Statistica Sinica Preprint No: SS-2025-0315

Title	Balancing Covariates in Survey Experiments
Manuscript ID	SS-2025-0315
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0315
Complete List of Authors	Pengfei Tian, Jiyang Ren and Yingying Ma
Corresponding Authors	Yingying Ma
E-mails	mayingying_11@163.com
Notice: Accepted author version.	

Statistica Sinica

BALANCING COVARIATES IN SURVEY EXPERIMENTS

Pengfei Tian[1], Jiyang Ren[1], and Yingying Ma[2]

[1] *Tsinghua University and* [2] *Beihang University*

Abstract: The survey experiment is widely used in economics and social sciences to evaluate the effects of treatments or programs. In a standard population-based survey experiment, the experimenter randomly draws experimental units from a target population of interest and then randomly assigns the sampled units to treatment or control conditions to explore the treatment effect of an intervention. Simple random sampling and treatment assignment can balance covariates on average. However, covariate imbalance often exists in finite samples. To address the imbalance issue, we study a stratified approach to balance covariates in a survey experiment. A stratified rejective sampling and rerandomization design is further proposed to enhance the covariate balance. We develop a design-based asymptotic theory for the widely used stratified difference-in-means estimator of the average treatment effect under the proposed design. In particular, we show that it is consistent and asymptotically a convolution of a normal distribution and two truncated normal distributions. This limiting distribution is more concentrated at the true average treatment effect than that under the existing experimental designs. Moreover, we propose a covariate adjustment method in the analysis stage, which can further improve the estimation efficiency. Numerical studies demonstrate the validity and improved efficiency of the proposed method.

Key words and phrases: Blocking, covariate adjustment, design-based inference, stratification, rerandomization.

1. Introduction

The survey experiment, also referred to as the experiment embedded within a survey, has been viewed as the gold standard for estimating the treatment effect for a target population of interest (Mutz, 2011). Following Mutz (2011), the survey experiment in this article is the abbreviation for a population-based survey experiment. It contains two stages: "survey" means the sampling stage, i.e., the experimenter randomly samples units from the target population, and "experiment" means the treatment assignment stage, i.e., the experimenter randomly assigns the sampled units to the treatment or control condition to explore the effect of an intervention. In the past decades, the survey experiment has gained increasing popularity in many fields, such as political science, education, and economics, because it is easy to implement and clear to distinguish cause and effect for the target population. For example, in the 1980s, the US Department of Labor conducted the Pennsylvania Reemployment Bonus survey experiment to test the incentive effects for unemployment insurance (Bilias, 2000). There is also another example that researchers at Boston University conducted a survey experiment to investigate the influence of public opinion on citizen perception (Dancey and Sheagley, 2018). The survey experiment provides grounded inferences about real-world behavior based on a representative sample (Gaines et al., 2007) and can effectively solve the problem of insufficient external validity of traditional one-stage randomized experiments (Mutz, 2011), which lack the stage of random sampling of a subset of units into the experiments.

Simple random sampling and complete randomization in a survey experiment balance observed and unobserved confounding factors on average and justify simple comparisons of average outcomes among the treatment and control groups. However, covariate imbalance happens not only between the sampled experimental units and the overall population of interest but also between treatment and control groups (Yang et al., 2023). In the survey sampling literature, stratified random sampling is widely used to balance covariates, generate a more representative sample, and increase the efficiency of a sample design concerning cost and precision (Imbens and Lancaster, 1996). Stratification or blocking is also used to balance covariates in the treatment assignment stage (Fisher, 1926; Imbens and Rubin, 2015). According to a recent survey (Lin et al., 2015), stratification has been prevalently used in randomized clinical trials. In recent studies, researchers introduced a finely stratified design approach grounded in the semi-parametric efficiency principle and super-population framework (Bai et al., 2022; Cytrynbaum, 2023; Tabord-Meehan, 2023).

Design-based or finite-population asymptotic results for a general stratification strategy in the survey experiment have not been fully explored. The design-based inference only considers the randomness of sampling and treatment assignment, with potential outcomes and covariates being fixed. It can be seen as a form of conditional inference within the super-population framework, i.e., drawing causal inferences conditional on the potential outcomes and covariates. The design-based inference can be traced back to Fisher and Neyman (Splawa-Neyman, 1923; Fisher, 1935) and is gaining increasing popularity in causal inference theory and practice (Lin, 2013; Imbens and Rubin, 2015; Li and Ding, 2017; Liu and Yang, 2020; Lu et al., 2023).

Our first contribution is to establish the design-based asymptotic theory for the stratified difference-in-means estimator of the average treatment effect under the stratified randomized survey experiment. Specifically, we show that this estimator is consistent and asymptotically normal, and that its asymptotic variance is usually smaller than that under a completely randomized survey experiment. Hence, stratification can improve efficiency. We also obtain the optimal stratum-specific sampling proportion and treated proportion to minimize the asymptotic variance of the stratified difference-in-means estimator. Note that we consider a general asymptotic regime that allows the sampling fraction to tend to zero, the proportions of treated units to vary across strata, as well as the number of strata or the associated stratum sizes to diverge. Thus, our theoretical results cover the scenarios of a few large strata, many small strata, and a combination thereof. To establish a central limit theorem (CLT) for our estimator, we develop a Hájek coupling technique that reduces the estimator to a sum of independent random variables. Our setting introduces additional challenges due to the presence of two sources of design randomness, sampling and treatment assignment, which are not jointly handled in existing work.

To address this, we recast the joint design as a unified one-stage within-stratum permutation that partitions units into treated, control, and unsampled groups. Unlike standard one-stage asymptotic theory, which requires group proportions to be bounded away from zero and one, our framework allows the sampling fraction to vanish. We further develop a refined large-/small-strata decomposition, building on Bickel and Freedman (1984), and introduce a threshold that ensures uniform control of the coupling remainders across strata. This enables us to establish a general CLT under diverging strata and vanishing sampling fractions.

Rejective sampling (Fuller, 2009) and rerandomization (Morgan and Rubin, 2012; Li and Ding, 2020) are more general approaches to balance covariates. This approach needs covariates information from another survey or previous research to calibrate sampling or assignment procedure to decrease the chance imbalance. Recently, Yang et al. (2023) proposed a rejective sampling and rerandomized experimental design to avoid covariate imbalance at both the sampling and treatment assignment stages in the survey experiment. They also took into account a stratified design but provided that the sampling ratios and propensity scores are asymptotically the same across strata. Wang et al. (2023) considered the combination of stratification and rerandomization as suggested by Donald Rubin and verified that such a design can achieve better computational and efficiency performance, but their discussion is only for one-stage experiments. Stratification, rejective sampling, and rerandomization are easy to interpret and implement, and intuitively, applying these three methods together can further reduce the covariate imbalance and improve the computational and estimation efficiency. However, limited studies have addressed the statistical properties of the combination in the survey experiment under general settings, allowing the diverging strata number or both large strata and small strata to exist.

Our second contribution is to introduce an innovative Stratified Rejective Sampling and ReRandomized (SRSRR) experimental design and provide a covariate adjustment procedure within the SRSRR design's analysis stage. This design inherits the advantages of traditional stratification in exploring subpopulation properties and organizational convenience (Cochran, 1977), and simultaneously enhances covariate balance and improves the efficiency of treatment effect estimation. We establish the design-based asymptotic theory for the stratified difference-in-means estimator under the proposed SRSRR experiment and show that its limiting distribution is no longer normal but a convolution of a normal distribution and two truncated normal distributions. This limiting distribution is more concentrated at the true average treatment effect than that under the stratified randomized survey experiment. Thus, it verifies that SRSRR can further improve the estimation efficiency for treatment effects. Moreover, we provide a conservative estimator for the asymptotic distribution to facilitate valid inferences. Lastly, we propose a covariate adjustment method to further improve the estimation efficiencies. The validity and improved efficiency of the proposed methods are demonstrated through numerical studies.

2. Stratified rejective sampling and rerandomization

2.1 Stratified population-based survey experiment

To eliminate the potential risk of chance imbalance and conduct the experiment more conveniently, researchers often stratify the population according to important covariates that are predictive of the outcomes and conduct a stratified randomized survey experiment. Let N denote the size of a finite target population. Units are partitioned into K_N strata indexed by $k = 1, \ldots, K_N$, where stratum k contains $N_{[k]}$ units and $\sum_{k=1}^{K_N} N_{[k]} = N$. Throughout the paper, we use a subscript $[k]$ to denote stratum-specific quantities. The stratified randomized survey experiment consists of two stages: (1) sampling stage: randomly sample n units from the target population by stratified random sampling without replacement with $n_{[k]}$ units sampled in stratum k; and (2) treatment assignment stage: the sampled units are randomly assigned to the treatment and control groups using stratified randomization. Within each stratum k, $n_{[k]1}$ units are randomly selected to receive the treatment, and the remaining $n_{[k]0} = n_{[k]} - n_{[k]1}$ units are designated for the control group. The treatment assignments across strata are independent, and the total sample size n, the sample size $n_{[k]}$ in stratum k, and the number of treated units $n_{[k]1}$ in stratum k are fixed and satisfy $\sum_{k=1}^{K_N} n_{[k]} = n$ with $2 \leq n_{[k]1} \leq n_{[k]} - 2$ for $k = 1, \ldots, K_N$.

