Supplementary Material for "Balancing covariates in survey experiments"

Section S1 contains additional theoretical results on the mean and covariance of $\sqrt{n}(\hat{\tau}-\tau, \hat{\tau}_X^{\mathrm{T}}, \hat{\delta}_W^{\mathrm{T}})^{\mathrm{T}}$ under the stratified randomized survey experiment and the optimal sampling and treated proportions under the SRSRR experiment with and without covariate adjustment.

Section S2 provides the names of covariates used in the Cooperative Congressional Election Study.

Section S3 provides tables for numerical results in the main text.

Section S4 provides additional simulation results to show the finite sample advantage for design-stage rerandomization over analysis-stage regression adjustment.

Section S5 shows the equivalence between two stratified rejective sampling and rerandomization procedures.

Section S6 contains theoretical proofs for Theorem 1–6, Corollary 1–2 in the main text, and Proposition S1, Theorem S1–S2 in Section S1, Theorem S3, Corollary S1 in Section S5 in the Supplementary Material.

S1 Additional theoretical results

First, we provide notations in Tables S1 and S2.

S1.1 Mean and covariance of $\sqrt{n}(\hat{\tau}-\tau, \hat{\tau}_X^{\mathrm{T}}, \hat{\delta}_W^{\mathrm{T}})^{\mathrm{T}}$

Recall the definition of Section 3, we establish the first two moments of $\sqrt{n}(\hat{\tau}-\tau, \hat{\tau}_X^{\mathrm{T}}, \hat{\delta}_W^{\mathrm{T}})^{\mathrm{T}}$ as follows.

Proposition S1. *Under the stratified randomized survey experiment,* $\sqrt{n}(\hat{\tau}-\tau, \hat{\tau}_X^{\mathrm{T}}, \hat{\delta}_W^{\mathrm{T}})^{\mathrm{T}}$

Table S1: Notation Table

N, n, K_N	Population size; sample size; strata number
$N_{[k]}, n_{[k]}$	Size of stratum k in the population; sample size in stratum k
$\Pi_{[k]}, \pi_{[k]}$	Population proportion of stratum k; sample proportion of stratum k
$f, f_{[k]}$	Sampling ratio in the population; sampling ratio in the k-th stratum
$n_{[k]1}, n_{[k]0}$	Numbers of treated and control units in stratum k
n_1, n_0	Total numbers of treated and control units: $n_1 = \sum_{k=1}^{K_N} n_{[k]1}$, $n_0 = \sum_{k=1}^{K_N} n_{[k]0}$
$e_{[k]1}, e_{[k]0}$	Stratum-k treatment and control assignment probabilities
$\mathcal{S}$	Index set of sampled units
Z_i, T_i	Sampling indicator of unit i; treatment indicator of unit i
$Y_i(1), Y_i(0), Y_i$	Potential outcomes under treatment and control; observed outcome
$\tau_i, \tau_{[k]}, \tau$	Unit-level effect; stratum-k average effect; overall average effect
$\bar{Y}_{[k]1}, \bar{Y}_{[k]0}$	Stratum-k sample mean outcomes (treatment/control)
$\bar{Y}_1, \bar{Y}_0$	Overall sample mean outcomes (treatment/control)
$\hat{\tau}_{[k]}, \hat{\tau}$	Stratum-specific estimator; unadjusted overall estimator
$W_i \in \mathbb{R}^{J_1}, X_i \in \mathbb{R}^{J_2}$	Covariates at the sampling stage; covariates at the treatment stage
$\bar{W}, \bar{W}_{\mathcal{S}}$	Population mean and weighted sample mean of covariate W
$\bar{W}_{[k]}, \bar{W}_{[k]\mathcal{S}}$	Stratum k mean and stratum k sample mean of covariate W
$\hat{\delta}_W$	Sampling-stage covariate mean difference based on W_i
$\bar{X}_1, \bar{X}_0$	Weighted sample mean of the treated units and control units
$\bar{X}_{[k]1}, \bar{X}_{[k]0}$	Stratum-k sample means of X_i in treatment and control groups
$\hat{\tau}_X$	Covariate mean difference: $\hat{\tau}_X = \bar{X}_t - \bar{X}_c$
$S^2_{[k]W}$	Stratum-k population covariance matrix of W_i
$\bar{X}_{[k]\mathcal{S}}, S^2_{[k]X\mid\mathcal{S}}$	Stratum-k mean vector and covariance matrix of X_i conditional on $\mathcal{S}$
M_S, a_S	Sampling-stage Mahalanobis distance and its acceptance threshold
M_T, a_T	Assignment-stage Mahalanobis distance and its acceptance threshold
q_ξ	ξ-th quantile of the standard normal distribution

