection the asymptotic sampling properties of $\hat{\tau}_*$ $(* = \text{N}, \text{F}, \text{L})$ under the three ReP schemes in Definition 3. The results demonstrate the multiple benefits of ReP in strengthening causal conclusions from experimental data, and elucidate the advantage of Lin (2013)'s method for convenient and efficient inference under ReP.

Let $S_x^2 = (N-1)^{-1}\sum_{i=1}^N x_i x_i^{\mathrm{T}}$ denote the finite-population covariance of the centered $(x_i)_{i=1}^N$. Condition 1 below gives the standard regularity conditions for design-based finite-population asymptotic analysis; see Li and Ding (2017) for a review.

Condition 1. *As $N \to \infty$, (i) $e_q = N_q/N$ has a limit in $(0,1)$ for $q = 0, 1$, (ii) the first two finite-population moments of $\{Y_i(0), Y_i(1), x_i\}_{i=1}^N$ have finite limits; S_x^2 and its limit are both nonsingular, and (iii) $N^{-1}\sum_{i=1}^N Y_i^4(q) = O(1)$ for $q = 0, 1$; $N^{-1}\sum_{i=1}^N \|x_i\|_4^4 = O(1)$.*

Let γ_q be the coefficient vector of x_i from $\texttt{lm}\{Y_i(q) \sim 1 + x_i\}$ over $i = 1, \dots, N$. This is a theoretical fit with $\{Y_i(q)\}_{i=1}^N$ only partially observable depending on the treatment assignment. Condition 1 ensures that e_q, γ_q, and S_x^2 all have finite limits as N tends to infinity. For notational simplicity, we will use the same symbols to denote their respective limits when no confusion would arise. Lemma 1 below follows from Zhao and Ding (2021) and states the asymptotic distributions of $\hat{\tau}_*$ $(* = \text{N}, \text{F}, \text{L})$ under complete randomization. This provides the baseline for evaluating the efficiency gains by ReP.

Lemma 1. *Under complete randomization and Condition 1, we have*

$$\sqrt{N}\begin{pmatrix} \hat{\tau}_* - \tau \\ \hat{\tau}_x \end{pmatrix} \rightsquigarrow \mathcal{N}\left\{0_{J+1}, \begin{pmatrix} v_* & c_*^{\mathrm{T}} \\ c_* & v_x \end{pmatrix}\right\} \qquad (* = \text{N}, \text{F}, \text{L})$$

with $v_x = (e_0 e_1)^{-1} S_x^2$,

$$c_{\text{N}} = S_x^2(e_0^{-1}\gamma_0 + e_1^{-1}\gamma_1), \quad c_{\text{F}} = S_x^2(e_1^{-1} - e_0^{-1})(\gamma_1 - \gamma_0), \quad c_{\text{L}} = 0_J,$$

and $v_* - v_{\text{L}} = c_*^{\mathrm{T}} v_x^{-1} c_* \geq 0$ for $* = \text{N}, \text{F}, \text{L}$. We give the explicit expressions of v_* $(* = \text{N}, \text{F}, \text{L})$ in the Supplementary Material.

Recall α_j $(j = 1, \dots, J)$ and α_0 as the thresholds for the marginal and joint rules. Let a_0 be the $(1-\alpha_0)$th quantile of the χ_J^2 distribution. Let a_j be the $(1-\alpha_j/2)$th quantile of the standard normal distribution, vectorized as $a = (a_1, \dots, a_J)^{\mathrm{T}}$. Let $\epsilon \sim \mathcal{N}(0,1)$ be a standard normal random variable. Let

$$\mathcal{L} \sim \epsilon_0 \mid \{\|\epsilon_0\|_2^2 \leq a_0\}, \quad \mathcal{T}_{\text{t}} \sim \epsilon_{\text{t}} \mid \{|\epsilon_{\text{t}}| \leq a\}, \quad \mathcal{T}_{\text{t}}' \sim \epsilon_{\text{t}} \mid \{|\epsilon_{\text{t}}| \leq a,\ \|\epsilon_{\text{t}}\|_{\mathcal{M}} \leq a_0\} \tag{1}$$

be three truncated normal random vectors independent of ϵ, with $\epsilon_0 \sim \mathcal{N}(0_J, I_J)$ and $\epsilon_{\text{t}} \sim \mathcal{N}\{0_J, D(v_x)\}$. Proposition 1 below gives the asymptotic sampling distributions of $\hat{\tau}_*$ $(* = \text{N}, \text{F}, \text{L})$ under the three ReP schemes in Definition 3. For comparison, we also include the results under ReM to highlight the connection (Zhao and Ding, 2021). Let $\mathcal{A}_{\text{rem}} = \{\|\hat{\tau}_x\|_{\mathcal{M}} \leq a_0\}$ denote the acceptance criterion under ReM with threshold a_0. Let $\hat{\theta} \mid \mathcal{A}$ represent the distribution of $\hat{\theta}$ under rerandomization with acceptance criterion $\mathcal{A}$.

Proposition 1. *Assume Condition 1 and recall the notation in Lemma 1 and Eq. (1). Then*

$$\begin{aligned}
\sqrt{N}(\hat{\tau}_* - \tau) \mid \mathcal{A}_{\text{rem}} &\rightsquigarrow v_{\text{L}}^{1/2}\epsilon + c_*^{\mathrm{T}} v_x^{-1/2}\mathcal{L}, \\
\sqrt{N}(\hat{\tau}_* - \tau) \mid \mathcal{A}_{\text{t,jt}} &\rightsquigarrow v_{\text{L}}^{1/2}\epsilon + c_*^{\mathrm{T}} v_x^{-1/2}\mathcal{L}, \\
\sqrt{N}(\hat{\tau}_* - \tau) \mid \mathcal{A}_{\text{t,mg}} &\rightsquigarrow v_{\text{L}}^{1/2}\epsilon + c_*^{\mathrm{T}} v_x^{-1}\sigma(v_x)\mathcal{T}_{\text{t}}, \\
\sqrt{N}(\hat{\tau}_* - \tau) \mid \mathcal{A}_{\text{t,cs}} &\rightsquigarrow v_{\text{L}}^{1/2}\epsilon + c_*^{\mathrm{T}} v_x^{-1}\sigma(v_x)\mathcal{T}_{\text{t}}'
\end{aligned}$$

for $* = \text{N}, \text{F}$, whereas $\sqrt{N}(\hat{\tau}_{\text{L}} - \tau) \mid \mathcal{A} \rightsquigarrow \mathcal{N}(0, v_{\text{L}})$ for all $\mathcal{A} \in \{\mathcal{A}_{\text{rem}}, \mathcal{A}_{\text{t,jt}}, \mathcal{A}_{\text{t,mg}}, \mathcal{A}_{\text{t,cs}}\}$.

Proposition 1 has two implications. First, all three estimators remain consistent under all four rerandomization schemes, with the joint rule being asymptotically equivalent to ReM. Second, the asymptotic distributions of $\hat{\tau}_{\text{L}}$ remain the same as that under complete randomization in Lemma 1, whereas those of $\hat{\tau}_{\text{N}}$ and $\hat{\tau}_{\text{F}}$ change to convolutions of normal and truncated normal when their respective c_*'s are not 0_J. We show in Lemma S4 in the Supplementary Material that $\mathcal{L} \succeq \epsilon_0$ and $\mathcal{T}_{\text{t}}, \mathcal{T}_{\text{t}}' \succeq \epsilon_{\text{t}}$ by the Gaussian correlation inequality. This provides the basis for quantifying the impact of ReP on the asymptotic efficiency of each $\hat{\tau}_*$, as well as the asymptotic relative efficiency of $\hat{\tau}_*$ across $* = \text{N}, \text{F}, \text{L}$ We formalize the intuition in Theorem 1 below.

Theorem 1. *Assume Condition 1 and define $\rho(J, a_0) = \mathbb{P}(\chi_{J+2} \leq a_0)/\mathbb{P}(\chi_J \leq a_0)$ with $\rho(J, a_0) < 1$. For all $\mathcal{A} \in \{\mathcal{A}_{\text{rem}}, \mathcal{A}_{\text{t,jt}}, \mathcal{A}_{\text{t,mg}}, \mathcal{A}_{\text{t,cs}}\}$, we have*

(i)

$$(\hat{\tau}_x \mid \mathcal{A}) \succeq_\infty \hat{\tau}_x \tag{2}$$

$$\text{with} \quad \frac{\mathbb{E}_{\text{a}}(\|\hat{\tau}_x\|_2^2 \mid \mathcal{A}_{\text{t,jt}})}{\mathbb{E}_{\text{a}}(\|\hat{\tau}_x\|_2^2)} = \frac{\mathbb{E}_{\text{a}}(\|\hat{\tau}_x\|_2^2 \mid \mathcal{A}_{\text{rem}})}{\mathbb{E}_{\text{a}}(\|\hat{\tau}_x\|_2^2)} = \rho(J, a_0);$$

(ii)

$$(\hat{\tau}_{\mathrm{N}} \mid \mathcal{A}) \succeq_{\infty} \hat{\tau}_{\mathrm{N}}, \quad (\hat{\tau}_{\mathrm{F}} \mid \mathcal{A}) \succeq_{\infty} \hat{\tau}_{\mathrm{F}}, \quad (\hat{\tau}_{\mathrm{L}} \mid \mathcal{A}) \dot{\sim} \hat{\tau}_{\mathrm{L}},$$
$$(\hat{\tau}_{\mathrm{L}} \mid \mathcal{A}) \succeq_{\infty} (\hat{\tau}_{*} \mid \mathcal{A}) \quad \text{for} \quad * = \mathrm{N}, \mathrm{F}, \tag{3}$$

with $(\hat{\tau}_{\mathrm{N}} \mid \mathcal{A}) \dot{\sim} \hat{\tau}_{\mathrm{N}} \dot{\sim} \hat{\tau}_{\mathrm{L}}$ if and only if $c_{\mathrm{N}} = 0_J$ and $(\hat{\tau}_{\mathrm{F}} \mid \mathcal{A}) \dot{\sim} \hat{\tau}_{\mathrm{F}} \dot{\sim} \hat{\tau}_{\mathrm{L}}$ if and only if $c_{\mathrm{F}} = 0_J$;

(iii) for $* \in \{\mathrm{N}, \mathrm{F}, \mathrm{L}\}$, the asymptotic conditional bias of $\hat{\tau}_*$ given $\hat{\tau}_x$ satisfies

$$\frac{\mathbb{E}_{\mathrm{a}}\left[\left\{\mathbb{E}_{\mathrm{a}}(\hat{\tau}_* - \tau \mid \hat{\tau}_x, \mathcal{A})\right\}^2\right]}{\mathbb{E}_{\mathrm{a}}\left[\left\{\mathbb{E}_{\mathrm{a}}(\hat{\tau}_* - \tau \mid \hat{\tau}_x)\right\}^2\right]} \leq 1 \tag{4}$$

with $$\frac{\mathbb{E}_{\mathrm{a}}\left[\left\{\mathbb{E}_{\mathrm{a}}(\hat{\tau}_* - \tau \mid \hat{\tau}_x, \mathcal{A}_{\mathrm{t,jt}})\right\}^2\right]}{\mathbb{E}_{\mathrm{a}}\left[\left\{\mathbb{E}_{\mathrm{a}}(\hat{\tau}_* - \tau \mid \hat{\tau}_x)\right\}^2\right]} = \frac{\mathbb{E}_{\mathrm{a}}\left[\left\{\mathbb{E}_{\mathrm{a}}(\hat{\tau}_* - \tau \mid \hat{\tau}_x, \mathcal{A}_{\mathrm{rem}})\right\}^2\right]}{\mathbb{E}_{\mathrm{a}}\left[\left\{\mathbb{E}_{\mathrm{a}}(\hat{\tau}_* - \tau \mid \hat{\tau}_x)\right\}^2\right]} = \rho(J, a_0);$$

(iv) for $* \neq ** \in \{\mathrm{N}, \mathrm{F}, \mathrm{L}\}$,

$$\frac{\mathbb{E}_{\mathrm{a}}\left\{(\hat{\tau}_* - \hat{\tau}_{**})^2 \mid \mathcal{A}\right\}}{\mathbb{E}_{\mathrm{a}}\left\{(\hat{\tau}_* - \hat{\tau}_{**})^2\right\}} \leq 1 \tag{5}$$

with $$\frac{\mathbb{E}_{\mathrm{a}}\left\{(\hat{\tau}_* - \hat{\tau}_{**})^2 \mid \mathcal{A}_{\mathrm{t,jt}}\right\}}{\mathbb{E}_{\mathrm{a}}\left\{(\hat{\tau}_* - \hat{\tau}_{**})^2\right\}} = \frac{\mathbb{E}_{\mathrm{a}}\left\{(\hat{\tau}_* - \hat{\tau}_{**})^2 \mid \mathcal{A}_{\mathrm{rem}}\right\}}{\mathbb{E}_{\mathrm{a}}\left\{(\hat{\tau}_* - \hat{\tau}_{**})^2\right\}} = \rho(J, a_0).$$

For a random quantity $\hat{\theta}$, $(\hat{\theta} \mid \mathcal{A}) \succeq_{\infty} \hat{\theta}$ implies that rerandomization increases the asymptotic peakedness of $\hat{\theta}$, whereas $(\hat{\theta} \mid \mathcal{A}) \dot{\sim} \hat{\theta}$ implies that rerandomization has no effect asymptotically. The implications of Theorem 1 are hence threefold. First, Theorem 1(i) establishes the utility of the two-sample t-test-based ReP to improve covariate balance in terms of the asymptotic distribution of $\hat{\tau}_x$. This gives another measure of improved covariate balance in addition to the self-evident improvement in the realized allocation. Second, Theorem 1(ii) shows the utility of ReP in improving the asymptotic efficiency of $\hat{\tau}_{\mathrm{N}}$ and $\hat{\tau}_{\mathrm{F}}$, and ensures the asymptotic efficiency of $\hat{\tau}_{\mathrm{L}}$ over $\hat{\tau}_{\mathrm{N}}$ and $\hat{\tau}_{\mathrm{F}}$ under all three ReP schemes with the asymptotic efficiency unaffected by rerandomization. We thus recommend using $\hat{\tau}_{\mathrm{L}}$ for inference under ReP, with details given in Section 3.3. Third, Theorem 1(iii)–(iv) are direct consequences of Theorem 1(i) and illustrate the utility of ReP in reducing conditional biases and improving coherence across $\hat{\tau}_*$ $(* = \mathrm{N}, \mathrm{F}, \mathrm{L})$. In particular, we use $\mathbb{E}_{\mathrm{a}}(\hat{\tau}_* - \tau \mid \hat{\tau}_x, \mathcal{A})$ and $\mathbb{E}_{\mathrm{a}}(\hat{\tau}_* - \tau \mid \hat{\tau}_x)$ to measure the asymptotic conditional biases of $\hat{\tau}_*$ given $\hat{\tau}_x$ under rerandomization and complete randomization, respectively, in Theorem 1(iii), and use $\mathbb{E}_{\mathrm{a}}\{(\hat{\tau}_* - \hat{\tau}_{**})^2 \mid \mathcal{A}\}$ and $\mathbb{E}_{\mathrm{a}}\{(\hat{\tau}_* - \hat{\tau}_{**})^2\}$ to measure the coherence between estimators $\hat{\tau}_*$ and $\hat{\tau}_{**}$ in Theorem 1(iv).

These implications together illustrate the value of ReP: despite having no effect on the asymptotic efficiency of $\hat{\tau}_{\mathrm{L}}$, ReP promotes not only covariate balance between treatment groups but also more coherent inferences across different estimators. The combination of ReP and Lin (2013)'s estimator therefore results in both covariate balance and efficient inference. The existing theory on rerandomization focuses on the efficiency gain by ReM analogous to Theorem 1(ii). Here we give more comprehensive results about the multiple benefits of ReP.

The asymptotic equivalence of $(\hat{\tau}_{\mathrm{L}} \mid \mathcal{A})$ and $\hat{\tau}_{\mathrm{L}}$ for $\mathcal{A} \in \{\mathcal{A}_{\mathrm{rem}}, \mathcal{A}_{\mathrm{t,jt}}, \mathcal{A}_{\mathrm{t,mg}}, \mathcal{A}_{\mathrm{t,cs}}\}$ in Proposition 1 and Theorem 1(ii) is no coincidence but the consequence of $\hat{\tau}_{\mathrm{L}}$ being asymptotically independent of $\hat{\tau}_x$ under complete randomization (c.f. Lemma 1). Balance criteria based on $\hat{\tau}_x$ thus have no effect on $\hat{\tau}_{\mathrm{L}}$ asymptotically, with $\mathcal{A}_{\mathrm{rem}}, \mathcal{A}_{\mathrm{t,jt}}, \mathcal{A}_{\mathrm{t,mg}}$, and $\mathcal{A}_{\mathrm{t,cs}}$ all being special cases. The same argument underpins the asymptotic equivalence of $\hat{\tau}_*$ and $(\hat{\tau}_* \mid \mathcal{A})$ for $* \in \{\mathrm{N}, \mathrm{F}\}$ when $c_* = 0_J$ under special configurations of the potential outcomes; an example is $c_{\mathrm{F}} = 0_J$ when the individual treatment effects τ_i are constant across all i's. The resulting $\hat{\tau}_*$ $(* \in \{\mathrm{N}, \mathrm{F}\})$ is asymptotically identically distributed as $\hat{\tau}_{\mathrm{L}}$ under complete randomization by Lemma 1, with the asymptotic sampling distribution unaffected by rerandomization.

More generally, the linear projection of $\hat{\tau}_*$ on $\hat{\tau}_x$ equals $\mathrm{proj}(\hat{\tau}_* \mid \hat{\tau}_x) = \tau + c_*^{\mathrm{T}} v_x^{-1} \hat{\tau}_x$ with regard to the asymptotic distribution under complete randomization in Lemma 1, and is asymptotically independent of the corresponding residual, denoted by $\mathrm{res}(\hat{\tau}_* \mid \hat{\tau}_x) = \hat{\tau}_* - \mathrm{proj}(\hat{\tau}_* \mid \hat{\tau}_x) = \hat{\tau}_* - \tau - c_*^{\mathrm{T}} v_x^{-1} \hat{\tau}_x$. This ensures

$$\hat{\tau}_* = \mathrm{proj}(\hat{\tau}_* \mid \hat{\tau}_x) + \mathrm{res}(\hat{\tau}_* \mid \hat{\tau}_x) = \tau + c_*^{\mathrm{T}} v_x^{-1} \hat{\tau}_x + \mathrm{res}(\hat{\tau}_* \mid \hat{\tau}_x), \tag{6}$$

where $\mathrm{res}(\hat{\tau}_* \mid \hat{\tau}_x)$ satisfies $\sqrt{N}\mathrm{res}(\hat{\tau}_* \mid \hat{\tau}_x) \rightsquigarrow \mathcal{N}(0, v_{\mathrm{L}})$ and is asymptotically independent of $\hat{\tau}_x$. Balance criteria based on $\hat{\tau}_x$ can thus only affect the $c_*^{\mathrm{T}} v_x^{-1} \hat{\tau}_x$ part in Eq. (6) asymptotically, and turn it into a truncated normal with greater peakedness when $c_* \neq 0_J$. This gives the intuition behind Proposition 1 and Theorem 1.

3.3. Wald-type inference

Proposition 1 and Theorem 1 together establish the asymptotic distributions and relative efficiency of $\hat{\tau}_*$ $(* = \mathrm{N}, \mathrm{F}, \mathrm{L})$ under the two-sample t-test-based ReP. The results provide two guidelines on subsequent Wald-type inference of the average treatment effect.

First, the Wald-type inference based on $\hat{\tau}_{\mathrm{L}}$ is asymptotically the most efficient and can be conducted using the same normal approximation as under complete randomization. Specifically, let $\hat{\mathrm{se}}_{\mathrm{L}}$ denote the EHW standard error of $\hat{\tau}_{\mathrm{L}}$ from the same OLS fit. Lin (2013) and Li et al. (2018, Lemma A16) ensured that it is asymptotically appropriate for estimating the true standard error of $\hat{\tau}_{\mathrm{L}}$

Table 1
Nine ReP schemes under the treatment-control experiment.

Rule	Model option: $\dagger = \mathrm{t}, \mathrm{lm}, \mathrm{logit}$
marginal (mg)	$p_{j,\dagger} \geq \alpha_j \ (j = 1, \dots, J)$
joint (jt)	$p_{0,\dagger} \geq \alpha_0$
consensus (cs)	$p_{j,\dagger} \geq \alpha_j \ (j = 0, 1, \dots, J)$

under both complete randomization and the three ReP schemes in Definition 3, justifying the Wald-type inference based on $(\hat{\tau}_{\mathrm{L}}, \hat{\mathrm{se}}_{\mathrm{L}})$ and normal approximation. This illustrates the advantage of ReP for allowing for convenient regression-assisted analysis by the interacted regression.

Second, ReP narrows the asymptotic sampling distributions of $\hat{\tau}_{\mathrm{N}}$ and $\hat{\tau}_{\mathrm{F}}$ in general, rendering standard inference procedures based on their EHW standard errors and normal approximation overconservative. As an illustration, denote by $\hat{\mathrm{se}}_*$ the EHW standard error of $\hat{\tau}_*$ for $* = \mathrm{N}, \mathrm{F}$ and by $z_{1-\alpha/2}$ the $100(1-\alpha/2)\%$ quantile of standard normal. The standard $100(1-\alpha)\%$ confidence interval based on normal approximation equals $\hat{\tau}_* \pm z_{1-\alpha/2} \times \hat{\mathrm{se}}_*$ for $* = \mathrm{N}, \mathrm{F}$, and is overconservative under ReP. Rerandomization-specific sampling distributions are thus necessary for better-calibrated inference based on $\hat{\tau}_*$ $(* = \mathrm{N}, \mathrm{F})$. Recall $(\hat{\gamma}_{\mathrm{L},0}, \hat{\gamma}_{\mathrm{L},1})$ as the sample analogs of (γ_0, γ_1) from Section 2.2. With v_{L} and c_* being the only unknowns in the asymptotic distributions of $\hat{\tau}_*$ $(* = \mathrm{N}, \mathrm{F})$ in Proposition 1, we can estimate them using $\hat{v}_{\mathrm{L}} = N\hat{\mathrm{se}}_{\mathrm{L}}^2$ and the sample analogs $\hat{c}_{\mathrm{N}} = S_x^2(e_0^{-1}\hat{\gamma}_{\mathrm{L},0} + e_1^{-1}\hat{\gamma}_{\mathrm{L},1})$ and $\hat{c}_{\mathrm{F}} = S_x^2(e_1^{-1} - e_0^{-1})(\hat{\gamma}_{\mathrm{L},1} - \hat{\gamma}_{\mathrm{L},0})$, respectively, and conduct inference based on the resulting plug-in distributions (Li et al., 2018). Specifically, we can generate a large number of independent draws from the plug-in sampling distribution, and use the empirical quantiles to approximate the true quantiles for constructing confidence intervals. As an illustration, the plug-in distribution of $\hat{\tau}_{\mathrm{N}}$ under the marginal rule is $\sqrt{N}(\hat{\tau}_{\mathrm{N}} - \tau) \mid \mathcal{A}_{\mathrm{t,mg}} \rightsquigarrow \hat{v}_{\mathrm{L}}^{1/2}\epsilon + \hat{c}_{\mathrm{N}}^{\mathrm{T}} v_x^{-1}\sigma(v_x)\mathcal{T}_{\mathrm{t}}$. We can use a large number of independent draws of ϵ and $\mathcal{T}_{\mathrm{t}}$ to simulate the distribution of $\hat{v}_{\mathrm{L}}^{1/2}\epsilon + \hat{c}_{\mathrm{N}}^{\mathrm{T}} v_x^{-1}\sigma(v_x)\mathcal{T}_{\mathrm{t}}$. Let $\hat{q}_{1-\alpha/2}$ denote the $100(1-\alpha/2)\%$ quantile of this empirical distribution. Then $\hat{\tau}_{\mathrm{N}} \pm q_{1-\alpha/2}/\sqrt{N}$ gives an approximate $100(1-\alpha)\%$ confidence interval of τ. This modification mitigates the overconservativeness of the Wald-type inference based on normal approximation at the cost of additional computational efforts. This, by contrast, illustrates the convenience of Lin (2013)'s method for efficient and well-calibrated inference under ReP.

4. ReP based on linear and logistic regressions

4.1. Linear and logistic regressions for assessing covariate balance

The two-sample t-tests measure covariate balance by the difference in covariate means and are numerically equivalent to a component-wise regression of x_i on $(1, Z_i)$, assessing how x_i varies with different values of Z_i. The idea of the propensity score (Rosenbaum and Rubin, 1983), on the other hand, motivates an alternative measure of covariate balance by assessing how Z_i varies with x_i.

Consider the linear regression of Z_i on $(1, x_i)$, denoted by $\mathtt{lm}(Z_i \sim 1 + x_{i1} + \cdots + x_{iJ})$. Let $\hat{\beta}_j$ denote the coefficient of the jth covariate x_{ij} for $j = 1, \dots, J$. The magnitude of $\hat{\beta}_j$ gives an intuitive measure of the influence of covariate j on the treatment assignment, with a well-balanced assignment expected to have all $\hat{\beta}_j$'s close to zero; see, e.g., de Mel et al. (2009, Table 1) and Kuziemko et al. (2015, Table 3) for balance tables based on these regression outputs. This motivates three linear regression-based ReP schemes under the marginal, joint, and consensus rules, respectively.

To begin with, denote by $p_{j,\mathrm{lm}}$ the p-value associated with $\hat{\beta}_j$ from standard software packages. The marginal rule accepts a randomization if and only if $p_{j,\mathrm{lm}} \geq \alpha_j$ for all $j = 1, \dots, J$ for some prespecified thresholds $\alpha_j \in (0, 1)$.

Alternatively, let $p_{0,\mathrm{lm}}$ be the p-value from the F-test of $\mathtt{lm}(Z_i \sim 1 + x_{i1} + \cdots + x_{iJ})$ against the empty model $\mathtt{lm}(Z_i \sim 1)$. It is a standard output of linear regression by most software packages and provides a summary measure of the magnitudes of all $\hat{\beta}_j$'s. The joint rule then accepts a randomization if and only if $p_{0,\mathrm{lm}} \geq \alpha_0$ for some prespecified threshold $\alpha_0 \in (0, 1)$. This is the recommendation by Gerber and Green (2012).

The consensus rule, accordingly, accepts a randomization if and only if it is acceptable under both marginal and joint rules with $p_{j,\mathrm{lm}} \geq \alpha_j$ for all $j = 0, 1, \dots, J$. This extends the three two-sample t-test-based criteria in Definition 3 to the linear regression of Z_i on $(1, x_i)$.

One concern with the above approach based on $\mathtt{lm}(Z_i \sim 1 + x_{i1} + \cdots + x_{iJ})$ is that linear regression is not intended for binary responses like Z_i. An immediate alternative is to consider the logistic regression of Z_i on $(1, x_i)$, denoted by $\mathtt{logit}(Z_i \sim 1 + x_{i1} + \cdots + x_{iJ})$, instead and form acceptance criteria based on p-values from its MLE fit (Hansen and Bowers, 2008).

Specifically, let $p_{j,\mathrm{logit}}$ be the p-value associated with the coefficient of x_{ij} from $\mathtt{logit}(Z_i \sim 1 + x_{i1} + \cdots + x_{iJ})$ for $j = 1, \dots, J$, and let $p_{0,\mathrm{logit}}$ be the p-value from the likelihood ratio test (LRT) of $\mathtt{logit}(Z_i \sim 1 + x_{i1} + \cdots + x_{iJ})$ against the empty model $\mathtt{logit}(Z_i \sim 1)$. They are all standard outputs of logistic regression by most software packages, and allow us to form the marginal, joint, and consensus criteria in identical ways as those based on $\{p_{j,\dagger} : j = 0, 1, \dots, J\}$ for $\dagger = \mathrm{t}, \mathrm{lm}$.

This defines in total nine ReP schemes, as the combinations of three *model options*—the two-sample t-tests of x_i ("t"), the linear regression of Z_i on $(1, x_i)$ ("lm"), and the logistic regression of Z_i on $(1, x_i)$ ("logit")—and the marginal ("mg"), joint ("jt"), and consensus ("cs") rules, summarized in Table 1. We extend below the results under the two-sample t-test-based schemes to the regression-based variants.

4.2. Asymptotic theory

We derive in this subsection the asymptotic sampling properties of $\hat{\tau}_*$ ($* = \text{N}, \text{F}, \text{L}$) under the six linear or logistic regression-based ReP schemes in Table 1. Echoing the comments after Theorem 1, the result illustrates the utility of the six regression-based ReP schemes in improving covariate balance, reducing conditional biases, and promoting coherence across different estimators, and establishes the asymptotic efficiency of $\hat{\tau}_{\text{L}}$ under all six schemes with the asymptotic sampling distribution unaffected by ReP.

To this end, we first introduce an additional regularity condition that underpins the design-based properties of logistic regression. Let

$$\pi(\tilde{x}_i, \theta) = \frac{\exp(\tilde{x}_i^{\mathrm{T}}\theta)}{1 + \exp(\tilde{x}_i^{\mathrm{T}}\theta)} \qquad \text{for} \quad \tilde{x}_i = (1, x_i^{\mathrm{T}})^{\mathrm{T}} \text{ and } \theta \in \mathbb{R}^{J+1}.$$

Then $H(\theta) = -N^{-1}\sum_{i=1}^N \pi(\tilde{x}_i, \theta)\{1 - \pi(\tilde{x}_i, \theta)\}\tilde{x}_i\tilde{x}_i^{\mathrm{T}}$ gives the Hessian matrix of the log-likelihood function under the logistic model scaled by N^{-1}; we suppress the dependence of $H(\theta)$ on N.

Condition 2. *As $N \to \infty$, $H(\theta)$ converges to a negative-definite matrix $H_\infty(\theta) < 0$ for all $\theta \in \mathbb{R}^{J+1}$, and the convergence is uniform over θ on any compact set $\Theta \subset \mathbb{R}^{J+1}$.*

To gain intuition about the uniform convergence assumption in Condition 2, consider a superpopulation working model where the x_i's are independent and identically distributed with finite second moment. The uniform law of large numbers ensures that $H(\theta)$ converges to some $H_\infty(\theta)$ almost surely on $\mathbb{R}^{J+1}$, with the convergence being uniform on any compact set $\Theta \subset \mathbb{R}^{J+1}$ (Newey and McFadden, 1994; Ferguson, 1996, Chapter 16). Accordingly, if the finite population is indeed an independent and identically distributed random sample from some superpopulation with finite second moment, then the uniform convergence holds. This suggests the mildness of our assumption on uniform convergence. In addition, we can show that $H(\theta) \le 0$ for all $\theta \in \mathbb{R}^{J+1}$. This suggests the mildness of the assumption on the negative definiteness of $H_\infty(\theta)$.