We use $\mathcal{S}$ to denote the set of sampled units. For each unit i ($i = 1, \ldots, N$), let Z_i be the sampling indicator; $Z_i = 1$ if unit i is sampled and $Z_i = 0$ otherwise. For $i \in \mathcal{S}$, let T_i be a binary treatment assignment indicator with $T_i = 1$ if unit i is assigned to the treatment group and $T_i = 0$ otherwise. Define $\Pi_{[k]} = N_{[k]}/N$ as the proportion of stratum size in the target population and $\pi_{[k]} = n_{[k]}/n$ as the proportion of stratum size in the sample. In practice, we often employ the proportionate stratified sampling with $\pi_{[k]} = \Pi_{[k]}$ for $k = 1, \ldots, K_N$ (Cochran, 1977).

We adopt the potential outcomes framework to define the treatment effect. For unit i, let $Y_i(1)$ and $Y_i(0)$ denote the potential outcomes under treatment and control, respectively. The unit-level treatment effect is $\tau_i = Y_i(1) - Y_i(0)$, and the average treatment effect is $\tau = N^{-1} \sum_{k=1}^{K_N} \sum_{i \in [k]} \tau_i = \sum_{k=1}^{K_N} \Pi_{[k]} \tau_{[k]}$, where $\tau_{[k]} = N_{[k]}^{-1} \sum_{i \in [k]} \tau_i$ is the average treatment effect in stratum k and $i \in [k]$ indexes unit i in stratum k. The observed outcome is $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$ for $i \in \mathcal{S}$. The difference in the sample means of the outcomes between the treatment and control groups within that stratum is an unbiased estimator for $\tau_{[k]}$:

$$\hat{\tau}_{[k]} = \bar{Y}_{[k]1} - \bar{Y}_{[k]0} = n_{[k]1}^{-1} \sum_{i \in [k]} Z_i T_i Y_i - n_{[k]0}^{-1} \sum_{i \in [k]} Z_i (1 - T_i) Y_i.$$

By plugging in, we further obtain an unbiased estimator for τ as $\hat{\tau} = \bar{Y}_1 - \bar{Y}_0 = \sum_{k=1}^{K_N} \Pi_{[k]} \hat{\tau}_{[k]}$, where $\bar{Y}_1 = \sum_{k=1}^{K_N} \Pi_{[k]} \bar{Y}_{[k]1}$ and $\bar{Y}_0 = \sum_{k=1}^{K_N} \Pi_{[k]} \bar{Y}_{[k]0}$.

2.2 Stratified rejective sampling and rerandomization

Stratified random sampling and stratified randomization can balance discrete covariates that are most predictive of the outcomes. However, additional covariates aside from the stratification variables may remain unbalanced. In the sampling stage, if these covariates are not well-balanced such that the sampled units are not representative, the estimation accuracy on the target population decreases (Banerjee et al., 2017). In the treatment assignment stage, the covariate imbalance between the treatment and control groups may lead to large variability and conditional bias. Thus, balancing additional covariates at both stages is desirable. Suppose that we collect covariates $W_i \in \mathbb{R}^{J_1}$ at the sampling stage from another survey or previous research. After the sampling stage, only a small proportion of the population is sampled, and more covariates of the sampled units, denoted by $X_i \in \mathbb{R}^{J_2}$, can be collected at the treatment assignment stage. Next, we introduce how to use SRSRR to balance W_i and X_i. SRSRR consists of the following two stages:

Stratified rejective sampling. In this stage, to make the sampled units more representative of the target population, we can reject the samples that result in covariate imbalance and repeat the stratified random sampling procedure until we obtain a sample with satisfactory covariate balance between the sampled units and the population (Fuller, 2009). Let $\bar{W} = N^{-1}\sum_{i=1}^{N} W_i = \sum_{k=1}^{K_N} \Pi_{[k]}\bar{W}_{[k]}$ and $\bar{W}_{\mathcal{S}} = \sum_{k=1}^{K_N} \Pi_{[k]}\bar{W}_{[k]\mathcal{S}}$ denote the population mean and weighted sample mean, respectively, where $\bar{W}_{[k]} = N_{[k]}^{-1}\sum_{i\in[k]} W_i$ and $\bar{W}_{[k]\mathcal{S}} = n_{[k]}^{-1}\sum_{i\in[k]} Z_i W_i$ are the stratum-specific population mean and sample mean, respectively. We can use the Mahalanobis distance between $\bar{W}$ and $\bar{W}_{\mathcal{S}}$ to measure the covariate imbalance, which has the advantage of being affinely invariant (Morgan and Rubin, 2012). The Mahalanobis distance is defined as

$$\mathrm{M}_S = (\bar{W}_{\mathcal{S}} - \bar{W})^{\mathrm{T}}\Big\{\sum_{k=1}^{K_N} \Pi_{[k]}^2\Big(\frac{1}{n_{[k]}} - \frac{1}{N_{[k]}}\Big) S_{[k]W}^2\Big\}^{-1}(\bar{W}_{\mathcal{S}} - \bar{W}),$$

where $S_{[k]W}^2 = (N_{[k]} - 1)^{-1}\sum_{i\in[k]}(W_i - \bar{W}_{[k]})(W_i - \bar{W}_{[k]})^{\mathrm{T}}$ is the stratum-specific population covariance of W_i. The acceptability of the sample set is determined by the condition $\mathrm{M}_S \le a_S$, where $a_S > 0$ is a pre-specified fixed threshold. We repeat the stratified sampling procedure until the sample set satisfies this acceptability criterion. We can choose a_S such that the acceptance probability achieves a given level.

Stratified rerandomization. After sampling, we balance the covariates of the sampled units between treatment and control groups by stratified rerandomization (Wang et al., 2023). Specifically, for the sampled units, let $\bar{X}_1 = \sum_{k=1}^{K_N} \Pi_{[k]}\bar{X}_{[k]1}$ and $\bar{X}_0 = \sum_{k=1}^{K_N} \Pi_{[k]}\bar{X}_{[k]0}$ be the weighted sample means of the covariates X_i under the treatment and control conditions, respectively, where $\bar{X}_{[k]1} = n_{[k]1}^{-1}\sum_{i\in[k]} Z_i T_i X_i$ and $\bar{X}_{[k]0} = n_{[k]0}^{-1}\sum_{i\in[k]} Z_i(1 - T_i)X_i$ are the stratum-specific sample means of X_i under treatment and control in stratum k. Let $\hat{\tau}_X = \bar{X}_1 - \bar{X}_0$. The Mahalanobis distance between $\bar{X}_1$ and $\bar{X}_0$ given the sample $\mathcal{S}$ is defined as

$$\mathrm{M}_T = \hat{\tau}_X^{\mathrm{T}}\mathrm{cov}(\hat{\tau}_X \mid \mathcal{S})^{-1}\hat{\tau}_X = (\bar{X}_1 - \bar{X}_0)^{\mathrm{T}}\Big\{\sum_{k=1}^{K_N} \Pi_{[k]}^2\frac{n_{[k]}}{n_{[k]1}n_{[k]0}} S_{[k]X|\mathcal{S}}^2\Big\}^{-1}(\bar{X}_1 - \bar{X}_0),$$

where $S^2_{[k]X|\mathcal{S}} = (n_{[k]} - 1)^{-1} \sum_{i\in[k]} Z_i(X_i - \bar{X}_{[k]\mathcal{S}})(X_i - \bar{X}_{[k]\mathcal{S}})^{\mathrm{T}}$ is the covariance of X_i conditional on the sample set $\mathcal{S}$ with $\bar{X}_{[k]\mathcal{S}} = n_{[k]}^{-1} \sum_{i\in[k]} Z_i X_i$. The treatment assignment is acceptable if and only if $\mathrm{M}_T \leq a_T$ for a pre-specified fixed $a_T > 0$. We discard the undesired treatment assignments and repeat the stratified randomization until the treatment assignment is acceptable. To ensure covariate balance and powerful design-based inference, we can choose a_T such that the probability of a treatment assignment being acceptable achieves a given level, such as 0.01 or 0.001 as suggested by Morgan and Rubin (2012).

Remark 1. While we implement stratified rerandomization using the Mahalanobis distance, a large class of alternative rerandomization methods can be accommodated by replacing M_T with a general quadratic-form balance metric. Specifically, many rerandomization criteria can be written as an ellipsoidal acceptance rule $Q_A(\hat{\tau}_X) = \hat{\tau}_X^\top A\, \hat{\tau}_X \leq a_T$ with $A \succeq 0$, which includes tiered rerandomization (Morgan and Rubin, 2015), ridge rerandomization (Branson and Shao, 2021), PCA rerandomization (Zhang et al., 2024), and joint-test based acceptance rules that can be expressed via quadratic forms (Zhao and Ding, 2024). Recent work by Schindl and Branson (2024) develops a general framework for rerandomization with ellipsoidal (quadratic-form) acceptance regions, characterizes the resulting covariance reduction along the covariates' eigen-directions, and provides guidance and optimality results for selecting A in practice. In particular, the Mahalanobis choice maximizes the total variance reduction across covariate directions, whereas the Euclidean choice $A = I$ minimizes the Frobenius norm of the covariance matrix after quadratic form rerandomization. It would be interesting to generalize these optimality results to the SRSRR framework.

When $K = 1$, our SRSRR framework reduces to Yang et al. (2023); when $f = 1$, it reduces to Wang et al. (2023). More broadly, SRSRR clarifies the division of labor across stages: stratification and rejective sampling operate at the sampling stage to improve representativeness and efficiency, whereas rerandomization operates at the assignment stage to improve covariate balance; thus, the two tools are complementary rather than interchangeable. This separation also provides a computational benefit, since stratification shrinks the assignment space for rerandomization (Schultzberg and Johansson, 2022; Wang et al., 2023).