Table S2: Notation Table (Continued)

e_1, e_0	Overall assignment proportions: $e_1 = n_1/n$, $e_0 = n_0/n$
$\bar{Y}_{[k]}(t), \bar{X}_{[k]}$	Stratum-k finite-population means of $Y_i(t)$ and X_i
$S^2_{[k]t}, S^2_{[k]\tau}$	Stratum-k finite-population variances of $Y_i(t)$ and τ_i
$S^2_{[k]X}, S_{[k]X,t}$	Stratum-k covariance of X_i and covariance between X_i and $Y_i(t)$
$S_{[k]W,t}, S_{[k]W,\tau}$	Stratum-k covariances between W_i and $Y_i(t)$, and between W_i and τ_i
$V_{[k]}, V$	Stratum-k covariance block matrix and $V = \sum_{k=1}^{K_N} \Pi^2_{[k]} \pi^{-1}_{[k]} V_{[k]}$
R^2_W, R^2_X	Squared multiple correlations: $R^2_W = (V_{\tau W} V^{-1}_{WW} V_{W\tau})/V_{\tau\tau}$, $R^2_X = (V_{\tau X} V^{-1}_{XX} V_{X\tau})/V_{\tau\tau}$
$L_{J,a}$	truncated normal random variable $L_{J,a} \sim D_1 \mid D^{\mathrm{T}} D \le a$ with $D \sim \mathcal{N}(0, I_J)$
$\nu_{J,a}$	variance of truncated normal random variable $\nu_{J,a} = \mathrm{var}(L_{J,a}) = \mathrm{pr}(\chi^2_{J+2} \le a)/\mathrm{pr}(\chi^2_J \le a)$
$E_i \in \mathbb{R}^{J_3}, C_i \in \mathbb{R}^{J_4}$	Analysis-stage covariates: E_i observed for all N units and C_i observed for sampled units
$\hat{\delta}_E$	Sampling-stage covariate mean difference based on E_i (analogous to $\hat{\delta}_W$)
$\hat{\tau}_C$	Covariate mean difference between treatment and control based on C_i (analogous to $\hat{\tau}_X$)
$\hat{\tau}_{\mathrm{adj}}$	Linearly adjusted estimator: $\hat{\tau}_{\mathrm{adj}} = \hat{\tau} - \beta^{\mathrm{T}} \hat{\tau}_C - \gamma^{\mathrm{T}} \hat{\delta}_E$
$\beta_{\mathrm{opt}}, \gamma_{\mathrm{opt}}$	Optimal coefficients: $\beta_{\mathrm{opt}} = V^{-1}_{CC} V_{C\tau}$, $\gamma_{\mathrm{opt}} = V^{-1}_{EE} V_{E\tau}$
$\hat{\beta}, \hat{\gamma}$	Plug-in estimators of β_{opt} and γ_{opt}
V_{CC}, V_{EE}	Asymptotic covariance matrices of $\hat{\tau}_C$ and $\hat{\delta}_E$
$V_{C\tau}, V_{E\tau}$	Cross-covariances between $\hat{\tau}_C$ and $\hat{\tau}$, and between $\hat{\delta}_E$ and $\hat{\tau}$
R^2_C, R^2_E	Squared multiple correlations: $R^2_C = (V_{\tau C} V^{-1}_{CC} V_{C\tau})/V_{\tau\tau}$, $R^2_E = (V_{\tau E} V^{-1}_{EE} V_{E\tau})/V_{\tau\tau}$
$\hat{R}^2_C, \hat{R}^2_E$	Plug-in estimators of R^2_C and R^2_E
$\hat{V}_{CC}$	Plug-in estimator of V_{CC} (e.g., $\hat{V}_{CC} = \sum_{k=1}^{K_N} \Pi^2_{[k]} \pi^{-1}_{[k]} (e_{[k]1} e_{[k]0})^{-1} S^2_{[k]C\mid\mathcal{S}}$)
SRSRR	Stratified rejective sampling and rerandomized survey experiment (Our method)
SRSE	Stratified randomized survey experiment
SRSE-S	SRSE with rejective sampling only ($a_T = \infty$)
SRSE-R	SRSE with rerandomization only ($a_S = \infty$)
CRSE	Completely randomized survey experiment (no stratification)
CRSE-S	CRSE with rejective sampling only
CRSE-R	CRSE with rerandomization only
RRSE	Rejective-sampling and rerandomized survey experiment without stratification

has mean zero and covariance

$$V = \begin{pmatrix} V_{\tau\tau} & V_{\tau X} & V_{\tau W} \\ V_{X\tau} & V_{XX} & V_{XW} \\ V_{W\tau} & V_{WX} & V_{WW} \end{pmatrix} = \sum_{k=1}^{K_N} \Pi_{[k]}^2 \pi_{[k]}^{-1} V_{[k]}.$$

S1.2 Without covariate adjustment

In the SRSRR experiment, it is important to determine the optimal stratum-specific sampling proportion $f_{[k]}$ and treated proportion $e_{[k]1}$, which minimizes the asymptotic variance of $\hat{\tau}$. The below theorem solves this problem.