For $\dagger = \text{lm}, \text{logit}$, let $\mathcal{A}_{\dagger,\text{jt}} = \{p_{0,\dagger} \ge \alpha_0\}$, $\mathcal{A}_{\dagger,\text{mg}} = \{p_{j,\dagger} \ge \alpha_j \text{ for all } j = 1, \ldots, J\}$, and $\mathcal{A}_{\dagger,\text{cs}} = \mathcal{A}_{\dagger,\text{jt}} \cap \mathcal{A}_{\dagger,\text{mg}} = \{p_{j,\dagger} \ge \alpha_j \text{ for all } j = 0, 1, \ldots, J\}$ denote the acceptance criteria under the six regression-based ReP schemes. Recall the definitions of v_x, ϵ, $\mathcal{L}$, a_0, and $a = (a_1, \ldots, a_J)^{\mathrm{T}}$ from Section 3.2. Let

$$\mathcal{T}_{\text{lm}} \sim \epsilon_{\text{lm}} \mid \{|\epsilon_{\text{lm}}| \le a\}, \qquad \mathcal{T}'_{\text{lm}} \sim \epsilon_{\text{lm}} \mid \{|\epsilon_{\text{lm}}| \le a,\ \|\epsilon_{\text{lm}}\|_{\mathcal{M}} \le a_0\} \tag{7}$$

be two truncated normal random vectors independent of ϵ, with $\epsilon_{\text{lm}} \sim \mathcal{N}\{0_J, D(v_x^{-1})\}$. The Gaussian correlation inequality ensures that $\mathcal{T}_{\text{lm}}, \mathcal{T}'_{\text{lm}} \succeq \epsilon_{\text{lm}}$; see Lemma S4 in the Supplementary Material. This underlies the improved asymptotic efficiency of $\hat{\tau}_{\text{N}}$ and $\hat{\tau}_{\text{F}}$ under regression-based ReP. We state the details in Proposition 2 and Theorem 2 below.

Proposition 2. *Assume Condition 1 for $\dagger = \text{lm}$ and Conditions 1–2 for $\dagger = \text{logit}$. Recall the notation in Lemma 1, Eq. (1), and Eq. (7). For $\dagger \in \{\text{lm}, \text{logit}\}$, we have*

$$\begin{aligned}
\sqrt{N}(\hat{\tau}_* - \tau) \mid \mathcal{A}_{\dagger,\text{jt}} &\rightsquigarrow v_{\text{L}}^{1/2}\epsilon + c_*^{\mathrm{T}} v_x^{-1/2}\mathcal{L}, \\
\sqrt{N}(\hat{\tau}_* - \tau) \mid \mathcal{A}_{\dagger,\text{mg}} &\rightsquigarrow v_{\text{L}}^{1/2}\epsilon + c_*^{\mathrm{T}} \sigma(v_x^{-1})\mathcal{T}_{\text{lm}}, \\
\sqrt{N}(\hat{\tau}_* - \tau) \mid \mathcal{A}_{\dagger,\text{cs}} &\rightsquigarrow v_{\text{L}}^{1/2}\epsilon + c_*^{\mathrm{T}} \sigma(v_x^{-1})\mathcal{T}'_{\text{lm}},
\end{aligned}$$

for $* = \text{N}, \text{F}$, whereas $\sqrt{N}(\hat{\tau}_{\text{L}} - \tau) \mid \mathcal{A} \rightsquigarrow \mathcal{N}(0, v_{\text{L}})$ for all $\mathcal{A} \in \{\mathcal{A}_{\dagger,\diamond} : \dagger = \text{lm}, \text{logit};\ \diamond = \text{jt}, \text{mg}, \text{cs}\}$.

Theorem 2. *Theorem 1 holds for all $\mathcal{A} \in \{\mathcal{A}_{\dagger,\diamond} : \dagger = \text{lm}, \text{logit};\ \diamond = \text{jt}, \text{mg}, \text{cs}\}$.*

All comments after Proposition 1 and Theorem 1 extend here with no need of modification. Proposition 2 gives the asymptotic sampling distributions of $\hat{\tau}_*$ ($* = \text{N}, \text{F}, \text{L}$) under the six regression-based ReP schemes, and establishes the asymptotic equivalence of the linear and logistic regression model options under all three rules. As a direct implication of Proposition 2, Theorem 2 highlights the utility of the regression-based ReP in improving covariate balance, reducing conditional biases, and promoting coherence between adjusted and unadjusted analyses, and ensures the asymptotic efficiency of $\hat{\tau}_{\text{L}}$ under all six schemes. The interacted regression is thus our recommendation for subsequent inference under the linear and logistic regression-based ReP as well, with all discussion in Section 3.3 extending here verbatim.

Juxtapose Proposition 2 with Proposition 1. The three joint criteria are asymptotically equivalent to not only each other but also ReM with threshold a_0. This is no coincidence but the consequence of the test statistics used by these criteria all being asymptotically equivalent to $\|\hat{\tau}_x\|_{\mathcal{M}}$; see Remark S2 in the Supplementary Material for details. The marginal criteria based on the linear and logistic regressions, on the other hand, differ from that based on the two-sample t-tests even asymptotically. The difference is nevertheless immaterial based on simulation evidence.

Echoing the comments at the end of Section 2.1, we view the linear and logistic regressions as purely numeric procedures based on OLS or MLE for computing the p-values and estimators, and invoke none of the underlying modeling assumptions in evaluating the outputs. The results in Proposition 2 and Theorem 2 therefore hold regardless of how well (i) the linear and logistic models underlying $\texttt{lm}(Z_i \sim 1 + x_i)$ and $\texttt{logit}(Z_i \sim 1 + x_i)$ represent the true treatment assignment mechanism and (ii) the linear models underlying $\texttt{lm}(Y_i \sim 1 + Z_i)$, $\texttt{lm}(Y_i \sim 1 + Z_i + x_i)$, and $\texttt{lm}(Y_i \sim 1 + Z_i + x_i + Z_i x_i)$ represent the true outcome model. This

Table 2
Four ReP schemes under multi-armed experiments.

Rule	F-test	Multinomial logistic regression
marginal	$p_{j,\mathrm{f}} \geq \alpha_j$ for all $j = 1, \dots, J$	$p_{qj,\mathrm{logit}} \geq \alpha_{qj}$ for all qj
joint	n.a.	$p_{0,\mathrm{logit}} \geq \alpha_0$
consensus	n.a.	$p_{qj,\mathrm{logit}} \geq \alpha_{qj}$ for all qj, $p_{0,\mathrm{logit}} \geq \alpha_0$

concludes our discussion on ReP under the treatment-control experiment. The criteria based on two-sample t-tests are arguably the most straightforward, making the discussion on the regression-based variants seem to be of theoretical interest only. The logistic regression nevertheless provides a key stepping stone for extending the current results to experiments with more than two treatment arms. We give the details in the next section.

5. ReP in multi-armed experiments

5.1. Basic setting and covariate balance criteria

Multi-armed experiments enable comparisons of more than two treatment levels simultaneously, and are intrinsic to applications with multiple factors of interest. To conduct rerandomization in such settings, a straightforward option is to check balance for all pairs of treatment arms, and accept a randomization if and only if all pairwise comparisons pass the balance check (Morgan, 2011). Depending on the number of treatment arms in question, however, this may result in a large number of pairwise comparisons and become unwieldy in practice. A more practical alternative is to use a test that directly measures the balance across all treatment arms.

To this end, the covariate-wise F-test provides a natural way of extending the marginal two-sample t-test to more than two treatment arms, measuring the balance of individual covariates across all treatment arms simultaneously. See, e.g., de Mel et al. (2013) and Dupas and Robinson (2013) for balance tables based on covariate-wise F-tests. The multinomial logistic regression, on the other hand, is a straightforward extension of the logistic regression and provides a way to measure both covariate-wise and overall balances across all treatment arms by the idea of the propensity score. See Gerber et al. (2009) for an example of balance check based on the multinomial logistic regression. We formalize below their extensions to ReP.

Consider a multi-armed experiment with $Q > 2$ treatment levels, indexed by $q \in \mathcal{Q} = \{1, \dots, Q\}$, and a study population of N units, indexed by $i = 1, \dots, N$. Renew $x_i = (x_{i1}, \dots, x_{iJ})^{\mathrm{T}}$ as the centered covariate vector and $Z_i \in \mathcal{Q} = \{1, \dots, Q\}$ as the initial treatment assignment of unit i. For $j = 1, \dots, J$, let $p_{j,\mathrm{f}}$ denote the p-value from the marginal F-test on covariate j based on $(x_{ij}, Z_i)_{i=1}^N$. The marginal F-test-based criterion for ReP accepts a randomization if and only if $p_{j,\mathrm{f}} \geq \alpha_j$ for all $j = 1, \dots, J$ for some prespecified thresholds $\alpha_j \in (0, 1)$. Let $\mathcal{I}_{iq} = 1(Z_i = q)$ denote the indicator of treatment level q. The $p_{j,\mathrm{f}}$ can also be computed as the p-value from the F-test of $\texttt{lm}(x_{ij} \sim 1 + \mathcal{I}_{i1} + \dots + \mathcal{I}_{i,Q-1})$ against the empty model $\texttt{lm}(x_{ij} \sim 1)$.

The multinomial logistic regression, on the other hand, accommodates the marginal, joint, and consensus rules for ReP together via one MLE fit. Renew $\texttt{logit}(Z_i \sim 1 + x_{i1} + \dots + x_{iJ})$ as the multinomial logistic regression of $Z_i \in \mathcal{Q}$ on $(1, x_i)$ over $i = 1, \dots, N$. Assume without loss of generality level Q as the reference level. The MLE fit of $\texttt{logit}(Z_i \sim 1 + x_{i1} + \dots + x_{iJ})$ yields one coefficient of x_{ij} for each non-reference level $q \in \mathcal{Q}_+ = \{1, \dots, Q-1\}$, denoted by $\tilde{\beta}_{qj}$. We use the subscript $+$ to signify quantities associated with the non-reference levels. Let $p_{qj,\mathrm{logit}}$ be the p-value associated with $\tilde{\beta}_{qj}$ from standard software packages. The marginal rule accepts a randomization if and only if $p_{qj,\mathrm{logit}} \geq \alpha_{qj}$ for all $q \in \mathcal{Q}_+$ and $j = 1, \dots, J$ for some prespecified thresholds $\alpha_{qj} \in (0, 1)$.

Alternatively, let $p_{0,\mathrm{logit}}$ be the p-value from the LRT of $\texttt{logit}(Z_i \sim 1 + x_{i1} + \dots + x_{iJ})$ against the empty model $\texttt{logit}(Z_i \sim 1)$. It is a standard output of multinomial logistic regression from most software packages and gives a summary measure of the magnitudes of $\tilde{\beta}_{qj}$'s as a whole. The joint rule then accepts a randomization if and only if $p_{0,\mathrm{logit}} \geq \alpha_0$ for some prespecified threshold $\alpha_0 \in (0, 1)$. The consensus rule, accordingly, accepts a randomization if and only if it is acceptable under both the marginal and joint rules. This defines three additional criteria for conducting ReP under multi-armed experiments. We summarize the definitions in Table 2.

Other criteria can be formed based on tests for multivariate analysis of variance (Morgan, 2011) or linear regression of $\mathcal{I}_{iq} = 1(Z_i = q)$ on $(1, x_i)$. These alternatives in general involve more technical subtleties and can be unwieldy in practice. To save space, we will focus on ReP based on marginal F-tests and multinomial logistic regression in the main paper due to their practical convenience, and relegate details on the alternative criteria to the Supplementary Material.

5.2. Treatment effects and regression estimators

We next define the average treatment effect and regression estimators under multi-armed experiments, extending the notation and definitions in Section 2. Renew $Y_i(q) \in \mathbb{R}$ as the potential outcome of unit i if assigned to treatment level $q \in \mathcal{Q} = \{1, \dots, Q\}$. The observed outcome equals $Y_i = \sum_{q \in \mathcal{Q}} \mathcal{I}_{iq} Y_i(q)$ with $\mathcal{I}_{iq} = 1(Z_i = q)$. Renew $\bar{Y}(q) = N^{-1} \sum_{i=1}^N Y_i(q)$ as the average potential outcome under treatment level $q \in \mathcal{Q}$, vectorized as $\bar{Y} = (\bar{Y}(1), \dots, \bar{Y}(Q))^{\mathrm{T}} \in \mathbb{R}^Q$. The goal is to estimate the finite-population average treatment effect

$$\tau = G\bar{Y}$$

for some prespecified contrast matrix G with all row sums equal to zero. The $\tau = \bar{Y}(1) - \bar{Y}(0)$ under the treatment-control experiment is a special case with $\bar{Y} = (\bar{Y}(1), \bar{Y}(0))^{\mathrm{T}}$ and $G = (1, -1)$.

Define

$$\text{N}: \ \texttt{lm}(Y_i \sim \mathcal{I}_{i1} + \cdots + \mathcal{I}_{iQ}),$$
$$\text{F}: \ \texttt{lm}(Y_i \sim \mathcal{I}_{i1} + \cdots + \mathcal{I}_{iQ} + x_i),$$
$$\text{L}: \ \texttt{lm}(Y_i \sim \mathcal{I}_{i1} + \cdots + \mathcal{I}_{iQ} + \mathcal{I}_{i1} x_i + \cdots + \mathcal{I}_{iQ} x_i)$$

as the unadjusted, additive, and interacted linear regressions of Y_i on $\{\mathcal{I}_{iq} : q \in \mathcal{Q}\}$ and x_i, respectively, indexed by $* = \text{N}$ (unadjusted), F (additive), and L (interacted). Let

$$\hat{Y}_* = (\hat{Y}_*(1), \ldots, \hat{Y}_*(Q))^{\mathrm{T}} \quad (* = \text{N}, \text{F}, \text{L})$$

denote the coefficient vectors of $(\mathcal{I}_{i1}, \ldots, \mathcal{I}_{iQ})^{\mathrm{T}}$ from these three regressions, respectively. They are consistent for estimating $\bar{Y}$ under complete randomization (Lu, 2016; Zhao and Ding, 2023), and allow us to estimate $\tau = G\bar{Y}$ by

$$\hat{\tau}_* = G\hat{Y}_* \qquad (* = \text{N}, \text{F}, \text{L}).$$

Of interest are the validity and relative efficiency of $\hat{\tau}_*$'s under ReP. We give the details in the following.

5.3. Asymptotic theory

We present in this subsection the asymptotic theory of ReP under multi-armed experiments. Assume throughout that the initial allocation is obtained by complete randomization. The experimenter assigns completely at random $N_q > 0$ units to level $q \in \mathcal{Q}$ with $\sum_{q \in \mathcal{Q}} N_q = N$, and accepts the allocation if and only if the assignments satisfy the prespecified covariate balance criterion.

5.3.1. Baseline efficiency under complete randomization

Recall the definitions of $\hat{x}(q) = N_q^{-1} \sum_{i: Z_i = q} x_i$, $e_q = N_q / N$, γ_q, and Condition 1 in Sections 2–3 under the treatment-control experiment. Renew them for multi-armed experiments with $q \in \mathcal{Q} = \{1, \ldots, Q\}$. Let $\hat{x} = (\hat{x}(1)^{\mathrm{T}}, \ldots, \hat{x}(Q)^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{JQ}$ and $\gamma_{\text{F}} = \sum_{q \in \mathcal{Q}} e_q \gamma_q$. Lemma 2 below follows from Zhao and Ding (2023) and states the asymptotic distributions of $\hat{Y}_*$ $(* = \text{N}, \text{F}, \text{L})$ under complete randomization. The results ensure $\hat{\tau}_{\text{L}} \succeq_\infty \hat{\tau}_{\text{N}}, \hat{\tau}_{\text{F}}$ under complete randomization and provide the baseline for evaluating the efficiency gains under ReP. Let $1_{Q \times Q}$ denote the $Q \times Q$ matrix of all ones.

Lemma 2. *Under complete randomization and the multi-armed version of Condition 1, we have*

$$\sqrt{N} \begin{pmatrix} \hat{Y}_* - \bar{Y} \\ \hat{x} \end{pmatrix} \rightsquigarrow \mathcal{N} \left\{ 0_{Q+JQ}, \begin{pmatrix} V_* & \Gamma_* V_x \\ V_x \Gamma_*^{\mathrm{T}} & V_x \end{pmatrix} \right\} \qquad (* = \text{N}, \text{F}, \text{L})$$

with $V_x = N \text{cov}(\hat{x}) = \{\text{diag}(e_q^{-1})_{q \in \mathcal{Q}} - 1_{Q \times Q}\} \otimes S_x^2$,

$$\Gamma_{\text{N}} = \text{diag}(\gamma_q^{\mathrm{T}})_{q \in \mathcal{Q}}, \quad \Gamma_{\text{F}} = \text{diag}\{(\gamma_q - \gamma_{\text{F}})^{\mathrm{T}}\}_{q \in \mathcal{Q}}, \quad \Gamma_{\text{L}} = 0_{Q \times JQ},$$

and $V_* = V_{\text{L}} + \Gamma_* V_x \Gamma_*^{\mathrm{T}} \geq V_{\text{L}}$ *for* $* = \text{N}, \text{F}, \text{L}$. *We give the explicit expressions of* V_**'s in the Supplementary Material.*

5.3.2. Rep based on marginal F-tests

Let $\mathcal{A}_{\text{f}} = \{p_{j,\text{f}} \geq \alpha_j \text{ for all } j = 1, \ldots, J\}$ denote the acceptance criterion under ReP based on the marginal F-tests. Renew $\hat{\tau}_x = (G_x \otimes I_J)\hat{x}$, where G_x is a prespecified contrast matrix with all row sums equal to zero. It defines a general measure of the difference in $\{\hat{x}(q) : q \in \mathcal{Q}\}$, extending $\hat{\tau}_x = \hat{x}(1) - \hat{x}(0)$ under the treatment-control experiment to multi-armed experiments (Zhao and Ding, 2023).

Theorem 3. *Assume the multi-armed version of Condition 1.*

(i) Eqs. (2)–(3) hold for $\hat{\tau}_x = (G_x \otimes I_J)\hat{x}$, $\hat{\tau}_* = G\hat{Y}_*$ $(* = \text{N}, \text{F}, \text{L})$, *and* $\mathcal{A} = \mathcal{A}_{\text{f}}$ *for arbitrary contrast matrices* G *and* G_x. *In particular,* $(\hat{\tau}_{\text{N}} \mid \mathcal{A}_{\text{f}}) \dot{\sim} \hat{\tau}_{\text{N}} \dot{\sim} \hat{\tau}_{\text{L}}$ *if* $\Gamma_{\text{N}} = 0_{Q \times JQ}$ *and* $(\hat{\tau}_{\text{F}} \mid \mathcal{A}_{\text{f}}) \dot{\sim} \hat{\tau}_{\text{F}} \dot{\sim} \hat{\tau}_{\text{L}}$ *if* $\Gamma_{\text{F}} = 0_{Q \times JQ}$.

(ii) Analogous to Eqs. (4)–(5), for $* \neq ** \in \{\text{N}, \text{F}, \text{L}\}$, *we have*

$$\frac{\mathbb{E}_{\text{a}} \left[\left\| \mathbb{E}_{\text{a}}(\hat{\tau}_* - \tau \mid \hat{\tau}_x, \mathcal{A}_{\text{f}}) \right\|_2^2 \right]}{\mathbb{E}_{\text{a}} \left[\left\| \mathbb{E}_{\text{a}}(\hat{\tau}_* - \tau \mid \hat{\tau}_x) \right\|_2^2 \right]} \leq 1, \qquad \frac{\mathbb{E}_{\text{a}} \left\{ \left\| \hat{\tau}_* - \hat{\tau}_{**} \right\|_2^2 \mid \mathcal{A}_{\text{f}} \right\}}{\mathbb{E}_{\text{a}} \left\{ \left\| \hat{\tau}_* - \hat{\tau}_{**} \right\|_2^2 \right\}} \leq 1.$$

Echoing the comments after Theorems 1–2, Theorem 3 illustrates the utility of the marginal F-test-based ReP for improving covariate balance, reducing conditional biases, and promoting coherence across different estimators under multi-armed experiments, and ensures the asymptotic efficiency of $\hat{\tau}_{\text{L}}$ with identical asymptotic sampling distribution as under complete randomization. Subsequent inference can thus be conducted based on $\hat{\tau}_{\text{L}}$ and its EHW covariance in full parallel with the discussion in Section 3.3. We relegate the details about the asymptotic distributions and inference to Section S1 of the Supplementary Material.

5.3.3. Rep based on the multinomial logistic regression

We now address the ReP schemes based on the multinomial logistic regression of Z_i on $(1, x_i)$. Recall $\tilde{x}_i = (1, x_i^{\mathrm{T}})^{\mathrm{T}}$. For $\theta = (\theta_1^{\mathrm{T}}, \dots, \theta_{Q-1}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{(J+1)(Q-1)}$ with $\theta_q \in \mathbb{R}^{J+1}$, let

$$\pi_q(\tilde{x}_i, \theta) = \frac{\exp(\tilde{x}_i^{\mathrm{T}}\theta_q)}{1 + \sum_{q' \in \mathcal{Q}_+} \exp(\tilde{x}_i^{\mathrm{T}}\theta_{q'})} \qquad (q \in \mathcal{Q}_+),$$

and let $H(\theta) = (H_{qq'}(\theta))_{q,q' \in \mathcal{Q}_+}$ with $H_{qq'}(\theta) = N^{-1}\sum_{i=1}^N \pi_q(\tilde{x}_i, \theta)\{\pi_{q'}(\tilde{x}_i, \theta) - 1(q = q')\}\tilde{x}_i\tilde{x}_i^{\mathrm{T}}$. Then $H(\theta) \le 0$ gives the Hessian matrix of the scaled log-likelihood function under the multinomial logistic model. Condition 3 below extends Condition 2 to multi-armed experiments and states the uniform convergence requirement on $H(\theta)$.

Condition 3. *As $N \to \infty$, $H(\theta)$ converges to a negative-definite matrix $H_\infty(\theta) < 0$ for all $\theta \in \mathbb{R}^{(J+1)(Q-1)}$, and the convergence is uniform over θ on any compact set $\Theta \subset \mathbb{R}^{(J+1)(Q-1)}$.*

Let $\mathcal{A}_{\mathrm{logit,jt}} = \{p_{0,\mathrm{logit}} \ge \alpha_0\}$, $\mathcal{A}_{\mathrm{logit,mg}} = \{p_{qj,\mathrm{logit}} \ge \alpha_{qj}$ for all $qj\}$, and $\mathcal{A}_{\mathrm{logit,cs}} = \mathcal{A}_{\mathrm{logit,jt}} \cap \mathcal{A}_{\mathrm{logit,mg}} = \{p_{0,\mathrm{logit}} \ge \alpha_0;\ p_{qj,\mathrm{logit}} \ge \alpha_{qj}$ for all $qj\}$ denote the acceptance criteria under the joint, marginal, and consensus rules, respectively.

Theorem 4. *Assume Condition 3 and the multi-armed version of Condition 1. Theorem 3 holds if we replace all $\mathcal{A}_{\mathrm{f}}$ with $\mathcal{A}_{\mathrm{logit},\diamond}$ for $\diamond = \mathrm{jt, mg, cs}$.*

All comments after Theorem 3 extend here after changing the "marginal F-test-based" to "multinomial logistic regression-based". We omit the details to avoid repetition. We give the explicit forms of the asymptotic sampling distributions of $\hat{\tau}_*$ ($* = \mathrm{N, F, L}$) in Proposition S2 of the Supplementary Material. Echoing the comments after Theorem 2, all results in Proposition S2 and Theorem 4 are design-based and hold regardless of how well the models corresponding to the multinomial logistic and linear regressions in rerandomization and analysis represent the true data-generating processes. The proof of Proposition S2 further introduces a novel technical result on the asymptotic equivalence of the LRT and the Wald test for logistic and multinomial logistic regressions from the design-based perspective. We relegate the details to Theorem S1 in the Supplementary Material.

6. ReP in stratified randomized experiments

We now extend the results to stratified randomized experiments. Due to space limitations, we focus on the treatment-control experiment in stating the results. Extension to multi-armed experiments is similar and omitted. Consider N units in K strata of sizes $N_{[k]}$ ($k = 1, \dots, K;\ \sum_{k=1}^K N_{[k]} = N$). Stratified randomization conducts an independent complete randomization in each stratum, and randomly assigns $N_{[k]z}$ units to treatment level z in stratum k ($k = 1, \dots, K;\ z = 1, 0;\ N_{[k]1} + N_{[k]0} = N_{[k]}$). Building on Sections 3–4, we can define one covariate balance criterion within each stratum, and accept an allocation if and only if all strata satisfy the corresponding stratum-wise balance criteria. Let $\mathcal{A}_{[k]}$ denote the acceptance criterion in stratum k. The overall acceptance criterion is then

$$\mathcal{A} = \mathcal{A}_{[1]} \cap \mathcal{A}_{[2]} \cap \cdots \cap \mathcal{A}_{[K]}.$$

Remark 1. There are other ways to define the balance criterion under stratified experiments. For example, we can run a global test to quantify the covariate imbalance across all strata; see, e.g., Cai et al. (2015, Tables A1 and A2) and see Wang et al. (2021) for an analog based on the Mahalanobis distance. We can also define the balance criterion based on both the global and stratum-wise tests. We leave the corresponding theory for ReP to future research.

Assume $\mathcal{A}_{[k]} \in \{\mathcal{A}_{\dagger,\diamond} : \dagger = \mathrm{t, lm, logit},\ \diamond = \mathrm{jt, mg, cs}\}$ throughout the rest of this section. For $k = 1, \dots, K$, denote by $\{i \in [k]\}$ the set of units in stratum k, $\pi_{[k]} = N_{[k]}/N$ the relative size of stratum k, and $\tau_{[k]} = N_{[k]}{}^{-1}\sum_{i \in [k]}\{Y_i(1) - Y_i(0)\}$ the stratum-wise average treatment effect. The finite-population average treatment effect equals

$$\tau = N^{-1}\sum_{i=1}^N \{Y_i(1) - Y_i(0)\} = \sum_{k=1}^K N^{-1}\sum_{i \in [k]}\{Y_i(1) - Y_i(0)\} = \sum_{k=1}^K \pi_{[k]}\tau_{[k]}.$$

Let $\hat{\tau}_{*[k]}$ and $\hat{\mathrm{se}}_{*[k]}$ denote the basic estimator and EHW standard error obtained from stratum k, where $*$ can be N, F, and L. With a slight abuse of notation, renew $\hat{\tau}_* = \sum_{k=1}^K \pi_{[k]}\hat{\tau}_{*[k]}$ as a point estimate of τ and $\hat{\mathrm{se}}_*^2 = \sum_{k=1}^K \pi_{[k]}^2\hat{\mathrm{se}}_{*[k]}^2$ as its squared EHW standard error under stratified randomization. This abuse of notation causes little confusion because $\hat{\tau}_*$ and $\hat{\mathrm{se}}_*$ reduce to their definitions under complete randomization when $K = 1$. To compute $\hat{\tau}_*$ and $\hat{\mathrm{se}}_*$ from one global regression for each $* = \mathrm{N, F, L}$, let $S_{ik} = 1(i \in [k])$ denote the indicator of unit i being in stratum k, and let $S'_{ik} = S_{ik} - \pi_{[k]}$ denote the centered version of S_{ik} with $\sum_{i=1}^N S'_{ik} = 0$. Without loss of generality, let K be the reference level and let $S'_i = (S'_{i1}, \dots, S'_{i,K-1})^{\mathrm{T}}$ be the concatenation of S'_{ik} for $k = 1, \dots, K - 1$. Then $\hat{\tau}_*$ and $\hat{\mathrm{se}}_*$ ($* = \mathrm{N, F, L}$) are the coefficients of Z_i and the corresponding robust standard errors in

$$\begin{aligned}
&\mathrm{N}: \ \mathtt{lm}(Y_i \sim 1 + Z_i + S'_i + S'_i Z_i), \\
&\mathrm{F}: \ \mathtt{lm}(Y_i \sim 1 + Z_i + x_i + S'_i + S'_i Z_i + S'_i \otimes x_i), \qquad (8)\\
&\mathrm{L}: \ \mathtt{lm}(Y_i \sim 1 + Z_i + x_i + Z_i x_i + S'_i + S'_i Z_i + S'_i \otimes x_i + S'_i \otimes Z_i x_i),
\end{aligned}$$

respectively, where $\otimes$ denotes the Kronecker product with $S_i'\otimes x_i = (S_{i1}'x_i^{\mathrm{T}}, \dots, S_{i,K-1}'x_i^{\mathrm{T}})^{\mathrm{T}}$ and $S_i'\otimes Z_i x_i = (S_{i1}'Z_i x_i^{\mathrm{T}}, \dots, S_{i,K-1}'Z_i x_i^{\mathrm{T}})^{\mathrm{T}}$.

Let $\hat\tau_{x[k]}$ denote the difference in covariate means in stratum k ($k = 1, \dots, K$). Theorem 5 below parallels Theorems 1–2 and summarizes the asymptotic sampling properties of $\hat\tau_*$ under rerandomized stratified experiments.

Theorem 5. *Assume that Conditions 1–2 hold for all strata and $\pi_{[k]}$ has a limit in $(0,1)$ for all k. For $\mathcal{A} = \mathcal{A}_{[1]} \cap \mathcal{A}_{[2]} \cap \cdots \cap \mathcal{A}_{[K]}$, where $\mathcal{A}_{[k]} \in \{\mathcal{A}_{\dagger,\diamond} : \dagger = \mathrm{t}, \mathrm{lm}, \mathrm{logit}, \ \diamond = \mathrm{jt}, \mathrm{mg}, \mathrm{cs}\}$, we have*

(i) $(\hat\tau_{x[k]} \mid \mathcal{A}) \succeq_\infty \hat\tau_{x[k]}$ *for* $k = 1, \dots, K$;

(ii)

$$(\hat\tau_{\mathrm{N}} \mid \mathcal{A}) \succeq_\infty \hat\tau_{\mathrm{N}}, \quad (\hat\tau_{\mathrm{F}} \mid \mathcal{A}) \succeq_\infty \hat\tau_{\mathrm{F}}, \quad (\hat\tau_{\mathrm{L}} \mid \mathcal{A}) \stackrel{\cdot}{\sim} \hat\tau_{\mathrm{L}},$$
$$(\hat\tau_{\mathrm{L}} \mid \mathcal{A}) \succeq_\infty (\hat\tau_* \mid \mathcal{A}) \quad \text{for} \quad * = \mathrm{N}, \mathrm{F};$$

(iii) for $* \in \{\mathrm{N}, \mathrm{F}, \mathrm{L}\}$, the asymptotic conditional bias of $\hat\tau_*$ given $\{\hat\tau_{x[k]}\}_{k=1}^K$ satisfies

$$\frac{\mathbb{E}_{\mathrm{a}}\left[\left\{\mathbb{E}_{\mathrm{a}}(\hat\tau_* - \tau \mid \{\hat\tau_{x[k]}\}_{k=1}^K, \mathcal{A})\right\}^2\right]}{\mathbb{E}_{\mathrm{a}}\left[\left\{\mathbb{E}_{\mathrm{a}}(\hat\tau_* - \tau \mid \{\hat\tau_{x[k]}\}_{k=1}^K)\right\}^2\right]} \le 1;$$

(iv) for $* \neq ** \in \{\mathrm{N}, \mathrm{F}, \mathrm{L}\}$,

$$\frac{\mathbb{E}_{\mathrm{a}}\left\{(\hat\tau_* - \hat\tau_{**})^2 \mid \mathcal{A}\right\}}{\mathbb{E}_{\mathrm{a}}\left\{(\hat\tau_* - \hat\tau_{**})^2\right\}} \le 1.$$

All comments after Theorems 1–2 and in Section 3.3 extend here after minimal modification. From Theorem 5(ii), the Wald-type inference based on $\hat\tau_{\mathrm{L}}$ is asymptotically the most efficient and can be conducted using the same normal approximation based on $\hat{\mathrm{se}}_{\mathrm{L}}$ as under complete randomization. On the other hand, ReP narrows the asymptotic distributions of $\hat\tau_{\mathrm{N}}$ and $\hat\tau_{\mathrm{F}}$ compared with complete randomization. Standard inference procedures based on normal approximation are hence overconservative. From Theorem 5(iii)–(iv), ReP reduces the conditional biases of the three estimators and improves their coherence in terms of mean squared difference.