3. Design-based asymptotic theory

In this section, we investigate the asymptotic distribution of $\hat{\tau}$ under the stratified randomized survey experiment (SRSE) and the SRSRR experiment, respectively.

3.1 Asymptotic properties under SRSE

We introduce some notations first. Let $n_1 = \sum_{k=1}^{K_N} n_{[k]1}$ and $n_0 = \sum_{k=1}^{K_N} n_{[k]0}$ be the total numbers of units assigned to the treatment and control groups, respectively. Let $f = n/N$, $e_1 = n_1/n$, and $e_0 = n_0/n$ be the total proportions of sampled, treated, and control units, respectively. Accordingly, for $k = 1, \ldots, K_N$, let $f_{[k]} = n_{[k]}/N_{[k]}$, $e_{[k]1} = n_{[k]1}/n_{[k]}$, and $e_{[k]0} = n_{[k]0}/n_{[k]}$ be the proportions of sampled, treated, and control units in stratum k, respectively. Within stratum k, denote $\bar{Y}_{[k]}(t) = (1/N_{[k]})\sum_{i\in[k]} Y_i(t)$ for $t = 0, 1$ and $\bar{X}_{[k]} = (1/N_{[k]})\sum_{i\in[k]} X_i$ as the stratum-specific finite-population means of $Y_i(t)$ and X_i, respectively. Define $S^2_{[k]t} = (N_{[k]} - 1)^{-1}\sum_{i\in[k]}\{Y_i(t) - \bar{Y}_{[k]}(t)\}^2$ and $S^2_{[k]\tau} = (N_{[k]} - 1)^{-1}\sum_{i\in[k]}(\tau_i - \tau_{[k]})^2$ as the stratum-specific finite-population variances of $Y_i(t)$ and τ_i. Similarly, we use $S^2_{[k]X} = (N_{[k]} - 1)^{-1}\sum_{i\in[k]}(X_i - \bar{X}_{[k]})(X_i - \bar{X}_{[k]})^{\mathrm{T}}$ and $S_{[k]X,t} = S^{\mathrm{T}}_{[k]t,X} = (N_{[k]} - 1)^{-1}\sum_{i\in[k]}(X_i - \bar{X}_{[k]})\{Y_i(t) - \bar{Y}_{[k]}(t)\}$ to denote the stratum-specific finite-population covariance of X_i and covariance between X_i and $Y_i(t)$, respectively. Substituting X_i with W_i, we can analogously define $S^2_{[k]W}$ and $S_{[k]W,t} = S^{\mathrm{T}}_{[k]t,W}$. Let $S_{[k]W,\tau} = S^{\mathrm{T}}_{[k]\tau,W} = (N_{[k]} - 1)^{-1}\sum_{i\in[k]}(W_i - \bar{W}_{[k]})(\tau_i - \tau_{[k]})$. Denote I_k as a $k \times k$ identity matrix and $0_{J_1\times J_2}$ as a $J_1 \times J_2$ zero matrix. Let $\|\cdot\|_\infty$ and $\|\cdot\|_2$ be the ℓ_∞ and ℓ_2 norms of a vector, respectively. Define $V_{[k]}$ as the matrix given by

$$\begin{pmatrix} e^{-1}_{[k]1}S^2_{[k]1} + e^{-1}_{[k]0}S^2_{[k]0} - f_{[k]}S^2_{[k]\tau} & e^{-1}_{[k]1}S_{[k]1,X} + e^{-1}_{[k]0}S_{[k]0,X} & (1 - f_{[k]})S_{[k]\tau,W} \\ e^{-1}_{[k]1}S_{[k]X,1} + e^{-1}_{[k]0}S_{[k]X,0} & (e_{[k]1}e_{[k]0})^{-1}S^2_{[k]X} & 0_{J_2\times J_1} \\ (1 - f_{[k]})S_{[k]W,\tau} & 0_{J_1\times J_2} & (1 - f_{[k]})S^2_{[k]W} \end{pmatrix}.$$

Proposition S1 in the Supplementary Material shows that

$$V := \mathrm{cov}\big\{\sqrt{n}(\hat{\tau}-\tau, \hat{\tau}_X^{\mathrm{T}}, \hat{\delta}_W^{\mathrm{T}})^{\mathrm{T}}\big\} = \begin{pmatrix} V_{\tau\tau} & V_{\tau X} & V_{\tau W} \\ V_{X\tau} & V_{XX} & V_{XW} \\ V_{W\tau} & V_{WX} & V_{WW} \end{pmatrix} = \sum_{k=1}^{K_N} \Pi_{[k]}^2 \pi_{[k]}^{-1} V_{[k]}.$$

To infer τ based on $\hat{\tau}$, we need to further determine its asymptotic distribution. Let $\mathcal{M}_L$ be the set of finite-population quantities satisfying the maximum squared distance requirement and bounded stratum-specific second-moment conditions: $\mathcal{M}_L = \big\{(a_1,\ldots,a_N) : \max_{k=1,\ldots,K_N} \max_{i\in[k]} n^{-1}\|a_i-\bar{a}_{[k]}\|_\infty^2 \to 0, \max_{k=1,\ldots,K_N} N_{[k]}^{-1}\sum_{i\in[k]}\|a_i-\bar{a}_{[k]}\|_\infty^2 \le L\big\}$, where $\bar{a}_{[k]} = (1/N_{[k]})\sum_{i\in[k]} a_i$. We assume that the dimensions of covariates X and W are fixed, and we need Condition 1 below to derive the joint asymptotic normality of $\sqrt{n}(\hat{\tau}-\tau, \hat{\tau}_X^{\mathrm{T}}, \hat{\delta}_W^{\mathrm{T}})$.

Condition 1. For $t = 0, 1$, as $n \to \infty$, assume that

(i) there exist constants $c_1, c_2 \in (0,1)$ and $c_3 > 0$, such that for $k = 1,\ldots,K_N$, $c_1 \le e_{[k]1} \le 1-c_1$, $f_{[k]} \le c_2$, and $f/f_{[k]} \le c_3$;

(ii) f has a limit in $[0,1]$;

(iii) $\sum_{k=1}^{K_N}\Pi_{[k]}^2\pi_{[k]}^{-1}e_{[k]t}^{-1}S_{[k]t}^2$, $\sum_{k=1}^{K_N}\Pi_{[k]}^2\pi_{[k]}^{-1}e_{[k]t}^{-1}S_{[k]X,t}$, $\sum_{k=1}^{K_N}\Pi_{[k]}^2\pi_{[k]}^{-1}(1-f_{[k]})S_{[k]W,t}$, $\sum_{k=1}^{K_N}\Pi_{[k]}^2\pi_{[k]}^{-1}f_{[k]}S_{[k]\tau}^2$, $\sum_{k=1}^{K_N}\Pi_{[k]}^2\pi_{[k]}^{-1}(e_{[k]1}e_{[k]0})^{-1}S_{[k]X}^2$, and $\sum_{k=1}^{K_N}\Pi_{[k]}^2\pi_{[k]}^{-1}(1-f_{[k]})S_{[k]W}^2$ have finite limits and the limits of V_{XX} and V_{WW} are invertible;

(iv) there exists a constant $L > 0$ independent of N, such that $\{Y_i(t)\}_{i=1}^N$, $\{X_i\}_{i=1}^N$, $\{W_i\}_{i=1}^N \in \mathcal{M}_L$.

Condition 1(i) requires that the proportion of treated units within each stratum, $e_{[k]1}$, is bounded away from zero and one. For simplicity, Condition 1(i) assumes that $f_{[k]}$ is bounded away from one. Our method and theory can be extended to the case $f_{[k]} = 1$ for some k, though the notation and theoretical statements become more involved. Condition 1(i) also requires that $f_{[k]}$ is not negligible relative to f. Otherwise, some stratum information dominates relative to other strata, and it is impossible to infer the overall treatment effect. Condition 1(i)-(ii) together allow both the total sampling proportion f and the stratum-specific sampling proportion $f_{[k]}$ to tend to zero, which are natural requirements, and similar conditions can be found in the literature (Li and Ding, 2017; Yang et al., 2023). Condition 1(iii) ensures that the covariance V has a finite limit. For notation simplicity, we still use the same notation to denote the limit of V when no confusion would arise. When $f_{[k]} = f$ and $e_{[k]t} = e_t$ for all $k = 1, \ldots, K$ and $t \in \{0, 1\}$, Condition 1(iii) reduces to requiring that weighted sums $e_t^{-1} \sum_{k=1}^{K_N} \Pi_{[k]} S^2_{[k]t}$, $e_t^{-1} \sum_{k=1}^{K_N} \Pi_{[k]} S^2_{[k]X,t}$, $\sum_{k=1}^{K_N} \Pi_{[k]} S^2_{[k]W,t}$, $\sum_{k=1}^{K_N} \Pi_{[k]} S^2_{[k]\tau}$ have finite limits and that $\sum_{k=1}^{K_N} \Pi_{[k]} S^2_{[k]X}$, $\sum_{k=1}^{K_N} \Pi_{[k]} S^2_{[k]W}$ converge to finite, positive-definite limits. Condition 1(iv) is a bounded moment type condition on the potential outcomes and covariates. In particular, the condition $\max_{k=1,\ldots,K_N} \max_{i\in[k]} n^{-1}\{Y_i(1) - \bar{Y}_{[k]}(1)\}^2 \to 0$, requires that the stratum-specific deviations of the potential outcomes from their stratum-specific mean should not be very large. It is a typical condition used in the finite-population asymptotic theory (Li and Ding, 2017; Li et al., 2018; Yang et al., 2023). Moreover, a sufficient condition for the array a_i's ($a_i = Y_i(1), Y_i(0), X_i$, or W_i) to belong to $\mathcal{M}_L$ is that there exist constants $L' > 0$ and $\delta > 0$ such that $\max_{k=1,\ldots,K} N_{[k]}^{-1} \sum_{i\in[k]} ||a_i - \bar{a}_{[k]}||_\infty^{2+\delta} \leq L'$ with $f^2 n^\delta \to \infty$. Lighter tails of the a_i's correspond to larger values of δ, which in turn permit a faster decay rate for f. We then have the following theorem.