Theorem S1. *The asymptotic variance of $\hat{\tau}$ under the SRSRR experiment is minimized if (i) for given treated proportion $e_{[k]1}$ and total sampling proportion f, we have*

$$f_{[k]}/f = A_k \Big/ \Big(\sum_{k'=1}^{K_N} \Pi_{[k']} A_{k'} \Big),$$

where

$$\begin{aligned} A_k = &\Big\{ e_{[k]1}^{-1} S_{[k]1}^2 + e_{[k]0}^{-1} S_{[k]0}^2 - 2(1-\nu_{J_1,a_S}) S_{[k]\tau W} V_{WW}^{-1} V_{W\tau} \\ &+ (1-\nu_{J_1,a_S}) V_{\tau W} V_{WW}^{-1} S_{[k]W}^2 V_{WW}^{-1} V_{W\tau} \\ &- 2(1-\nu_{J_2,a_T})(e_{[k]1}^{-1} S_{[k]1,X} + e_{[k]0}^{-1} S_{[k]0,X}) V_{XX}^{-1} V_{X\tau} \\ &+ (1-\nu_{J_2,a_T})(e_{[k]1} e_{[k]0})^{-1} V_{\tau X} V_{XX}^{-1} S_{[k]X}^2 V_{XX}^{-1} V_{X\tau} \Big\}^{1/2}, \end{aligned}$$

and (ii) for given sampling proportion $f_{[k]}$, we have

$$e_{[k]1} = (a_{[k]1})/(a_{[k]1} + a_{[k]0}),$$

where $a_{[k]t} = (|S^2_{[k]t} + (1-\nu_{J_2,a_T})V_{\tau X}V^{-1}_{XX}S^2_{[k]X}V^{-1}_{XX}V_{X\tau} - 2(1-\nu_{J_2,a_T})S_{[k]t,X}V^{-1}_{XX}V_{X\tau}|)^{1/2}$ *for* $t = 0, 1$.

Similar to Theorem 2, we can derive the optimal sampling proportion and treated proportion based on domain knowledge or prior studies. However, the difference lies in that the computation of $V_{\tau\tau}$, $V_{\tau W}$, and V_{WW} relies on the sampling proportion $f_{[k]}$, while $V_{\tau X}$ and V_{XX} depend on the treated proportion $e_{[k]1}$. We can use the following iterative process to obtain the optimal sampling proportion and treated proportion.

We use superscripts $[\cdot]^{(m)}$ to represent the value of $[\cdot]$ in the mth iteration. Based on Theorem S1, we can update $f_{[k]}$ and $e_{[k]1}$ as $f^{(m+1)}_{[k]} = fA^{(m)}_k/(\sum_{k=1}^{K_N}\Pi_{[k]}A^{(m)}_k)$ and $e^{(m+1)}_{[k]t} = a^{(m)}_{[k]t}/(a^{(m)}_{[k]1} + a^{(m)}_{[k]0})$ for $t = 0, 1$. Accordingly, we can iteratively update $A^{(m+1)}_k$, $a^{(m+1)}_{[k]1}$, and $a^{(m+1)}_{[k]0}$ until convergence is achieved. We recommend the selection of initial values of $f^{(0)}_{[k]}$ and $e^{(0)}_{[k]1}$ via the results of Theorem 2.

S1.3 With covariate adjustment

In the SRSRR experiment, after covariate adjustment, we can determine the optimal stratum-specific sampling proportion $f_{[k]}$ and treated proportion $e_{[k]1}$ via minimizing the asymptotic variance of $\hat{\tau}_{\mathrm{adj}}$. The following theorem solves this problem.

Theorem S2. *The asymptotic variance of* $\hat{\tau}_{\mathrm{adj}}$ *under the SRSRR experiment (or the stratified randomized survey experiment) is minimized if (i) for given treated* *proportion* $e_{[k]1}$ *and total sampling proportion* f*, we have*

$$f_{[k]}/f = B_k \Big/ \Big(\sum_{k=1}^{K_N} \Pi_{[k]} B_k \Big),$$

where

$$\begin{aligned} B_k =& \Big\{ e_{[k]1}^{-1} S_{[k]1}^2 + e_{[k]0}^{-1} S_{[k]0}^2 - 2S_{[k]\tau E} V_{EE}^{-1} V_{E\tau} + V_{\tau E} V_{EE}^{-1} S_{[k]E}^2 V_{EE}^{-1} V_{E\tau} \\ & - 2(e_{[k]1}^{-1} S_{[k]1,C} + e_{[k]0}^{-1} S_{[k]0,C}) V_{CC}^{-1} V_{C\tau} + (e_{[k]1} e_{[k]0})^{-1} V_{\tau C} V_{CC}^{-1} S_{[k]C}^2 V_{CC}^{-1} V_{C\tau} \Big\}^{1/2}, \end{aligned}$$

and (ii) for given sampling proportion $f_{[k]}$*, we have*

$$e_{[k]1} = b_{[k]1}/(b_{[k]1} + b_{[k]0}),$$

where $b_{[k]t} = (|S_{[k]t}^2 + V_{\tau C} V_{CC}^{-1} S_{[k]C}^2 V_{CC}^{-1} V_{C\tau} - 2S_{[k]t,C} V_{CC}^{-1} V_{C\tau}|)^{1/2}$ *for* $t = 0, 1$*.*

Similar to Theorem S1, the computation of $V_{\tau\tau}$, $V_{\tau E}$, and V_{EE} relies on the sampling proportion $f_{[k]}$, while $V_{\tau C}$ and V_{CC} depends on the treated proportion $e_{[k]1}$. By Theorem S2, we can update the values of $f_{[k]}$ and $e_{[k]1}$ as follows: $f_{[k]}^{(m+1)} = fB_k^{(m)}/(\sum_{k=1}^{K_N} \Pi_{[k]} B_k^{(m)})$ and $e_{[k]t}^{(m+1)} = b_{[k]t}^{(m)}/(b_{[k]1}^{(m)} + b_{[k]0}^{(m)})$, for $t = 0, 1$. Subsequently, we update $B_k^{(m+1)}$ and $b_{[k]1}^{(m+1)}, b_{[k]0}^{(m+1)}$ using the updated values of $f_{[k]}^{(m+1)}$, $e_{[k]1}^{(m+1)}$, and $e_{[k]0}^{(m+1)}$. This process continues until convergence is achieved. We recommend using Theorem 2 to obtain initial values for $f_{[k]}^{(0)}$ and $e_{[k]1}^{(0)}$.