Remark 2. To recover $\hat\tau_*$ and $\hat{\mathrm{se}}_*$, the three regression specifications in Eq. (8) require the interactions between S_i' and all regressors in the unadjusted, additive, and interacted regressions that we used for completely randomized experiments; c.f. Section 2.1. A more common formulation of the unadjusted and additive regressions under stratified randomized experiments is $\mathtt{lm}(Y_i \sim 1 + Z_i + S_i')$ and $\mathtt{lm}(Y_i \sim 1 + Z_i + x_i + S_i')$, without the interaction terms. See, e.g., Bugni et al. (2018) and Ding (2021a). Denote by $\tilde\tau_{\mathrm{N}}$ and $\tilde\tau_{\mathrm{F}}$ the coefficients of Z_i from these two more commonly seen specifications. Denote by $e_{[k]} = N_{[k]1}/N_{[k]}$ the proportion of treatment in stratum k for $k = 1, \dots, K$. Proposition 3 below shows that $\tilde\tau_{\mathrm{N}}$ and $\tilde\tau_{\mathrm{F}}$ are in general inconsistent for estimating τ unless $e_{[k]}(1 - e_{[k]})$'s have the same limit across all k. We hence focus on $\hat\tau_*$ ($* = \mathrm{N}, \mathrm{F}, \mathrm{L}$) in this paper and relegate the theory of $\tilde\tau_{\mathrm{N}}$ and $\tilde\tau_{\mathrm{F}}$ to future research.

Proposition 3. *Assume that Conditions 1–2 hold for all strata and $\pi_{[k]}$ has a limit in $(0,1)$ for all k. Then $\tilde\tau_* = \sum_{k=1}^K \omega_{[k]}\tau_{[k]} + o_{\mathbb{P}}(1)$ for $* = \mathrm{N}, \mathrm{F}$, where $\omega_{[k]} = \pi_{[k]}e_{[k]}(1 - e_{[k]}) / \sum_{k'=1}^K \pi_{[k']}e_{[k']}(1 - e_{[k']})$. We have $\omega_{[k]} = \pi_{[k]}$ for all k if and only if $e_{[k]}(1 - e_{[k]})$'s are constant across all k.*

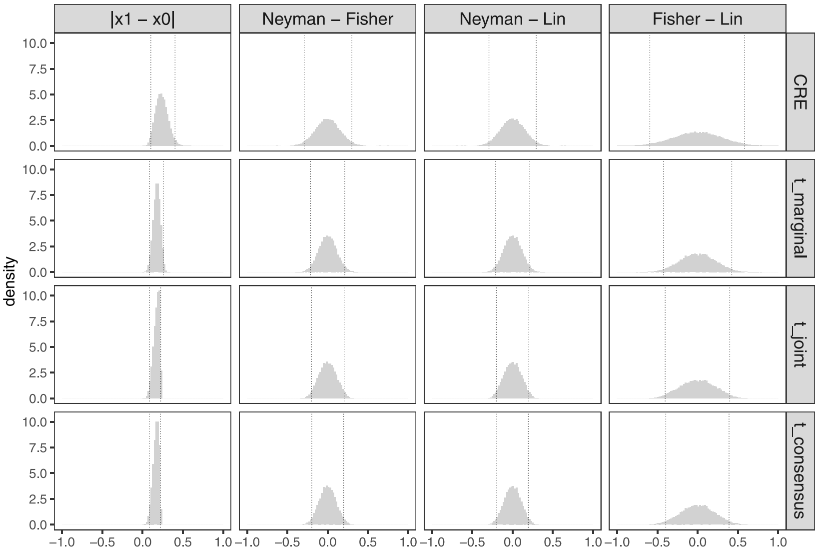
7. Numerical examples

We now illustrate the finite-sample properties of ReP by simulation. The results are coherent with the asymptotic theory in Sections 3–5, featuring improved covariate balances and overall efficiency of $\hat\tau_{\mathrm{L}}$ over $\hat\tau_{\mathrm{N}}$ and $\hat\tau_{\mathrm{F}}$.

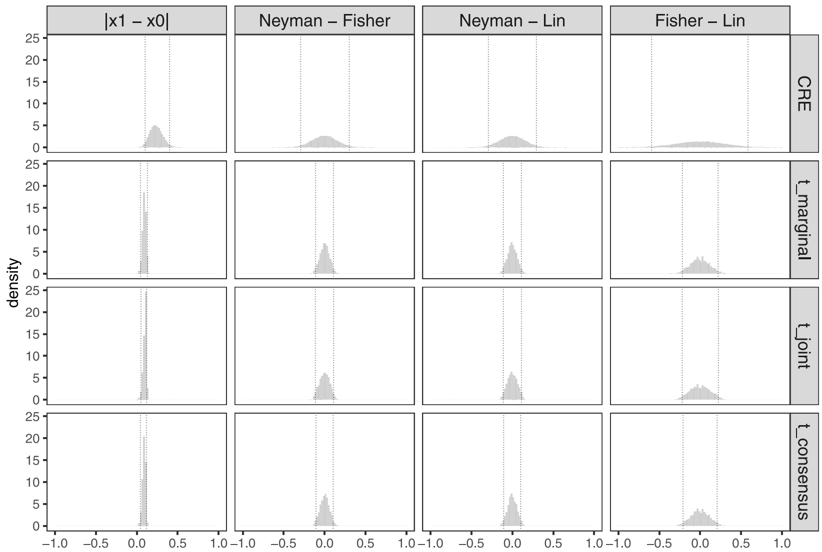
Consider a treatment-control experiment with $N = 500$ units, indexed by $i = 1, \dots, N$, and treatment arm sizes $(N_0, N_1) = (400, 100)$. For each i, we draw a $J = 5$ dimensional covariate vector $x_i = (x_{i1}, \dots, x_{i5})^{\mathrm{T}}$ with x_{ij} as independent Uniform$(-1, 1)$, and generate the potential outcomes as $Y_i(0) \sim \mathcal{N}(-\sum_{j=1}^5 x_{ij}^3, 0.1^2)$ and $Y_i(1) \sim \mathcal{N}(\sum_{j=1}^5 x_{ij}^3, 0.4^2)$. We center the $Y_i(0)$'s and $Y_i(1)$'s respectively to ensure $\tau = \bar Y(1) - \bar Y(0) = 0$, and fix $\{Y_i(0), Y_i(1), x_i\}_{i=1}^N$ in the simulation. For each iteration, we draw a random permutation of N_1 1's and N_0 0's to obtain the initial allocation under complete randomization.

Fig. 1 shows the distributions of $\|\hat\tau_x\|_2 = \|\hat x(1) - \hat x(0)\|_2$, $\hat\tau_{\mathrm{N}} - \hat\tau_{\mathrm{F}}$, $\hat\tau_{\mathrm{N}} - \hat\tau_{\mathrm{L}}$, and $\hat\tau_{\mathrm{F}} - \hat\tau_{\mathrm{L}}$ under complete randomization and the three two-sample t-test-based ReP schemes over 50000 independent initial allocations. The results under complete randomization are summarized over all 50000 allocations, whereas those under ReP are summarized over the subsets of allocations that satisfy the respective balance criteria. We vary the thresholds for the marginal rule from $\alpha_j = 0.15$ to $\alpha_j = 0.5$ for $j = 1, \dots, J$, and choose α_0 accordingly to ensure that the joint rule has approximately the same acceptance rate as the marginal rule. The message is coherent across different rules and thresholds: ReP reduces the difference in covariate means and the differences across different estimators, both in line with the theoretical results in Theorem 1. Compare sub-plots (a) and (b) under different values of α_j's. More stringent thresholds result in greater reduction in the differences when everything else stays the same.

Fig. 2 shows the distributions of $\hat\tau_*$ ($* = \mathrm{N}, \mathrm{F}, \mathrm{L}$). The message is coherent across different rules and thresholds: ReP improves the efficiency of $\hat\tau_{\mathrm{N}}$ and $\hat\tau_{\mathrm{F}}$ but leaves that of $\hat\tau_{\mathrm{L}}$ unchanged, both in line with Theorem 1. Compare sub-plots (a) and (b) under different values of α_j's. Increasing the thresholds improves the efficiency of $\hat\tau_*$ ($* = \mathrm{N}, \mathrm{F}$) when everything else stays the same.



(a) $\alpha_j = 0.15$ for $j = 1, \dots, J$, and $\alpha_0 = 0.55$.

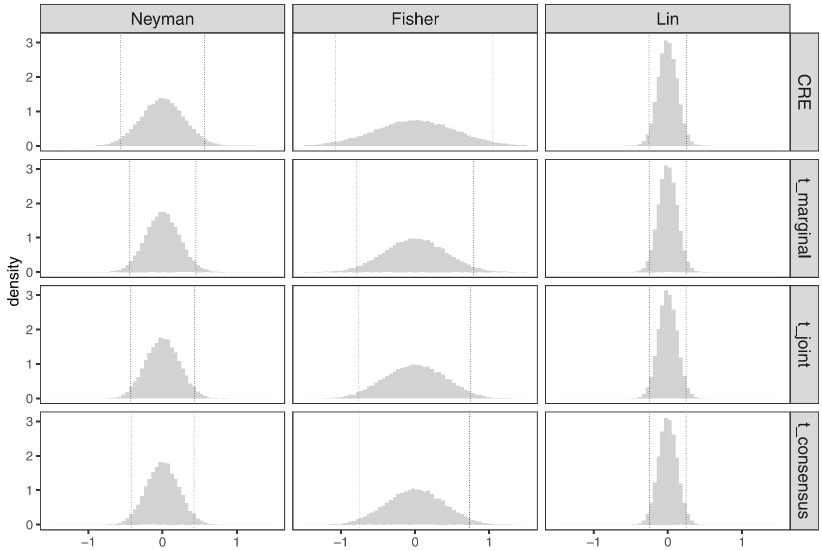


(b) $\alpha_j = 0.50$ for $j = 1, \dots, J$, and $\alpha_0 = 0.95$.

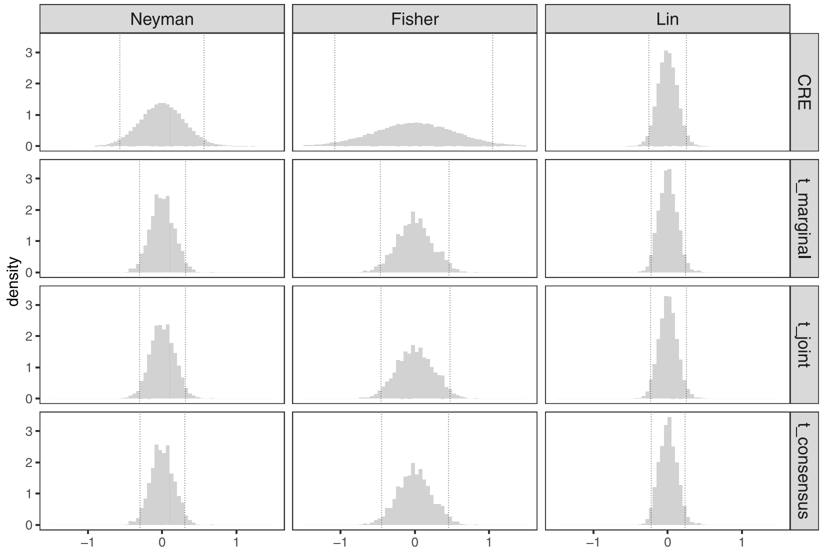
Fig. 1. Distributions of $\|\hat{\tau}_x\|_2 = \|\hat{x}(1) - \hat{x}(0)\|_2$, $\hat{\tau}_N - \hat{\tau}_F$, $\hat{\tau}_N - \hat{\tau}_L$, and $\hat{\tau}_F - \hat{\tau}_L$ under complete randomization and the three two-sample t-test-based ReP schemes over 50000 independent initial allocations. The results under complete randomization, labeled as "CRE", are summarized over all 50000 allocations, whereas those under ReP, labeled as "t_marginal", "t_joint", and "t_consensus", respectively, are summarized over the subsets of allocations that satisfy the respective balance criteria. The vertical lines correspond to the 0.025 and 0.975 empirical quantiles, respectively.

8. Discussion

ReP provides a tool for improving covariate balance in randomized experiments. We examined thirteen ReP schemes for treatment-control and multi-armed experiments, and quantified their theoretical properties from the design-based perspective. The theory clarifies three important issues regarding causal inference under ReP. First, the estimator from the interacted regression is asymptotically the most efficient under all ReP schemes examined, with the asymptotic sampling distribution remaining unchanged by ReP. We can thus conduct inference based on this estimator and its EHW standard error or covariance via identical procedure as that under complete randomization. Second, ReP improves the asymptotic efficiency of the estimators from the



(a) $\alpha_j = 0.15$ for $j = 1, \dots, J$, and $\alpha_0 = 0.55$.



(b) $\alpha_j = 0.50$ for $j = 1, \dots, J$, and $\alpha_0 = 0.95$.

Fig. 2. Distributions of $\hat{\tau}_*$ ($* = \text{N}, \text{F}, \text{L}$) under complete randomization and the three two-sample t-test-based ReP schemes over 50000 independent initial allocations. The results under complete randomization, labeled as "CRE", are summarized over all 50000 allocations, whereas those under ReP, labeled as "t_marginal", "t_joint", and "t_consensus", respectively, are summarized over the subsets of allocations that satisfy the respective balance criteria. The true τ is 0. The vertical lines correspond to the 0.025 and 0.975 empirical quantiles, respectively.

unadjusted and additive regressions relative to complete randomization, necessitating rerandomization-specific inference to avoid overconservativeness. Third, ReP reduces conditional biases of the three estimators and ensures more coherent inferences across them. These results illustrate the value of ReP for strengthening causal conclusions from experimental data, and highlight the value of the interacted adjustment for convenient and efficient inference under ReP.

We focused on the thirteen criteria in Tables 1 and 2 because of their conceptual straightforwardness and connections with the commonly used balance tests. The variety of other test options for balance check promises a spectrum of alternative schemes for conducting ReP, catering to the needs of different studies. We give the details in the Supplementary Material.

We focused on interval estimation based on asymptotic distributions for large-sample inference under ReP. Alternatively, the Fisher randomization test provides a way to conduct finite-sample exact inference of the sharp null hypothesis of zero treatment effects for all units. In addition, with properly chosen test statistics, the Fisher randomization test is also asymptotically valid for testing the weak null hypothesis of zero average treatment effect. Zhao and Ding (2021) established the theory of the Fisher randomization test for testing both sharp and weak nulls under ReM. All results therein extend to ReP with minimal modification.

We focused on inference under the finite-population, design-based framework. All results extend to superpopulation model-assisted inference with minimal modification. In particular, we need to modify the EHW standard error from the interacted regression to account for the additional variability in centering the covariates; see Zhao and Ding (2021, Section S1.2) for details.

Declaration of competing interest

Thank you for considering our submission to Journal of Econometrics. We declare that we have no actual, potential or perceived conflict of interest in relation to submission.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2024.105724.

References

Ashraf, N., Berry, J., Shapiro, J.M., 2010. Can higher prices stimulate product use? Evidence from a field experiment in zambia. Amer. Econ. Rev. 100, 2383–2413.
Bai, Y.H., Shaikh, A.M., Romano, J.P., 2022. Inference in experiments with matched pairs. J. Amer. Statist. Assoc. 117, 1726–1737.
Bailey, R.A., 1987. Restricted randomization: A practical example. J. Amer. Statist. Assoc. 82, 712–719.
Branson, Z., Dasgupta, T., 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.
Branson, Z., Shao, S., 2021. Ridge rerandomization: An experimental design strategy in the presence of covariate collinearity. J. Statist. Plann. Inference 211, 287–314.
Bruhn, M., McKenzie, D., 2009. In pursuit of balance: Randomization in practice in development field experiments. Am. Econ. J. Appl. Econ. 1, 200–232.
Bugni, F.A., Canay, I.A., Shaikh, A.M., 2018. Inference under covariate-adaptive randomization. J. Amer. Statist. Assoc. 113, 1784–1796.
Bugni, F.A., Canay, I.A., Shaikh, A.M., 2019. Inference under covariate-adaptive randomization with multiple treatments. Quant. Econ. 10, 1747–1785.
Cai, J., Janvry, A.D., Sadoulet, E., 2015. Social networks and the decision to insure. Am. Econ. J. Appl. Econ. 7, 81–108.
Cox, D.R., 1982. Randomization and concomitant variables in the design of experiments. In: Kallianpur, G., Krishnaiah, P.R., Ghosh, J.K. (Eds.), Statistics and Probability: Essays in Honor of C. R. Rao. North-Holland, Amsterdam, pp. 197–202.
de Mel, S., McKenzie, D., Woodruff, C., 2009. Are women more credit constrained? Experimental evidence on gender and microenterprise returns. Am. Econ. J. Appl. Econ. 1, 1–32.
de Mel, S., McKenzie, D., Woodruff, C., 2013. The demand for, and consequences of, formalization among informal firms in Sri Lanka. Am. Econ. J. Appl. Econ. 5, 122–150.
Ding, P., 2021a. The Frisch–Waugh–Lovell theorem for standard errors. Statist. Probab. Lett. 168, 108945.
Ding, P., 2021b. Two seemingly paradoxical results in linear models: the variance inflation factor and the analysis of covariance. J. Causal Inference 9, 1–8.
Dupas, P., Robinson, J., 2013. Why don't the poor save more? Evidence from health savings experiments. Amer. Econ. Rev. 103, 1138–1171.
Ferguson, T.S., 1996. A course in large sample theory, first ed. Chapman and Hall, London.
Fisher, R.A., 1935. The design of experiments, first ed. Oliver and Boyd, Edinburgh, London.
Fogarty, C.B., 2018. Regression-assisted inference for the average treatment effect in paired experiments. Biometrika 105, 994–1000.
Freedman, D.A., 2008a. On regression adjustments to experimental data. Adv. Appl. Math. 40, 180–193.
Freedman, D.A., 2008b. Randomization does not justify logistic regression. Statist. Sci. 23, 237–249.
Gerber, A.S., Green, D.P., 2012. Field Experiments: Design, Analysis, and Interpretation. W. W. Norton and Company, New York.
Gerber, A.S., Karlan, D., Bergan, D., 2009. Does the media matter? A field experiment measuring the effect of newspapers on voting behavior and political opinions. Am. Econ. J. Appl. Econ. 1, 35–52.
Glennerster, R., Takavarasha, K., 2014. Running Randomized Evaluations. Princeton University Press, Princeton.
Guo, K., Basse, G., 2023. The generalized Oaxaca–Blinder estimator. J. Amer. Statist. Assoc. 118, 524–536.
Hansen, B.B., Bowers, J., 2008. Covariate balance in simple, stratified and clustered comparative studies. Statist. Sci. 23, 219–236.
Higgins, M.J., Sävje, F., Sekhon, J.S., 2016. Improving massive experiments with threshold blocking. Proc. Natl. Acad. Sci. 113, 7369–7376.
Imai, K., King, G., Nall, C., 2009. The essential role of pair matching in cluster-randomized experiments, with application to the Mexican universal health insurance evaluation. Statist. Sci. 24, 29–53.
Imbens, G.W., Rubin, D.B., 2015. Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. Cambridge University Press, Cambridge.
Johansson, P., Rubin, D.B., Schultzberg, M., 2021. On optimal rerandomization designs. J. R. Stat. Soc. Ser. B Stat. Methodol. 83, 395–403.
Johansson, P., Schultzberg, M., 2022. Rerandomization: A complement or substitute for stratification in randomized experiments? J. Statist. Plann. Inference 218, 43–58.
Kuziemko, I., Norton, M.I., Saez, E., Stantcheva, S., 2015. How elastic are preferences for redistribution? Evidence from randomized survey experiments. Am. Econ. Rev. 105, 1478–1508.
Li, X., Ding, P., 2017. General forms of finite population central limit theorems with applications to causal inference. J. Amer. Statist. Assoc. 112, 1759–1769.
Li, X., Ding, P., 2020. Rerandomization and regression adjustment. J. R. Stat. Soc. Ser. B Stat. Methodol. 82, 241–268.
Li, X., Ding, P., Rubin, D.B., 2018. Asymptotic theory of rerandomization in treatment–control experiments. Proc. Natl. Acad. Sci. 115, 9157–9162.
Li, X., Ding, P., Rubin, D.B., 2020. Rerandomization in 2^K factorial experiments. Ann. Statist. 48, 43–63.
Lin, W., 2013. Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. Ann. Appl. Stat. 7, 295–318.
Liu, H., Yang, Y., 2020. Regression-adjusted average treatment effect estimates in stratified randomized experiments. Biometrika 107, 935–948.
Lu, J., 2016. Covariate adjustment in randomization-based causal inference for 2^K factorial designs. Statist. Probab. Lett. 119, 11–20.
Miratrix, L., Sekhon, J., Yu, B., 2013. Adjusting treatment effect estimates by post-stratification in randomized experiments. J. R. Stat. Soc. Ser. B Stat. Methodol. 75, 369–396.
Morgan, K.L., 2011. Rerandomization to Improve Covariate Balance in Randomized Experiments (Ph.D. thesis).
Morgan, K.L., Rubin, D.B., 2012. Rerandomization to improve covariate balance in experiments. Ann. Statist. 40, 1263–1282.

Morgan, K.L., Rubin, D.B., 2015. Rerandomization to balance tiers of covariates. J. Amer. Statist. Assoc. 110, 1412–1421.

Negi, A., Wooldridge, J.M., 2021. Revisiting regression adjustment in experiments with heterogeneous treatment effects. Econometric Rev. 40, 504–534.

Newey, W.K., McFadden, D.L., 1994. Large sample estimation and hypothesis testing. In: Engle, R.F., McFadden, D.L. (Eds.), In: Handbook of Econometrics, vol. 4, Elsevier, Amsterdam, pp. 2111–2245.

Neyman, J., 1923. On the application of probability theory to agricultural experiments. Statist. Sci. 5, 465–472.

Pashley, N.E., Miratrix, L.W., 2021. Insights on variance estimation for blocked and matched pairs designs. J. Educ. Behav. Stat. 46, 271–296.

Rosenbaum, P.R., Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70, 41–55.

Royen, T., 2014. A simple proof of the Gaussian correlation conjecture extended to multivariate gamma distributions. Far East J. Theor. Stat. 48, 139–145.

Sherman, S., 1955. A theorem on convex sets with applications. Ann. Math. Stat. 26, 763–767.

Wang, X., Wang, T., Liu, H., 2021. Rerandomization in stratified randomized experiments. J. Amer. Statist. Assoc. 118, 1295–1304.

Ye, T., Shao, J., Yi, Y., Zhao, Q., 2023. Toward better practice of covariate adjustment in analyzing randomized clinical trials. J. Amer. Statist. Assoc. 118, 2370–2382.

Zhao, A., Ding, P., 2021. Covariate-adjusted Fisher randomization tests for the average treatment effect. J. Econometrics 225, 278–294.

Zhao, A., Ding, P., 2023. Covariate adjustment in multiarmed, possibly factorial experiments. J. R. Stat. Soc. Ser. B Stat. Methodol. 85, 1–23.

Supplementary Material

Section S1 gives the additional results for ReP in multi-armed experiments.

Section S2 presents extensions to alternative covariate balance criteria for ReP.

Section S3 reviews the test statistics that underlie the p-values we use to form ReP.

Section S4 states the key lemmas for proving the results in the main paper. In particular, Theorem S1 is a novel technical result, and formalizes the design-based properties of the MLE outputs from logistic and multinomial logistic regressions. The result establishes the asymptotic equivalence of the LRT and the Wald test for logistic and multinomial logistic regressions from the design-based perspective.

Section S5 gives the proofs of the results in the main paper.

Section S7 gives the proof of Theorem S1.

Notation. Assume centered covariates with $\bar{x} = N^{-1}\sum_{i=1}^{N} x_i = 0_J$ throughout to simplify the presentation. For two sequences of random vectors $\{A_N\}_{N=1}^{\infty}$ and $\{B_N\}_{N=1}^{\infty}$ with $A_N \rightsquigarrow A$ and $B_N \rightsquigarrow B$ in $\mathbb{R}^m$, write $A_N \succeq_\infty B_N$ if $A \succeq B$, and write $A_N \dot{\sim} B_N$ if A and B have the same distribution. Definition 2 in the main paper is a special case of this definition of $\succeq_\infty$ and $\dot{\sim}$, with $\hat{\theta}_1 \dot{\sim} \hat{\theta}_2$ and $\hat{\theta}_1 \succeq_\infty \hat{\theta}_2$ being abbreviations of $\sqrt{N}(\hat{\theta}_1 - \theta) \dot{\sim} \sqrt{N}(\hat{\theta}_2 - \theta)$ and $\sqrt{N}(\hat{\theta}_1 - \theta) \succeq_\infty \sqrt{N}(\hat{\theta}_2 - \theta)$, respectively, when the meaning of θ is clear from the context.

S1. Additional results for ReP with multiple arms

S1.1. Asymptotic sampling distributions of $\hat{\tau}_*$ ($* =$ N, F, L)

Proposition S1 below gives the asymptotic distributions of $\hat{\tau}_*$ ($* = \text{N}, \text{F}, \text{L}$) under ReP based on the marginal F-tests with $\mathcal{A}_\text{f} = \{p_{j,\text{f}} \geq \alpha_j \text{ for all } j = 1, \dots, J\}$. Renew $\epsilon \sim \mathcal{N}(0_Q, I_Q)$ as a $Q \times 1$ standard normal random vector. Let $\mathcal{T}_\text{f} \sim \epsilon_\text{f} \mid \{\sum_{q\in\mathcal{Q}} e_q \epsilon_{\text{f},qj}^2 \leq a_j' S_{x,j}^2 \text{ for all } j = 1, \dots, J\}$ be a truncated normal random vector independent of ϵ, where $\epsilon_\text{f} = (\epsilon_{\text{f},qj})_{q\in\mathcal{Q};\ j=1,\dots,J} \sim \mathcal{N}(0_{JQ}, V_x)$, $S_{x,j}^2 = (N-1)^{-1}\sum_{i=1}^{N} x_{ij}x_{ij}^\text{T}$, and a_j' denotes the $(1-\alpha_j)$th quantile of the χ^2_{Q-1} distribution. Lemma S4 ensures $\mathcal{T}_\text{f} \succeq \epsilon_\text{f}$ by the Gaussian correlation inequality.

Proposition S1. Assume the multi-armed version of Condition 1. Recall the notation in Lemma 2. Then

$$\sqrt{N}(\hat{Y}_* - \bar{Y}) \mid \mathcal{A}_\text{f} \rightsquigarrow V_\text{L}^{1/2}\epsilon + \Gamma_*\mathcal{T}_\text{f} \quad (* = \text{N}, \text{F}),$$
$$\sqrt{N}(\hat{Y}_\text{L} - \bar{Y}) \mid \mathcal{A}_\text{f} \rightsquigarrow \mathcal{N}(0_Q, V_\text{L}).$$

Proposition S2 below gives the asymptotic distributions of $\hat{\tau}_*$ ($* = \text{N}, \text{F}, \text{L}$) under ReP based on the multinomial logistic regression with $\mathcal{A}_{\text{logit,jt}} = \{p_{0,\text{logit}} \geq \alpha_0\}$, $\mathcal{A}_{\text{logit,mg}} = \{p_{qj,\text{logit}} \geq \alpha_{qj} \text{ for all } qj\}$, and $\mathcal{A}_{\text{logit,cs}} = \mathcal{A}_{\text{logit,jt}} \cap \mathcal{A}_{\text{logit,mg}} = \{p_{0,\text{logit}} \geq \alpha_0;\ p_{qj,\text{logit}} \geq \alpha_{qj} \text{ for all } qj\}$. Let a_{qj} be the $(1-\alpha_{qj}/2)$th quantile of the standard normal distribution for $q \in \mathcal{Q}_+$ and $j = 1, \ldots, J$. Without introducing new notation, renew a_0 as the $(1-\alpha_0)$th quantile of the $\chi^2_{J(Q-1)}$ distribution, and renew $a = (a_{qj})_{q\in\mathcal{Q}_+;\ j=1,\ldots,J}$ as the vectorization of a_{qj} in lexicographical order. The definitions of a_0 and a reduce to those in (1) with $(\alpha_j, a_j) = (\alpha_{1j}, a_{1j})$ when $Q = 2$.

Recall that $\epsilon \sim \mathcal{N}(0_Q, I_Q)$. Let

$$\mathcal{L} \sim \epsilon_0 \mid \{\|\epsilon_0\|_2^2 \leq a_0\}, \quad \mathcal{T}_{\text{logit}} \sim \epsilon_{\text{logit}} \mid \{|\epsilon_{\text{logit}}| \leq a\}, \quad \mathcal{T}'_{\text{logit}} \sim \epsilon_{\text{logit}} \mid \{|\epsilon_{\text{logit}}| \leq a,\ \|\epsilon_{\text{logit}}\|_{\mathcal{M}} \leq a_0\}$$

be three truncated normal random vectors independent of ϵ, with $\epsilon_0 \sim \mathcal{N}(0_{J(Q-1)}, I_{J(Q-1)})$ and $\epsilon_{\text{logit}} \sim \mathcal{N}\{0_{J(Q-1)}, D(V_\Psi)\}$. We have $\mathcal{L} \succeq \epsilon_0$ and $\mathcal{T}_{\text{logit}}, \mathcal{T}'_{\text{logit}} \succeq \epsilon_{\text{logit}}$ by the Gaussian correlation inequality; see Lemma S4 in the Supplementary Material.

Recall the definition of Γ_* ($* = \text{N}, \text{F}, \text{L}$) from Lemma 2. Let $\Gamma'_* = \Gamma_*\{(I_{Q-1}, -e_Q^{-1}e_+)^{\text{T}} \otimes I_J\}$ for $* = \text{N}, \text{F}, \text{L}$ with $e_+ = (e_1, \ldots, e_{Q-1})^{\text{T}}$; the subscript $+$ signifies quantities associated with the non-reference levels. Let $\Psi = \{\Phi^{-1} \operatorname{diag}(e_+)\} \otimes (S_x^2)^{-1}$ with $\operatorname{diag}(e_+) = \operatorname{diag}(e_q)_{q\in\mathcal{Q}_+}$ and $\Phi = \operatorname{diag}(e_+) - e_+e_+^{\text{T}}$. Let

$$V_{x+} = N\operatorname{cov}(\hat{x}_+) = (R_+^{-1} - 1_{(Q-1)\times(Q-1)}) \otimes S_x^2, \qquad V_\Psi = N\operatorname{cov}(\Psi\hat{x}_+) = \Psi V_{x+}\Psi^{\text{T}}$$

with $\hat{x}_+ = (\hat{x}(1)^{\text{T}}, \ldots, \hat{x}(Q-1)^{\text{T}})^{\text{T}}$ and $\operatorname{cov}(\cdot)$ denoting covariance under complete randomization.

Let $\kappa = (I_{Q-1}, -e_Q^{-1}e_+)^{\text{T}} \otimes I_J$ such that $\Gamma'_* = \Gamma_*\kappa$ for $* = \text{N}, \text{F}, \text{L}$. It follows from $0_J = \bar{x} = \sum_{q\in\mathcal{Q}} e_q\hat{x}(q)$ that

$$\hat{x} = \kappa\hat{x}_+, \qquad \Gamma_*\hat{x} = \Gamma'_*\hat{x}_+ \qquad (* = \text{N}, \text{F}, \text{L}). \tag{S1}$$

This gives the intuition behind the definition of Γ'_*.