Theorem 1. *Under Condition 1 and the stratified randomized survey experiment, we have* $\sqrt{n}(\hat{\tau} - \tau, \hat{\tau}_X^{\mathrm{T}}, \hat{\delta}_W^{\mathrm{T}})^{\mathrm{T}} \xrightarrow{d} \mathcal{N}(0, V)$.

Theorem 1 provides a normal approximation for the distribution of $\hat{\tau}$ under the stratified randomized survey experiment, which can be used to make design-based inferences for the average treatment effect if one can estimate the asymptotic variance consistently or conservatively. Moreover, it provides the theoretical basis for deriving the asymptotic distribution of $\hat{\tau}$ under the SRSRR experiment.

Theorem 1 requires very weak conditions on the number of strata, stratum sizes, and the proportion of treated units in each stratum. The asymptotic regime is very general and covers a wide range of cases, such as a diverging number of strata with fixed stratum sizes, diverging stratum sizes with a fixed number of strata, diverging numbers of strata and stratum sizes, or some combination thereof, along with varied proportions of treated units across strata. Theorem 1 also allows us to compare the asymptotic efficiency of $\hat{\tau}$ under the completely randomized and stratified randomized survey experiments. In the completely randomized survey experiment, the commonly-used average treatment effect estimator is the standard difference-in-means estimator $\hat{\tau}_{\mathrm{C}} = (1/n_1)\sum_{i=1}^{N} Z_i T_i Y_i - (1/n_0)\sum_{i=1}^{N} Z_i(1-T_i)Y_i$. To make a fair comparison and guarantee that $\hat{\tau} = \hat{\tau}_{\mathrm{C}}$, we consider a scenario with equal sampling and treated proportions across all strata, i.e., $e_{[k]1} = e_1$ and $f_{[k]} = f$. This uniformity ensures that each unit has the same probability to be sampled and the same probability to be assigned to the treatment as a non-stratification strategy. Consequently, the improvement in estimation efficiency solely arises from the division of units into distinct strata. Let $V_{\tau\tau,\mathrm{C}}$ denote the variance of $\hat{\tau}_{\mathrm{C}}$ under the completely randomized survey experiment. We can deduce Corollary 1 below.

Corollary 1. *Under Condition 1, if $e_{[k]1} = e_1$ and $f_{[k]} = f$, for $k = 1, \ldots, K_N$, we have $V_{\tau\tau,\mathrm{C}} - V_{\tau\tau} = \frac{N}{N-1}\sum_{k=1}^{K_N} \Pi_{[k]} d_{[k]} - \frac{1}{N-1}\sum_{k=1}^{K_N}(1-\Pi_{[k]})V_{[k]\tau\tau}$, where $d_{[k]} = [(e_0/e_1)^{1/2}\{\bar{Y}_{[k]}(1) - \bar{Y}(1)\} + (e_1/e_0)^{1/2}\{\bar{Y}_{[k]}(0) - \bar{Y}(0)\}]^2 + (1-f)(\tau_{[k]} - \tau)^2$.*

Remark 2. When $f = 1$, Corollary 1 represents the comparison of the asymptotic variance of $\hat{\tau}$ under the one-stage completely randomized experiment and stratified randomized experiment. We focus on the case that $f < 1$. The first term in the difference of the asymptotic variance, $N(N-1)^{-1}\sum_{k=1}^{K_N} \Pi_{[k]} d_{[k]} \geq 0$, measures a weighted between-strata variation, which equals to zero if and only if $\bar{Y}_{[k]}(t) = \bar{Y}(t)$ for all $k = 1, \ldots, K_N$ and $t = 0, 1$. It corresponds to the case that the stratification variable is not predictive at all. Otherwise, this term and its limit are usually larger than zero. The second term, $(N-1)^{-1}\sum_{k=1}^{K_N}(1-\Pi_{[k]})V_{[k]\tau\tau}$, measures a weighted within-strata variation. Condition 1 implies that $(N-1)^{-1}\sum_{k=1}^{K_N}(1-\Pi_{[k]})V_{[k]\tau\tau} \leq (4L/c_1)(K_N-1)/(N-1)$, which tends to zero if $K_N/N \to 0$. Thus, stratification improves the asymptotic efficiency as long as the number of strata is not too large. If $K_N/N \nrightarrow 0$, the second term may not be negligible. However, the between-strata variation is usually larger than the within-strata variation. Hence, stratification can lead to a discernible boost in efficiency for most cases.

In the stratified randomized survey experiment, it is crucial to determine the optimal stratum-specific sampling proportion $f_{[k]}$ and treated proportion $e_{[k]1}$, which can be obtained via minimizing the asymptotic variance of $\hat{\tau}$. Theorem 2 below solves this problem.

Theorem 2. *The asymptotic variance of $\hat{\tau}$ under the stratified randomized survey experiment is minimized if (i) for given treated proportion $e_{[k]1}$ and total sampling proportion f, we have $f_{[k]}/f = \sqrt{e_{[k]1}^{-1}S_{[k]1}^2 + e_{[k]0}^{-1}S_{[k]0}^2}/(\sum_{k'=1}^{K_N}\Pi_{[k']}\sqrt{e_{[k']1}^{-1}S_{[k']1}^2 + e_{[k']0}^{-1}S_{[k']0}^2})$, or (ii) for a given sampling proportion $f_{[k]}$, we have $e_{[k]1} = \sqrt{S_{[k]1}^2}/(\sqrt{S_{[k]1}^2}+\sqrt{S_{[k]0}^2})$.*

If the potential outcomes are homogeneous across strata such that $e_{[k]1}^{-1}S_{[k]1}^2 + e_{[k]0}^{-1}S_{[k]0}^2$ is a constant across k, then the optimal $f_{[k]} = f$ and $\pi_{[k]} = \Pi_{[k]}$. That is, the proportionate stratified sampling is optimal. Moreover, if the stratum-specific variances of $Y_i(1)$ and $Y_i(0)$ are the same, then the balanced design (i.e., $e_{[k]1} = 0.5$) is optimal. Generally, if we can estimate the stratum-specific variance of the potential outcomes $Y_i(1)$ and $Y_i(0)$ by domain knowledge or prior studies, then we can choose the stratum-specific sampling proportion and treated proportion by Theorem 2 to achieve the smallest asymptotic variance of $\hat{\tau}$. Cytrynbaum (2023) discussed the optimal stratification methods under budget constraints in survey experiments, but they considered the super-population framework.

3.2 Asymptotic properties under SRSRR

The asymptotic distribution of $\hat{\tau}$ under the SRSRR experiment depends on the squared multiple correlations between $\hat{\tau}$ and $\hat{\delta}_W$ and that between $\hat{\tau}$ and $\hat{\tau}_X$ defined by $R_W^2 = (V_{\tau W} V_{WW}^{-1} V_{W\tau})/V_{\tau\tau}$ and $R_X^2 = (V_{\tau X} V_{XX}^{-1} V_{X\tau})/V_{\tau\tau}$. Under Condition 1, R_W^2 and R_X^2 have limiting values and we will still use the same notation to denote their limiting values for notation simplicity when no confusion would arise.

Theorem 1 gives an asymptotic theory under the stratified randomized survey experiment conducted without rejective sampling and rerandomization. When these two techniques apply in the experiment, the asymptotic distribution of $\sqrt{n}(\hat{\tau} - \tau)$ will no longer be normal generally. Instead, it becomes a convolution of a normal distribution and two truncated normal distributions; see Theorem 3 below for the details. Let ε be a standard normal random variable and $L_{J,a} \sim D_1 \mid D^{\mathrm{T}} D \leq a$ be a truncated normal random variable with $D = (D_1, D_2, ..., D_J)^{\mathrm{T}} \sim \mathcal{N}(0, I_J)$. As shown by Li et al. (2018), $\mathrm{var}(L_{J,a}) = \nu_{J,a} = \mathrm{pr}(\chi^2_{J+2} \leq a)/\mathrm{pr}(\chi^2_J \leq a) \in (0, 1)$, where χ^2_J

is a chi-squared distribution with J degrees of freedom. Moreover, $L_{J,a}$ is unimodal, symmetric, and more concentrated at zero than the normal random variable with the same variance. Theorem 3 below describes the asymptotic distribution of $\hat{\tau}$ under the SRSRR experiment.