S2 Covariates used in the Cooperative Congressional Election Study

The race variable is used to form four strata and is not included in the covariate vectors below. The covariates used in different stages are as follows:

1. Sampling stage, W_i: age, gender, whether the highest level of education is college or higher;

2. Treatment assignment stage, additional covariates in X_i: whether family annual income is less than 60000, whether the individual believes the economy has gotten worse last year;

3. Analysis stage, additional covariates in C_i: whether the individual is liberal or moderate, whether party identification is democrat, whether the individual follows the news and public affairs most of the time.

We present the squared multiple correlations of the three generated datasets in Table S3 below.

Table S3: Squared multiple correlations of three generated datasets

shrinkage	$R_W^2 = R_E^2$	R_X^2	R_C^2
0.1	0.384	0.254	0.368
0.2	0.399	0.174	0.342
0.3	0.331	0.195	0.455

S3 Tables for numerical results in the main text

Table S4: Simulation results for Case 1

Design	Str	Rej-Sam	ReR	Estimator	Bias($\times 10^2$)	SD	RMSE	Length	CP(%)
SRSE	✓			$\hat{\tau}$	-0.317	0.137	0.137	0.547	95.4
SRSE-S	✓	✓		$\hat{\tau}$	-0.544	0.132	0.132	0.522	95.2
SRSE-R	✓		✓	$\hat{\tau}$	-0.224	0.107	0.107	0.443	96.1
SRSRR	✓	✓	✓	$\hat{\tau}$	0.306	0.096	0.096	0.411	97.1
SRSE	✓			$\hat{\tau}_{\text{adj}}$	-0.010	0.034	0.034	0.169	98.5
SRSE-S	✓	✓		$\hat{\tau}_{\text{adj}}$	-0.050	0.033	0.033	0.168	98.2
SRSE-R	✓		✓	$\hat{\tau}_{\text{adj}}$	-0.062	0.034	0.034	0.169	98.0
SRSRR	✓	✓	✓	$\hat{\tau}_{\text{adj}}$	0.092	0.034	0.034	0.170	98.1
CRSE				$\hat{\tau}$	-0.483	0.150	0.150	0.592	95.2
CRSE-S		✓		$\hat{\tau}$	-0.437	0.138	0.138	0.570	95.6
CRSE-R			✓	$\hat{\tau}$	-0.359	0.129	0.129	0.497	94.0
RRSE		✓	✓	$\hat{\tau}$	0.179	0.123	0.123	0.470	94.1
CRSE				$\hat{\tau}_{\text{adj}}$	-0.134	0.074	0.074	0.283	94.8
CRSE-S		✓		$\hat{\tau}_{\text{adj}}$	0.041	0.069	0.069	0.284	95.5
CRSE-R			✓	$\hat{\tau}_{\text{adj}}$	-0.273	0.073	0.073	0.284	94.8
RRSE		✓	✓	$\hat{\tau}_{\text{adj}}$	0.093	0.072	0.072	0.283	95.3

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.

Table S5: Simulation results for Case 3

Design	Str	Rej-Sam	ReR	Estimator	Bias($\times 10^2$)	SD	RMSE	Length	CP(%)
SRSE	✓			$\hat{\tau}$	0.250	0.146	0.146	0.608	96.1
SRSE-S	✓	✓		$\hat{\tau}$	0.401	0.136	0.136	0.567	96.3
SRSE-R	✓		✓	$\hat{\tau}$	-0.107	0.135	0.135	0.570	96.4
SRSRR	✓	✓	✓	$\hat{\tau}$	-0.153	0.128	0.128	0.529	94.9
SRSE	✓			$\hat{\tau}_{\mathrm{adj}}$	-0.258	0.110	0.110	0.464	97.1
SRSE-S	✓	✓		$\hat{\tau}_{\mathrm{adj}}$	0.484	0.111	0.111	0.464	96.9
SRSE-R	✓		✓	$\hat{\tau}_{\mathrm{adj}}$	0.349	0.107	0.107	0.464	96.8
SRSRR	✓	✓	✓	$\hat{\tau}_{\mathrm{adj}}$	0.077	0.111	0.111	0.464	95.9
CRSE				$\hat{\tau}$	-0.302	0.183	0.183	0.727	94.9
CRSE-S		✓		$\hat{\tau}$	-0.467	0.185	0.185	0.731	94.7
CRSE-R			✓	$\hat{\tau}$	0.268	0.143	0.143	0.568	95.2
RRSE		✓	✓	$\hat{\tau}$	-0.422	0.147	0.147	0.566	94.6
CRSE				$\hat{\tau}_{\mathrm{adj}}$	-0.072	0.136	0.136	0.544	95.8
CRSE-S		✓		$\hat{\tau}_{\mathrm{adj}}$	-0.222	0.140	0.140	0.543	95.3
CRSE-R			✓	$\hat{\tau}_{\mathrm{adj}}$	0.236	0.138	0.138	0.543	94.9
RRSE		✓	✓	$\hat{\tau}_{\mathrm{adj}}$	-0.516	0.142	0.142	0.543	94.5

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.