Proposition S2. Assume Condition 3 and the multi-armed version of Condition 1. Recall the notation in Lemma 2. Then

$$\begin{aligned}
\sqrt{N}(\hat{Y}_* - \bar{Y}) \mid \mathcal{A}_{\text{logit,jt}} &\rightsquigarrow V_{\text{L}}^{1/2}\epsilon + \Gamma'_* V_{x+}^{1/2}\mathcal{L}, \\
\sqrt{N}(\hat{Y}_* - \bar{Y}) \mid \mathcal{A}_{\text{logit,mg}} &\rightsquigarrow V_{\text{L}}^{1/2}\epsilon + \Gamma'_*\Psi^{-1}\sigma(V_\Psi)\mathcal{T}_{\text{logit}}, \\
\sqrt{N}(\hat{Y}_* - \bar{Y}) \mid \mathcal{A}_{\text{logit,cs}} &\rightsquigarrow V_{\text{L}}^{1/2}\epsilon + \Gamma'_*\Psi^{-1}\sigma(V_\Psi)\mathcal{T}'_{\text{logit}}
\end{aligned}$$

for $* = \text{N}, \text{F}$, whereas $\sqrt{N}(\hat{Y}_{\text{L}} - \bar{Y}) \mid \mathcal{A}_{\text{logit},\diamond} \rightsquigarrow \mathcal{N}(0_Q, V_{\text{L}})$ for $\diamond = \text{jt}, \text{mg}, \text{cs}$.

S1.2. Wald-type inference

Recall from Theorems 3 and 4 that $\hat{\tau}_{\text{L}}$ is asymptotically the most efficient under ReP with multiple arms, with $\sqrt{N}(\hat{Y}_{\text{L}} - \bar{Y}) \mid \mathcal{A} \rightsquigarrow \mathcal{N}(0_Q, V_{\text{L}})$ for all $\mathcal{A} \in \{\mathcal{A}_{\text{f}}, \mathcal{A}_{\text{logit},\diamond} : \diamond = \text{jt}, \text{mg}, \text{cs}\}$. This suggests subsequent inference based on $\hat{\tau}_{\text{L}}$ and its EHW covariance by normal approximation.

Specifically, let $\hat{V}_{\text{L}}'$ be the EHW covariance of $\hat{Y}_{\text{L}}$ from the same OLS fit. Zhao and Ding (2023) showed that it is asymptotically appropriate for estimating the true sampling covariance under complete randomization; see Lemma S2 in Section S4.1. The same reasoning as in Li et al. (2018, Lemma A16) ensures that the asymptotic appropriateness extends to ReP based on the marginal F-tests as well. This, together with the asymptotic normality of $\hat{\tau}_{\text{L}}$ from Lemma 2 and Proposition S1 in the main paper, justifies the Wald-type inference of τ based on $(\hat{\tau}_{\text{L}}, G\hat{V}_{\text{L}}'G^{\text{T}})$ under both complete randomization and ReP. The Fisher randomization test can be conducted similarly using $\hat{\tau}_{\text{L}}^{\text{T}}(G\hat{V}_{\text{L}}'G^{\text{T}})^{-1}\hat{\tau}_{\text{L}}$ as the test statistic for both the strong and weak null hypotheses (Wu and Ding 2021). This illustrates the convenience of the interacted regression for inference under general experiments.

The asymptotic sampling distribution of $\hat{\tau}_*$ ($* = \text{N}, \text{F}$), on the other hand, is altered by ReP into a convolution of independent normal and truncated normal when $\Gamma_* \neq 0$, resulting in greater peakedness than that under complete randomization. Inference based on the usual normal approximation, as a result, is overly conservative, deterring statistically significant findings. This, again, illustrates the value of the interacted regression for convenient and efficient inference under ReP for general experiments.

S1.3. Extensions of the joint t-test and linear regression

The marginal F-tests give an immediate extension of the marginal two-sample t-tests to more than two treatment arms. The range of tests commonly used in multivariate analysis of variance, on the other hand, provide reasonable substitutes to the Hotelling's T^2 test under the joint rule (Morgan 2011). Common choices of test statistics include Wilks' Λ, the Lawley–Hotelling trace, the Pillai–Bartlett trace, and Roy's largest root. One complication is that the distributions of these test statistics are not well studied under the design-based inference. Morgan (2011) recommended using the Fisher randomization test to generate an empirical distribution for the test statistic of choice. This is sound in theory but can become unwieldy in practice.

Generalization of the linear regression-based criteria can be accomplished by a dichotomization of the treatment assignment variable. Recall that $\mathcal{I}_{iq} = 1(Z_i = q)$ denotes the indicator for assignment to treatment arm $q \in \mathcal{Q} = \{1, \ldots, Q\}$ in a general experiment. We can conduct balance checks based on Q

separate linear regressions as lm($\mathcal{I}_{iq} \sim 1 + x_{i1} + \cdots + x_{iJ}$) over $i = 1, \ldots, N$ for $q = 1, \ldots, Q$. The process yields one joint and J marginal p-values for each $q \in \mathcal{Q}$, measuring the influence of x_{ij}'s on assignment to treatment level q. We can form acceptance criteria accordingly based on whether some or all of them exceed some prespecified thresholds. Despite the conceptual straightforwardness of this approach, however, it requires additional data transformations, and gives only measures of covariate balance for the Q treatment levels separately.

S1.4. Covariate-wise and treatment-wise p-values

The multinomial logistic regression of Z_i on $(1, x_i)$ enables a variety of ways to conduct ReP under general experiments. We formed the marginal rule based on $(p_{qj,\text{logit}})_{q \in \mathcal{Q}_+;\ j=1,\ldots,J}$, with one p-value for each estimated coefficient $\tilde{\beta}_{qj}$, corresponding to the correlation between covariate j and assignment to treatment level $q \in \mathcal{Q}_+$. Alternatively, most standard software packages also allow us to test the overall effect of a given covariate $j \in \{1, \ldots, J\}$ across all treatment levels; see, e.g., the `test` command in stata. Denote by $p_{\cdot j,\text{logit}}$ the resulting p-value associated with the overall effect of covariate j. We can also form the balance criterion using the $p_{\cdot j,\text{logit}}$'s, and accept a randomization if and only if $p_{\cdot j,\text{logit}} \geq \alpha_j$ for all $j = 1, \ldots, J$. The resulting criterion parallels the marginal rules under the treatment-control experiment, and yields analogous results with $\hat{\tau}_{\text{L}}$ being our recommendation.

More generally, we can arrange the $\tilde{\beta}_{qj}$'s into a matrix, with rows corresponding to the $Q - 1$ non-reference treatment levels and columns corresponding to the J covariates:

Non-reference	Covariate					Treatment-wise
treatment level	1	$\ldots$	j	$\ldots$	J	p-value
1	$\tilde{\beta}_{11}$	$\ldots$	$\tilde{\beta}_{1j}$	$\ldots$	$\tilde{\beta}_{1J}$	$p_{1\cdot,\text{logit}}$
$\vdots$						$\vdots$
q	$\tilde{\beta}_{q1}$	$\ldots$	$\tilde{\beta}_{qj}$	$\ldots$	$\tilde{\beta}_{qJ}$	$p_{q\cdot,\text{logit}}$
$\vdots$						$\vdots$
$Q-1$	$\tilde{\beta}_{Q-1,1}$	$\ldots$	$\tilde{\beta}_{Q-1,j}$	$\ldots$	$\tilde{\beta}_{Q-1,J}$	$p_{Q-1,\cdot,\text{logit}}$
Covariate-wise p-value	$p_{\cdot 1,\text{logit}}$	$\ldots$	$p_{\cdot j,\text{logit}}$	$\ldots$	$p_{\cdot J,\text{logit}}$	$p_{0,\text{logit}}$

The $p_{qj,\text{logit}}$'s, $p_{\cdot j,\text{logit}}$'s, and $p_{0,\text{logit}}$ then correspond to the cells, columns, and the entire matrix, respectively, measuring the deviations of the corresponding $\tilde{\beta}_{qj}$'s from 0.

By symmetry, we can also conduct one Wald test for each row of the matrix, $\tilde{\boldsymbol{\beta}}_q = (\tilde{\beta}_{q1}, \ldots, \tilde{\beta}_{qJ})^{\text{T}}$, and accept a randomization if and only if the resulting treatment-wise p-values, denoted by $p_{q\cdot,\text{logit}}$ for

$q \in \mathcal{Q}_+$, satisfy some prespecified criterion. The magnitude of $p_{q\cdot,\text{logit}}$ intuitively reflects the covariate balance between the treatment level $q \in \mathcal{Q}_+$ and the reference level Q. The acceptance rule based on $\{p_{q\cdot,\text{logit}} : q \in \mathcal{Q}_+\}$ hence involves $Q-1$ pairwise comparisons of the non-reference levels to the reference level, simplifying the approach that consults all pairwise comparisons.

This yields four types of p-values, namely $p_{qj,\text{logit}}$, $p_{\cdot j,\text{logit}}$, $p_{q\cdot,\text{logit}}$, and $p_{0,\text{logit}}$, summarized in the first row of Table S1. They measure the covariate balance at the treatment-covariate pair, covariate, treatment, and overall levels, respectively, and provide the ingredients for defining a whole spectrum of balance criteria under ReP. The marginal, joint, and consensus rules in Table 2 use $\{p_{qj,\text{logit}} : q \in \mathcal{Q}_+;\ j = 1, \dots, J\}$, $p_{0,\text{logit}}$, and their union to form the acceptance criteria, respectively, but the choice can be general. A key consideration is that the joint p-value $p_{0,\text{logit}}$ is invariant to non-degenerate transformation of the covariate vector, in the sense of $x_i' = Ax_i$ for some nonsingular $J \times J$ matrix A, whereas the treatment-covariate-wise, covariate-wise, and treatment-wise p-values in general are not unless A is diagonal. Emphases on specific covariates or treatment levels, on the other hand, justify the use of covariate- or treatment-wise p-values, respectively.

The same discussion extends to the p-values from (i) the treatment-wise regressions $\texttt{lm}(\mathcal{I}_{iq} \sim 1 + x_{i1} + \cdots + x_{iJ})$ over $i = 1, \dots, N$ for $q \in \mathcal{Q}$; (ii) the covariate-wise regressions $\texttt{lm}(x_{ij} \sim 1 + \mathcal{I}_{i1} + \cdots + \mathcal{I}_{i,Q-1})$ over $i = 1, \dots, N$ for $j \in \{1, \dots, J\}$; and (iii) the two-sample t-test of $\{x_{ij} : Z_i = q\}$ and $\{x_{ij} : Z_i = Q\}$ for each pair of $(q, j) \in \mathcal{Q}_+ \times \{1, \dots, J\}$, respectively.

Specifically, recall that the treatment-wise regression $\texttt{lm}(\mathcal{I}_{iq} \sim 1 + x_{i1} + \cdots + x_{iJ})$ extends the linear regression of Z_i on $(1, x_i)$ under the treatment-control experiment to general experiments, measuring the influence of covariates on assignment to treatment level $q \in \mathcal{Q}$. Denote by $\hat{\boldsymbol{\beta}}_q = (\hat{\beta}_{q1}, \dots, \hat{\beta}_{qJ})^{\mathrm{T}}$ the coefficient vector of $(x_{i1}, \dots, x_{iJ})^{\mathrm{T}}$ from the OLS fit. It yields two types of p-values, namely

(i) the marginal p-value associated with each individual $\hat{\beta}_{qj}$, denoted by $p_{qj,\text{lm}}$; and

(ii) the treatment-wise p-value from the F-test of $\texttt{lm}(\mathcal{I}_{iq} \sim 1 + x_{i1} + \cdots + x_{iJ})$ against $\texttt{lm}(\mathcal{I}_{iq} \sim 1)$, denoted by $p_{q\cdot,\text{lm}}$.

They are analogous to the $p_{qj,\text{logit}}$'s and $p_{q\cdot,\text{logit}}$'s from the multinomial logistic regression, respectively, and allow us to form balance criteria like (i) $p_{qj,\text{lm}} \geq \alpha_{qj}$ for all qj, or (ii) $p_{q\cdot,\text{lm}} \geq \alpha_q$ for all $q \in \mathcal{Q}$. See the second row of Table S1.

Next, recall $\texttt{lm}(x_{ij} \sim 1 + \mathcal{I}_{i1} + \cdots + \mathcal{I}_{i,Q-1})$ as a regression formulation for computing the $p_{j,\text{f}}$ from the marginal F-test of $(x_{ij}, Z_i)_{i=1}^N$. The resulting fit, in addition to yielding $p_{j,\text{f}}$ as the p-value from the F-test of $\texttt{lm}(x_{ij} \sim 1 + \mathcal{I}_{i1} + \cdots + \mathcal{I}_{i,Q-1})$ against $\texttt{lm}(x_{ij} \sim 1)$, also yields one marginal p-value for the coefficient of each $\mathcal{I}_{iq}$ $(q \in \mathcal{Q}_+)$, denoted by $p_{qj,\text{f}}$. The two types of p-values are analogous to the $p_{qj,\text{logit}}$'s and $p_{\cdot j,\text{logit}}$'s,

Table S1: Four strategies for forming the treatment-covariate-wise, covariate-wise, treatment-wise, and joint p-values under general experiments. Let $\mathcal{J} = \{1, \ldots, J\}$

	Treatment-covariate-wise (q, j)	Covariate-wise (j)	Treatment-wise (q)	Joint
$\texttt{logit}(Z_i \sim 1 + x_{i1} + \cdots + x_{iJ})$ over $i = 1, \ldots, N$	$p_{qj,\text{logit}}$	$p_{\cdot j,\text{logit}}$	$p_{q\cdot,\text{logit}}$	$p_{0,\text{logit}}$
$\texttt{lm}(\mathcal{I}_{iq} \sim 1 + x_{i1} + \cdots + x_{iJ})$ over $i = 1, \ldots, N$ for $q \in \mathcal{Q}$	$p_{qj,\text{lm}}$	n.a.	$p_{q\cdot,\text{lm}}$	n.a.
$\texttt{lm}(x_{ij} \sim 1 + \mathcal{I}_{i1} + \cdots + \mathcal{I}_{i,Q-1})$ over $i = 1, \ldots, N$ for $j \in \mathcal{J}$	$p_{qj,\text{f}}$	$p_{j,\text{f}}$	n.a.	n.a.
Two-sample t-test of $\{x_{ij} : Z_i = q\}$ and $\{x_{ij} : Z_i = Q\}$ for $j \in \mathcal{J}$ and $q \in \mathcal{Q}_+$	$p_{qj,\text{t}}$	n.a.	n.a.	n.a.

respectively, and allow us to form balance criteria accordingly. The $p_{j,\text{t}}$'s under the treatment-control experiment are a special case with $Q = 2$ and $\mathcal{I}_{i1} = Z_i$. See also Barrera-Osorio et al. (2011, Table 2) and Dupas and Robinson (2013, Tables 1 and A1) for applications that use the $p_{qj,\text{f}}$'s for balance check.

Last but not least, we can conduct one two-sample t-test of $\{x_{ij} : Z_i = q\}$ and $\{x_{ij} : Z_i = Q\}$ for each pair of $(q, j) \in \mathcal{Q}_+ \times \{1, \ldots, J\}$, comparing the balance of covariate j between treatment groups $q \in \mathcal{Q}_+$ and Q. Denote by $p_{qj,\text{t}}$ the resulting two-sided p-value. It can be implemented by fitting $\texttt{lm}(x_{ij} \sim 1 + \mathcal{I}_{iq})$ over $\{i : Z_i \in \{q, Q\}\}$, and allows us to form balance criteria like $p_{qj,\text{t}} \geq \alpha_{qj}$ for all qj for some prespecified thresholds $\alpha_{qj} \in (0, 1)$.

This yields four strategies for forming treatment-covariate-wise, covariate-wise, treatment-wise, and joint p-values under general experiments, summarized in Table S1. The multinomial logistic regression accommodates all four types of p-values via one MLE fit, and is hence our recommendation in general.

S2. Alternative covariate balance criteria

S2.1. Rerandomization with tiers of covariates

When covariates vary in a priori importance, Morgan and Rubin (2015) proposed rerandomizing based on Mahalanobis distance within tiers of covariate importance, imposing more stringent criteria for covariates that are thought to be more important. Extension of ReP to such settings is straightforward under the marginal rules by setting the covariate-wise thresholds according to the importance. To construct joint rules for ReP with tiers of covariates, we can conduct one joint test for each tier of covariates, and set the tier-wise thresholds according to the importance. The consensus rules then follow as the intersections of the corresponding marginal and joint rules. These ReP schemes complicate the asymptotic distributions of $\hat{\tau}_*$ ($* = \text{N}, \text{F}$) but keep that of $\hat{\tau}_\text{L}$ unchanged. We recommend the same analysis based on $\hat{\tau}_\text{L}$.

S2.2. Alternative joint tests from regression

The discussion so far assumed default tests from standard software packages. As a result, we conducted an F-test to construct the joint criterion under the linear regression model option, namely $\mathcal{A}_{\text{lm,jt}} = \{p_{0,\text{lm}} \geq \alpha_0\}$, and conducted an LRT to construct the joint criterion under the logistic and multinomial logistic regression model options, namely $\mathcal{A}_{\text{logit,jt}} = \{p_{0,\text{logit}} \geq \alpha_0\}$, respectively. The Wald test, on the other hand, enables definition of the joint criteria in a unified way.

Specifically, recall $\texttt{lm}(Z_i \sim 1 + x_i)$ and $\texttt{logit}(Z_i \sim 1 + x_i)$ as the linear and logistic regressions under the treatment-control experiment. Let $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_J)^\text{T}$ and $\tilde{\beta} = (\tilde{\beta}_1, \ldots, \tilde{\beta}_J)^\text{T}$ be the coefficient vectors of x_i from the OLS and MLE fits, respectively, with $\hat{V}$ and $\tilde{V}$ as the corresponding estimated covariances. The Wald tests of $\hat{\beta}$ and $\tilde{\beta}$ compare $W_{\text{lm}} = \hat{\beta}^\text{T}\hat{V}^{-1}\hat{\beta}$ and $W_{\text{logit}} = \tilde{\beta}^\text{T}\tilde{V}^{-1}\tilde{\beta}$ against the χ^2_J distribution, and define two alternatives to the F-test and LRT, respectively, measuring the magnitudes of $(\hat{\beta}_j)_{j=1}^J$ and $(\tilde{\beta}_j)_{j=1}^J$ as a whole. All results in Proposition 2 and Theorem 2 on $\mathcal{A}_{\text{lm,jt}}$ and $\mathcal{A}_{\text{logit,jt}}$ extend verbatim to the resulting ReP schemes, with $\hat{\tau}_\text{L}$ being our recommendation.

Likewise for all results in Proposition S2 and Theorem 4 on $\mathcal{A}_{\text{logit,jt}}$ to extend verbatim to the ReP based on the Wald test of $\tilde{\beta} = (\tilde{\beta}_{qj})_{q \in \mathcal{Q}_+;\ j=1,\ldots,J}$ under the general experiment, with $\hat{\tau}_\text{L}$ being our recommendation.

See Lemmas S1, S6, S9 and Theorem S1 in Section S4 for the proof of the asymptotic equivalence of the Wald test to the F-test and LRT, respectively. See also de Mel et al. (2009, Table 1) for an application of the Wald test to balance check.

S2.3. EHW standard errors for balance test

We assumed the default p-values from standard software packages for forming the balance criteria. The standard error and covariance involved in their computation are hence in general the classic standard errors and covariances derived under homoskedasticity. Alternatively, we can form the test statistics using the EHW robust standard errors and covariances as in the analysis stage, and compute the p-values accordingly. We give the explicit forms of the resulting test statistics in Remark S1 in Section S3.5, and show in Remark S2 in Section S4.4 their equivalence with the classic counterparts as N tends to infinity. All results thus extend to ReP based on the robustly studentized test statistics, with $\hat{\tau}_\text{L}$ being our recommendation.

S2.4. p-values from other standard statistical tests

The discussion so far concerned p-values from the linear, logistic, and multinomial logistic regressions in standard software packages. Alternatively, the probit and multinomial probit regressions provide two intuitive variants to the logistic and multinomial logistic regressions of Z_i on $(1, x_i)$, respectively, accommodating marginal, joint, and consensus rules via one MLE fit. We conjecture that the results are analogous, and leave the technical details to future work.

In addition, the two-sample Kolmogorov–Smirnov test and the chi-square test of independence exemplify alternative test choices for forming covariate balance criteria under the treatment-control and general experiments, respectively. See Gerber et al. (2009) and Chen et al. (2010) for their applications to covariate balance check. We leave the theory on their properties for rerandomization to future work.

S3. Test statistics and equivalent forms for acceptance criteria in Tables 1 and 2

We review in this section the test statistics that underlie the p-values we use to form the ReP schemes in Tables 1 and 2, respectively. To avoid repetition, we treat the logistic regression for treatment-control experiments as a special case of the multinomial logistic regression with $Q = 2$ and level 2 relabeled as level 0.

Assume α_j $(j = 1, \dots, J)$ and α_0 as the thresholds for the marginal and joint criteria under the treatment-control experiment, respectively.

Assume α_j $(j = 1, \dots, J)$, α_{qj} $(q \in \mathcal{Q}_+;\ j = 1, \dots, J)$, and α_0 as the thresholds for the marginal F-tests and the marginal and joint tests based on the multinomial logistic regression, respectively, under the general experiment.

S3.1. Two-sample t-tests

Marginal tests. Let $\hat{\tau}_{x,j} = \hat{x}_j(1) - \hat{x}_j(0)$, where $\hat{x}_j(q) = N_q^{-1} \sum_{i:Z_i=q} x_{ij}$, be the difference in means of the jth covariate, equaling the jth component of $\hat{\tau}_x$. The pooled standard error for $\hat{\tau}_{x,j}$ equals

$$\hat{\text{se}}_j = \sqrt{\frac{(N_1 - 1)\hat{S}^2_{x,j}(1) + (N_0 - 1)\hat{S}^2_{x,j}(0)}{N - 2} \left(\frac{1}{N_1} + \frac{1}{N_0}\right)},$$

with $\hat{S}^2_{x,j}(q) = (N_q - 1)^{-1} \sum_{i:Z_i=q} \{x_{ij} - \hat{x}_j(q)\}^2$ for $q = 0, 1$. The two-sample t-test uses

$$t_{j,\text{t}} = \hat{\tau}_{x,j} / \hat{\text{se}}_j$$

as the test statistic, and computes $p_{j,\mathrm{t}}$ based on the t_{N-2} distribution as

$$p_{j,\mathrm{t}} = \mathbb{P}(|A| \geq |t_{j,\mathrm{t}}|), \qquad \text{where} \quad A \sim t_{N-2}.$$

Let $T_\mathrm{t} = (t_{1,\mathrm{t}}, \dots, t_{J,\mathrm{t}})^\mathrm{T}$. The marginal criterion equals

$$\begin{aligned} \mathcal{A}_{\mathrm{t,mg}} &= \{p_{j,\mathrm{t}} \geq \alpha_j \text{ for all } j = 1, \dots, J\} \\ &= \{|t_{j,\mathrm{t}}| \leq a_{j,\mathrm{t}} \text{ for all } j = 1, \dots, J\} \\ &= \{|T_\mathrm{t}| \leq a_\mathrm{t}\}, \end{aligned} \tag{S2}$$

with $a_\mathrm{t} = (a_{1,\mathrm{t}}, \dots, a_{J,\mathrm{t}})^\mathrm{T}$ and $a_{j,\mathrm{t}}$ denoting the $(1-\alpha_j/2)$th quantile of the t_{N-2} distribution. Numerically, $\hat{\tau}_{x,j}$, $\hat{\mathrm{se}}_j$, $t_{j,\mathrm{t}}$, and $p_{j,\mathrm{t}}$ equal the coefficient, classic standard error, t-value, and p-value associated with Z_i from the OLS fit of $\texttt{lm}(x_{ij} \sim 1 + Z_i)$ over $i = 1, \dots, N$, respectively. This gives an alternative implementation of the marginal t-tests via OLS.

Joint test. Recall $W_\mathrm{t} = \hat{\tau}_x^\mathrm{T} \hat{\Omega}^{-1} \hat{\tau}_x$ as the test statistic for the joint two-sample t-test. The pooled estimated covariance equals

$$\hat{\Omega} = \frac{(N_1 - 1)\hat{S}_x^2(1) + (N_0 - 1)\hat{S}_x^2(0)}{N - 2} \left(\frac{1}{N_1} + \frac{1}{N_0} \right), \tag{S3}$$

with $\hat{S}_x^2(q) = (N_q - 1)^{-1} \sum_{i:Z_i=q} \{x_i - \hat{x}(q)\}\{x_i - \hat{x}(q)\}^\mathrm{T}$ for $q = 0, 1$. Assume the joint Wald test for computing $p_{0,\mathrm{t}}$, with χ^2_J as the reference distribution. The acceptance criterion equals

$$\mathcal{A}_{\mathrm{t,jt}} = \{p_{0,\mathrm{t}} \geq \alpha_0\} = \{W_\mathrm{t} \leq a_0\},$$

where a_0 denotes the $(1 - \alpha_0)$th quantile of the χ^2_J distribution.

S3.2. Linear regression

Marginal tests. Let $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_J)^\mathrm{T}$ denote the coefficient vector of $x_i = (x_{i1}, \dots, x_{iJ})^\mathrm{T}$ from $\texttt{lm}(Z_i \sim 1 + x_{i1} + \cdots + x_{iJ})$. Let $\hat{V}$ be the associated estimated covariance, with $\hat{V}_{jj}$ as the (j,j)th element for $j = 1, \dots, J$. The marginal test of $\hat{\beta}_j$ takes

$$t_{j,\mathrm{lm}} = \hat{\beta}_j / \hat{V}_{jj}^{1/2}$$

as the test-statistic, and computes $p_{j,\text{lm}}$ based on the t_{N-1-J} distribution as

$$p_{j,\text{lm}} = \mathbb{P}(|A| \geq |t_{j,\text{lm}}|), \qquad \text{where} \;\; A \sim t_{N-1-J}.$$

Let $T_{\text{lm}} = (t_{1,\text{lm}}, \dots, t_{J,\text{lm}})^{\mathrm{T}}$. The marginal criterion equals

$$\begin{aligned} \mathcal{A}_{\text{lm,mg}} &= \{p_{j,\text{lm}} \geq \alpha_j \text{ for all } j = 1, \dots, J\} \\ &= \{|t_{j,\text{lm}}| \leq a_{j,\text{lm}} \text{ for all } j = 1, \dots, J\} \\ &= \{|T_{\text{lm}}| \leq a_{\text{lm}}\}, \end{aligned} \tag{S4}$$

with $a_{\text{lm}} = (a_{1,\text{lm}}, \dots, a_{J,\text{lm}})^{\mathrm{T}}$ and $a_{j,\text{lm}}$ denoting the $(1 - \alpha_j/2)$th quantile of the t_{N-1-J} distribution.

Joint test. The F-test for linear regression compares $\texttt{lm}(Z_i \sim 1 + x_{i1} + \cdots + x_{iJ})$ against the empty model $\texttt{lm}(Z_i \sim 1)$. Let RSS_0 and RSS_1 denote the residual sums of squares from the null and full regressions, respectively. The test statistic equals

$$F = \frac{(\text{RSS}_0 - \text{RSS}_1)/J}{\text{RSS}_1/(N-1-J)},$$

and is compared against the $F_{J,N-1-J}$ distribution to compute $p_{0,\text{lm}}$ as

$$p_{0,\text{lm}} = \mathbb{P}(A \geq F), \qquad \text{where} \;\; A \sim F_{J,N-1-J}.$$

The acceptance criterion equals

$$\mathcal{A}_{\text{lm,jt}} = \{p_{0,\text{lm}} \geq \alpha_0\} = \{F \leq f_{J,N-1-J}\},$$

where $f_{J,N-1-J}$ denotes the $(1 - \alpha_0)$th quantile of the $F_{J,N-1-J}$ distribution.

Whereas the F-test is the default joint test for linear regression returned by most software packages, we can also compute the joint p-value by a Wald test with test statistic $W_{\text{lm}} = \hat{\beta}^{\mathrm{T}} \hat{V}^{-1} \hat{\beta}$. The resulting p-value equals

$$p'_{0,\text{lm}} = \mathbb{P}(A \geq W_{\text{lm}}), \qquad \text{where} \;\; A \sim \chi^2_J,$$

with $\{p'_{0,\text{lm}} \geq \alpha_0\} = \{W_{\text{lm}} \leq a_0\}$.

Lemma S1 below gives the numeric correspondence between F and W_{lm}, and underpins the asymptotic equivalence between $p_{0,\text{lm}}$ and $p'_{0,\text{lm}}$ for constructing the joint criterion from linear regression; see Lemma

S6. For two sequences of events $(\mathcal{A}_N)_{N=1}^{\infty}$ and $(\mathcal{B}_N)_{N=1}^{\infty}$, write $\mathcal{A}_N \doteq \mathcal{B}_N$ if $\mathbb{P}(\mathcal{A}_N \backslash \mathcal{B}_N) = o(1)$ and $\mathbb{P}(\mathcal{B}_N \backslash \mathcal{A}_N) = o(1)$.

Lemma S1. $F = J^{-1} W_{\mathrm{lm}}$ and $f_{J,N-1-J} = J^{-1} a_0 + o(1)$ such that

$$\mathcal{A}_{\mathrm{lm,jt}} = \{p_{0,\mathrm{lm}} \geq \alpha_0\} = \{F \leq f_{J,N-1-J}\} = \{W_{\mathrm{lm}} \leq a_{\mathrm{f}}\} \doteq \{W_{\mathrm{lm}} \leq a_0\} = \{p'_{0,\mathrm{lm}} \geq \alpha_0\},$$

where $a_{\mathrm{f}} = J \cdot f_{J,N-1-J} = a_0 + o(1)$.

S3.3. Multinomial logistic regression

Recall level Q as the reference level when fitting $\texttt{logit}(Z_i \sim 1 + x_{i1} + \cdots + x_{iJ})$ by MLE. The fitting algorithm assumes that $(x_i, Z_i)_{i=1}^N$ are independent samples from

$$Z_i \mid x_i \ \sim \ \text{multinomial}\{1; (\pi_1(x_i), \ldots, \pi_Q(x_i))\},$$

with

$$\log \frac{\pi_q(x_i)}{\pi_Q(x_i)} = \beta_{q0} + x_i^{\mathrm{T}} \boldsymbol{\beta}_q \quad \text{for } q \in \mathcal{Q}_+ = \{1, \ldots, Q-1\} \text{ and } \boldsymbol{\beta}_q = (\beta_{q1}, \ldots, \beta_{qJ})^{\mathrm{T}}. \tag{S5}$$

Let $\tilde{\boldsymbol{\beta}}_q = (\tilde{\beta}_{q1}, \ldots, \tilde{\beta}_{qJ})^{\mathrm{T}}$ be the MLE of $\boldsymbol{\beta}_q$ for $q \in \mathcal{Q}_+$. Let $\tilde{\beta} = (\tilde{\boldsymbol{\beta}}_1^{\mathrm{T}}, \ldots, \tilde{\boldsymbol{\beta}}_{Q-1}^{\mathrm{T}})^{\mathrm{T}} = (\tilde{\beta}_{qj})_{q \in \mathcal{Q}_+;\ j=1,\ldots,J} \in \mathbb{R}^{J(Q-1)}$, with $\tilde{V} = (\tilde{V}_{qj,q'j'})$ as the estimated covariance from the same MLE fit. The notation simplifies to $\tilde{\beta} = (\tilde{\beta}_1, \ldots, \tilde{\beta}_J)^{\mathrm{T}}$ and $\tilde{V} = (\tilde{V}_{jj'})_{j,j'=1,\ldots,J}$ under the treatment-control experiment with $Q = 2$, $\mathcal{Q}_+ = \{1\}$, $\tilde{\beta} = \tilde{\boldsymbol{\beta}}_1$, and $(\tilde{\beta}_j, \tilde{V}_{jj'}) = (\tilde{\beta}_{1j}, \tilde{V}_{1j,1j'})$.