Theorem 3. *Under Condition 1 and the SRSRR experiment, we have*

$$\sqrt{n}\{\hat{\tau} - \tau\} \xrightarrow{d} V_{\tau\tau}^{1/2}\Big\{\sqrt{1 - R_W^2 - R_X^2} \cdot \varepsilon + \sqrt{R_W^2} \cdot L_{J_1,a_S} + \sqrt{R_X^2} \cdot L_{J_2,a_T}\Big\},$$

where ε, L_{J_1,a_S}, *and* L_{J_2,a_T} *are independent.*

Remark 3. Another approach is to reject sampling and treatment assignment until both $M_S \leq a_S$ and $M_T \leq a_T$ are satisfied. This method needs more computational cost but is asymptotically equivalent to our method; see Section S5 in the Supplementary Material. Wang and Li (2022) considered an asymptotic regime with a diminishing threshold and diverging covariate dimension. Extending their analysis, when $\log(\mathrm{pr}(M_S \leq a_S)^{-1})/J_1 \to \infty$, the truncated normal term L_{J_1,a_S} is asymptotically negligible. A similar result applies to L_{J_2,a_T}.

Theorem 3 implies that the asymptotic distribution of $\hat{\tau}$ is no longer normal, but by Li et al. (2018), this limiting distribution is still unimodal and symmetric around zero. Moreover, SRSRR improves the efficiency compared to the standard stratified randomized survey experiment. The results are summarized in Corollary 2 below.

Corollary 2. *Under Condition 1, the percentage reduction in the asymptotic variance of $\hat{\tau}$ under the SRSRR experiment compared to that under the stratified randomized survey experiment is* $[100 \times \{(1-\nu_{J_1,a_S})R_W^2 + (1-\nu_{J_2,a_T})R_X^2\}]\%$. *Moreover, the asymptotic distribution of $\hat{\tau}$ under the SRSRR experiment is more concentrated at τ than that under the stratified randomized survey experiment.*

Similar to Theorem 2, we can determine the optimal stratum-specific sampling proportion $f_{[k]}$ and treated proportion $e_{[k]1}$ to minimize the asymptotic variance of $\hat{\tau}$ under the SRSRR experiment. However, since the squared multiple correlations R_W^2 and R_X^2 depend on $f_{[k]}$ and $e_{[k]1}$, the optimizer does not have a closed-form solution. In the Supplementary Material, we provide an algorithm to approximate the optimal $f_{[k]}$ and $e_{[k]1}$.

3.3 Conservative variance and distribution estimators

To infer the average treatment effect under the SRSRR experiment, we need to estimate the asymptotic distribution of $\hat{\tau}$. Intuitively, we can use the sample variance and covariance to estimate the corresponding population quantities. Specifically, let $s_{[k]1}^2 = (n_{[k]1}-1)^{-1}\sum_{i\in[k]} Z_iT_i(Y_i-\bar{Y}_{[k]1})^2$ and $s_{[k]0}^2 = (n_{[k]0}-1)^{-1}\sum_{i\in[k]} Z_i(1-T_i)(Y_i-\bar{Y}_{[k]0})^2$ be the stratum-specific variances of $Y_i(1)$ and $Y_i(0)$ under the treatment and control, respectively. Denote $s_{[k]X,1} = (n_{[k]1}-1)^{-1}\sum_{i\in[k]} Z_iT_i(X_i-\bar{X}_{[k]1})(Y_i-\bar{Y}_{[k]1})$ and $s_{[k]X,0} = (n_{[k]0}-1)^{-1}\sum_{i\in[k]} Z_i(1-T_i)(X_i-\bar{X}_{[k]0})(Y_i-\bar{Y}_{[k]0})$ as the stratum-specific covariances of $Y_i(1)$ and $Y_i(0)$ with X_i under treatment and control, respectively. Similarly, we define $s_{[k]W,1} = (n_{[k]1} - 1)^{-1} \sum_{i\in[k]} Z_i T_i (W_i - \bar{W}_{[k]1})(Y_i - \bar{Y}_{[k]1})$ and $s_{[k]W,0} = (n_{[k]0} - 1)^{-1} \sum_{i\in[k]} Z_i (1 - T_i)(W_i - \bar{W}_{[k]0})(Y_i - \bar{Y}_{[k]0})$.

Note that the term $S^2_{[k]\tau}$, i.e., the stratum-specific finite-population variance of τ_i, is not estimable because we cannot observe $Y_i(1)$ and $Y_i(0)$ simultaneously for any unit. Thus, we need to estimate $V_{\tau\tau}$ in Theorem 3 by a conservative estimator: $\hat{V}_{\tau\tau} = \sum_{k=1}^{K_N} \Pi^2_{[k]} \pi^{-1}_{[k]} (e^{-1}_{[k]1} s^2_{[k]1} + e^{-1}_{[k]0} s^2_{[k]0})$. The covariate W_i is often available for all units by preliminary surveys. Thus, we do not need to estimate V_{WW}. In contrast, the covariate X_i is often available only for the sampled units. Accordingly, we can estimate V_{XX} by $\hat{V}_{XX} = \sum_{k=1}^{K_N} \Pi^2_{[k]} \pi^{-1}_{[k]} (e_{[k]1} e_{[k]0})^{-1} S^2_{[k]X|\mathcal{S}}$. The other elements of V can be estimated by $\hat{V}_{X\tau} = \hat{V}^{\mathrm{T}}_{\tau X} = \sum_{k=1}^{K_N} \Pi^2_{[k]} \pi^{-1}_{[k]} (e^{-1}_{[k]1} s_{[k]X,1} + e^{-1}_{[k]0} s_{[k]X,0})$ and $\hat{V}_{W\tau} = \hat{V}^{\mathrm{T}}_{\tau W} = \sum_{k=1}^{K_N} \Pi^2_{[k]} \pi^{-1}_{[k]} (1 - f_{[k]})(s_{[k]W,1} - s_{[k]W,0})$. The squared multiple correlations can also be estimated by a plug-in method $\hat{R}^2_W = \hat{V}_{\tau W} V^{-1}_{WW} \hat{V}_{W\tau} / \hat{V}_{\tau\tau}$ and $\hat{R}^2_X = \hat{V}_{\tau X} \hat{V}^{-1}_{XX} \hat{V}_{X\tau} / \hat{V}_{\tau\tau}$.

Then, the asymptotic distribution in Theorem 3 can be conservatively estimated by $\hat{V}^{1/2}_{\tau\tau} \{ \sqrt{1 - \hat{R}^2_W - \hat{R}^2_X} \cdot \varepsilon + \sqrt{\hat{R}^2_W} \cdot L_{J_1, a_S} + \sqrt{\hat{R}^2_X} \cdot L_{J_2, a_T} \}$. Let $\nu_\xi(\hat{V}_{\tau\tau}, \hat{R}^2_W, \hat{R}^2_X)$ be the ξ-th quantile of the above distribution. For a given $\alpha \in (0, 1)$, we have the following theorem.

Theorem 4. *Under Condition 1 and the SRSRR experiment,* $\hat{V}_{\tau\tau} - V_{\tau\tau} = \sum_{k=1}^{K_N} \Pi^2_{[k]} \pi^{-1}_{[k]} f_{[k]} S^2_{[k]\tau} + o_p(1)$, *and* $\left[\hat{\tau} - n^{-1/2} \nu_{1-\alpha/2}(\hat{V}_{\tau\tau}, \hat{R}^2_W, \hat{R}^2_X),\ \hat{\tau} + n^{-1/2} \nu_{1-\alpha/2}(\hat{V}_{\tau\tau}, \hat{R}^2_W, \hat{R}^2_X) \right]$

is an asymptotic conservative $1-\alpha$ *confidence interval for* τ*.*

Remark 4. For finely stratified randomized experiments, $n_{[k]1} = 1$ or $n_{[k]0} = 1$. In this case, we can use the variance estimation method proposed by Pashley and Miratrix (2021) and Wang and Li (2022).

Remark 5. As an alternative variance estimator, one may use the causal bootstrap of Imbens and Menzel (2021) and its stratified extension in Yu et al. (2025). Conceptually, the procedure first constructs a finite-population "science table" by imputing the missing potential outcomes via rank-preserving imputation, and then repeatedly replays the sampling and treatment-assignment mechanisms to generate bootstrap replicates $\hat{\tau}^{*1}, \ldots, \hat{\tau}^{*B}$. The variance $V_{\tau\tau}$ can be estimated by the empirical variance of $\{\sqrt{n}\hat{\tau}^{*b}\}_{b=1}^{B}$, which may have better finite-sample performance than the Neyman-type conservative variance estimator $\hat{V}_{\tau\tau}$.

Theorem 4 ensures that the covariance estimator is asymptotically conservative; it is consistent if and only if the unit level treatment effect is constant within each stratum, i.e., $S^2_{[k]\tau} = 0$ for $k = 1, \ldots, K_N$.

4. Regression adjustment

Regression or covariate adjustment is widely used in the analysis stage of randomized experiments to improve the estimation efficiency (see, e.g., Lin, 2013; Bloniarz et al., 2016; Liu and Yang, 2020; Zhu et al., 2025; Liu et al., 2024; Zhao and Ding, 2022).

Rerandomization and regression adjustment can be combined to further improve the efficiency (Li and Ding, 2020; Wang et al., 2023; Lu et al., 2023). In this section, we propose a covariate adjustment method under the SRSRR experiment.

In the analysis stage, more covariates may be collected for both the population and the sampled units. Assume that we have collected covariates $E_i \in \mathbb{R}^{J_3}$ for all N units and $C_i \in \mathbb{R}^{J_4}$ for the sampled units in the analysis stage. As shown by Li et al. (2018) and Yang et al. (2023), if the designer and analyzer do not communicate well such that some of the covariates used in the design stage are not used in the analysis stage, the covariate-adjusted estimator may degrade the efficiency. Therefore, to avoid the efficiency loss, it is reasonable to assume that more covariates are used in the analysis stage, i.e., $E_i \supset W_i$ and $C_i \supset X_i$.