Table S6: Results for the CCES data under Scenario 1 (the first dataset)

Design	Str	Rej-Sam	ReR	Estimator	Bias($\times 10^3$)	SD($\times 10^3$)	RMSE($\times 10^3$)	Length($\times 10^3$)	CP(%)
SRSE	✓			$\hat{\tau}$	-0.12	5.22	5.22	20.46	95.6
SRSE-S	✓	✓		$\hat{\tau}$	0.00	3.98	3.98	16.08	96.0
SRSE-R	✓		✓	$\hat{\tau}$	-0.06	4.69	4.68	17.80	94.1
SRSRR	✓	✓	✓	$\hat{\tau}$	-0.00	3.12	3.12	12.55	95.0
SRSE	✓			$\hat{\tau}_{\text{adj}}$	0.15	2.61	2.61	10.23	94.7
SRSE-S	✓	✓		$\hat{\tau}_{\text{adj}}$	0.06	2.64	2.64	10.24	94.3
SRSE-R	✓		✓	$\hat{\tau}_{\text{adj}}$	0.00	2.67	2.67	10.26	94.9
SRSRR	✓	✓	✓	$\hat{\tau}_{\text{adj}}$	0.02	2.59	2.59	10.21	95.2
CRSE				$\hat{\tau}$	0.16	5.80	5.79	22.77	94.1
CRSE-S		✓		$\hat{\tau}$	0.35	4.63	4.65	18.85	96.5
CRSE-R			✓	$\hat{\tau}$	0.13	5.21	5.21	20.38	94.7
RRSE		✓	✓	$\hat{\tau}$	0.09	4.08	4.08	15.93	95.9
CRSE				$\hat{\tau}_{\text{adj}}$	0.14	3.52	3.52	13.88	94.4
CRSE-S		✓		$\hat{\tau}_{\text{adj}}$	0.20	3.48	3.48	13.84	95.2
CRSE-R			✓	$\hat{\tau}_{\text{adj}}$	0.16	3.61	3.61	13.90	95.1
RRSE		✓	✓	$\hat{\tau}_{\text{adj}}$	0.09	3.52	3.52	13.88	96.3

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.

Table S7: Results for the CCES data under Scenario 2 (the second dataset)

Design	Str	Rej-Sam	ReR	Estimator	Bias($\times 10^3$)	SD($\times 10^3$)	RMSE($\times 10^3$)	Length($\times 10^3$)	CP(%)
SRSE	✓			$\hat{\tau}$	-0.16	5.18	5.18	20.06	95.4
SRSE-S	✓	✓		$\hat{\tau}$	0.04	3.90	3.90	15.58	94.7
SRSE-R	✓		✓	$\hat{\tau}$	-0.11	4.81	4.81	18.33	94.6
SRSRR	✓	✓	✓	$\hat{\tau}$	-0.10	3.36	3.36	13.24	95.3
SRSE	✓			$\hat{\tau}_{\text{adj}}$	0.15	2.61	2.61	10.23	94.4
SRSE-S	✓	✓		$\hat{\tau}_{\text{adj}}$	0.06	2.64	2.64	10.24	94.6
SRSE-R	✓		✓	$\hat{\tau}_{\text{adj}}$	-0.00	2.66	2.66	10.26	94.8
SRSRR	✓	✓	✓	$\hat{\tau}_{\text{adj}}$	0.02	2.59	2.59	10.21	95.1
CRSE				$\hat{\tau}$	0.05	5.65	5.64	22.30	94.9
CRSE-S		✓		$\hat{\tau}$	0.32	4.43	4.44	18.28	96.8
CRSE-R			✓	$\hat{\tau}$	0.12	5.28	5.27	20.65	94.2
RRSE		✓	✓	$\hat{\tau}$	0.04	4.13	4.12	16.24	95.7
CRSE				$\hat{\tau}_{\text{adj}}$	0.15	3.46	3.46	13.65	94.2
CRSE-S		✓		$\hat{\tau}_{\text{adj}}$	0.19	3.42	3.43	13.61	95.3
CRSE-R			✓	$\hat{\tau}_{\text{adj}}$	0.15	3.56	3.57	13.67	94.9
RRSE		✓	✓	$\hat{\tau}_{\text{adj}}$	0.10	3.46	3.46	13.65	95.9

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.