Marginal tests. The marginal test of $\tilde{\beta}_{qj}$ uses

$$t_{qj,\mathrm{logit}} = \tilde{\beta}_{qj} / \tilde{V}_{qj,qj}^{1/2}$$

as the test statistic, and computes $p_{qj,\mathrm{logit}}$ as

$$p_{qj,\mathrm{logit}} = \mathbb{P}(|A| \geq |t_{qj,\mathrm{logit}}|), \qquad \text{where} \ \ A \sim \mathcal{N}(0,1).$$

Let $T_{\mathrm{logit}} = (t_{qj,\mathrm{logit}})_{q \in \mathcal{Q}_+;\ j=1,\ldots,J} \in \mathbb{R}^{J(Q-1)}$ in lexicographical order of qj. The marginal criterion equals

$$\begin{aligned} \mathcal{A}_{\mathrm{logit,mg}} &= \{p_{qj,\mathrm{logit}} \geq \alpha_{qj} \text{ for all } q \in \mathcal{Q}_+ \text{ and } j = 1, \ldots, J\} \\ &= \{|t_{qj,\mathrm{logit}}| \leq a_{qj} \text{ for all } q \in \mathcal{Q}_+ \text{ and } j = 1, \ldots, J\} \\ &= \{|T_{\mathrm{logit}}| \leq a\}, \end{aligned}$$

with $a = (a_{qj})_{q\in\mathcal{Q}_+;\ j=1,\dots,J} \in \mathbb{R}^{J(Q-1)}$ and a_{qj} denoting the $(1-\alpha_{qj}/2)$th quantile of the standard normal distribution.

When $Q = 2$, we have $(\tilde\beta_j, \tilde V_{jj}, p_{j,\text{logit}}, \alpha_j) = (\tilde\beta_{1j}, \tilde V_{1j,1j}, p_{1j,\text{logit}}, \alpha_{1j})$ for $j = 1, \dots, J$. The definition of a simplifies to $a = (a_1, \dots, a_J)^{\mathrm{T}} \in \mathbb{R}^J$, with a_j equaling the $(1-\alpha_j/2)$th quantile of the standard normal distribution. The definition of T_{logit} simplifies to $T_{\text{logit}} = (t_{1,\text{logit}}, \dots, t_{J,\text{logit}})^{\mathrm{T}}$ with

$$t_{j,\text{logit}} = \tilde\beta_j / \tilde V_{jj}^{1/2}.$$

The marginal criterion equals

$$\begin{aligned}\mathcal{A}_{\text{logit,mg}} &= \{p_{j,\text{logit}} \ge \alpha_j \text{ for all } j = 1, \dots, J\} \\ &= \{|t_{j,\text{logit}}| \le a_j \text{ for all } j = 1, \dots, J\} \\ &= \{|T_{\text{logit}}| \le a\}.\end{aligned} \tag{S6}$$

Joint test. The LRT for multinomial logistic regression tests $\texttt{logit}(Z_i \sim 1 + x_i)$ against the empty model $\texttt{logit}(Z_i \sim 1)$. Denote by λ_{LRT} the resulting test statistic, with the explicit form given in (S30). Then $p_{0,\text{logit}}$ is computed based on the $\chi^2_{J(Q-1)}$ distribution as

$$p_{0,\text{logit}} = \mathbb{P}(A \ge \lambda_{\text{LRT}}), \qquad \text{where } A \sim \chi^2_{J(Q-1)}.$$

The joint criterion equals

$$\mathcal{A}_{\text{logit,jt}} = \{p_{0,\text{logit}} \ge \alpha_0\} = \{\lambda_{\text{LRT}} \le a_0\},$$

where a_0 denotes the $(1-\alpha_0)$th quantile of the $\chi^2_{J(Q-1)}$ distribution. The definition of a_0 reduces to that under the treatment-control experiment, namely the $(1-\alpha_0)$th quantile of the χ^2_J distribution, with $Q = 2$.

Whereas the LRT is the default joint test for multinomial logistic regression returned by most software packages, we can also compute the joint p-value via a joint Wald test with test statistic $W_{\text{logit}} = \tilde\beta^{\mathrm{T}} \tilde V^{-1} \tilde\beta$. The resulting p-value equals

$$p'_{0,\text{logit}} = \mathbb{P}(A \ge W_{\text{logit}}), \qquad \text{where } A \sim \chi^2_{J(Q-1)},$$

with $\{p'_{0,\text{logit}} \ge \alpha_0\} = \{W_{\text{logit}} \le a_0\}$. Theorem S1 in Section S4 ensures $\lambda_{\text{LRT}} - W_{\text{logit}} = o_{\mathbb{P}}(1)$. This underpins the asymptotic equivalence between the LRT and the Wald test for constructing the joint criterion from multinomial logistic regression; see Lemma S6.

S3.4. Marginal F-tests under the general experiment

Renew $\hat{x}_j(q) = N_q^{-1} \sum_{i:Z_i=q} x_{ij}$ for $q \in \mathcal{Q} = \{1, \ldots, Q\}$. The F-test of covariate j uses

$$F_j = \frac{\sum_{q \in \mathcal{Q}} N_q \hat{x}_j^2(q)/(Q-1)}{\sum_{q \in \mathcal{Q}} \sum_{i:Z_i=q} \{x_{ij} - \hat{x}_j(q)\}^2/(N-Q)} \tag{S7}$$

as the test statistic, and compares it against the $F_{Q-1,N-Q}$ distribution to compute $p_{j,\mathrm{f}}$ as

$$p_{j,\mathrm{f}} = \mathbb{P}(A \geq F_j), \qquad \text{where} \quad A \sim F_{Q-1,N-Q}.$$

The marginal criterion equals

$$\mathcal{A}_{\mathrm{f,mg}} = \{p_{j,\mathrm{f}} \geq \alpha_j \text{ for all } j = 1, \ldots, J\} = \{F_j \leq a_{j,\mathrm{f}} \text{ for all } j = 1, \ldots, J\}, \tag{S8}$$

where $a_{j,\mathrm{f}}$ denotes the $(1-\alpha_j)$th quantile of the $F_{Q-1,N-Q}$ distribution.

S3.5. Unification

Table S2 summarizes the regression realizations, test statistics, and reference distributions for the thirteen ReP schemes in Tables 1 and 2. Table S3 summarizes the alternative expressions of the nine criteria in Table 1 for treatment-control experiments in terms of the test statistics.

In particular, recall $T_\dagger = (t_{1,\dagger}, \ldots, t_{J,\dagger})^{\mathrm{T}}$ as the vector of the marginal t-statistics for $\dagger = \mathrm{t}, \mathrm{lm}, \mathrm{logit}$ under the treatment-control experiment. Direct comparison shows that $\hat{\Omega}_{jj} = \hat{\mathrm{se}}_j^2$ for $j = 1, \ldots, J$, and allows us to unify the marginal criteria in (S2), (S4), and (S6) as

$$\mathcal{A}_{\dagger,\mathrm{mg}} \;=\; \{|T_\dagger| \leq a_\dagger\} \qquad (\dagger = \mathrm{t}, \mathrm{lm}, \mathrm{logit}),$$

with $a_{\mathrm{logit}} = a$ and

$$\begin{aligned} T_{\mathrm{t}} &= \mathrm{diag}(\hat{\mathrm{se}}_j^{-1})_{j=1}^J \hat{\tau}_x &&= \sigma(\hat{\Omega})^{-1}\hat{\tau}_x, \\ T_{\mathrm{lm}} &= \mathrm{diag}(\hat{V}_{jj}^{-1/2})_{j=1}^J \hat{\beta} &&= \sigma(\hat{V})^{-1}\hat{\beta}, \\ T_{\mathrm{logit}} &= \mathrm{diag}(\tilde{V}_{jj}^{-1/2})_{j=1}^J \tilde{\beta} &&= \sigma(\tilde{V})^{-1}\tilde{\beta}. \end{aligned}$$

The expressions in the last column of Table S3 also extend to the criteria based on the multinomial logistic regression for general experiments with renewed definitions of $T_{\mathrm{logit}} = (t_{qj,\mathrm{logit}})_{q \in \mathcal{Q}_+;\, j=1,\ldots,J}$,

Table S2: Regression realizations, test statistics, and reference distributions for the thirteen ReP schemes in Tables 1 and 2. Let $\mathcal{I}_{i,+} = (\mathcal{I}_{i1}, \ldots, \mathcal{I}_{i,Q-1})^{\mathrm{T}}$ for $i = 1, \ldots, N$.

Q	Model option	Regression realization	Test statistics and reference distributions					
			Marginal		Joint (default)		Joint (Wald)	
2	t	$\texttt{lm}(x_{ij} \sim 1 + Z_i)$	$\hat{\tau}_{x,j}/\hat{\mathrm{se}}_j$	t_{N-2}			$\hat{\tau}_x^{\mathrm{T}}\hat{\Omega}^{-1}\hat{\tau}_x$	χ^2_J
	lm	$\texttt{lm}(Z_i \sim 1 + x_i)$	$\hat{\beta}_j/\hat{V}_{jj}^{1/2}$	t_{N-1-J}	F	$F_{J,N-1-J}$	$\hat{\beta}^{\mathrm{T}}\hat{V}^{-1}\hat{\beta}$	χ^2_J
	logit	$\texttt{logit}(Z_i \sim 1 + x_i)$	$\tilde{\beta}_j/\tilde{V}_{jj}^{1/2}$	$\mathcal{N}(0,1)$	λ_{LRT}	χ^2_J	$\tilde{\beta}^{\mathrm{T}}\tilde{V}^{-1}\tilde{\beta}$	χ^2_J
≥ 2	f	$\texttt{lm}(x_{ij} \sim 1 + \mathcal{I}_{i,+})$	F_j	$F_{Q-1,N-Q}$				
	logit	$\texttt{logit}(Z_i \sim 1 + x_i)$	$\tilde{\beta}_{qj}/\tilde{V}_{qj,qj}^{1/2}$	$\mathcal{N}(0,1)$	λ_{LRT}	$\chi^2_{J(Q-1)}$	$\tilde{\beta}^{\mathrm{T}}\tilde{V}^{-1}\tilde{\beta}$	$\chi^2_{J(Q-1)}$

$a = (a_{qj})_{q\in\mathcal{Q}_+;\ j=1,\ldots,J}$, and a_0.

Table S3: Acceptance criteria for $\dagger = \mathrm{t}, \mathrm{lm}, \mathrm{logit}$ under the treatment-control experiment.

	by p-values	t	lm	logit
$\mathcal{A}_{\dagger,\mathrm{mg}}$	$p_{j,\dagger} \geq \alpha_j\ (j = 1, \ldots, J)$	$\lvert T_{\mathrm{t}}\rvert \leq a_{\mathrm{t}}$	$\lvert T_{\mathrm{lm}}\rvert \leq a_{\mathrm{lm}}$	$\lvert T_{\mathrm{logit}}\rvert \leq a$
$\mathcal{A}_{\dagger,\mathrm{jt}}$	$p_{0,\dagger} \geq \alpha_0$	$W_{\mathrm{t}} \leq a_0$	$W_{\mathrm{lm}} \leq a_{\mathrm{f}}$	$\lambda_{\text{LRT}} \leq a_0$
$\mathcal{A}_{\dagger,\mathrm{cs}}$	$p_{j,\dagger} \geq \alpha_j\ (j = 0, 1, \ldots, J)$	$\lvert T_{\mathrm{t}}\rvert \leq a_{\mathrm{t}},\ W_{\mathrm{t}} \leq a_0$	$\lvert T_{\mathrm{lm}}\rvert \leq a_{\mathrm{lm}},\ W_{\mathrm{lm}} \leq a_{\mathrm{f}}$	$\lvert T_{\mathrm{logit}}\rvert \leq a,\ \lambda_{\text{LRT}} \leq a_0$

The test statistics satisfy

$$\begin{aligned} T_{\mathrm{t}} &= \sigma(\hat{\Omega})^{-1}\hat{\tau}_x, & T_{\mathrm{lm}} &= \sigma(\hat{V})^{-1}\hat{\beta}, & T_{\mathrm{logit}} &= \sigma(\tilde{V})^{-1}\tilde{\beta}, \\ W_{\mathrm{t}} &= \hat{\tau}_x^{\mathrm{T}}\hat{\Omega}^{-1}\hat{\tau}_x, & W_{\mathrm{lm}} &= \hat{\beta}^{\mathrm{T}}\hat{V}^{-1}\hat{\beta}, & W_{\mathrm{logit}} &= \tilde{\beta}^{\mathrm{T}}\tilde{V}^{-1}\tilde{\beta}, \end{aligned}$$

with $a_\dagger = a + o(1)$ ($\dagger = \mathrm{t}, \mathrm{lm}$) and $a_{\mathrm{f}} = a_0 + o(1)$.

Remark S1. Recall $\hat{\mathrm{se}}_j$ as the classic standard error of $\hat{\tau}_{x,j}$ from $\texttt{lm}(x_{ij} \sim 1+Z_i)$. Echoing the discussion in Section S2.3, we can replace it with the EHW standard error from the same OLS fit, denoted by $\hat{\mathrm{se}}'_j$, and conduct the marginal two-sample t-test based on the robustly studentized t-statistic $t'_{j,\mathrm{t}} = \hat{\tau}_{x,j}/\hat{\mathrm{se}}'_j$ for $j = 1, \ldots, J$.

Extensions to other criteria are straightforward by replacing the $\hat{\Omega}$, $\hat{V}$, and $\tilde{V}$ in Tables S2 and S3 with their respective heteroskedasticity-robust counterparts. In particular, the EHW counterparts of $\hat{V}$ and $\tilde{V}$ can be obtained as direct outputs from the same linear, logistic, and multinomial logistic regressions, respectively. The robust counterpart of $\hat{\Omega}$ can be computed as $\hat{\Omega}' = N_1^{-1}\hat{S}^2_x(1) + N_0^{-1}\hat{S}^2_x(0)$, recalling $\hat{S}^2_x(q)$ as the sample covariance of $\{x_i : Z_i = q\}$ for $q = 0, 1$. The resulting robust variant of the Hotelling's T^2 statistic, namely $W'_{\mathrm{t}} = \hat{\tau}_x^{\mathrm{T}}\hat{\Omega}'\hat{\tau}_x$, defines the multivariate Behrens–Fisher T^2 statistic. We show in Remark S2 in Section S4.4 below the asymptotic equivalence between the classic and robust test statistics for defining ReP.

S4. Lemmas

We give in this section the key lemmas for quantifying the asymptotic sampling properties of $\hat\tau_*$ ($* = \text{N}, \text{F}, \text{L}$) under ReP.

S4.1. Asymptotic theory under complete randomization

We review in this subsection the theory of regression adjustment under complete randomization. Assume a general experiment with $\mathcal{Q} = \{1, \dots, Q\}$ throughout. The treatment-control experiment is a special case with $Q = 2$ and level 2 relabeled as level 0.

Recall γ_q as the coefficient vector of x_i from $\texttt{lm}\{Y_i(q) \sim 1 + x_i\}$ over $i = 1, \dots, N$. Let $\gamma_\text{F} = \sum_{q\in\mathcal{Q}} e_q \gamma_q$, and let $S_* = (S_{*,qq'})_{q,q'\in\mathcal{Q}}$ ($* = \text{N}, \text{F}, \text{L}$) be the finite-population covariance matrices of $Y_{i,\text{N}}(q) = Y_i(q)$, $Y_{i,\text{F}}(q) = Y_i(q) - x_i^\text{T}\gamma_\text{F}$, and $Y_{i,\text{L}}(q) = Y_i(q) - x_i^\text{T}\gamma_q$, respectively, with $S_{*,qq'} = (N-1)^{-1}\sum_{i=1}^N \{Y_{i,*}(q) - \bar Y(q)\}\{Y_{i,*}(q') - \bar Y(q')\}$. Let

$$V_* = \text{diag}(S_{*,qq}/e_q)_{q\in\mathcal{Q}} - S_* \qquad (* = \text{N}, \text{F}, \text{L}).$$

Condition 1 ensures that e_q, γ_q, γ_F, S_x^2, and V_* all have finite limits as N tends to infinity. For notational simplicity, we will use the same symbols to denote their respective limits when no confusion would arise. Recall that $\hat x = (\hat x(1)^\text{T}, \dots, \hat x(Q)^\text{T})^\text{T}$, with $\hat x(q) = N_q^{-1}\sum_{i:Z_i=q} x_i$ for $q \in \mathcal{Q}$. Let $\hat V_*'$ be the EHW covariance of $\hat Y_*$ from the same OLS fit. Lemma S2 follows from Zhao and Ding (2023), and clarifies the design-based properties of $\hat Y_*$ and $\hat V_*'$.

Lemma S2. Assume a completely randomized general experiment and Condition 1. For $* = \text{N}, \text{F}, \text{L}$,

(i)

$$\sqrt{N}\begin{pmatrix}\hat Y_* - \bar Y \\ \hat x\end{pmatrix} \rightsquigarrow \mathcal{N}\left\{0_{Q+JQ}, \begin{pmatrix} V_* & \Gamma_* V_x \\ V_x\Gamma_*^\text{T} & V_x\end{pmatrix}\right\},$$

with $V_x = N\text{cov}(\hat x) = \{\text{diag}(e_q^{-1})_{q\in\mathcal{Q}} - 1_{Q\times Q}\} \otimes S_x^2$, $V_* = V_\text{L} + \Gamma_* V_x \Gamma_*^\text{T} \geq V_\text{L}$ for

$$\Gamma_\text{N} = \text{diag}(\gamma_q^\text{T})_{q\in\mathcal{Q}}, \qquad \Gamma_\text{F} = \text{diag}\{(\gamma_q - \gamma_\text{F})^\text{T}\}_{q\in\mathcal{Q}}, \qquad \Gamma_\text{L} = 0_{Q\times JQ},$$

and hence $\hat Y_\text{L} \succeq_\infty \hat Y_\text{N}, \hat Y_\text{F}$;

(ii) $N\hat V_*' - V_* = S_* + o_\mathbb{P}(1)$ with $S_* \geq 0$.

Lemma S2(i) ensures the consistency and asymptotic normality of $\hat{\tau}_* = G\hat{Y}_*$ for estimating $\tau = GY$ under complete randomization. Lemma S2(ii) ensures the asymptotic appropriateness of the EHW covariance for estimating the true sampling covariance, and thereby justifies the Wald-type inference of τ based on $(\hat{\tau}_*, G\hat{V}_*'G^{\mathrm{T}})$ and normal approximation.

The theory under the treatment-control experiment then follows as a special case with $\hat{\tau}_* = \hat{Y}_*(1) - \hat{Y}_*(0)$ by the invariance of OLS to non-degenerate transformation of the regressor vector. Let $\hat{\mathrm{se}}_*$ be the EHW standard error of $\hat{\tau}_*$ from the same OLS fit.

Corollary S1. Assume a completely randomized treatment-control experiment and Condition 1. For $* = \mathrm{N}, \mathrm{F}, \mathrm{L}$,

(i)

$$\sqrt{N}\begin{pmatrix} \hat{\tau}_* - \tau \\ \hat{\tau}_x \end{pmatrix} \rightsquigarrow \mathcal{N}\left\{0_{J+1}, \begin{pmatrix} v_* & c_*^{\mathrm{T}} \\ c_* & v_x \end{pmatrix}\right\},$$

with $v_* = (-1,1)V_*(-1,1)^{\mathrm{T}}$, $v_x = (e_0e_1)^{-1}S_x^2$, and

$$c_{\mathrm{N}} = S_x^2(e_0^{-1}\gamma_0 + e_1^{-1}\gamma_1), \qquad c_{\mathrm{F}} = S_x^2(e_1^{-1} - e_0^{-1})(\gamma_1 - \gamma_0), \qquad c_{\mathrm{L}} = 0_J$$

satisfying $v_* - v_{\mathrm{L}} = c_*^{\mathrm{T}} v_x^{-1} c_* \geq 0$;

(ii) $N(\hat{\mathrm{se}}_*)^2 - v_* = (-1,1)S_*(-1,1)^{\mathrm{T}} + o_{\mathbb{P}}(1)$ with $(-1,1)S_*(-1,1)^{\mathrm{T}} \geq 0$.

Parallel to the comments after Lemma S2, Corollary S1(i) states the consistency and asymptotic normality of $\hat{\tau}_*$ for estimating $\tau = \bar{Y}(1) - \bar{Y}(0)$ under complete randomization, and ensures the asymptotic efficiency of $\hat{\tau}_{\mathrm{L}}$ over $\hat{\tau}_{\mathrm{N}}$ and $\hat{\tau}_{\mathrm{F}}$. Corollary S1(ii) justifies the Wald-type inference based on $(\hat{\tau}_*, \hat{\mathrm{se}}_*)$ and normal approximation.

S4.2. Peakedness

Lemma S3 below states the celebrated Gaussian correlation inequality, with the recent breakthrough proof due to Royen (2014); see also Latała and Matlak (2017).

Lemma S3 (Gaussian correlation inequality)**.** Let μ be an m-dimensional Gaussian probability measure on $\mathbb{R}^m$, that is, μ is a multivariate normal distribution, centered at the origin. Then $\mu(\mathcal{C}_1 \cap \mathcal{C}_2) \geq \mu(\mathcal{C}_1)\mu(\mathcal{C}_2)$ for all convex sets $\mathcal{C}_1, \mathcal{C}_2 \subset \mathbb{R}^m$ that are symmetric about the origin.

Lemma S3 immediately implies Corollary S2 below, which states that a mean-zero Gaussian measure restricted to a symmetric convex set is more peaked than the original unrestricted measure.

Corollary S2. Let $\epsilon \sim \mathcal{N}(0_m, \Sigma)$. Then $\epsilon \mid \{\epsilon \in \mathcal{C}\} \succeq \epsilon$ for arbitrary convex set $\mathcal{C} \subset \mathbb{R}^m$ that is symmetric about the origin.

Proof of Corollary S2. Let $\mathcal{T} \sim \epsilon \mid \{\epsilon \in \mathcal{C}\}$. The result follows from

$$\mathbb{P}(\mathcal{T} \in \mathcal{C}_1) = \mathbb{P}(\epsilon \in \mathcal{C}_1 \mid \epsilon \in \mathcal{C}) = \frac{\mathbb{P}(\epsilon \in \mathcal{C}_1,\ \epsilon \in \mathcal{C})}{\mathbb{P}(\epsilon \in \mathcal{C})} \geq \mathbb{P}(\epsilon \in \mathcal{C}_1)$$

for arbitrary symmetric convex set $\mathcal{C}_1 \subset \mathbb{R}^m$ by Lemma S3. □

Recall that

$$\begin{aligned}
\mathcal{T}_{\mathrm{t}} &\sim \epsilon_{\mathrm{t}} \mid \{|\epsilon_{\mathrm{t}}| \leq a\}, & \mathcal{T}'_{\mathrm{t}} &\sim \epsilon_{\mathrm{t}} \mid \{|\epsilon_{\mathrm{t}}| \leq a,\ \|\epsilon_{\mathrm{t}}\|_{\mathcal{M}} \leq a_0\},\\
\mathcal{T}_{\mathrm{lm}} &\sim \epsilon_{\mathrm{lm}} \mid \{|\epsilon_{\mathrm{lm}}| \leq a\}, & \mathcal{T}'_{\mathrm{lm}} &\sim \epsilon_{\mathrm{lm}} \mid \{|\epsilon_{\mathrm{lm}}| \leq a,\ \|\epsilon_{\mathrm{lm}}\|_{\mathcal{M}} \leq a_0\},\\
\mathcal{T}_{\mathrm{logit}} &\sim \epsilon_{\mathrm{logit}} \mid \{|\epsilon_{\mathrm{logit}}| \leq a\}, & \mathcal{T}'_{\mathrm{logit}} &\sim \epsilon_{\mathrm{logit}} \mid \{|\epsilon_{\mathrm{logit}}| \leq a,\ \|\epsilon_{\mathrm{logit}}\|_{\mathcal{M}} \leq a_0\},\\
\mathcal{L} &\sim \epsilon_0 \mid \{\|\epsilon_0\|_2^2 \leq a_0\}, & \mathcal{T}_{\mathrm{f}} &\sim \epsilon_{\mathrm{f}} \mid \{\textstyle\sum_{q\in\mathcal{Q}} e_q \epsilon_{\mathrm{f},qj}^2 \leq a'_j S_{x,j}^2 \text{ for all } j = 1,\dots,J\},
\end{aligned}$$

with $\epsilon_{\mathrm{t}} \sim \mathcal{N}\{0_J, D(v_x)\}$, $\epsilon_{\mathrm{lm}} \sim \mathcal{N}\{0_J, D(v_x^{-1})\}$, $\epsilon_{\mathrm{logit}} \sim \mathcal{N}\{0_{J(Q-1)}, D(V_\Psi)\}$, $\epsilon_0 \sim \mathcal{N}(0_{J(Q-1)}, I_{J(Q-1)})$, and $\epsilon_{\mathrm{f}} \sim \mathcal{N}(0_{JQ}, V_x)$.

Lemma S4. $\mathcal{T}_*, \mathcal{T}'_* \succeq \epsilon_*$ for $* = \mathrm{t}, \mathrm{lm}, \mathrm{logit}$, $\mathcal{L} \succeq \epsilon_0$, and $\mathcal{T}_{\mathrm{f}} \succeq \epsilon_{\mathrm{f}}$.

Proof of Lemma S4. The results follow from Corollary S2 and the convexity of $\{u \in \mathbb{R}^m : |u| \leq a\}$, $\{u \in \mathbb{R}^m : |u| \leq a,\ \|u\|_{\mathcal{M}} \leq a_0\}$, $\{u \in \mathbb{R}^m : \|u\|_2^2 \leq a_0\}$, and $\{u = (u_{qj})_{q\in\mathcal{Q};\ j=1,\dots,J} \in \mathbb{R}^{JQ} : \sum_{q\in\mathcal{Q}} e_q u_{qj}^2 \leq a'_j S_{x,j}^2 \text{ for all } j = 1,\dots,J\}$.

□

Lemma S5 below reviews three classical results in probability for comparing peakedness between random vectors. Lemma S5(i)–(ii) are proved by Dharmadhikari and Joag-Dev (1988, Lemma 7.2 and Theorem 7.5), and Li et al. (2020) used them. Lemma S5(iii) is from Sherman (1955, Lemma 3) and underpins the results for stratified experiments.

Lemma S5. (i) If two $m \times 1$ symmetric random vectors A and B satisfy $A \succeq B$, then $CA \succeq CB$ for any matrix C with compatible dimensions.

(ii) Let A, B_1, and B_2 be three independent $m \times 1$ symmetric random vectors. If A is normal and $B_1 \succeq B_2$, then $A + B_1 \succeq A + B_2$.

(iii) Let A_1, A_2, B_1, B_2 be continuous random variables such that (a) A_1 and A_2 are independent, B_1 and B_2 are independent and (b) A_i is more peaked than B_i for $i = 1, 2$. Then $A_1 + A_2$ is more peaked than $B_1 + B_2$.

S4.3. Design-based properties of multinomial logistic regression

Theorem S1 below is a novel technical result, and clarifies the design-based properties of the MLE outputs from the logistic and multinomial logistic regressions, respectively. The result ensures the asymptotic equivalence between the LRT and the Wald test in terms of both the test statistics and the corresponding p-values. We relegate the proof to Section S7.

Recall that $\hat{x}_+ = (\hat{x}(1)^{\mathrm{T}}, \dots, \hat{x}(Q-1)^{\mathrm{T}})^{\mathrm{T}}$, with $V_{x+} = N\mathrm{cov}(\hat{x}_+)$ denoting its scaled sampling covariance under complete randomization. Recall that $V_\Psi = N\mathrm{cov}(\Psi\hat{x}_+) = \Psi V_{x+}\Psi^{\mathrm{T}}$, with $\Psi = \{\Phi^{-1}\,\mathrm{diag}(e_+)\} \otimes (S_x^2)^{-1}$, $e_+ = (e_1, \dots, e_{Q-1})^{\mathrm{T}}$, and $\Phi = \mathrm{diag}(e_+) - e_+e_+^{\mathrm{T}}$.

Theorem S1. Consider a completely randomized experiment with $Q \geq 2$ treatment arms. Under Conditions 1 and 3, we have

$$\sqrt{N}(\tilde{\beta} - \Psi\hat{x}_+) = o_{\mathbb{P}}(1), \qquad \sqrt{N}\tilde{\beta} \rightsquigarrow \mathcal{N}(0_{J(Q-1)}, V_\Psi), \qquad N\tilde{V} = V_\Psi + o_{\mathbb{P}}(1),$$
$$\lambda_{\mathrm{LRT}} - N\hat{x}_+^{\mathrm{T}}V_{x+}^{-1}\hat{x}_+ = o_{\mathbb{P}}(1), \qquad W_{\mathrm{logit}} - N\hat{x}_+^{\mathrm{T}}V_{x+}^{-1}\hat{x}_+ = o_{\mathbb{P}}(1), \qquad \lambda_{\mathrm{LRT}} - W_{\mathrm{logit}} = o_{\mathbb{P}}(1),$$
$$\lambda_{\mathrm{LRT}} \rightsquigarrow \chi^2_{J(Q-1)}, \qquad W_{\mathrm{logit}} \rightsquigarrow \chi^2_{J(Q-1)}.$$

For a treatment-control experiment with $Q = 2$ and the reference level $q = 2$ relabeled as 0, the results simplify to

$$\sqrt{N}\{\tilde{\beta} - (S_x^2)^{-1}\hat{\tau}_x\} = o_{\mathbb{P}}(1), \qquad \sqrt{N}\tilde{\beta} \rightsquigarrow \mathcal{N}\{0_J, (e_0e_1)^{-1}(S_x^2)^{-1}\},$$
$$N\tilde{V} = (e_0e_1)^{-1}(S_x^2)^{-1} + o_{\mathbb{P}}(1).$$

S4.4. Weak convergence

Lemma S6 below underpins the asymptotic equivalence between balance criteria based on asymptotically equivalent thresholds or test statistics. The proof follows from standard probability calculation and is thus omitted.

Lemma S6. Let $(A_N)_{N=1}^\infty$ be a sequence of $m \times 1$ random vectors. Let $(B_N)_{N=1}^\infty$ and $(B'_N)_{N=1}^\infty$ be two sequences of random variables with $B_N - B'_N = o_{\mathbb{P}}(1)$ and $(A_N, B_N)_{N=1}^\infty$ having a continuous limiting distribution, represented by (A, B). Let $(b_N)_{N=1}^\infty$ be a sequence of constants with a finite limit, $b_\infty = \lim_{N\to\infty} b_N < \infty$. Then

(i) $A_N \mid \{B_N \le b_N\} \dot\sim A_N \mid \{B_N \le b_\infty\}$ provided $\mathbb{P}(B \le b_\infty) > 0$;

(ii) $A_N \mid \{B_N \le b\} \dot\sim A_N \mid \{B'_N \le b\}$ for arbitrary fixed $b \in \mathbb{R}$ that satisfies $\mathbb{P}(B \le b) > 0$.

Definition S1 below extends the notion of rerandomization with general covariate balance criterion (ReG) from Li et al. (2018).

Definition S1. Let $\phi(B, C)$ be a binary *covariate balance indicator function*, where $\phi(\cdot,\cdot)$ is a binary indicator function and (B, C) are two statistics computed from the data. An ReG accepts a randomization if $\phi(B, C) = 1$.