Substituting W_i (or X_i) with E_i (or C_i), we can further obtain quantities related to E_i (or C_i), such as $\hat{\delta}_E$, V_{EE}, and R_E^2. As $\hat{\delta}_E$ and $\hat{\tau}_C$ characterize the covariate imbalance in the sampling and treatment assignment stages, we consider the linearly adjusted estimator $\hat{\tau}_{\text{adj}} = \hat{\tau} - \beta^{\mathrm{T}}\hat{\tau}_C - \gamma^{\mathrm{T}}\hat{\delta}_E$, where β and γ are vectors adjusting for the remaining covariate imbalance. The adjusted vectors corresponding to the most efficient (i.e., smallest asymptotic variance) linearly adjusted estimator are given by $(\beta_{\text{opt}}, \gamma_{\text{opt}}) = \arg\min_{\beta,\gamma} E(\hat{\tau} - \beta^{\mathrm{T}}\hat{\tau}_C - \gamma^{\mathrm{T}}\hat{\delta}_E - \tau)^2$. Recall that V_{CC}, $V_{C\tau}$, V_{EE}, and $V_{E\tau}$ are defined analogously to V_{XX}, $V_{X\tau}$, V_{WW}, and $V_{W\tau}$. By Proposition S1, we have $\beta_{\text{opt}} = V_{CC}^{-1}V_{C\tau}$ and $\gamma_{\text{opt}} = V_{EE}^{-1}V_{E\tau}$. In practice, β_{opt} and γ_{opt} can be consistently estimated by

$$\hat{\beta} = \hat{V}_{CC}^{-1}\sum_{k=1}^{K_N}\Pi_{[k]}^2\pi_{[k]}^{-1}(e_{[k]1}^{-1}s_{[k]C,1}+e_{[k]0}^{-1}s_{[k]C,0}),$$
$$\hat{\gamma} = V_{EE}^{-1}\sum_{k=1}^{K_N}\Pi_{[k]}^2\pi_{[k]}^{-1}(1-f_{[k]})(s_{[k]E,1}-s_{[k]E,0}),$$

where $\hat{V}_{CC}=\sum_{k=1}^{K_N}\Pi_{[k]}^2\pi_{[k]}^{-1}(e_{[k]1}e_{[k]0})^{-1}S_{[k]C|\mathcal{S}}^2$. Then, we obtain a linearly adjusted estimator $\hat{\tau}_{\text{adj}}=\hat{\tau}-\hat{\beta}^{\mathrm{T}}\hat{\tau}_C-\hat{\gamma}^{\mathrm{T}}\hat{\delta}_E$. Next, we demonstrate that $\hat{\tau}_{\text{adj}}$ offers further efficiency improvement compared to $\hat{\tau}$ in both the stratified randomized survey experiment and the SRSRR experiment. For this purpose, we require a regularity Condition 2 on E_i and C_i below, which is analogous to Condition 1 on W_i and X_i.

Condition 2. As $n\to\infty$, (i) the following finite limits exist for $t=0,1$: $\sum_{k=1}^{K_N}\Pi_{[k]}^2\pi_{[k]}^{-1}e_{[k]t}^{-1}S_{[k]C,t}$, $\sum_{k=1}^{K_N}\Pi_{[k]}^2\pi_{[k]}^{-1}(1-f_{[k]})S_{[k]E,t}$, $\sum_{k=1}^{K_N}\Pi_{[k]}^2\pi_{[k]}^{-1}(e_{[k]1}e_{[k]0})^{-1}S_{[k]C}^2$, and $\sum_{k=1}^{K_N}\Pi_{[k]}^2\pi_{[k]}^{-1}(1-f_{[k]})S_{[k]E}^2$ and the limits of V_{CC} and V_{EE} are invertible; (ii) there exists a constant $L>0$ independent of N, such that $\{C_i\}_{i=1}^N$, $\{E_i\}_{i=1}^N\in\mathcal{M}_L$.

Theorem 5. *Under Condition 1–2 and the SRSRR experiment (or the stratified randomized survey experiment), we have* $\sqrt{n}(\hat{\tau}_{\text{adj}}-\tau)\xrightarrow{d}\mathcal{N}(0,(1-R_E^2-R_C^2)V_{\tau\tau})$. *Moreover,* $\hat{\tau}_{\text{adj}}$ *has the smallest asymptotic variance among the class of linearly adjusted estimators* $\{\hat{\tau}-\beta^{\mathrm{T}}\hat{\tau}_C-\gamma^{\mathrm{T}}\hat{\delta}_E,\ \beta\in\mathbb{R}^{J_4},\ \gamma\in\mathbb{R}^{J_3}\}$.

Theorem 5 implies that the asymptotic distribution of $\hat{\tau}_{\text{adj}}$ under the stratified randomized survey experiment ($a_S=a_T=\infty$) and SRSRR experiment are both normal and it is optimal in the class of linearly adjusted estimators if we use more covariate information at the analysis stage. As $\hat{\tau}$ is also a linearly adjusted estimator with $\beta = 0$ and $\gamma = 0$, $\hat{\tau}_{\text{adj}}$ is asymptotically more efficient (no larger asymptotic variance) than the unadjusted estimator $\hat{\tau}$ under the stratified randomized survey experiment. Under the SRSRR experiment, the normal component in the asymptotic distribution of $\hat{\tau}$ is $\mathcal{N}(0, (1 - R_W^2 - R_X^2)V_{\tau\tau})$. Since $E_i \supset W_i$ and $C_i \supset X_i$, we have $R_W^2 \leq R_E^2$ and $R_X^2 \leq R_C^2$. Accordingly, $\hat{\tau}_{\text{adj}}$ is asymptotically more efficient than $\hat{\tau}$ under the SRSRR experiment.

To estimate the asymptotic variance of $\hat{\tau}_{\text{adj}}$, recall that $V_{\tau\tau}$ can be conservatively estimated by $\hat{V}_{\tau\tau} = \sum_{k=1}^{K_N} \Pi_{[k]}^2 \pi_{[k]}^{-1} (e_{[k]1}^{-1} s_{[k]1}^2 + e_{[k]0}^{-1} s_{[k]0}^2)$. Thus, we only need to estimate R_E^2 and R_C^2. By plugging-in, their estimators are $\hat{R}_E^2 = \hat{V}_{\tau E} V_{EE}^{-1} \hat{V}_{E\tau} / \hat{V}_{\tau\tau}$ and $\hat{R}_C^2 = \hat{V}_{\tau C} \hat{V}_{CC}^{-1} \hat{V}_{C\tau} / \hat{V}_{\tau\tau}$. Denote q_ξ as the ξ-th quantile of a standard normal distribution. Theorem 6 below justifies the Wald type inference of τ based on $\hat{\tau}_{\text{adj}}$ under the stratified randomized survey experiment and SRSRR experiment, respectively.

Theorem 6. *Under Condition 1–2 and the SRSRR experiment (or the stratified randomized survey experiment), the confidence interval* $[\hat{\tau}_{\text{adj}} - n^{-1/2} q_{1-\alpha/2} \hat{V}_{\tau\tau}^{1/2} \sqrt{1 - \hat{R}_E^2 - \hat{R}_C^2},\ \hat{\tau}_{\text{adj}} + n^{-1/2} q_{1-\alpha/2} \hat{V}_{\tau\tau}^{1/2} \sqrt{1 - \hat{R}_E^2 - \hat{R}_C^2}]$ *has an asymptotic coverage rate greater than or equal to* $1 - \alpha$*, and it is asymptotically shorter than, or at least as short as, the confidence interval based on* $\hat{\tau}$ *in Theorem 4.*

Theorem 6 implies that the confidence interval based on $\hat{\tau}_{\text{adj}}$ is asymptotically conservative and $\hat{\tau}_{\text{adj}}$ improves the inference efficiency compared to $\hat{\tau}$. This conclusion holds for either the stratified randomized survey experiment ($a_S = a_T = \infty$) or SRSRR experiment.

5. Numerical studies

5.1 Synthetic data

In this section, we conduct a simulation study to evaluate the finite-sample performance of $\hat{\tau}$ and $\hat{\tau}_{\text{adj}}$ under the stratified randomized survey experiment (SRSE) and SRSRR experiment. We consider three scenarios for the number of strata K_N and stratum size $N_{[k]}$: (**Case 1**) small strata only with $K_N = 200$ and each stratum of size $N_{[k]} = 200$ for $1 \leq k \leq K_N$; (**Case 2**) small strata together with two large strata, i.e., $N_{[k]} = 200$ for $1 \leq k \leq K_N - 2$ and $N_{[k]} = 2000$ for $K_N - 1 \leq k \leq K_N$ with $K_N = 22$, and (**Case 3**) two large strata with $K_N = 2$ and $N_{[k]} = 4000$ for $1 \leq k \leq K_N$.