Table S8: Results for the CCES data under Scenario 3 (the third dataset)

Design	Str	Rej-Sam	ReR	Estimator	Bias($\times 10^3$)	SD($\times 10^3$)	RMSE($\times 10^3$)	Length($\times 10^3$)	CP(%)
SRSE	✓			$\hat{\tau}$	-0.21	5.74	5.74	22.01	94.7
SRSE-S	✓	✓		$\hat{\tau}$	0.09	4.61	4.61	18.00	94.2
SRSE-R	✓		✓	$\hat{\tau}$	0.17	5.17	5.17	19.87	94.0
SRSRR	✓	✓	✓	$\hat{\tau}$	-0.19	3.97	3.97	15.29	94.7
SRSE	✓			$\hat{\tau}_{\text{adj}}$	0.15	2.61	2.61	10.21	94.2
SRSE-S	✓	✓		$\hat{\tau}_{\text{adj}}$	0.06	2.64	2.64	10.22	94.3
SRSE-R	✓		✓	$\hat{\tau}_{\text{adj}}$	-0.01	2.66	2.66	10.24	94.9
SRSRR	✓	✓	✓	$\hat{\tau}_{\text{adj}}$	0.02	2.59	2.59	10.19	95.5
CRSE				$\hat{\tau}$	-0.05	6.06	6.06	23.97	95.3
CRSE-S		✓		$\hat{\tau}$	0.30	4.98	4.99	20.28	96.5
CRSE-R			✓	$\hat{\tau}$	0.11	5.58	5.58	21.88	94.9
RRSE		✓	✓	$\hat{\tau}$	-0.16	4.48	4.48	17.76	95.6
CRSE				$\hat{\tau}_{\text{adj}}$	0.15	3.42	3.42	13.49	94.0
CRSE-S		✓		$\hat{\tau}_{\text{adj}}$	0.17	3.39	3.39	13.44	95.0
CRSE-R			✓	$\hat{\tau}_{\text{adj}}$	0.15	3.54	3.54	13.51	94.9
RRSE		✓	✓	$\hat{\tau}_{\text{adj}}$	0.11	3.42	3.42	13.48	95.7

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.

S4 Additional simulation results

Although $\hat{\tau}_{\mathrm{adj}}$ has the same asymptotic distribution under SRSE and SRSRR, its performance for a small sample size could be different. In this section, we conduct a simulation to compare the performance of $\hat{\tau}_{\mathrm{adj}}$ under SRSE and SRSRR for a sample size of $n = 80$. Specifically, the potential outcomes and covariates are generated in the same way as those in Case 1 in the main text, except that we set $K_N = 4$ with $N_{[k]} = 40$, $f = 0.5$, $n_{[k]} = 20$, $n_{[k]1} = 10$ for $k = 1, \ldots, K_N$. Table S9 shows the results. We can see that the SD and RMSE of $\hat{\tau}_{\mathrm{adj}}$ under SRSRR are approximately 17% smaller than those under SRSE. Therefore, rejective sampling and rerandomization have advantages when the sample size is small. Furthermore, we report the SD and RMSE of the oracle-adjusted estimator $\hat{\tau}_{\mathrm{ora}} = \hat{\tau} - \beta_{\mathrm{opt}}^{\mathrm{T}} \hat{\tau}_C - \gamma_{\mathrm{opt}}^{\mathrm{T}} \hat{\delta}_E$, whose performances are similar under two designs. A more balanced design usually decreases $\hat{\tau}_C$ and further decreases the gap between $\hat{\tau}_{\mathrm{ora}}$ and $\hat{\tau}_{\mathrm{adj}}$.

Table S9: Simulation results for $\hat{\tau}_{\mathrm{adj}}$ when $n = 80$

Design	Estimator	Bias($\times 10^2$)	SD	RMSE	CI-length	CP(%)
SRSE	$\hat{\tau}_{\mathrm{adj}}$	-2.000	0.350	0.350	2.091	95.7
	$\hat{\tau}_{\mathrm{ora}}$	-0.173	0.203	0.203	2.092	97.5
SRSRR	$\hat{\tau}_{\mathrm{adj}}$	-0.183	0.291	0.291	2.076	96.5
	$\hat{\tau}_{\mathrm{ora}}$	0.061	0.201	0.201	2.076	97.4

SRSE, stratified randomized survey experiment; SRSRR, stratified rejective sampling and rerandomized survey experiment.

S5 Asymptotic equivalence between single-stage and two-stage rerandomized survey experiments

In this section, we extend the result of Yang et al. (2023, B2) to a stratified version. Single-stage stratified rejective sampling and rerandomized experiment refers to stratified survey experiments conducted only when sampling and treatment assignment satisfying $M_S \leq a_S, M_T \leq a_T$ simultaneously, which is denoted by $\widetilde{\mathrm{SRSRR}}$. The two-stage version refers to the one discussed in the main text, which means that we first conduct the sampling if $M_S \leq a_S$ is satisfied and then conduct the treatment assignment if $M_T \leq a_T$ is satisfied. Two-stage stratified rejective sampling and rerandomized survey experiment is denoted as SRSRR.