The definition of ReG is general and includes all nine criteria in Table S3 as special cases. Table S4 below summarizes the covariate balance indicator functions for ReM and the nine ReP schemes in Table S3 under the treatment-control experiment, respectively. As an illustration, $(B, C) = (\sqrt{N}\hat\tau_x, N\hat\Omega)$ under the two-sample t-test model option, with $\phi(\cdot,\cdot)$ equaling

$$\phi(u, v) = \begin{cases} 1\{|\sigma(v)^{-1}u| \le a_{\mathrm{t}}\} & \text{under the marginal rule;} \\ 1(u^{\mathrm{T}}v^{-1}u \le a_0) & \text{under the joint rule;} \\ 1\{|\sigma(v)^{-1}u| \le a_{\mathrm{t}},\ u^{\mathrm{T}}v^{-1}u \le a_0\} & \text{under the consensus rule.} \end{cases}$$

The resulting covariate balance indicator functions equal

$$\phi(B, C) = \begin{cases} 1\{|\sigma(\hat\Omega)^{-1}\hat\tau_x| \le a_{\mathrm{t}}\} & \text{under the marginal rule;} \\ 1(\hat\tau_x^{\mathrm{T}}\hat\Omega^{-1}\hat\tau_x \le a_0) & \text{under the joint rule;} \\ 1\{|\sigma(\hat\Omega)^{-1}\hat\tau_x| \le a_{\mathrm{t}},\ \hat\tau_x^{\mathrm{T}}\hat\Omega^{-1}\hat\tau_x \le a_0\} & \text{under the consensus rule} \end{cases}$$

given $\phi(u, v) = \phi(u/\sqrt{N}, v/N)$ in all three cases.

Lemma S7 below is a generalization of Li et al. (2018, Proposition A1), and gives the asymptotic distribution of arbitrary random elements under ReG. To this end, Condition S1 below imposes some smoothness constraints on the associated ϕ to prevent the acceptance region from being a set of measure zero. All covariate balance criteria in Table S4 satisfy Condition S1.

Condition S1. The binary indicator function $\phi(\cdot,\cdot)$ satisfies: (i) $\phi(\cdot,\cdot)$ is almost surely continuous; (ii) for $u \sim \mathcal{N}(0_J, v_0)$, we have $\mathbb{P}\{\phi(u, v_0) = 1\} > 0$ for all $v_0 > 0$, and $\mathrm{cov}\{u \mid \phi(u, v_0) = 1\}$ is a continuous function of v_0.

Lemma S7 (Weak convergence under ReG)**.** Assume $(\phi_N)_{N=1}^\infty$ as a sequence of binary indicator functions under Condition S1 that converges to ϕ. For a sequence of random elements $(A_N, B_N, C_N)_{N=1}^\infty$ that

Table S4: The covariate balance indicator functions and the corresponding (B, C) for ReM and the nine ReP schemes in Table S3 under the treatment-control experiment. The joint criterion under the "logit" model option is given in terms of the asymptotically equivalent Wald statistic to highlight the analogy across different model options.

	ReM	t
(B, C)	$(\sqrt{N}\hat{\tau}_x, N\text{cov}(\hat{\tau}_x))$	$(\sqrt{N}\hat{\tau}_x, N\hat{\Omega})$
marginal	n.a.	$1\{\lvert\sigma(\hat{\Omega})^{-1}\hat{\tau}_x\rvert \le a_\text{t}\}$
joint	$1\{\hat{\tau}_x^\text{T}\text{cov}(\hat{\tau}_x)^{-1}\hat{\tau}_x \le a_0\}$	$1(\hat{\tau}_x^\text{T}\hat{\Omega}^{-1}\hat{\tau}_x \le a_0)$
consensus	n.a.	$1\{\lvert\sigma(\hat{\Omega})^{-1}\hat{\tau}_x\rvert \le a_\text{t},\ \hat{\tau}_x^\text{T}\hat{\Omega}^{-1}\hat{\tau}_x \le a_0\}$
	lm	logit
(B, C)	$(\sqrt{N}\hat{\beta}, N\hat{V})$	$(\sqrt{N}\tilde{\beta}, N\tilde{V})$
marginal	$1\{\lvert\sigma(\hat{V})^{-1}\hat{\beta}\rvert \le a_\text{lm}\}$	$1\{\lvert\sigma(\tilde{V})^{-1}\tilde{\beta}\rvert \le a\}$
joint	$1(\hat{\beta}^\text{T}\hat{V}^{-1}\hat{\beta} \le a_\text{f})$	$1(\tilde{\beta}^\text{T}\tilde{V}^{-1}\tilde{\beta} \le a_0)$
consensus	$1\{\lvert\sigma(\hat{V})^{-1}\hat{\beta}\rvert \le a_\text{lm},\ \hat{\beta}^\text{T}\hat{V}^{-1}\hat{\beta} \le a_\text{f}\}$	$1\{\lvert\sigma(\tilde{V})^{-1}\tilde{\beta}\rvert \le a,\ \tilde{\beta}^\text{T}\tilde{V}^{-1}\tilde{\beta} \le a_0\}$

satisfies $(A_N, B_N, C_N) \rightsquigarrow (A, B, C)$ as $N \to \infty$, we have

$$(A_N, B_N) \mid \{\phi_N(B_N, C_N) = 1\} \quad \rightsquigarrow \quad (A, B) \mid \{\phi(B, C) = 1\}$$

in the sense that, for any continuity set $\mathcal{S}$ of $(A, B) \mid \{\phi(B, C) = 1\}$,

$$\mathbb{P}\{(A_N, B_N) \in \mathcal{S} \mid \phi_N(B_N, C_N) = 1\} = \mathbb{P}\{(A, B) \in \mathcal{S} \mid \phi(B, C) = 1\} + o(1).$$

Lemma S8 below gives the asymptotic joint distributions of $\hat{\tau}_*$ and the elements that determine the covariate balance measures in Table S4. The result provides the basis for verifying Propositions 1 and 2 in a unified way.

Lemma S8. Assume a completely randomized treatment-control experiment and Conditions 1 and 2. For $* = \text{N}, \text{F}, \text{L}$, we have

$$\begin{aligned}
\left(\sqrt{N}(\hat{\tau}_* - \tau), \sqrt{N}\hat{\tau}_x, N\text{cov}(\hat{\tau}_x)\right) &\quad \rightsquigarrow \quad (A_*, B, v_x),\\
\left(\sqrt{N}(\hat{\tau}_* - \tau), \sqrt{N}\hat{\tau}_x, N\hat{\Omega}\right) &\quad \rightsquigarrow \quad (A_*, B, v_x),\\
\left(\sqrt{N}(\hat{\tau}_* - \tau), \sqrt{N}\hat{\beta}, N\hat{V}\right) &\quad \rightsquigarrow \quad (A_*, v_x^{-1}B, v_x^{-1}),\\
\left(\sqrt{N}(\hat{\tau}_* - \tau), \sqrt{N}\tilde{\beta}, N\tilde{V}\right) &\quad \rightsquigarrow \quad \left(A_*, (S_x^2)^{-1}B, (e_0e_1)^{-1}(S_x^2)^{-1}\right)
\end{aligned}$$

with

$$\begin{pmatrix} A_* \\ B \end{pmatrix} \sim \mathcal{N}\left\{0_{J+1}, \begin{pmatrix} v_* & c_*^{\mathrm{T}} \\ c_* & v_x \end{pmatrix}\right\}.$$

Proof of Lemma S8. The result on ReM follows from Corollary S1. The result on the "logit" model option follows from Theorem S1. We verify below the results on the two-sample t-test ("t") and linear regression ("lm") model options, respectively.

Result on the two-sample t-test model option. The joint distribution of $\sqrt{N}(\hat{\tau}_* - \tau, \hat{\tau}_x)$ follows from Corollary S1. The probability limit of $N\hat{\Omega}$ follows from $N\hat{\Omega} = (e_0 e_1)^{-1} S_x^2 + o_{\mathbb{P}}(1) = v_x + o_{\mathbb{P}}(1)$ by (S3) and $\hat{S}_x^2(q) = S_x^2 + o_{\mathbb{P}}(1)$ under complete randomization and Condition 1.

Result on the linear regression model option. The Frisch–Waugh–Lovell theorem ensures that

$$\hat{\beta} = \frac{N_1}{N-1}(S_x^2)^{-1}\hat{x}(1) = \frac{N_0 N_1}{(N-1)N}(S_x^2)^{-1}\hat{\tau}_x = \frac{N}{N-1} v_x^{-1}\hat{\tau}_x \tag{S9}$$

is a non-degenerate linear transformation of $\hat{\tau}_x$. The joint distribution of $\sqrt{N}(\hat{\tau}_* - \tau, \hat{\beta})$ then follows from Corollary S1.

Further let

$$S_Z^2 = \frac{1}{N-1}\sum_{i=1}^N (Z_i - \bar{Z})^2 = \frac{N}{N-1} e_0 e_1, \quad S_{xZ} = \frac{1}{N-1}\sum_{i=1}^N x_i (Z_i - \bar{Z}) = \frac{N}{N-1} e_1 \hat{x}(1),$$

with $\bar{Z} = N^{-1}\sum_{i=1}^N Z_i = e_1$. The probability limit of $\hat{V}$ follows from

$$N\hat{V} = \hat{\sigma}^2 \left(\sum_{i=1}^N x_i x_i^{\mathrm{T}}\right)^{-1} = \frac{N}{N-1}\hat{\sigma}^2 (S_x^2)^{-1}$$

with

$$\hat{\sigma}^2 = \frac{1}{N-1-J}\sum_{i=1}^N (Z_i - \bar{Z} - x_i^{\mathrm{T}}\hat{\beta})^2 = \frac{N-1}{N-1-J}(S_Z^2 + \hat{\beta}^{\mathrm{T}} S_x^2 \hat{\beta} - 2\hat{\beta}^{\mathrm{T}} S_{xZ}) = e_0 e_1 + o_{\mathbb{P}}(1).$$

□

Lemma S9 gives the asymptotic joint distributions of $(\hat{Y}_*, \hat{x}_+)$ and $(\hat{Y}_*, \tilde{\beta}, \tilde{V})$ under the general experiment, respectively, analogous to Lemma S8.

Lemma S9. Assume a completely randomized general experiment and Conditions 1 and 3. Let $A \sim \mathcal{N}(0_Q, V_{\mathrm{L}})$, $B \sim \mathcal{N}(0_{JQ}, V_x)$, and $B' \sim \mathcal{N}(0_{J(Q-1)}, V_{x+})$ be independent normal random vectors. For

$* = \text{N}, \text{F}, \text{L}$, we have

(i) $\sqrt{N}(\hat{Y}_* - \bar{Y}, \hat{x}) \rightsquigarrow (A + \Gamma_* B, B)$, $\sqrt{N}(\hat{Y}_* - \bar{Y}, \hat{x}_+) \rightsquigarrow (A + \Gamma'_* B', B')$;

(ii) $(\sqrt{N}(\hat{Y}_* - \bar{Y}), \sqrt{N}\tilde{\beta}, N\tilde{V}) \rightsquigarrow (A + \Gamma'_* B', \Psi B', V_\Psi)$;

(iii) $\hat{Y}_* \mid \{\lambda_{\text{LRT}} \leq a_0\} \ \dot{\sim} \ \hat{Y}_* \mid \{W_{\text{logit}} \leq a_0\}$.

Proof of Lemma S9. The result on $\sqrt{N}(\hat{Y}_* - \bar{Y}, \hat{x})$ follows from Lemma S2. This, together with (S1), further ensures $\sqrt{N}(\hat{Y}_* - \bar{Y} - \Gamma'_* \hat{x}_+, \hat{x}_+) \rightsquigarrow (A, B')$, and hence $\sqrt{N}(\hat{Y}_* - \bar{Y}, \hat{x}_+) \rightsquigarrow (A + \Gamma'_* B', B')$ for $* = \text{N}, \text{F}, \text{L}$.

The result on $(\sqrt{N}(\hat{Y}_* - \bar{Y}), \sqrt{N}\tilde{\beta}, N\tilde{V})$ then follows from $\sqrt{N}(\tilde{\beta} - \Psi \hat{x}_+) = o_{\mathbb{P}}(1)$ and $N\tilde{V} = V_\Psi + o_{\mathbb{P}}(1)$ by Theorem S1.

The asymptotic equivalence between $\hat{Y}_* \mid \{\lambda_{\text{LRT}} \leq a_0\}$ and $\hat{Y}_* \mid \{W_{\text{logit}} \leq a_0\}$ follows from $\lambda_{\text{LRT}} - N\hat{x}_+^{\mathrm{T}} V_{x+}^{-1} \hat{x}_+ = o_{\mathbb{P}}(1)$ and $W_{\text{logit}} - N\hat{x}_+^{\mathrm{T}} V_{x+}^{-1} \hat{x}_+ = o_{\mathbb{P}}(1)$ by Theorem S1 and Lemma S6. □

Remark S2. Lemma S8 and its proof also imply some of the comments we made in the main text.

First, Lemma S8, together with (S9) and $\sqrt{N}\{\tilde{\beta} - (S_x^2)^{-1}\hat{\tau}_x\} = o_{\mathbb{P}}(1)$ from Theorem S1, ensures the asymptotic equivalence between $\|\hat{\tau}_x\|_{\mathcal{M}} = \hat{\tau}_x^{\mathrm{T}} \text{cov}(\hat{\tau}_x)^{-1} \hat{\tau}_x$, $W_{\text{t}} = \hat{\tau}_x^{\mathrm{T}} \hat{\Omega}^{-1} \hat{\tau}_x$, $W_{\text{lm}} = \hat{\beta}^{\mathrm{T}} \hat{V}^{-1} \hat{\beta}$, and $W_{\text{logit}} = \tilde{\beta}^{\mathrm{T}} \tilde{V}^{-1} \tilde{\beta}$ in the sense that $W_\dagger - \|\hat{\tau}_x\|_{\mathcal{M}} = o_{\mathbb{P}}(1)$ for $\dagger = \text{t}, \text{lm}, \text{logit}$. This elucidates the asymptotic equivalence between ReM and the joint criteria under the treatment-control experiment by Lemma S6.

Next, recall $\hat{\text{se}}_j$ and $\hat{\text{se}}'_j$ as the classic and EHW standard errors of $\hat{\tau}_{x,j}$, respectively, from Remark S1. It follows from the Frisch–Waugh–Lovell theorem and Zhao and Ding (2021, Lemma S1) that

$$\hat{\text{se}}_j^2 = \frac{1}{N-2}\left(\frac{S_{x,j}^2}{S_Z^2} - \hat{\tau}_{x,j}^2\right), \qquad (\hat{\text{se}}'_j)^2 = \frac{(Z - 1_N \bar{Z})^{\mathrm{T}} \operatorname{diag}(\epsilon_{ij}^2)_{i=1}^N (Z - 1_N \bar{Z})}{\|Z - 1_N \bar{Z}\|_2^2}$$

with $Z = (Z_1, \ldots, Z_N)^{\mathrm{T}}$, $S_{x,j}^2 = (N-1)^{-1} \sum_{i=1}^N x_{ij}^2$, $\epsilon_{ij} = x_{ij} - Z_i \hat{\tau}_{x,j}$, and

$$N\hat{\text{se}}_j^2 = (e_0 e_1)^{-1} S_{x,j}^2 + o_{\mathbb{P}}(1), \qquad N(\hat{\text{se}}'_j)^2 = (e_0 e_1)^{-1} S_{x,j}^2 + o_{\mathbb{P}}(1).$$

This ensures $t'_{j,\text{t}} = \hat{\tau}_{x,j} / \hat{\text{se}}'_j = t_{j,\text{t}} + o_{\mathbb{P}}(1)$, and hence the asymptotic equivalence of the classic and EHW standard errors for constructing the marginal criterion under the two-sample t-tests. The results for other criteria are similar and thus omitted.

Importantly, the asymptotic equivalence between the classic and EHW standard errors does not hold for $\hat{\text{se}}'_*$ ($* = \text{N}, \text{F}, \text{L}$) and their classic counterparts based on the default outputs of $\texttt{lm}(Y_i \sim 1 + Z_i)$, $\texttt{lm}(Y_i \sim 1 + Z_i + x_i)$, and $\texttt{lm}(Y_i \sim 1 + Z_i + x_i + Z_i x_i)$ in general. Specifically, the classic standard errors of $\hat{\tau}_*$ ($* = \text{N}, \text{F}, \text{L}$) are not necessarily asymptotically conservative for estimating the true sampling variances, and can thus lead to invalid inferences. As a result, the use of EHW standard errors is immaterial for rerandomization yet crucial for analysis.

S5. Proofs of the results under complete randomization

S5.1. Asymptotic distributions in Propositions 1–2 and S1–S2

Proof of Propositions 1 and 2. Let

$$\begin{pmatrix} A_* \\ B \end{pmatrix} \sim \mathcal{N}\left\{0_{J+1}, \begin{pmatrix} v_* & c_*^{\mathrm{T}} \\ c_* & v_x \end{pmatrix}\right\}$$

with $\sqrt{N}(\hat{\tau}_* - \tau, \hat{\tau}_x^{\mathrm{T}})^{\mathrm{T}} \rightsquigarrow (A_*, B^{\mathrm{T}})^{\mathrm{T}}$ for $* = \mathrm{N}, \mathrm{F}, \mathrm{L}$. Recall the definitions of $\mathcal{A}_{\dagger,\diamond}$ for $\dagger = \mathrm{t}, \mathrm{lm}, \mathrm{logit}$ and $\diamond = \mathrm{jt}, \mathrm{mg}, \mathrm{cs}$ in terms of the test statistics from Table S3.

Two-sample t-tests. For $* = \mathrm{N}, \mathrm{F}, \mathrm{L}$, let $(A_N, B_N, C_N) = (\sqrt{N}(\hat{\tau}_* - \tau), \sqrt{N}\hat{\tau}_x, N\hat{\Omega})$ with $(A_N, B_N, C_N) \rightsquigarrow (A_*, B, v_x)$ by Lemma S8.

- Recall that $\mathcal{A}_{\mathrm{t,jt}} = \{p_{0,\mathrm{t}} \geq \alpha_0\} = \{W_{\mathrm{t}} \leq a_0\}$ under the joint rule. We have

$$\begin{aligned} \sqrt{N}(\hat{\tau}_* - \tau) \mid \mathcal{A}_{\mathrm{t,jt}} &= \sqrt{N}(\hat{\tau}_* - \tau) \mid \{\hat{\tau}_x^{\mathrm{T}}\hat{\Omega}^{-1}\hat{\tau}_x \leq a_0\} \\ &\rightsquigarrow A_* \mid \{B^{\mathrm{T}} v_x^{-1} B \leq a_0\} \\ &\sim A_* \mid \{\|B\|_{\mathcal{M}} \leq a_0\} \end{aligned}$$

  by applying Lemma S7 to (A_N, B_N, C_N) and $\phi_N(u, v) = \phi(u, v) = 1(u^{\mathrm{T}} v^{-1} u \leq a_0)$.

- Recall that $\mathcal{A}_{\mathrm{t,mg}} = \{p_{j,\mathrm{t}} \geq \alpha_j,\ j = 1, \ldots, J\} = \{|T_{\mathrm{t}}| \leq a_{\mathrm{t}}\}$ under the marginal rule. We have

$$\begin{aligned} \sqrt{N}(\hat{\tau}_* - \tau) \mid \mathcal{A}_{\mathrm{t,mg}} &= \sqrt{N}(\hat{\tau}_* - \tau) \mid \{|\sigma(\hat{\Omega})^{-1}\hat{\tau}_x| \leq a_{\mathrm{t}}\} \\ &\rightsquigarrow A_* \mid \{|\sigma(v_x)^{-1} B| \leq a\} \end{aligned}$$

  by applying Lemma S7 to (A_N, B_N, C_N), $\phi_N(u, v) = 1\{|\sigma(v)^{-1}u| \leq a_{\mathrm{t}}\}$, and $\phi(u, v) = 1\{|\sigma(v)^{-1}u| \leq a\}$.

- Recall that $\mathcal{A}_{\mathrm{t,cs}} = \{p_{j,\mathrm{t}} \geq \alpha_j,\ j = 0, 1, \ldots, J\} = \{W_{\mathrm{t}} \leq a_0,\ |T_{\mathrm{t}}| \leq a_{\mathrm{t}}\}$ under the consensus rule. We have

$$\sqrt{N}(\hat{\tau}_* - \tau) \mid \mathcal{A}_{\mathrm{t,cs}} = \sqrt{N}(\hat{\tau}_* - \tau) \mid \{\hat{\tau}_x^{\mathrm{T}}\hat{\Omega}^{-1}\hat{\tau}_x \leq a_0,\ |\sigma(\hat{\Omega})^{-1}\hat{\tau}_x| \leq a_{\mathrm{t}}\}$$

$$\rightsquigarrow \quad A_* \mid \{\|B\|_{\mathcal{M}} \le a_0,\ |\sigma(v_x)^{-1}B| \le a\}$$

by applying Lemma S7 to (A_N, B_N, C_N), $\phi_N(u,v) = 1(u^{\mathrm{T}}v^{-1}u \le a_0) \cdot 1\{|\sigma(v)^{-1}u| \le a_{\mathrm{t}}\}$, and $\phi(u,v) = 1(u^{\mathrm{T}}v^{-1}u \le a_0) \cdot 1\{|\sigma(v)^{-1}u| \le a\}$.

It thus suffices to compute

$$A_* \mid \{\|B\|_{\mathcal{M}} \le a_0\}, \quad A_* \mid \{|\sigma(v_x)^{-1}B| \le a\}, \quad A_* \mid \{\|B\|_{\mathcal{M}} \le a_0,\ |\sigma(v_x)^{-1}B| \le a\},$$

respectively.

To this end, write

$$A_* = (A_* - c_*^{\mathrm{T}} v_x^{-1} B) + c_*^{\mathrm{T}} v_x^{-1} B, \tag{S10}$$

with $A_* - c_*^{\mathrm{T}} v_x^{-1} B \sim \mathcal{N}(0, v_{\mathrm{L}})$ and independent of B. This ensures

$$\begin{aligned} A_* \mid \{\|B\|_{\mathcal{M}} \le a_0\} &\sim (A_* - c_*^{\mathrm{T}} v_x^{-1} B) + c_*^{\mathrm{T}} v_x^{-1} \left[B \mid \{\|B\|_{\mathcal{M}} \le a_0\}\right] \\ &\sim v_{\mathrm{L}}^{1/2}\epsilon + c_*^{\mathrm{T}} v_x^{-1}(v_x^{1/2}\mathcal{L}). \end{aligned} \tag{S11}$$

Likewise for the results under the marginal and consensus rules. In particular, let $\epsilon_{\mathrm{t}} = \sigma(v_x)^{-1}B \sim \mathcal{N}\{0, D(v_x)\}$ to write

$$B = \sigma(v_x)\epsilon_{\mathrm{t}}, \qquad \{|\sigma(v_x)^{-1}B| \le a\} = \{|\epsilon_{\mathrm{t}}| \le a\}.$$

This ensures

$$\begin{aligned} B \mid \{|\sigma(v_x)^{-1}B| \le a\} &= \sigma(v_x)\epsilon_{\mathrm{t}} \mid \{|\epsilon_{\mathrm{t}}| \le a\} \\ &\sim \sigma(v_x)\mathcal{T}_{\mathrm{t}}, \\ B \mid \{\|B\|_{\mathcal{M}} \le a_0,\ |\sigma(v_x)^{-1}B| \le a\} &= \sigma(v_x)\epsilon_{\mathrm{t}} \mid \{\|\epsilon_{\mathrm{t}}\|_{\mathcal{M}} \le a_0,\ |\epsilon_{\mathrm{t}}| \le a\} \\ &\sim \sigma(v_x)\mathcal{T}_{\mathrm{t}}', \end{aligned}$$

and thus

$$\begin{aligned} A_* \mid \{|\sigma(v_x)^{-1}B| \le a\} &\sim v_{\mathrm{L}}^{1/2}\epsilon + c_*^{\mathrm{T}} v_x^{-1}\sigma(v_x)\mathcal{T}_{\mathrm{t}}, \\ A_* \mid \{\|B\|_{\mathcal{M}} \le a_0,\ |\sigma(v_x)^{-1}B| \le a\} &\sim v_{\mathrm{L}}^{1/2}\epsilon + c_*^{\mathrm{T}} v_x^{-1}\sigma(v_x)\mathcal{T}_{\mathrm{t}}' \end{aligned}$$

by (S10).

Linear regression. For $* = \textsc{n}, \textsc{f}, \textsc{l}$, let $(A_N, B_N, C_N) = (\sqrt{N}(\hat\tau_* - \tau), \sqrt{N}\hat\beta, N\hat V)$ with $(A_N, B_N, C_N) \rightsquigarrow (A_*, v_x^{-1}B, v_x^{-1})$ by Lemma S8.

- Recall that $\mathcal{A}_{\mathrm{lm,jt}} = \{p_{0,\mathrm{lm}} \geq \alpha_0\} = \{W_{\mathrm{lm}} \leq a_{\mathrm{f}}\}$ under the joint rule by Lemma S1. We have

$$\begin{aligned}\sqrt{N}(\hat\tau_* - \tau) \mid \mathcal{A}_{\mathrm{lm,jt}} &= \sqrt{N}(\hat\tau_* - \tau) \mid \{\hat\beta^{\mathrm{T}}\hat V^{-1}\hat\beta \leq a_{\mathrm{f}}\} \\ &\rightsquigarrow A_* \mid \{(v_x^{-1}B)^{\mathrm{T}} v_x (v_x^{-1}B) \leq a_0\} \\ &\sim A_* \mid \{\|B\|_{\mathcal{M}} \leq a_0\}\end{aligned}$$

  by applying Lemma S7 to (A_N, B_N, C_N), $\phi_N(u,v) = 1(u^{\mathrm{T}}v^{-1}u \leq a_{\mathrm{f}})$, and $\phi(u,v) = 1(u^{\mathrm{T}}v^{-1}u \leq a_0)$.

- Recall that $\mathcal{A}_{\mathrm{lm,mg}} = \{p_{j,\mathrm{lm}} \geq \alpha_j,\ j = 1, \dots, J\} = \{|T_{\mathrm{lm}}| \leq a_{\mathrm{lm}}\}$ under the marginal rule. We have

$$\begin{aligned}\sqrt{N}(\hat\tau_* - \tau) \mid \mathcal{A}_{\mathrm{lm,mg}} &= \sqrt{N}(\hat\tau_* - \tau) \mid \{|\sigma(\hat V)^{-1}\hat\beta| \leq a_{\mathrm{lm}}\} \\ &\rightsquigarrow A_* \mid \{|\sigma(v_x^{-1})^{-1}v_x^{-1}B| \leq a\}\end{aligned}$$

  by applying Lemma S7 to (A_N, B_N, C_N), $\phi_N(u,v) = 1\{|\sigma(v)^{-1}u| \leq a_{\mathrm{lm}}\}$, and $\phi(u,v) = 1\{|\sigma(v)^{-1}u| \leq a\}$.

- Recall that $\mathcal{A}_{\mathrm{lm,cs}} = \{p_{j,\mathrm{lm}} \geq \alpha_j,\ j = 0, 1, \dots, J\} = \{W_{\mathrm{lm}} \leq a_{\mathrm{f}},\ |T_{\mathrm{lm}}| \leq a_{\mathrm{lm}}\}$ under the consensus rule. We have

$$\begin{aligned}\sqrt{N}(\hat\tau_* - \tau) \mid \mathcal{A}_{\mathrm{lm,cs}} &= \sqrt{N}(\hat\tau_* - \tau) \mid \{\hat\beta^{\mathrm{T}}\hat V^{-1}\hat\beta \leq a_{\mathrm{f}},\ |\sigma(\hat V)^{-1}\hat\beta| \leq a_{\mathrm{lm}}\} \\ &\rightsquigarrow A_* \mid \{(v_x^{-1}B)^{\mathrm{T}} v_x (v_x^{-1}B) \leq a_0,\ |\sigma(v_x^{-1})^{-1}v_x^{-1}B| \leq a\} \\ &\sim A_* \mid \{\|B\|_{\mathcal{M}} \leq a_0,\ |\sigma(v_x^{-1})^{-1}v_x^{-1}B| \leq a\}\end{aligned}$$

  by applying Lemma S7 to (A_N, B_N, C_N), $\phi_N(u,v) = 1(u^{\mathrm{T}}v^{-1}u \leq a_{\mathrm{f}}) \cdot 1\{|\sigma(v)^{-1}u| \leq a_{\mathrm{lm}}\}$, and $\phi(u,v) = 1(u^{\mathrm{T}}v^{-1}u \leq a_0) \cdot 1\{|\sigma(v)^{-1}u| \leq a\}$.

It thus suffices to compute $A_* \mid \{\|B\|_{\mathcal{M}} \leq a_0\}$, $A_* \mid \{|\sigma(v_x^{-1})^{-1}v_x^{-1}B| \leq a\}$, and $A_* \mid \{\|B\|_{\mathcal{M}} \leq a_0,\ |\sigma(v_x^{-1})^{-1}v_x^{-1}B| \leq a\}$, respectively.

The distribution of $A_* \mid \{\|B\|_{\mathcal{M}} \leq a_0\}$ is given by (S11). For $A_* \mid \{|\sigma(v_x^{-1})^{-1}v_x^{-1}B| \leq a\}$ and $A_* \mid \{\|B\|_{\mathcal{M}} \leq a_0,\ |\sigma(v_x^{-1})^{-1}v_x^{-1}B| \leq a\}$, let $\epsilon_{\mathrm{lm}} = \sigma(v_x^{-1})^{-1}v_x^{-1}B \sim \mathcal{N}\{0, D(v_x^{-1})\}$ to write

$$B = v_x\sigma(v_x^{-1})\epsilon_{\mathrm{lm}}, \qquad \{|\sigma(v_x^{-1})^{-1}v_x^{-1}B| \leq a\} = \{|\epsilon_{\mathrm{lm}}| \leq a\}.$$

This ensures

$$
\begin{aligned}
B \mid \{|\sigma(v_x^{-1})^{-1} v_x^{-1} B| \le a\} &= v_x \sigma(v_x^{-1}) \epsilon_{\mathrm{lm}} \mid \{|\epsilon_{\mathrm{lm}}| \le a\} \\
&\sim v_x \sigma(v_x^{-1}) \mathcal{T}_{\mathrm{lm}}, \\
B \mid \{\|B\|_{\mathcal{M}} \le a_0,\ |\sigma(v_x^{-1})^{-1} v_x^{-1} B| \le a\} &= v_x \sigma(v_x^{-1}) \epsilon_{\mathrm{lm}} \mid \{\|\epsilon_{\mathrm{lm}}\|_{\mathcal{M}} \le a_0,\ |\epsilon_{\mathrm{lm}}| \le a\} \\
&\sim v_x \sigma(v_x^{-1}) \mathcal{T}'_{\mathrm{lm}},
\end{aligned}
$$

and thus

$$
\begin{aligned}
A_* \mid \{|\sigma(v_x^{-1})^{-1} v_x^{-1} B| \le a\} &\sim v_{\mathrm{L}}^{1/2} \epsilon + c_*^{\mathrm{T}} \sigma(v_x^{-1}) \mathcal{T}_{\mathrm{lm}}, \\
A_* \mid \{\|B\|_{\mathcal{M}} \le a_0,\ |\sigma(v_x^{-1})^{-1} v_x^{-1} B| \le a\} &\sim v_{\mathrm{L}}^{1/2} \epsilon + c_*^{\mathrm{T}} \sigma(v_x^{-1}) \mathcal{T}'_{\mathrm{lm}}
\end{aligned}
$$

by (S10).

Logistic regression. For $* = \textsc{n}, \textsc{f}, \textsc{l}$, let $(A_N, B_N, C_N) = (\sqrt{N}(\hat\tau_* - \tau), \sqrt{N}\tilde\beta, N\tilde V)$ with $(A_N, B_N, C_N) \rightsquigarrow (A_*, (S_x^2)^{-1} B, (e_0 e_1)^{-1} (S_x^2)^{-1})$ by Lemma S8. We verify below that $(\hat\tau_* \mid \mathcal{A}_{\mathrm{logit},\diamond}) \mathrel{\dot\sim} (\hat\tau_* \mid \mathcal{A}_{\mathrm{lm},\diamond})$ for $\diamond = \mathrm{jt}, \mathrm{mg}, \mathrm{cs}$.