We independently generate the covariate $C_i = (C_{i1}, C_{i2}, C_{i3}, C_{i4})^{\mathrm{T}} \sim \mathcal{N}(0, \Sigma)$ for $1 \leq i \leq N$, where $\Sigma = (\sigma_{jk})$ with $\sigma_{jk} = 0.5^{|j-k|}$, $1 \leq j, k \leq 4$. In the design stage, let $W_i = C_{i1}$ and $X_i = (C_{i1}, C_{i2})^{\mathrm{T}}$. While in the analysis stage, we set $E_i = (C_{i1}, C_{i2}, C_{i3})^{\mathrm{T}}$. The potential outcome is generated from the following random effect model:

$$Y_i(t) = \sum_{j=1}^{4} C_{ij}\beta_{t1,j} + \exp\left(\sum_{j=1}^{4} C_{ij}\beta_{t2,j}\right) + D_{[k]} + \varepsilon_i(t), \quad t \in \{0, 1\}, \quad i \in [k].$$

For $1 \leq j \leq 4$, we generate $\beta_{11,j} \sim t_3$, $\beta_{12,j} \sim 0.1t_3$, $\beta_{01,j} \sim \beta_{11,j} + t_3$ and $\beta_{02,j} \sim$

$\beta_{12,j} + 0.1t_3$, where t_3 represents the t distribution with three degrees of freedom. We further generate $D_{[k]} \sim t_3$ for $1 \le k \le K_N$ and $\varepsilon_i(t) \sim \mathcal{N}(0, \sigma^2)$ for $t = 0, 1$. We choose σ^2 such that the signal-to-noise ratio is fixed at 10. In Cases 1 and 2, we consider homogeneous strata with $(\beta_{t1,j}, \beta_{t2,j})$ being the same across all strata. In Case 3, we consider two large heterogeneous strata and the coefficients $(\beta_{t1,j}, \beta_{t2,j})$'s are generated independently for each stratum.

The covariates and potential outcomes are generated once and then kept fixed. We repeat the sampling and treatment assignment procedure 10^3 times to approximate the distributions of $\hat{\tau}$ and $\hat{\tau}_{\text{adj}}$ and compute the bias, standard deviation (SD), root mean squared error (RMSE), empirical coverage probability (CP), and mean confidence interval length (Length) of 95% confidence intervals.

We consider the proportionate stratified sampling with $\Pi_{[k]} = \pi_{[k]}$ for $k = 1, \ldots, K_N$ and set $f = n/N = 1/10$. We choose the threshold a_S in the stratified rejective sampling to make the asymptotic accept rate $p_S = \text{pr}(\mathrm{M}_S \le a_S) = 0.01$. In the treatment assignment stage, within each stratum, half of the sampled units are randomly assigned to the treatment group, and the rest are assigned to the control group. We choose the threshold a_T such that the asymptotic accept rate in the stratified rerandomization is equal to $p_T = \text{pr}(\mathrm{M}_T \le a_T) = 0.01$. Next, we set $a_S = \infty$, which means that the stratified rejective sampling is not used, and denote the associated stratified randomized survey experiment with stratified rerandomization only by SRSE-R for short. Similarly, we set $a_T = \infty$, which means that the stratified rerandomization is not used in the experiment, and denote the associated stratified randomized survey experiment with stratified rejective sampling only by SRSE-S for short.

Table 1 shows the results for Case 2, which features many small strata along with two large strata. The results for Cases 1 and 3 are presented in Tables S4 and S5 of the Supplementary Material, and lead to similar conclusions.

First, the bias is negligible across all estimators and designs, so the RMSE is nearly identical to the SD.

Second, all design- and analysis-stage strategies yield substantial efficiency gains. For the unadjusted estimator $\hat{\tau}$, stratification reduces the SD from 0.160 under CRSE to 0.137 under SRSE, a reduction of 14.4%. Building on SRSE, rejective sampling further reduces the SD from 0.137 to 0.134, a modest gain of 2.2%, whereas rerandomization reduces it from 0.137 to 0.119, corresponding to a 13.1% reduction. When stratification, rejective sampling, and rerandomization are combined, the SD further decreases to 0.111 under SRSRR, representing a 19.0% reduction relative to SRSE and a 30.6% reduction relative to the baseline design CRSE. In addition, regression adjustment leads to further efficiency gains across all designs, reducing the SD by about 21% to 49% relative to the corresponding unadjusted estimator.

Third, for the regression-adjusted estimator $\hat{\tau}_{\text{adj}}$, the differences among SRSE,

Table 1: Simulation results for Case 2

Design	Str	Rej-Sam	ReR	Estimator	Bias($\times 10^2$)	SD	RMSE	Length	CP(%)
SRSE	✓			$\hat{\tau}$	0.029	0.137	0.137	0.552	96.0
SRSE-S	✓	✓		$\hat{\tau}$	0.314	0.134	0.134	0.536	95.9
SRSE-R	✓		✓	$\hat{\tau}$	-0.479	0.119	0.119	0.489	96.6
SRSRR	✓	✓	✓	$\hat{\tau}$	-0.074	0.111	0.111	0.466	96.6
SRSE	✓			$\hat{\tau}_{\text{adj}}$	0.039	0.072	0.072	0.300	95.5
SRSE-S	✓	✓		$\hat{\tau}_{\text{adj}}$	0.366	0.069	0.069	0.301	96.6
SRSE-R	✓		✓	$\hat{\tau}_{\text{adj}}$	-0.161	0.071	0.071	0.301	96.1
SRSRR	✓	✓	✓	$\hat{\tau}_{\text{adj}}$	-0.261	0.071	0.071	0.300	95.9
CRSE				$\hat{\tau}$	-0.720	0.160	0.160	0.630	95.5
CRSE-S		✓		$\hat{\tau}$	-1.062	0.149	0.150	0.617	95.7
CRSE-R			✓	$\hat{\tau}$	-0.275	0.148	0.148	0.578	95.1
RRSE		✓	✓	$\hat{\tau}$	-0.137	0.136	0.136	0.560	96.5
CRSE				$\hat{\tau}_{\text{adj}}$	-0.111	0.107	0.107	0.432	96.5
CRSE-S		✓		$\hat{\tau}_{\text{adj}}$	-0.601	0.107	0.107	0.432	95.4
CRSE-R			✓	$\hat{\tau}_{\text{adj}}$	0.665	0.109	0.109	0.432	94.7
RRSE		✓	✓	$\hat{\tau}_{\text{adj}}$	0.030	0.108	0.108	0.432	96.1

Str, stratification; Rej-Sam, rejective sampling; ReR, rerandomization; SRSE, stratified randomized survey experiment; SRSE-S, stratified randomized survey experiment with stratified rejective sampling; SRSE-R, stratified randomized survey experiment with rerandomization; SRSRR, stratified rejective sampling and rerandomized survey experiment; CRSE, completely randomized survey experiment; CRSE-S, completely randomized survey experiment with rejective sampling; CRSE-R, completely randomized survey experiment with rerandomization; RRSE, rejective sampling and rerandomized survey experiment.

SRSE-S, SRSE-R, and SRSRR are relatively small, suggesting that regression adjustment already captures a substantial part of the covariate imbalance. In particular, $\hat{\tau}_{\text{adj}}$ performs similarly under SRSE and SRSRR. However, the supplementary results show that when the sample size becomes smaller, the RMSE of $\hat{\tau}_{\text{adj}}$ under SRSRR can still be appreciably smaller than that under SRSE, by about 17%; see Table S9 in the Supplementary Material. This indicates that design-stage covariate balance may still provide additional gains beyond analysis-stage regression in finite samples.

Finally, the lengths of the confidence intervals exhibit similar patterns when stratification, rejective sampling, rerandomization, or regression adjustment is applied. Moreover, all methods produce confidence intervals with coverage probabilities close to or slightly above the nominal level 95%.

5.2 Cooperative congressional election study data

In this section, we consider the Cooperative Congressional Election Study (CCES) to illustrate the proposed design (SRSRR) and evaluate the performance of the unadjusted and covariate-adjusted average treatment effect estimators. CCES was a national stratified sample survey administered by YouGov (Christenson and Kriner, 2017, CCES). Many teams participated and pooled their studies into the whole survey experiment. We follow Yang et al. (2023) choosing the team survey of Boston University from CCES 2014 (Kriner, 2017; Kuriwaki, 2022). Participants were asked about their views on federal spending on scientific research. Researchers assigned about 1000 participants to two groups to ask their opinions on federal scientific research spending. Those participants assigned to the treatment group were provided with additional information about the current federal research budget: “each year, just over 1% of the federal budget is spent on scientific research”, while the control group was blind to this information. The outcome of interest is whether federal spending on scientific research should be increased, the same, or decreased, coded as 1, 2, and 3. We treat all the CCES 2014 participants as our target population with $N = 49452$. We stratify the population into four strata based on race: White, Black, Hispanic, and Other. In addition to the stratification variable race, eight pretreatment covariates are used at different stages of the SRSRR experiment. Specifically, three covariates are observed at the sampling stage, i.e., $W_i \in \mathbb{R}^3$. In the treatment assignment stage, we use five covariates, i.e., $X_i \in \mathbb{R}^5$ with $W_i \subset X_i$, which means that two additional covariates are observed at the treatment assignment stage. Moreover, all eight covariates are observed at the analysis stage with $C_i \in \mathbb{R}^8$. The covariate names are provided in the Supplementary Material.