Denote the acceptance probabilities for the sampling stage and assignment stage as p_S and p_T. The expected numbers of trials until acceptance for single-stage and two-stage SRSRR are $p_S^{-1}p_T^{-1}$ and $p_S^{-1}+p_T^{-1}$, respectively. Two-stage SRSRR decreases the computational cost significantly compared to single-stage. Denote $\mathrm{pr}(\cdot \mid \mathrm{SRSRR})$ and $\mathrm{pr}(\cdot \mid \widetilde{\mathrm{SRSRR}})$ as the conditional probability under SRSRR and $\widetilde{\mathrm{SRSRR}}$, respectively. We will prove the asymptotic equivalence between $\widetilde{\mathrm{SRSRR}}$ and SRSRR. We first introduce the total variation distance between the two designs,

$$
\begin{aligned}
&\mathrm{d}_{\mathrm{TV}}(\widetilde{\mathrm{SRSRR}}, \mathrm{SRSRR}) \\
&= \sup_{\mathcal{A} \subset \{0,1\}^N \times \{0,1\}^n} \left| \mathrm{pr}\big\{(Z, T_{\mathcal{S}}) \in \mathcal{A} \mid \widetilde{\mathrm{SRSRR}}\big\} - \mathrm{pr}\big\{(Z, T_{\mathcal{S}}) \in \mathcal{A} \mid \mathrm{SRSRR}\big\} \right|.
\end{aligned}
$$

Total variation distance measures the difference between the two designs and bounds many quantities of perturbation due to the two designs. Denote

$$\mathcal{M} = \{(Z, T_{\mathcal{S}}) : \text{ satisfying } M_S \le a_S, M_T \le a_T \text{ under SRSE}\}$$

as the set of all possible sampling and treatment assignment $(Z, T_{\mathcal{S}})$ under $\widetilde{\text{SRSRR}}$.

Theorem S3. *The total variation distance between the probability distributions of $(Z, T_{\mathcal{S}})$ under $\widetilde{SRSRR}$ and SRSRR, is bounded by*

$$\begin{aligned}
&d_{\text{TV}}(\widetilde{SRSRR}, SRSRR) \\
&= \sup_{\mathcal{A} \subset \{0,1\}^N \times \{0,1\}^n} \left| \text{pr}\big\{(Z, T_{\mathcal{S}}) \in \mathcal{A} \mid \widetilde{SRSRR}\big\} - \text{pr}\big\{(Z, T_{\mathcal{S}}) \in \mathcal{A} \mid SRSRR\big\} \right| \\
&\le 1(\mathcal{M} = \emptyset) + 1(\mathcal{M} \ne \emptyset) \cdot \frac{E\left|\text{pr}(M_T \le a_T \mid Z) - \text{pr}(M_T \le a_T \mid M_S \le a_S)\right|}{\text{pr}(M_T \le a_T, M_S \le a_S)}.
\end{aligned}$$

Theorem S3 bounds the total variation distance between the two designs. We can prove that the total variation distance tends to zero as shown in Corollary S1 below.

Corollary S1. *Under Condition 1, $d_{\text{TV}}(\widetilde{SRSRR}, SRSRR) \to 0$ as $n \to \infty$.*

The difference of conditional distributions of $\sqrt{n}(\hat{\tau} - \tau)$ or $\sqrt{n}(\hat{\tau}_{\text{adj}} - \tau)$ under SRSRR and $\widetilde{\text{SRSRR}}$ can be bounded by the total variation distance between SRSRR and $\widetilde{\text{SRSRR}}$. By Corollary S1, the difference tends to zero as $n \to \infty$. Therefore, the asymptotic distributions of $\sqrt{n}(\hat{\tau} - \tau)$ or $\sqrt{n}(\hat{\tau}_{\text{adj}} - \tau)$ are the same under the two designs.

S6 Proof of main results

S6.1 Proof of Theorem 1

Li and Ding (2017) have proved a non-stratified finite-population central limit theory (CLT) to establish the design-based theory for the difference-in-means estimator in the one-stage completely randomized experiment. Liu et al. (2024) extended the CLT to the one-stage stratified randomized experiment, assuming that the sampling proportion is one and the treated proportion within each stratum is bounded away from zero and one. Nevertheless, this requirement is unrealistic in the two-stage survey experiment. It is sometimes reasonable to allow the sampling proportion in the first stage close to zero. Hence, we need to handle the zero-limit situation for the sampling proportion in our proof. For this purpose, we will establish a new joint CLT for $\hat{\tau} = \bar{Y}_1 - \bar{Y}_0$, allowing Y_i to be a vector; see Lemma S1 below.

To provide a general result for both scalar and vector potential outcomes, we introduce $R_i(1) \in \mathbb{R}^d$ and $R_i(0) \in \mathbb{R}^d$ as d-dimensional potential outcomes. For example, we can directly set $R_i(1) = Y_i(1)$ and $R_i(0) = Y_i(0)$; we can also set $R_i(1) = (Y_i(1), W_i^{\mathrm{T}}, X_i^{\mathrm{T}})^{\mathrm{T}}$ and $R_i(0) = (Y_i(0), W_i^{\mathrm{T}}, X_i^{\mathrm{T}})^{\mathrm{T}}$. For $t = 0, 1$, define $\bar{R}_{[k]t}$, $\bar{R}_t$, $S^2_{[k]R(t)}$, τ_R, $\hat{\tau}_R$, and $S^2_{[k]\tau(R)}$ similarly to $\bar{Y}_{[k]t}$, $\bar{Y}_t$, $S^2_{[k]t}$, τ, $\hat{\tau}$, and $S^2_{[k]\tau}$, with Y_i replaced by R_i. Similar to the proof in Proposition S1, we can obtain

$$V_R = \mathrm{cov}(\sqrt{n}\hat{\tau}_R) = \mathrm{cov}\{\sqrt{n}(\bar{R}_1 - \bar{R}_0)\} = \sum_{k=1}^{K_N} \Pi_{[k]}^2 \pi_{[k]}^{-1} V_{[k]R},$$

where $V_{[k]R} = e_{[k]1}^{-1} S^2_{[k]R(1)} + e_{[k]0}^{-1} S^2_{[k]R(0)} - f_{[k]} S^2_{[k]\tau_R}$.