- Recall that $\mathcal{A}_{\mathrm{logit,jt}} = \{p_{0,\mathrm{logit}} \ge \alpha_0\} = \{\lambda_{\textsc{lrt}} \le a_0\}$ under the joint rule. By $\hat Y_* \mid \{\lambda_{\textsc{lrt}} \le a_0\} \mathrel{\dot\sim} \hat Y_* \mid \{W_{\mathrm{logit}} \le a_0\}$ from Lemma S9, we have

$$
\begin{aligned}
\hat\tau_* \mid \mathcal{A}_{\mathrm{logit,jt}} &= \hat\tau_* \mid \{\lambda_{\textsc{lrt}} \le a_0\} \\
&\mathrel{\dot\sim} \hat\tau_* \mid \{W_{\mathrm{logit}} \le a_0\} = \hat\tau_* \mid \{\tilde\beta^{\mathrm{T}} \tilde V^{-1} \tilde\beta \le a_0\},
\end{aligned}
\tag{S12}
$$

  with

$$
\begin{aligned}
\sqrt{N}(\hat\tau_* - \tau) \mid \{\tilde\beta^{\mathrm{T}} \tilde V^{-1} \tilde\beta \le a_0\} &\rightsquigarrow A_* \mid \{B^{\mathrm{T}} (S_x^2)^{-1} (e_0 e_1 S_x^2) (S_x^2)^{-1} B \le a_0\} \\
&\sim A_* \mid \{\|B\|_{\mathcal{M}} \le a_0\}
\end{aligned}
$$

  by applying Lemma S7 to (A_N, B_N, C_N) and $\phi_N(u,v) = \phi(u,v) = 1(u^{\mathrm{T}} v^{-1} u \le a_0)$. This ensures

$$
\sqrt{N}(\hat\tau_* - \tau) \mid \mathcal{A}_{\mathrm{logit,jt}} \rightsquigarrow A_* \mid \{\|B\|_{\mathcal{M}} \le a_0\},
$$

  identical to the limiting distribution of $\sqrt{N}(\hat\tau_* - \tau) \mid \mathcal{A}_{\mathrm{lm,jt}}$.

- Recall that $\mathcal{A}_{\mathrm{logit,mg}} = \{p_{j,\mathrm{logit}} \ge \alpha_j,\ j = 1, \dots, J\} = \{|T_{\mathrm{logit}}| \le a\}$ under the marginal rule. We have

$$\begin{aligned}\sqrt{N}(\hat{\tau}_* - \tau) \mid \mathcal{A}_{\text{logit,mg}} &= \sqrt{N}(\hat{\tau}_* - \tau) \mid \{|\sigma(\tilde{V})^{-1}\tilde{\beta}| \le a\} \\ &\rightsquigarrow A_* \mid \{|\sigma(v_x^{-1})^{-1} v_x^{-1} B| \le a\}\end{aligned}$$

by applying Lemma S7 to (A_N, B_N, C_N) and $\phi_N(u,v) = \phi(u,v) = 1\{|\sigma(v)^{-1}u| \le a\}$; in particular, the limit of $\{|\sigma(\tilde{V})^{-1}\tilde{\beta}| \le a\}$ follows from

$$\begin{aligned}\sigma(\tilde{V})^{-1}\tilde{\beta} &\rightsquigarrow \sigma\left((e_0e_1)^{-1}(S_x^2)^{-1}\right)(S_x^2)^{-1}B \\ &= \sigma\left((e_0e_1)^{-2}v_x^{-1}\right)(e_0e_1)v_x^{-1}B = \sigma(v_x^{-1})v_x^{-1}B\end{aligned}$$

given $v_x = (e_0e_1)^{-1}S_x^2$ and $\sigma((e_0e_1)^{-2}v_x^{-1}) = (e_0e_1)^{-1}\sigma(v_x^{-1})$. This is identical to the limiting distribution of $\sqrt{N}(\hat{\tau}_* - \tau) \mid \mathcal{A}_{\text{lm,mg}}$.

- Recall that $\mathcal{A}_{\text{logit,cs}} = \{p_{j,\text{logit}} \ge \alpha_j,\ j = 0, 1, \dots, J\} = \{\lambda_{\text{LRT}} \le a_0,\ |T_{\text{logit}}| \le a\}$ under the consensus rule. The same reasoning as in (S12) ensures

$$\begin{aligned}\hat{\tau}_* - \tau \mid \mathcal{A}_{\text{logit,cs}} &= \hat{\tau}_* - \tau \mid \{\lambda_{\text{LRT}} \le a_0,\ |\sigma(\tilde{V})^{-1}\tilde{\beta}| \le a\} \\ &\dot\sim \hat{\tau}_* - \tau \mid \{\tilde{\beta}^{\mathrm{T}}\tilde{V}^{-1}\tilde{\beta} \le a_0,\ |\sigma(\tilde{V})^{-1}\tilde{\beta}| \le a\},\end{aligned}$$

with

$$\begin{aligned}&\sqrt{N}(\hat{\tau}_* - \tau) \mid \{\tilde{\beta}^{\mathrm{T}}\tilde{V}^{-1}\tilde{\beta} \le a_0,\ |\sigma(\tilde{V})^{-1}\tilde{\beta}| \le a\} \\ \rightsquigarrow\ &A_* \mid \{\|B\|_{\mathcal{M}} \le a_0,\ |\sigma(v_x^{-1})^{-1}v_x^{-1}B| \le a\}\end{aligned}$$

by applying Lemma S7 to (A_N, B_N, C_N) and $\phi_N(u,v) = \phi(u,v) = 1(u^{\mathrm{T}}v^{-1}u \le a_0) \cdot 1\{|\sigma(v)^{-1}u| \le a\}$. This ensures

$$\sqrt{N}(\hat{\tau}_* - \tau) \mid \mathcal{A}_{\text{logit,cs}} \rightsquigarrow A_* \mid \{\|B\|_{\mathcal{M}} \le a_0,\ |\sigma(v_x^{-1})^{-1}v_x^{-1}B| \le a\},$$

identical to the limiting distribution of $\sqrt{N}(\hat{\tau}_* - \tau) \mid \mathcal{A}_{\text{lm,cs}}$.

□

Proof of Proposition S1. Recall $\mathcal{A}_{\text{f,mg}} = \{F_j \le a_{j,\text{f}} \text{ for all } j = 1, \dots, J\}$ from (S8) with the explicit forms of F_j in (S7). With $(N-1)S_{x,j}^2 = \sum_{i=1}^N x_{ij}^2 = \sum_{q\in\mathcal{Q}}\sum_{i:Z_i=q}\{x_{ij} - \hat{x}_j(q)\}^2 + \sum_{q\in\mathcal{Q}} N_q\hat{x}_j^2(q)$ by direct algebra, the expression of $\mathcal{A}_{\mathrm{f,mg}}$ can be simplified to

$$\mathcal{A}_{\mathrm{f,mg}} = \{F_j \leq a_{j,\mathrm{f}} \text{ for all } j\} = \left\{ \sum_{q\in\mathcal{Q}} e_q \{\sqrt{N}\hat{x}_j(q)\}^2 \leq a''_j \text{ for all } j \right\},$$

where

$$a''_j = \frac{(N-1)S^2_{x,j}}{(N-Q)/\{(Q-1)a_{j,\mathrm{f}}\}+1} = a'_j S^2_{x,j} + o(1)$$

by $a_{j,\mathrm{f}} = (Q-1)^{-1}a'_j + o(1)$.

Observe that $\sqrt{N}\hat{x} \rightsquigarrow \epsilon_{\mathrm{f}}$ and $\sqrt{N}\hat{x}_j(q) \rightsquigarrow \epsilon_{\mathrm{f},qj}$ by Lemma S2. The result follows from Lemmas S7 and S9 as

$$\begin{aligned}
& \sqrt{N}(\hat{Y}_* - \bar{Y}) \mid \mathcal{A}_{\mathrm{f,mg}} \\
= \;& \sqrt{N}(\hat{Y}_* - \bar{Y} - \Gamma_*\hat{x}) + \Gamma_*\sqrt{N}\hat{x} \;\middle|\; \left\{ \sum_{q\in\mathcal{Q}} e_q \{\sqrt{N}\hat{x}_j(q)\}^2 \leq a''_j \text{ for all } j \right\} \\
\rightsquigarrow \;& V_{\mathrm{L}}^{1/2}\epsilon + \Gamma_*\epsilon_{\mathrm{f}} \;\middle|\; \left\{ \sum_{q\in\mathcal{Q}} e_q \epsilon^2_{\mathrm{f},qj} \leq a'_j S^2_{x,j} \text{ for all } j \right\}.
\end{aligned}$$

□

Proof of Proposition S2. Renew a_0 as the $(1-\alpha_0)$th quantile of the $\chi^2_{J(Q-1)}$ distribution. Renew $a = (a_{qj})_{q\in\mathcal{Q}_+;\ j=1,\dots,J}$, where a_{qj} denotes the $(1-\alpha_{qj}/2)$th quantile of the standard normal distribution. The marginal, joint, and consensus criteria based on the multinomial logistic regression equal

$$\begin{aligned}
&\mathcal{A}_{\mathrm{logit,mg}} = \{|T_{\mathrm{logit}}| \leq a\}, \qquad \mathcal{A}_{\mathrm{logit,jt}} = \{\lambda_{\mathrm{LRT}} \leq a_0\}, \\
&\mathcal{A}_{\mathrm{logit,cs}} = \{\lambda_{\mathrm{LRT}} \leq a_0,\ |T_{\mathrm{logit}}| \leq a\},
\end{aligned} \tag{S13}$$

respectively, with $T_{\mathrm{logit}} = \mathrm{diag}(\tilde{V}^{-1/2}_{qj,qj})\tilde{\beta} = \sigma(\tilde{V})^{-1}\tilde{\beta}$ and $\lambda_{\mathrm{LRT}} - \tilde{\beta}^{\mathrm{T}}\tilde{V}^{-1}\tilde{\beta} = o_{\mathbb{P}}(1)$.

Let $A \sim \mathcal{N}(0_Q, V_{\mathrm{L}})$ and $B \sim \mathcal{N}(0_{J(Q-1)}, V_{x+})$ be two independent normal random vectors with $\mathrm{cov}(\Psi B) = V_\Psi$. For $* = \mathrm{N}, \mathrm{F}, \mathrm{L}$, let $(A_N, B_N, C_N) = (\sqrt{N}(\hat{Y}_* - \bar{Y}), \sqrt{N}\tilde{\beta}, N\tilde{V})$ with $(A_N, B_N, C_N) \rightsquigarrow (A + \Gamma'_* B, \Psi B, V_\Psi)$ by Lemma S9.

- Under the joint rule, Lemma S9 ensures

$$\hat{Y}_* \mid \mathcal{A}_{\mathrm{logit,jt}} \;=\; \hat{Y}_* \mid \{\lambda_{\mathrm{LRT}} \leq a_0\} \;\dot{\sim}\; \hat{Y}_* \mid \{\tilde{\beta}^{\mathrm{T}}\tilde{V}^{-1}\tilde{\beta} \leq a_0\} \tag{S14}$$

with

$$\begin{aligned}\sqrt{N}(\hat{Y}_* - \bar{Y}) \mid \{\tilde{\beta}^{\mathrm{T}}\tilde{V}^{-1}\tilde{\beta} \leq a_0\} &\rightsquigarrow A + \Gamma_*' B \mid \{(\Psi B)^{\mathrm{T}} V_\Psi^{-1}(\Psi B) \leq a_0\} \\ &\sim A + \Gamma_*' B \mid \{\|B\|_{\mathcal{M}} \leq a_0\}\end{aligned}$$

by applying Lemma S7 to (A_N, B_N, C_N) and $\phi_N(u,v) = \phi(u,v) = 1(u^{\mathrm{T}} v^{-1} u \leq a_0)$. This ensures

$$\sqrt{N}(\hat{Y}_* - \bar{Y}) \mid \mathcal{A}_{\text{logit,jt}} \rightsquigarrow A + \Gamma_*' B \mid \{\|B\|_{\mathcal{M}} \leq a_0\}.$$

- Under the marginal rule, we have

$$\begin{aligned}\sqrt{N}(\hat{Y}_* - \bar{Y}) \mid \mathcal{A}_{\text{logit,mg}} &= \sqrt{N}(\hat{Y}_* - \bar{Y}) \mid \{|\sigma(\tilde{V})^{-1}\tilde{\beta}| \leq a\} \\ &\rightsquigarrow A + \Gamma_*' B \mid \{|\sigma(V_\Psi)^{-1}\Psi B| \leq a\}\end{aligned}$$

by applying Lemma S7 to (A_N, B_N, C_N) and $\phi_N(u,v) = \phi(u,v) = 1\{|\sigma(v)^{-1}u| \leq a\}$.

- Under the consensus rule, we have

$$\begin{aligned}\hat{Y}_* \mid \mathcal{A}_{\text{logit,cs}} &= \hat{Y}_* \mid \{\lambda_{\text{LRT}} \leq a_0,\ |\sigma(\tilde{V})^{-1}\tilde{\beta}| \leq a\} \\ &\dot{\sim} \hat{Y}_* \mid \{\tilde{\beta}^{\mathrm{T}}\tilde{V}^{-1}\tilde{\beta} \leq a_0,\ |\sigma(\tilde{V})^{-1}\tilde{\beta}| \leq a\}\end{aligned}$$

by (S14) with

$$\begin{aligned}&\sqrt{N}(\hat{Y}_* - \bar{Y}) \mid \{\tilde{\beta}^{\mathrm{T}}\tilde{V}^{-1}\tilde{\beta} \leq a_0,\ |\sigma(\tilde{V})^{-1}\tilde{\beta}| \leq a\} \\ \rightsquigarrow\ & A + \Gamma_*' B \mid \{\|B\|_{\mathcal{M}} \leq a_0,\ |\sigma(V_\Psi)^{-1}\Psi B| \leq a\}\end{aligned}$$

by applying Lemma S7 to (A_N, B_N, C_N) and $\phi_N(u,v) = \phi(u,v) = 1(u^{\mathrm{T}} v^{-1} u \leq a_0) \cdot 1\{|\sigma(v)^{-1}u| \leq a\}$. This ensures

$$\sqrt{N}(\hat{Y}_* - \bar{Y}) \mid \mathcal{A}_{\text{logit,cs}} \rightsquigarrow A + \Gamma_*' B \mid \{\|B\|_{\mathcal{M}} \leq a_0,\ |\sigma(V_\Psi)^{-1}\Psi B| \leq a\}.$$

With $A \sim V_{\mathrm{L}}^{1/2}\epsilon$ and independent of B, it suffices to compute $B \mid \{\|B\|_{\mathcal{M}} \leq a_0\}$, $B \mid \{|\sigma(V_\Psi)^{-1}\Psi B| \leq a\}$, and $B \mid \{\|B\|_{\mathcal{M}} \leq a_0,\ |\sigma(V_\Psi)^{-1}\Psi B| \leq a\}$, respectively.

For the joint criterion, we have $B \mid \{\|B\|_{\mathcal{M}} \le a_0\} \sim V_{x+}^{1/2}\mathcal{L}$, and thus

$$A + \Gamma_*' B \mid \{\|B\|_{\mathcal{M}} \le a_0\} \quad \sim \quad V_{\mathrm{L}}^{1/2}\epsilon + \Gamma_*' V_{x+}^{1/2}\mathcal{L}.$$

For the marginal and consensus criteria, let $\epsilon_{\text{logit}} = \sigma(V_\Psi)^{-1}\Psi B \sim \mathcal{N}\{0, D(V_\Psi)\}$ to write

$$B = \Psi^{-1}\sigma(V_\Psi)\epsilon_{\text{logit}}, \qquad \{|\sigma(V_\Psi)^{-1}\Psi B| \le a\} = \{|\epsilon_{\text{logit}}| \le a\}.$$

This ensures

$$\begin{aligned} B \mid \{|\sigma(V_\Psi)^{-1}\Psi B| \le a\} &= \Psi^{-1}\sigma(V_\Psi)\epsilon_{\text{logit}} \mid \{|\epsilon_{\text{logit}}| \le a\} \\ &\sim \Psi^{-1}\sigma(V_\Psi)\mathcal{T}_{\text{logit}}, \\ B \mid \{\|B\|_{\mathcal{M}} \le a_0,\ |\sigma(V_\Psi)^{-1}\Psi B| \le a\} &= \Psi^{-1}\sigma(V_\Psi)\epsilon_{\text{logit}} \mid \{|\epsilon_{\text{logit}}| \le a,\ \|\epsilon_{\text{logit}}\|_{\mathcal{M}} \le a_0\} \\ &\sim \Psi^{-1}\sigma(V_\Psi)\mathcal{T}_{\text{logit}}', \end{aligned}$$

and thus

$$\begin{aligned} A + \Gamma_*' B \mid \{|\sigma(V_\Psi)^{-1}\Psi B| \le a\} &\sim V_{\mathrm{L}}^{1/2}\epsilon + \Gamma_*'\Psi^{-1}\sigma(V_\Psi)\mathcal{T}_{\text{logit}}, \\ A + \Gamma_*' B \mid \{\|B\|_{\mathcal{M}} \le a_0,\ |\sigma(V_\Psi)^{-1}\Psi B| \le a\} &\sim V_{\mathrm{L}}^{1/2}\epsilon + \Gamma_*'\Psi^{-1}\sigma(V_\Psi)\mathcal{T}_{\text{logit}}'. \end{aligned}$$

□

S5.2. Covariate balance and asymptotic relative efficiency in Theorems 1–4

Proof of Theorems 1–4. The asymptotic relative efficiency of $\hat{\tau}_*$ ($* = \mathrm{N}, \mathrm{F}, \mathrm{L}$) follows immediately from Propositions 1–2, Propositions S1–S2 and Lemmas S4–S5. We verify below the improved covariate balance under the ReP schemes based on the multinomial logistic regression. The results under the two-sample t-test-based, linear or logistic regression-based, and marginal F-test-based criteria are analogous and thus omitted. Note that the asymptotic conditional bias and the difference between different estimators are all functions of $\hat{\tau}_x$. The results associated with the covariance reduction factor $\rho(J, a_0)$ under the joint rules then follow from the results on ReM from Morgan and Rubin (2012) and Li et al. (2018).

Recall the acceptance criteria by test statistics from (S13). Recall from (S1) that $\hat{x} = \kappa\hat{x}_+$ and hence $\hat{\tau}_x = G_x\hat{x} = G_x\kappa\hat{x}_+$. Let $A \sim \mathcal{N}(0_{J(Q-1)}, V_{x+})$ with $\sqrt{N}\hat{x}_+ \rightsquigarrow A$ and $\sqrt{N}\hat{\tau}_x \rightsquigarrow G_x\kappa A$ by Lemma S9. The result follows from

$$\begin{aligned}
\sqrt{N}\hat{\tau}_x \mid \mathcal{A}_{\text{logit,jt}} &= G_x\kappa(\sqrt{N}\hat{x}_+) \mid \{\lambda_{\text{LRT}} \leq a_0\} \\
&\dot{\sim} G_x\kappa(\sqrt{N}\hat{x}_+) \mid \{N\hat{x}_+^{\mathrm{T}}V_{x+}^{-1}\hat{x}_+ \leq a_0\} \\
&\rightsquigarrow G_x\kappa A \mid \{A^{\mathrm{T}}V_{x+}^{-1}A \leq a_0\} \\
&\succeq G_x\kappa A, \\
\sqrt{N}\hat{\tau}_x \mid \mathcal{A}_{\text{logit,mg}} &= G_x\kappa(\sqrt{N}\hat{x}_+) \mid \{|\sigma(\tilde{V})^{-1}\tilde{\beta}| \leq a\} \\
&\dot{\sim} G_x\kappa(\sqrt{N}\hat{x}_+) \mid \{|\sigma(V_\Psi)^{-1}\sqrt{N}\Psi\hat{x}_+| \leq a\} \\
&\rightsquigarrow G_x\kappa A \mid \{|\sigma(V_\Psi)^{-1}\Psi A| \leq a\} \\
&\succeq G_x\kappa A, \\
\sqrt{N}\hat{\tau}_x \mid \mathcal{A}_{\text{logit,cs}} &= G_x\kappa(\sqrt{N}\hat{x}_+) \mid \{\lambda_{\text{LRT}} \leq a_0,\ |\sigma(\tilde{V})^{-1}\tilde{\beta}| \leq a\} \\
&\dot{\sim} G_x\kappa(\sqrt{N}\hat{x}_+) \mid \{N\hat{x}_+^{\mathrm{T}}V_{x+}^{-1}\hat{x}_+ \leq a_0,\ |\sigma(V_\Psi)^{-1}\sqrt{N}\Psi\hat{x}_+| \leq a\} \\
&\rightsquigarrow G_x\kappa A \mid \{A^{\mathrm{T}}V_{x+}^{-1}A \leq a_0,\ |\sigma(V_\Psi)^{-1}\Psi A| \leq a\} \\
&\succeq G_x\kappa A.
\end{aligned}$$

In particular, the three "$\dot{\sim}$" follow from $\lambda_{\text{LRT}} - N\hat{x}_+^{\mathrm{T}}V_{x+}^{-1}\hat{x}_+ = o_{\mathbb{P}}(1)$, $\sqrt{N}(\tilde{\beta} - \Psi\hat{x}_+) = o_{\mathbb{P}}(1)$, and $N\tilde{V} - V_\Psi = o_{\mathbb{P}}(1)$; the three "$\rightsquigarrow$" follow from Lemma S7; the three "$\succeq$" follow from Corollary S2. □

S6. Proofs of the results under stratified experiments

Proof of Theorem 5. Recall that $\hat{\tau}_* = \sum_{k=1}^{K} \pi_{[k]}\hat{\tau}_{*[k]}$ for $* = \text{N}, \text{F}, \text{L}$ under stratified experiments. Theorems 1–2 ensure the improved efficiency of $\hat{\tau}_{*[k]}$ and $\hat{\tau}_{x[k]}$ within each stratum. The results about $\hat{\tau}_*$ then follow from Lemma S5(i) and (iii). □

S6.1. Proof of Proposition 3

Lemma S10 below is a numeric result from Ding (2021, Theorem 5) that holds without any stochastic assumptions. It gives the key building block for proving Proposition 3.

Lemma S10. $\tilde{\tau}_{\text{N}} = \sum_{k=1}^{K} \omega_{[k]}\{\hat{Y}_{[k]}(1) - \hat{Y}_{[k]}(0)\}$, where $\hat{Y}_{[k]}(q)$ is the sample mean under treatment level $q \in \{0, 1\}$ in stratum k.

Proof of Proposition 3. The result about $\tilde{\tau}_{\text{N}}$ follows directly from Lemma S10. We verify below the result about $\tilde{\tau}_{\text{F}}$.

Let $S_i = (\,1(i \in [1]), \ldots, 1(i \in [K])\,)^{\mathrm{T}} \in \mathbb{R}^K$ to write $\texttt{lm}\{Y_i \sim Z_i + x_i + 1(i \in [1]) + \cdots + 1(i \in [K])\}$ as

$$\texttt{lm}(Y_i \sim Z_i + x_i + S_i). \tag{S15}$$

Let $\hat{\gamma}$ denote the coefficient vector of x_i from (S15). The definition of OLS ensures that $\tilde{\tau}_{\mathrm{F}}$ is the coefficient of Z_i from the OLS fit of

$$\texttt{lm}(Y_i - x_i^{\mathrm{T}}\hat{\gamma} \sim Z_i + S_i). \tag{S16}$$

Applying Lemma S10 to (S16) ensures

$$\tilde{\tau}_{\mathrm{F}} = \sum_{k=1}^{K} \omega_{[k]} \left[\left\{ \hat{Y}_{[k]}(1) - \hat{x}_{[k]}(1)^{\mathrm{T}}\hat{\gamma} \right\} - \left\{ \hat{Y}_{[k]}(0) - \hat{x}_{[k]}(0)^{\mathrm{T}}\hat{\gamma} \right\} \right] = \tilde{\tau}_{\mathrm{N}} - \sum_{k=1}^{K} \omega_{[k]} \left\{ \hat{x}_{[k]}(1) - \hat{x}_{[k]}(0) \right\}^{\mathrm{T}} \hat{\gamma}.$$

Observe that $\hat{x}_{[k]}(1) - \hat{x}_{[k]}(0) = o_{\mathbb{P}}(1)$. To verify the probability limit of $\tilde{\tau}_{\mathrm{F}}$, it suffices to show that $\hat{\gamma}$ has a finite probability limit. We verify this below.

Let $\tilde{Y}_i$ denote the residual of unit i from $\texttt{lm}(Y_i \sim Z_i + S_i)$. Let $\tilde{x}_i$ denote the residual of unit i from $\texttt{lm}(x_i \sim Z_i + S_i)$. By the Frisch–Waugh–Lovell theorem, $\hat{\gamma}$ equals the coefficient vector of $\tilde{x}_i$ from $\texttt{lm}(\tilde{Y}_i \sim \tilde{x}_i)$ as

$$\hat{\gamma} = (\tilde{X}^{\mathrm{T}}\tilde{X})^{-1}(\tilde{X}^{\mathrm{T}}\tilde{Y}) = \left(N^{-1} \sum_{i=1}^{N} \tilde{x}_i \tilde{x}_i^{\mathrm{T}} \right)^{-1} \left(N^{-1} \sum_{i=1}^{N} \tilde{x}_i \tilde{Y}_i \right),$$

where $\tilde{X} = (\tilde{x}_1, \ldots, \tilde{x}_N)^{\mathrm{T}}$ and $\tilde{Y} = (\tilde{Y}_1, \ldots, \tilde{Y}_N)^{\mathrm{T}}$. It thus suffices to show that $N^{-1}\sum_{i=1}^{N} \tilde{x}_i \tilde{x}_i^{\mathrm{T}}$ and $N^{-1}\sum_{i=1}^{N} \tilde{x}_i \tilde{Y}_i$ both have finite probability limits when Condition 1 holds for all strata. We show below the finite probability limit of $N^{-1}\sum_{i=1}^{N} \tilde{x}_i \tilde{x}_i^{\mathrm{T}}$. The proof for $N^{-1}\sum_{i=1}^{N} \tilde{x}_i \tilde{Y}_i$ is analogous and omitted.

From Lemma S10, the coefficient of Z_i from $\texttt{lm}(x_i \sim Z_i + S_i)$ equals

$$\tilde{\tau}_x = \sum_{k=1}^{K} \omega_{[k]} \left\{ \hat{x}_{[k]}(1) - \hat{x}_{[k]}(0) \right\},$$

where $\hat{x}_{[k]}(q) = N_{[k]q}^{-1} \sum_{i \in [k], Z_i = q} x_i$. Further recall $\hat{\tau}_{x[k]} = \hat{x}_{[k]}(1) - \hat{x}_{[k]}(0)$ as the difference in covariate means in stratum $k \in \{1, \ldots, K\}$. It follows from Ding (2021, Proof of Theorem 6) that

$$\tilde{x}_i = \begin{cases} x_i - \hat{x}_{[k]}(1) + (1 - e_{[k]})(\hat{\tau}_{x[k]} - \tilde{\tau}_x) & \text{if } Z_i = 1 \\ x_i - \hat{x}_{[k]}(0) - e_{[k]}(\hat{\tau}_{x[k]} - \tilde{\tau}_x) & \text{if } Z_i = 0 \end{cases} \quad \text{for } \ i \in [k],$$

with

$$\tilde{x}_i\tilde{x}_i^{\mathrm{T}} = \begin{cases} x_ix_i^{\mathrm{T}} + \left\{\hat{x}_{[k]}(1) - (1-e_{[k]})(\hat{\tau}_{x[k]} - \tilde{\tau}_x)\right\}\left\{\hat{x}_{[k]}(1) - (1-e_{[k]})(\hat{\tau}_{x[k]} - \tilde{\tau}_x)\right\}^{\mathrm{T}} & \\ -x_i\left\{\hat{x}_{[k]}(1) - (1-e_{[k]})(\hat{\tau}_{x[k]} - \tilde{\tau}_x)\right\}^{\mathrm{T}} - \left\{\hat{x}_{[k]}(1) - (1-e_{[k]})(\hat{\tau}_{x[k]} - \tilde{\tau}_x)\right\}x_i^{\mathrm{T}} & \text{if } Z_i = 1; \\ & \\ x_ix_i^{\mathrm{T}} + \left\{\hat{x}_{[k]}(0) + e_{[k]}(\hat{\tau}_{x[k]} - \tilde{\tau}_x)\right\}\left\{\hat{x}_{[k]}(0) + e_{[k]}(\hat{\tau}_{x[k]} - \tilde{\tau}_x)\right\}^{\mathrm{T}} & \\ -x_i\left\{\hat{x}_{[k]}(0) + e_{[k]}(\hat{\tau}_{x[k]} - \tilde{\tau}_x)\right\}^{\mathrm{T}} - \left\{\hat{x}_{[k]}(0) + e_{[k]}(\hat{\tau}_{x[k]} - \tilde{\tau}_x)\right\}x_i^{\mathrm{T}} & \text{if } Z_i = 0. \end{cases}$$

This ensures

$$\begin{aligned} \sum_{i\in[k]}\tilde{x}_i\tilde{x}_i^{\mathrm{T}} &= \sum_{i\in[k]}x_ix_i^{\mathrm{T}} \\ &\quad + N_{[k]1}\left\{\hat{x}_{[k]}(1) - (1-e_{[k]})(\hat{\tau}_{x[k]} - \tilde{\tau}_x)\right\}\left\{\hat{x}_{[k]}(1) - (1-e_{[k]})(\hat{\tau}_{x[k]} - \tilde{\tau}_x)\right\}^{\mathrm{T}} \\ &\quad - N_{[k]1}\hat{x}_{[k]}(1)\left\{\hat{x}_{[k]}(1) - (1-e_{[k]})(\hat{\tau}_{x[k]} - \tilde{\tau}_x)\right\}^{\mathrm{T}} - N_{[k]1}\left\{\hat{x}_{[k]}(1) - (1-e_{[k]})(\hat{\tau}_{x[k]} - \tilde{\tau}_x)\right\}\hat{x}_{[k]}(0)^{\mathrm{T}} \\ &\quad + N_{[k]0}\left\{\hat{x}_{[k]}(0) + e_{[k]}(\hat{\tau}_{x[k]} - \tilde{\tau}_x)\right\}\left\{\hat{x}_{[k]}(0) + e_{[k]}(\hat{\tau}_{x[k]} - \tilde{\tau}_x)\right\}^{\mathrm{T}} \\ &\quad - N_{[k]0}\hat{x}_{[k]}(0)\left\{\hat{x}_{[k]}(0) + e_{[k]}(\hat{\tau}_{x[k]} - \tilde{\tau}_x)\right\}^{\mathrm{T}} - N_{[k]0}\left\{\hat{x}_{[k]}(0) + e_{[k]}(\hat{\tau}_{x[k]} - \tilde{\tau}_x)\right\}\hat{x}_{[k]}(0)^{\mathrm{T}}. \end{aligned}$$

Let $\bar{x}_{[k]} = N_{[k]}{}^{-1}\sum_{i\in[k]}x_i$. It then follows from $\hat{x}_{[k]}(q) = \bar{x}_{[k]} + o_{\mathbb{P}}(1)$ $(q = 0, 1)$, $\tilde{\tau}_x = o_{\mathbb{P}}(1)$, and $\hat{\tau}_{x[k]} = o_{\mathbb{P}}(1)$ that when Condition 1 holds for all strata,

$$\begin{aligned} N_{[k]}{}^{-1}\sum_{i\in[k]}\tilde{x}_i\tilde{x}_i^{\mathrm{T}} &= N_{[k]}{}^{-1}\sum_{i\in[k]}x_ix_i^{\mathrm{T}} \\ &\quad + e_{[k]}\left\{\hat{x}_{[k]}(1) - (1-e_{[k]})(\hat{\tau}_{x[k]} - \tilde{\tau}_x)\right\}\left\{\hat{x}_{[k]}(1) - (1-e_{[k]})(\hat{\tau}_{x[k]} - \tilde{\tau}_x)\right\}^{\mathrm{T}} \\ &\quad - e_{[k]}\hat{x}_{[k]}(1)\left\{\hat{x}_{[k]}(1) - (1-e_{[k]})(\hat{\tau}_{x[k]} - \tilde{\tau}_x)\right\}^{\mathrm{T}} - e_{[k]}\left\{\hat{x}_{[k]}(1) - (1-e_{[k]})(\hat{\tau}_{x[k]} - \tilde{\tau}_x)\right\}\hat{x}_{[k]}(0)^{\mathrm{T}} \\ &\quad + (1-e_{[k]})\left\{\hat{x}_{[k]}(0) + e_{[k]}(\hat{\tau}_{x[k]} - \tilde{\tau}_x)\right\}\left\{\hat{x}_{[k]}(0) + e_{[k]}(\hat{\tau}_{x[k]} - \tilde{\tau}_x)\right\}^{\mathrm{T}} \\ &\quad - (1-e_{[k]})\hat{x}_{[k]}(0)\left\{\hat{x}_{[k]}(0) + e_{[k]}(\hat{\tau}_{x[k]} - \tilde{\tau}_x)\right\}^{\mathrm{T}} - (1-e_{[k]})\left\{\hat{x}_{[k]}(0) + e_{[k]}(\hat{\tau}_{x[k]} - \tilde{\tau}_x)\right\}\hat{x}_{[k]}(0)^{\mathrm{T}} \\ &= N_{[k]}{}^{-1}\sum_{i\in[k]}x_ix_i^{\mathrm{T}} \\ &\quad + e_{[k]}\bar{x}_{[k]}\bar{x}_{[k]}^{\mathrm{T}} - e_{[k]}\bar{x}_{[k]}\bar{x}_{[k]}^{\mathrm{T}} - e_{[k]}\bar{x}_{[k]}\bar{x}_{[k]}^{\mathrm{T}} \\ &\quad + (1-e_{[k]})\bar{x}_{[k]}\bar{x}_{[k]}^{\mathrm{T}} - (1-e_{[k]})\bar{x}_{[k]}\bar{x}_{[k]}^{\mathrm{T}} - (1-e_{[k]})\bar{x}_{[k]}\bar{x}_{[k]}^{\mathrm{T}} + o_{\mathbb{P}}(1) \\ &= S^2_{x[k]} + o_{\mathbb{P}}(1), \end{aligned}$$

where $S^2_{x[k]} = (N_{[k]} - 1)^{-1} \sum_{i \in [k]} (x_i - \bar{x}_{[k]})(x_i - \bar{x}_{[k]})^{\mathrm{T}}$. As a result,

$$N^{-1} \sum_{i=1}^{N} \tilde{x}_i \tilde{x}_i^{\mathrm{T}} = \sum_{k=1}^{K} \frac{N_{[k]}}{N} \left(N_{[k]}{}^{-1} \sum_{i \in [k]} \tilde{x}_i \tilde{x}_i^{\mathrm{T}} \right) = \sum_{k=1}^{K} \pi_{[k]} S^2_{x[k]} + o_{\mathbb{P}}(1)$$

has a finite probability limit when Condition 1 holds for all strata.