Next, we sample $n = 1000$ units with the same proportion across strata to conduct the experiment. Half of the sampled units are assigned to the treatment group and the rest to the control group. Because not all the potential outcomes are observed in the study, we cannot obtain the actual gains of the proposed design and the covariate-adjusted treatment effect estimator. So, we estimate the gains by generating a synthetic dataset with unobserved potential outcomes imputed by a linear model. Specifically, we fit a linear regression model of the observed outcome on the treatment indicator, all covariates, and their interactions, and then impute the unobserved potential outcomes using the fitted linear model. To make the experiment more realistic, we fine-tune the dataset and get three different datasets. Specifically, we first compute the individual treatment effect τ_i, and then shrink the potential outcomes of the control group toward the population means by a factor λ with $\lambda = 0.1, 0.2$ and 0.3, respectively. The potential outcomes under the control are defined as $Y_i^*(0) = \bar{Y}(0) + \lambda\{Y_i(0) - \bar{Y}(0)\}$. Accordingly, we redefine the potential outcomes under the treatment by $Y_i^*(1) = Y_i^*(0) + \tau_i$. The squared multiple correlations of the three generated datasets are presented in Table S3 in the Supplementary Material. The potential outcomes and covariates are fixed in the follow-up analysis. We set a_S and a_T to achieve an asymptotic acceptance probability of $p_S = p_T = 0.001$ for the SRSRR experiment, $a_T = \infty$ for SRSE-S, and $a_S = \infty$ for SRSE-R. We repeat the sampling and treatment assignment stages 1000 times to evaluate the finite-sample performance of $\hat{\tau}$ and $\hat{\tau}_{\text{adj}}$ for each dataset.

The results are summarized in Tables S6–S8 in the Supplementary Material. Across all three datasets, both SRSE-S and SRSE-R improve the efficiency of SRSE and their combination, SRSRR, performs the best among the four stratified designs. The unadjusted estimator $\hat{\tau}$ has the largest RMSE and longest CI length in the third dataset. This is mainly because the third dataset has the smallest R_W^2 and a relatively small R_X^2. Moreover, the covariate-adjusted estimator $\hat{\tau}_{\text{adj}}$ further improves the efficiency compared to the unadjusted estimator $\hat{\tau}$. It reduces the RMSE by $17.0\% - 54.5\%$ and CI length by $18.6\% - 53.6\%$, respectively.

Tables S6–S8 also include the results for rejective sampling and rerandomization without stratification, allowing comparison with the design of Yang et al. (2023). Across all three datasets, the corresponding non-stratified designs generally yield larger SDs, RMSEs, and longer confidence intervals for both $\hat{\tau}$ and $\hat{\tau}_{\text{adj}}$ than the stratified designs, providing further evidence that stratification improves the efficiency of survey experiments.

6. Discussion

The one-stage randomized experiment may lack external validity because of the distinction between the population and the sample. In practice, the survey experiment considered the data collection procedure and can address this issue. However, covariate imbalance often exists in the completely randomized survey experiment. In this work, we proposed a two-stage stratified rejective sampling and rerandomized (i.e., SRSRR) experimental design and further covariate adjustment method to balance covariates and improve efficiency. Additionally, we develop a design-based asymptotic theory for the stratified difference-in-means estimator under the proposed design, allowing heterogeneous strata regimes. Both theoretical and numerical results demonstrate that stratification and rerandomization can improve asymptotic efficiency. Finally, we discuss the optimal stratum-specific sampling proportion and treated proportion to achieve the smallest asymptotic variance. Our theory is purely design-based, and the validity of the resulting estimates and inference procedures does not require the correct specification of the underlying outcome model.

Practically, we recommend a staged workflow: first, stratify units using domain knowledge whenever possible; next, implement rejective sampling and rerandomization when feasible; and finally, apply regression adjustment to further reduce residual covariate imbalance and improve estimation precision.

Supplementary Materials

Supplementary Material includes additional theoretical results, additional simulation results, and proofs of all the theoretical results.

References

Bai, Y., J. P. Romano, and A. M. Shaikh (2022). Inference in experiments with matched pairs. *Journal of the American Statistical Association 117*(540), 1726–1737.

Banerjee, A. V., S. Chassang, and E. Snowberg (2017). Decision theoretic approaches to experiment design and external validity. In *Handbook of Economic Field Experiments*, Volume 1, pp. 141–174. Elsevier.

Bickel, P. J. and D. A. Freedman (1984). Asymptotic normality and the bootstrap in stratified sampling. *The Annals of Statistics 12*, 470–482.

Bilias, Y. (2000). Sequential testing of duration data: The case of the Pennsylvania 'reemployment bonus' experiment. *Journal of Applied Econometrics 15*(6), 575–594.

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

Branson, Z. and S. Shao (2021). Ridge rerandomization: An experimental design strategy in the presence of covariate collinearity. *Journal of Statistical Planning and Inference 211*, 287–314.

Christenson, D. P. and D. L. Kriner (2017). Constitutional qualms or politics as usual? The factors shaping public support for unilateral action. *American Journal of Political Science 61*(2), 335–349.

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

Cytrynbaum, M. (2023). Optimal stratification of survey experiments. *arXiv preprint arXiv:2111.08157*.

Dancey, L. and G. Sheagley (2018). Partisanship and perceptions of party-line voting in congress. *Political Research Quarterly 71*(1), 32–45.

Fisher, R. A. (1926). The arrangement of field experiments. *Journal of Ministry of Agriculture of Great Britain 33*, 503–513.

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

Fuller, W. A. (2009). Some design properties of a rejective sampling procedure. *Biometrika 96*(4), 933–944.

Gaines, B. J., J. H. Kuklinski, and P. J. Quirk (2007). The logic of the survey experiment reexamined.

Political Analysis 15(1), 1–20.

Imbens, G. and K. Menzel (2021). A causal bootstrap. *The Annals of Statistics 49*(3), 1460–1488.

Imbens, G. W. and T. Lancaster (1996). Efficient estimation and stratified sampling. *Journal of Econometrics 74*(2), 289–318.

Imbens, G. W. and D. B. Rubin (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press.

Kriner, D. (2017). CCES 2014, Team Module of Boston University (BOS).

Kuriwaki, S. (2022). Cumulative CCES Common Content.

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

Li, X. and P. Ding (2020). Rerandomization and regression adjustment. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 82*(1), 241–268.

Li, X., P. Ding, and D. B. Rubin (2018). Asymptotic theory of rerandomization in treatment-control experiments. *Proceedings of the National Academy of Sciences of the United States of America 115*(37), 9157–9162.

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

Lin, Y., M. Zhu, and Z. Su (2015). The pursuit of balance: an overview of covariate-adaptive randomization techniques in clinical trials. *Contemporary Clinical Trials 45*, 21–25.

Liu, H., J. Ren, and Y. Yang (2024). Randomization-based joint central limit theorem and efficient covariate adjustment in randomized block 2^K factorial experiments. *Journal of the American Statistical Association 119*(545), 136–150.

Liu, H. and Y. Yang (2020). Regression-adjusted average treatment effect estimators in stratified randomized experiments. *Biometrika 107*, 935–948.

Lu, X., T. Liu, H. Liu, and P. Ding (2023). Design-based theory for cluster rerandomization. *Biometrika 110*(2), 431–447.

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*(512), 1412–1421.

Mutz, D. C. (2011). *Population-based Survey Experiments*. Princeton, NJ, USA: Princeton University Press.

Pashley, N. E. and L. W. Miratrix (2021). Insights on variance estimation for blocked and matched pairs designs. *Journal of Educational and Behavioral Statistics 46*(3), 271–296.

Schindl, K. and Z. Branson (2024). A unified framework for rerandomization using quadratic forms. *arXiv preprint arXiv:2403.12815*.

Schultzberg, M. and P. Johansson (2022). Re-randomization: A complement or substitute for stratification in randomized experiments? *Journal of Statistical Planning and Inference 218*, 43–58.

Splawa-Neyman, J. (1923). On the application of probability theory to agricultural experiments. Essay on principles (with discussion). Section 9. Reprinted. *Statistical Science 5*, 465–472.

Tabord-Meehan, M. (2023). Stratification trees for adaptive randomisation in randomised controlled trials.

Review of Economic Studies 90(5), 2646–2673.

Wang, X., T. Wang, and H. Liu (2023). Rerandomization in stratified randomized experiments. *Journal of the American Statistical Association 118*(542), 1295–1304.

Wang, Y. and X. Li (2022). Rerandomization with diminishing covariate imbalance and diverging number of covariates. *The Annals of Statistics 50*(6), 3439–3465.

Yang, Z., T. Qu, and X. Li (2023). Rejective sampling, rerandomization, and regression adjustment in survey experiments. *Journal of the American Statistical Association 118*(542), 1207–1221.

Yu, H., K. Zhu, and H. Liu (2025). Sharp variance estimator and causal bootstrap in stratified randomized experiments. *Statistics in Medicine 44*(13-14), e70139.

Zhang, H., G. Yin, and D. B. Rubin (2024). PCA rerandomization. *Canadian Journal of Statistics 52*(1), 5–25.

Zhao, A. and P. Ding (2022). Reconciling design-based and model-based causal inferences for split-plot experiments. *The Annals of Statistics 50*, 1170–1192.

Zhao, A. and P. Ding (2024). No star is good news: A unified look at rerandomization based on p-values from covariate balance tests. *Journal of Econometrics 241*(1), 105724.

Zhu, K., H. Liu, and Y. Yang (2025). Design-based theory for lasso adjustment in randomized block experiments and rerandomized experiments. *Journal of Business & Economic Statistics 43*(3), 544–555.

Pengfei Tian

Qiuzhen College, Tsinghua University, Beijing, 100084, China

E-mail: tpf24@mails.tsinghua.edu.cn

Jiyang Ren

Department of Statistics and Data Science, Tsinghua University, Beijing, 100084, China

E-mail: rjy19@mails.tsinghua.edu.cn

Yingying Ma

School of Economics and Management, Beihang University, Beijing, 100191, China

E-mail: mayingying@buaa.edu.cn