□

S7. Proof of Theorem S1

We verify in this section the properties of the logistic and multinomial logistic regressions in Theorem S1. The results ensure the asymptotic equivalence of the LRT and the Wald test for logistic and multinomial logistic regressions from the design-based perspective. To this end, we first review some useful numeric facts about the multinomial logistic regression in Section S7.1, and then give the proof in Section S7.2.

Recall that $e_+ = (e_1, \dots, e_{Q-1})^{\mathrm{T}}$, $\mathrm{diag}(e_+) = \mathrm{diag}(e_q)_{q \in \mathcal{Q}_+}$, and $\Psi = \{\Phi^{-1} \mathrm{diag}(e_+)\} \otimes (S^2_x)^{-1}$, with $\Phi = \mathrm{diag}(e_+) - e_+ e_+^{\mathrm{T}}$. Let $R_+ = \mathrm{diag}(e_+)$ be a shorthand for $\mathrm{diag}(e_+)$ to write $\Psi = (\Phi^{-1} R_+) \otimes (S^2_x)^{-1}$ with

$$\Phi = R_+ - e_+ e_+^{\mathrm{T}} = \begin{pmatrix} e_1(1-e_1) & -e_1 e_2 & \dots & -e_1 e_{Q-1} \\ -e_2 e_1 & e_2(1-e_2) & \dots & -e_2 e_{Q-1} \\ \vdots & \vdots & & \vdots \\ -e_{Q-1} e_1 & -e_{Q-1} e_2 & \dots & e_{Q-1}(1-e_{Q-1}) \end{pmatrix}.$$

Recall that $\hat{x}_+ = (\hat{x}(1)^{\mathrm{T}}, \dots, \hat{x}(Q-1)^{\mathrm{T}})^{\mathrm{T}}$, with $V_{x+} = N \mathrm{cov}(\hat{x}_+)$. Then $V_{x+} = (R_+^{-1} - 1_{(Q-1)\times(Q-1)}) \otimes S^2_x$ equals the upper $J(Q-1) \times J(Q-1)$ submatrix of $V_x = \{\mathrm{diag}(e_q^{-1})_{q \in \mathcal{Q}} - 1_{Q \times Q}\} \otimes S^2_x$. This ensures

$$\begin{aligned} V_\Psi &= N \mathrm{cov}(\Psi \hat{x}_+) = \Psi V_{x+} \Psi^{\mathrm{T}} \qquad \text{(S17)} \\ &= \left\{(\Phi^{-1} R_+) \otimes (S^2_x)^{-1}\right\} \left\{(R_+^{-1} - 1_{(Q-1)\times(Q-1)}) \otimes S^2_x\right\} \left\{(\Phi^{-1} R_+)^{\mathrm{T}} \otimes (S^2_x)^{-1}\right\} \\ &= \left\{\Phi^{-1} R_+ (R_+^{-1} - 1_{(Q-1)\times(Q-1)}) R_+ \Phi^{-1}\right\} \otimes (S^2_x)^{-1} \\ &= \Phi^{-1} \otimes (S^2_x)^{-1} \end{aligned}$$

given $\Phi = R_+(R_+^{-1} - 1_{(Q-1)\times(Q-1)})R_+$. As a result, we have $\Psi = V_\Psi (R_+ \otimes I_J)$ and hence

$$(R_+ \otimes I_J) V_\Psi (R_+ \otimes I_J) = \Psi^{\mathrm{T}} V_\Psi^{-1} \Psi = V_{x+}^{-1}. \qquad \text{(S18)}$$

S7.1. Numeric facts about the multinomial logistic regression

Recall the multinomial logistic model from (S5). We have

$$\pi_q(x_i) = \pi_q(\tilde{x}_i, \theta) = \frac{\exp(\tilde{x}_i^{\mathrm{T}}\theta_q)}{1 + \sum_{q' \in \mathcal{Q}_+} \exp(\tilde{x}_i^{\mathrm{T}}\theta_{q'})} \qquad \text{for } \; q = 1, \ldots, Q, \tag{S19}$$

with $\tilde{x}_i = (1, x_i^{\mathrm{T}})^{\mathrm{T}}$, $\theta_Q = 0_{J+1}$, $\theta_q = (\beta_{q0}, \boldsymbol{\beta}_q^{\mathrm{T}})^{\mathrm{T}}$ for $q \in \mathcal{Q}_+$, and $\theta = (\theta_1^{\mathrm{T}}, \ldots, \theta_{Q-1}^{\mathrm{T}})^{\mathrm{T}}$. The scaled log-likelihood function of $(x_i, Z_i)_{i=1}^N$ equals

$$\begin{aligned} \bar{L}(\theta) &= N^{-1} \sum_{i=1}^N \log\{\pi_{Z_i}(x_i)\} && \text{(S20)} \\ &= N^{-1} \sum_{i=1}^N \left[\sum_{q \in \mathcal{Q}_+} \mathcal{I}_{iq} \tilde{x}_i^{\mathrm{T}} \theta_q - \log \left\{ 1 + \sum_{q \in \mathcal{Q}_+} \exp(\tilde{x}_i^{\mathrm{T}}\theta_q) \right\} \right]. \end{aligned}$$

The score function of $\bar{L}(\theta)$ equals

$$U(\theta) = \frac{\partial \bar{L}(\theta)}{\partial \theta} = \begin{pmatrix} U_1(\theta) \\ \vdots \\ U_{Q-1}(\theta) \end{pmatrix} \tag{S21}$$

with

$$U_q(\theta) = \frac{\partial \bar{L}(\theta)}{\partial \theta_q} = N^{-1} \sum_{i=1}^N \tilde{x}_i \{\mathcal{I}_{iq} - \pi_q(\tilde{x}_i, \theta)\} \qquad (q \in \mathcal{Q}_+).$$

The Hessian matrix of $\bar{L}(\theta)$ equals

$$H(\theta) = \frac{\partial^2 \bar{L}(\theta)}{\partial \theta \partial \theta^{\mathrm{T}}}, \tag{S22}$$

with the explicit form given by Condition 3.

S7.2. The proof

Lemmas S11 and S12 provide the basis for proving Theorem S1.

Lemma S11. (Rudin 1976, Theorem 7.17) Suppose $\{f_N(x)\}_{N=1}^\infty$ is a sequence of functions, differentiable on $[a, b]$ and such that $\{f_N(x_0)\}_{N=1}^\infty$ converges for some point x_0 on $[a, b]$. If the sequence of derivatives, $\{f_N'(x)\}_{N=1}^\infty$, converges uniformly on $[a, b]$, then $\{f_N(x)\}_{N=1}^\infty$ converges uniformly on $[a, b]$, to a function

$f(x)$, and

$$\lim_{N\to\infty} f_N'(x) = f'(x) \qquad (a \le x \le b).$$

Lemma S12. (Newey and McFadden 1994, Theorem 2.7) If there is a function $Q_0(\theta)$ such that (i) $Q_0(\theta)$ is uniquely maximized at θ_0; (ii) θ_0 is an element of the interior of a convex set Θ and $Q_N(\theta)$ is concave; and (iii) $Q_N(\theta) = Q_0(\theta) + o_{\mathbb{P}}(1)$ for all $\theta \in \Theta$, then the maximizer of $Q_N(\theta)$, denoted by $\hat{\theta}_N$, exists with probability approaching one and $\hat{\theta}_N = \theta_0 + o_{\mathbb{P}}(1)$.

Proof of Theorem S1. Zhao and Ding (2021, Lemma S5) ensures that

$$\hat{x}(q) = o(1) \qquad \text{for all } q \in \mathcal{Q} \tag{S23}$$

almost surely under Condition 1. For notational simplicity, we assume that (S23) is true in the following proof. The simplification does not affect the validity of the proof given all results concern either convergence in probability or convergence in distribution.

Let $\tilde{\boldsymbol{\beta}}_q$ and $\tilde{\theta}_q = (\tilde{\beta}_{q0}, \tilde{\boldsymbol{\beta}}_q^{\mathrm{T}})^{\mathrm{T}}$ be the MLEs of $\boldsymbol{\beta}_q$ and $\theta_q = (\beta_{q0}, \boldsymbol{\beta}_q^{\mathrm{T}})^{\mathrm{T}}$ for $q \in \mathcal{Q}_+$ in (S5), respectively, concatenated as $\tilde{\beta} = (\tilde{\boldsymbol{\beta}}_1^{\mathrm{T}}, \dots, \tilde{\boldsymbol{\beta}}_{Q-1}^{\mathrm{T}})^{\mathrm{T}}$ and $\tilde{\theta} = (\tilde{\theta}_1^{\mathrm{T}}, \dots, \tilde{\theta}_{Q-1}^{\mathrm{T}})^{\mathrm{T}}$.

Convergence of $\tilde{\theta}$. As a key intermediate result, we first verify

$$\tilde{\theta} = \theta^* + o_{\mathbb{P}}(1) \tag{S24}$$

for

$$\theta^* = ((\theta_1^*)^{\mathrm{T}}, \dots, (\theta_{Q-1}^*)^{\mathrm{T}})^{\mathrm{T}}, \qquad \text{where } \theta_q^* = (\beta_{q0}^*, 0_J^{\mathrm{T}})^{\mathrm{T}} \text{ with } \beta_{q0}^* = \log(e_q/e_Q).$$

By Lemma S12, it suffices to show that there exists a function $\bar{L}_\infty(\theta)$ such that (i) $\bar{L}_\infty(\theta)$ is uniquely maximized at θ^*; (ii) $\bar{L}(\theta)$ is concave on $\mathbb{R}^{(J+1)(Q-1)}$; and (iii) $\bar{L}(\theta) = \bar{L}_\infty(\theta) + o(1)$ for all $\theta \in \mathbb{R}^{(J+1)(Q-1)}$. We verify below these three sufficient conditions in the order of (iii) to (i) to (ii).

First, it follows from $\tilde{x}_i^{\mathrm{T}}\theta_q^* = \beta_{q0}^* = \log(e_q/e_Q)$ that $\pi_q(\tilde{x}_i, \theta^*) = e_q$ for all $q \in \mathcal{Q}$ and $i = 1, \dots, N$ by (S19). Plug this in the expressions of $\bar{L}(\theta)$, $U(\theta)$, and $H(\theta)$ from (S20), (S21), and Condition 3 to see

$$\bar{L}(\theta^*) \quad = \quad \sum_{q\in\mathcal{Q}} e_q \log(e_q),$$

$$
\begin{aligned}
U(\theta^*) &= \begin{pmatrix} U_1(\theta^*) \\ \vdots \\ U_{Q-1}(\theta^*) \end{pmatrix}, \quad \text{where} \quad U_q(\theta^*) = \begin{pmatrix} 0 \\ e_q \hat{x}(q) \end{pmatrix}, \qquad \text{(S25)} \\
H(\theta^*) &= -\Phi \otimes (\delta S_{\tilde{x}}^2) \quad \text{with} \quad H_{qq'}(\theta^*) = e_q\{e_{q'} - 1(q = q')\}(\delta S_{\tilde{x}}^2),
\end{aligned}
$$

where $\delta = 1 - N^{-1}$ and $S_{\tilde{x}}^2 = (N-1)^{-1} \sum_{i=1}^N \tilde{x}_i \tilde{x}_i^{\mathrm{T}} = \mathrm{diag}(\delta^{-1}, S_x^2)$. Under (S23), this ensures that $\bar{L}(\theta)$, $U(\theta)$, and $H(\theta)$ all converge pointwise at $\theta = \theta^*$, with

$$
H^* = H_\infty(\theta^*) = -\Phi \otimes \mathrm{diag}(1, S_x^2) < 0. \qquad \text{(S26)}
$$

In addition, Condition 3 ensures that $\partial U(\theta)/\partial\theta = H(\theta)$ converges uniformly to $H_\infty(\theta)$ on any compact set in $\mathbb{R}^{(J+1)(Q-1)}$. Let $U_{qj}(\theta) = \partial \bar{L}(\theta)/\partial \beta_{qj}$ be the (qj)th element of $U(\theta)$ for $q \in \mathcal{Q}_+$ and $j = 0, 1, \ldots, J$. Applying Lemma S11 component-wise to $f_N = U_{qj}$ ensures that there exists a function, denoted by $U_\infty(\theta)$, such that

$$
U(\theta) = U_\infty(\theta) + o(1), \qquad \partial U_\infty(\theta)/\partial\theta = H_\infty(\theta)
$$

for all $\theta \in \mathbb{R}^{(J+1)(Q-1)}$, and the convergence is uniform on any compact set in $\mathbb{R}^{(J+1)(Q-1)}$.

Sufficient condition (iii) then follows from applying Lemma S11 component-wise to $f_N = \bar{L}$, which ensures that there exists a function, denoted by $\bar{L}_\infty(\theta)$, such that

$$
\bar{L}(\theta) = \bar{L}_\infty(\theta) + o(1), \qquad \partial \bar{L}_\infty(\theta)/\partial\theta = U_\infty(\theta)
$$

for all $\theta \in \mathbb{R}^{(J+1)(Q-1)}$.

Sufficient condition (i) then follows from $\partial^2 \bar{L}_\infty(\theta)/\partial\theta\partial\theta^{\mathrm{T}} = H_\infty(\theta) < 0$ by Condition 3 and $U_\infty(\theta^*) = \lim_{N\to\infty} U(\theta^*) = 0$ by (S25).

For sufficient condition (ii), let $H_i(\theta) = (H_{i,qq'}(\theta))_{q,q' \in \mathcal{Q}_+}$ with

$$
H_{i,qq'}(\theta) = \pi_q(\tilde{x}_i, \theta)\{\pi_{q'}(\tilde{x}_i, \theta) - 1(q = q')\}\tilde{x}_i \tilde{x}_i^{\mathrm{T}}.
$$

Then $H(\theta) = N^{-1} \sum_{i=1}^N H_i(\theta)$ by the explicit form of $H(\theta)$ in Condition 3. Observe that $H_i(\theta) = -\Phi_i \otimes (\tilde{x}_i \tilde{x}_i^{\mathrm{T}})$, with $\Phi_i = (\Phi_{i,qq'})_{q,q' \in \mathcal{Q}_+}$ where $\Phi_{i,qq'} = \pi_q(\tilde{x}_i, \theta)\{1(q = q') - \pi_{q'}(\tilde{x}_i, \theta)\}$. It follows from $\Phi_i \geq 0$ that $H_i(\theta) \leq 0$ and hence $H(\theta) \leq 0$.

Asymptotic equivalence of $\tilde{\beta}$ and $\Psi\hat{x}_+$. We next verify $\sqrt{N}(\tilde{\beta} - \Psi\hat{x}_+) = o_{\mathbb{P}}(1)$. The proof follows from the same reasoning as that of Newey and McFadden (1994, Theorem 3.1) for independent and identically distributed samples.

Recall $U_{qj}(\theta) = \partial\bar{L}(\theta)/\partial\beta_{qj}$ as the (qj)th element of $U(\theta)$ for $q \in \mathcal{Q}_+$ and $j = 0, 1, \ldots, J$. Let $H_{qj}(\theta) = \partial U_{qj}(\theta)/\partial\theta \in \mathbb{R}^{(J+1)(Q-1)}$, with

$$H(\theta) = \partial U(\theta)/\partial\theta^{\mathrm{T}} = (H_{1,0}(\theta), H_{1,1}(\theta), \ldots, H_{Q-1,J}(\theta))^{\mathrm{T}}.$$

Expanding $U_{qj}(\theta)$ at θ^* yields

$$0 = U_{qj}(\tilde{\theta}) = U_{qj}(\theta^*) + \{H_{qj}(\theta'_{qj})\}^{\mathrm{T}}(\tilde{\theta} - \theta^*), \tag{S27}$$

where $\theta'_{qj} \in \mathbb{R}^{(J+1)(Q-1)}$ is a point on the line segment between $\tilde{\theta}$ and θ^*. That $\tilde{\theta} = \theta^* + o_{\mathbb{P}}(1)$ from (S24) ensures $\theta'_{qj} = \theta^* + o_{\mathbb{P}}(1)$ and hence $H_{qj}(\theta'_{qj}) = H_{qj}(\theta^*) + o_{\mathbb{P}}(1)$ for all $q \in \mathcal{Q}_+$ and $j = 0, 1, \ldots, J$.

Let H' be the matrix with rows $H_{qj}(\theta'_{qj})$ in lexicographical order of (qj). Then $H' = H^* + o_{\mathbb{P}}(1)$ with $H^* < 0$ and hence

$$1_{|H'|} = 1(H' \text{ is nonsingular}) = 1 + o_{\mathbb{P}}(1)$$

by Condition 3. Stacking (S27) in lexicographical order of (qj) yields $0 = U(\theta^*) + H'(\tilde{\theta} - \theta^*)$ and hence

$$1_{|H'|}\sqrt{N}(\tilde{\theta} - \theta^*) = -1_{|H'|}(H')^{-1}\sqrt{N}U(\theta^*).$$

This, together with $\sqrt{N}U(\theta^*)$ being asymptotically normal by (S25) and Lemma S2, ensures

$$\sqrt{N}(\tilde{\theta} - \theta^*) = O_P(1) \tag{S28}$$

and hence

$$\begin{aligned}
\sqrt{N}(\tilde{\theta} - \theta^*) &= 1_{|H'|}\sqrt{N}(\tilde{\theta} - \theta^*) + (1 - 1_{|H'|})\sqrt{N}(\tilde{\theta} - \theta^*) \\
&= -1_{|H'|}(H')^{-1}\sqrt{N}U(\theta^*) + (1 - 1_{|H'|})\sqrt{N}(\tilde{\theta} - \theta^*) \\
&= -(H^*)^{-1}\sqrt{N}U(\theta^*) + o_{\mathbb{P}}(1)
\end{aligned} \tag{S29}$$

by Slutsky's theorem. Observe that $(H^*)^{-1} = -\Phi^{-1} \otimes \mathrm{diag}\{1, (S_x^2)^{-1}\}$ from (S26). Removing the dimensions corresponding to $\{\beta_{q0} : q \in \mathcal{Q}_+\}$ in (S30) yields

$$\sqrt{N}\tilde{\beta} = \{\Phi^{-1} \otimes (S_x^2)^{-1}\}\sqrt{N}(R_+ \otimes I_J)\hat{x}_+ + o_{\mathbb{P}}(1) = \sqrt{N}\Psi\hat{x}_+ + o_{\mathbb{P}}(1)$$

by the explicit form of $U(\theta^*)$ from (S25).

Asymptotic normality of $\tilde\beta$. That $\sqrt{N}\tilde\beta \rightsquigarrow \mathcal{N}(0_{J(Q-1)}, V_\Psi)$ then follows from the asymptotic normality of $\hat x_+$ by Lemma S9 and Slutsky's theorem.

Convergence of $N\tilde V$. Let $\tilde H = H(\tilde\theta)$ be the value of $H(\theta)$ evaluated at the MLE. Then $\tilde V$ equals the submatrix of $(-N\tilde H)^{-1}$ after removing the rows and columns corresponding to the $Q-1$ intercepts, namely $\{\tilde\beta_{q0} : q \in \mathcal{Q}_+\}$. That $N\tilde V = V_\Psi + o_{\mathbb{P}}(1)$ follows from

$$\tilde H^{-1} = (H^*)^{-1} + o_{\mathbb{P}}(1) = -\Phi^{-1} \otimes \text{diag}\{1, (S_x^2)^{-1}\} + o_{\mathbb{P}}(1)$$

by (S24) and (S26), and the definition of the Kronecker product.

Asymptotic equivalence of λ_{LRT} and $N\hat x_+^{\mathrm{T}} V_{x+}^{-1}\hat x_+$. The LRT tests $\texttt{logit}(Z_i \sim 1 + x_i)$ against H_0 : $\texttt{logit}(Z_i \sim 1)$. Let $\Theta_0 = \{\theta = (\theta_1^{\mathrm{T}}, \dots, \theta_{Q-1}^{\mathrm{T}})^{\mathrm{T}} : \theta_q = (\beta_{q0}, 0_J^{\mathrm{T}})^{\mathrm{T}}\}$ be the restricted parameter space under H_0, with $\tilde\theta_0 \in \Theta_0$ as the MLE. The test statistic equals

$$\lambda_{\text{LRT}} = -2N\left\{\sup_{\theta\in\Theta_0} \bar L(\theta) - \sup_{\theta\in\mathbb{R}^{(J+1)(Q-1)}} \bar L(\theta)\right\} = 2N\{\bar L(\tilde\theta) - \bar L(\tilde\theta_0)\}.$$

For $\theta \in \Theta_0$, we have $\tilde x_i^{\mathrm{T}}\theta_q = \beta_{q0}$ such that (S20) reduces to $\bar L(\theta) = \sum_{q\in\mathcal{Q}} e_q \log(\pi_q)$, with

$$\pi_q = \frac{\exp(\beta_{q0})}{1 + \sum_{q'\in\mathcal{Q}_+}\exp(\beta_{q'0})} \qquad \text{for} \quad q \in \mathcal{Q}$$

denoting the identical value of $\pi_q(x_i)$ across $i = 1, \dots, N$; see (S19). The invariance of MLE to non-degenerate transformation of the parameters ensures that the MLEs of π_q and β_{q0} equal e_q and $\log(e_q/e_Q) = \beta_{q0}^*$, respectively, for $q \in \mathcal{Q}_+$. This ensures $\tilde\theta_0 = \theta^*$ and hence

$$\lambda_{\text{LRT}} = 2N\{\bar L(\tilde\theta) - \bar L(\theta^*)\}. \tag{S30}$$

We verify below $\lambda_{\text{LRT}} - N\hat x_+^{\mathrm{T}} V_{x+}^{-1}\hat x_+ = o_{\mathbb{P}}(1)$.

First, $\bar L(\theta^*) = \bar L(\tilde\theta) + 2^{-1}(\theta^* - \tilde\theta)^{\mathrm{T}} H(\theta')(\theta^* - \tilde\theta)$ for some θ' on the line segment of $\tilde\theta$ and θ^*. This, together with (S30), ensures

$$\lambda_{\text{LRT}} = -N(\tilde\theta - \theta^*)^{\mathrm{T}} H(\theta')(\tilde\theta - \theta^*) = -N(\tilde\theta - \theta^*)^{\mathrm{T}} H^*(\tilde\theta - \theta^*) + o_{\mathbb{P}}(1)$$

given $H(\theta') = H^* + o_{\mathbb{P}}(1)$ by (S24) and $\sqrt{N}(\tilde\theta - \theta^*) = O_P(1)$ by (S28).

Next, it follows from (S30) that

$$-N(\tilde{\theta}-\theta^*)^{\mathrm{T}}H^*(\tilde{\theta}-\theta^*) = -N\{U(\theta^*)\}^{\mathrm{T}}(H^*)^{-1}U(\theta^*) + o_{\mathbb{P}}(1).$$

The result then follows from

$$\begin{aligned} & -N\{U(\theta^*)\}^{\mathrm{T}}(H^*)^{-1}U(\theta^*) \\ = \;& N\big(e_1\hat{x}(1)^{\mathrm{T}}, \dots, e_{Q-1}\hat{x}(Q-1)^{\mathrm{T}}\big)\{\Phi^{-1}\otimes(S_x^2)^{-1}\}\begin{pmatrix} e_1\hat{x}(1) \\ \vdots \\ e_{Q-1}\hat{x}(Q-1)\end{pmatrix} \\ = \;& N\{\hat{x}_+^{\mathrm{T}}(R_+\otimes I_J)\}V_\Psi\{(R_+\otimes I_J)\hat{x}_+\} \\ = \;& N\hat{x}_+^{\mathrm{T}}V_{x+}^{-1}\hat{x}_+ \end{aligned}$$

by (S25)–(S26) and (S17)–(S18).

Asymptotic equivalence of W_{logit} and $N\hat{x}_+^{\mathrm{T}}V_{x+}^{-1}\hat{x}_+$. The result follows from

$$W_{\text{logit}} = \tilde{\beta}^{\mathrm{T}}\tilde{V}^{-1}\tilde{\beta} = N(\Psi\hat{x}_+)^{\mathrm{T}}V_\Psi^{-1}(\Psi\hat{x}_+) + o_{\mathbb{P}}(1) = N\hat{x}_+^{\mathrm{T}}V_{x+}^{-1}\hat{x}_+ + o_{\mathbb{P}}(1)$$

with $\sqrt{N}(\tilde{\beta}-\Psi\hat{x}_+) = o_{\mathbb{P}}(1)$, $N\tilde{V} - V_\Psi = o_{\mathbb{P}}(1)$ as we just proved and $\Psi^{\mathrm{T}}V_\Psi^{-1}\Psi = V_{x+}^{-1}$ by (S18).

Asymptotic distributions of λ_{LRT} and W_{logit}. The result follows from $N\hat{x}_+^{\mathrm{T}}V_{x+}^{-1}\hat{x}_+ \rightsquigarrow \chi^2_{J(Q-1)}$ by Lemma S9 and Slutsky's theorem.

Simplification under the treatment-control experiment. The result follows from $\hat{x}_+ = \hat{x}(1) = e_0\hat{\tau}_x$ and $\Psi = e_0^{-1}(S_x^2)^{-1}$ such that $\Psi\hat{x}_+ = e_0^{-1}(S_x^2)^{-1}\hat{x}(1) = (S_x^2)^{-1}\hat{\tau}_x$ and $V_\Psi = (e_0e_1)^{-1}(S_x^2)^{-1}$. □

References

Barrera-Osorio, F., M. Bertrand, L. L. Linden, and F. Perez-Calle (2011). Improving the design of conditional transfer programs: Evidence from a randomized education experiment in Colombia. *American Economic Journal: Applied Economics 3*, 167–95.

Chen, Y., F. M. Harper, J. Konstan, and S. X. Li (2010). Social comparisons and contributions to online communities: A field experiment on MovieLens. *American Economic Review 100*, 1358–1398.

de Mel, S., D. McKenzie, and C. Woodruff (2009). Are women more credit constrained? Experimental evidence on gender and microenterprise returns. *American Economic Journal: Applied Economics 1*, 1–32.

Dharmadhikari, S. and K. Joag-Dev (1988). *Unimodality, Convexity, and Applications*. San Diego, CA: Academic Press, Inc.

Ding, P. (2021). The Frisch–Waugh–Lovell theorem for standard errors. *Statistics & Probability Letters 168*, 108945.

Dupas, P. and J. Robinson (2013). Why don't the poor save more? Evidence from health savings experiments. *American Economic Review 103*, 1138–1171.

Gerber, A. S., D. Karlan, and D. Bergan (2009). Does the media matter? A field experiment measuring the effect of newspapers on voting behavior and political opinions. *American Economic Journal: Applied Economics 1*, 35–52.

Karlan, D. and M. A. McConnell (2014). Hey look at me: The effect of giving circles on giving. *Journal of Economic Behavior and Organization 106*, 402–412.

Latała, R. and D. Matlak (2017). *Royen's proof of the Gaussian correlation inequality*, pp. 197–202. Lecture Notes in Mathematics 2169. Springer.

Li, X., P. Ding, and D. B. Rubin (2018). Asymptotic theory of rerandomization in treatment–control experiments. *Proceedings of the National Academy of Sciences 115*, 9157–9162.

Li, X., P. Ding, and D. B. Rubin (2020). Rerandomization in 2^K factorial experiments. *The Annals of Statistics 48*, 43–63.

Morgan, K. L. (2011). *Rerandomization to Improve Covariate Balance in Randomized Experiments*. Ph. D. thesis.

Morgan, K. L. and D. B. Rubin (2012). Rerandomization to improve covariate balance in experiments. *The Annals of Statistics 40*, 1263–1282.

Morgan, K. L. and D. B. Rubin (2015). Rerandomization to balance tiers of covariates. *Journal of the American Statistical Association 110*, 1412–1421.

Newey, W. K. and D. L. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, Chapter 36, pp. 2111–2245. Amsterdam: Elsevier.

Royen, T. (2014). A simple proof of the Gaussian correlation conjecture extended to multivariate gamma distributions. *Far East Journal of Theoretical Statistics 48*, 139–145.

Rudin, W. (1976). *Principles of Mathematical Analysis* (3rd ed.). McGraw-Hill, New York.

Sherman, S. (1955). A theorem on convex sets with applications. *The Annals of Mathematical Statistics 26*, 763–767.

Wu, J. and P. Ding (2021). Randomization tests for weak null hypotheses in randomized experiments. *Journal of the American Statistical Association 116*, 1898–1913.

Zhao, A. and P. Ding (2021). Covariate-adjusted Fisher randomization tests for the average treatment effect. *Journal of Econometrics 225*, 278–294.

Zhao, A. and P. Ding (2023). Covariate adjustment in multiarmed, possibly factorial experiments. *Journal of the Royal Statistical Society Series B: Statistical Methodology 85*, 1–23.