Next, for the general d-dimensional potential outcomes $R_i(1)$ and $R_i(0)$, we need Condition S1 below to derive the asymptotic normality of $\sqrt{n}\hat{\tau}_R$.

Condition S1. *As $n \to \infty$, V_R has a finite limit and there exists a constant $L > 0$ independent of N, such that $\{R_i(1)\}_{i=1}^N$, $\{R_i(0)\}_{i=1}^N \in \mathcal{M}_L$.*

This is a condition for the general vector potential outcomes $R_i(1)$ and $R_i(0)$. Actually, when $R_i(1) = Y_i(1)$ and $R_i(0) = Y_i(0)$, this condition is implied by Condition 1(iii)-(iv). We still use V_R to denote its limit when no confusion would arise. Lemma S1 below establishes the asymptotic normality of $\sqrt{n}\hat{\tau}_R$.

Lemma S1. *Under Condition 1(i)–(ii) and Condition S1, we have*

$$\sqrt{n}(\hat{\tau}_R - \tau_R) \xrightarrow{d} \mathcal{N}(0, V_R).$$

Proof of Lemma S1. We prove Lemma S1 by using Hájek's coupling technique for obtaining the vector-form of the Wald–Wolfowitz–Hoeffding theorem for a linear (or bi-linear) rank statistic (Hájek, 1961).

It suffices for Lemma S1 to show that for any fixed vector $u \in \mathbb{R}^d$,

$$\sqrt{n}u^{\mathrm{T}}\{\bar{R}_1 - \bar{R}(1)\} - \sqrt{n}u^{\mathrm{T}}\{\bar{R}_0 - \bar{R}(0)\} \xrightarrow{d} \mathcal{N}(0, u^{\mathrm{T}}V_R u).$$

If $u^{\mathrm{T}}V_R u = 0$, the conclusion holds trivially. We then consider the case $u^{\mathrm{T}}V_R u > 0$. Let $u_1 = u$ and $u_0 = -u$. Denote $H_n = \sum_{q=0}^{1} \sqrt{n}u_q^{\mathrm{T}}\{\bar{R}_q - \bar{R}(q)\}$, which further equals to

$$\begin{aligned}
&\sum_{q=0}^{1}\sqrt{n}\sum_{k=1}^{K_N}\Pi_{[k]}\frac{1}{n_{[k]q}}\sum_{i\in[k]}Z_iI(T_i=q)u_q^{\mathrm{T}}\{R_i(q)-\bar{R}_{[k]}(q)\}\\
=&\sum_{k=1}^{K_N}\sum_{i\in[k]}\sum_{q=0}^{1}Z_iI(T_i=q)\cdot\frac{1}{\sqrt{n}}\cdot\frac{nN_{[k]}}{Nn_{[k]q}}A_i(q),
\end{aligned}$$

where $A_i(q) = u_q^{\mathrm{T}}\{R_i(q) - \bar{R}_{[k]}(q)\}$, for $i \in [k]$. Without loss of generality, suppose that the units are ordered strata-by-strata. Then, $i \in [k]$ means $\sum_{k'=1}^{k-1} N_{[k']} < i \leq \sum_{k'=1}^{k} N_{[k']}$. Let

$$b_i(q) = \begin{cases} \frac{1}{\sqrt{n}}\cdot\frac{nN_{[k]}}{Nn_{[k]q}}, & \sum_{k'=1}^{k-1} N_{[k']} + \sum_{q'=0}^{q-1} n_{[k]q'} < i \leq \sum_{k'=1}^{k-1} N_{[k']} + \sum_{q'=0}^{q} n_{[k]q'}, \\ 0, & \text{otherwise.} \end{cases}$$

Here, $\sum_{k'=1}^{0} N_{[k']} = 0$ and $\sum_{q'=0}^{-1} n_{[k]q'} = 0$. Let $(G_i, i \in [k])$ be the random partition of $[k]$, which means $(G_i, i \in [k])$ takes any permutation of $[k]$ with probability $1/N_{[k]}!$. Hence H_n has the same distribution as

$$H_n' = \sum_{k=1}^{K_N}\sum_{i\in[k]}\big\{A_i(1)b_{G_i}(1) + A_i(0)b_{G_i}(0)\big\}.$$

Next, we borrow ideas from Hájek (1961) to derive the asymptotic normality of H_n'. Let U_i i.i.d. $\sim U(0,1)$, the uniform distribution on $(0,1)$. Within block k, denote $b(\lambda, q)$ as the quantile function of $\{b_i(q),\ i \in [k]\}$.

Definition 1 (Hájek (1961))**.** *The quantile function of N real numbers, $\{c_i, i = 1, \ldots, N\}$, is defined as $c(\lambda) = c_{[i]}$ for $(i-1)/N < \lambda \leq i/N$ and $1 \leq i \leq N$, where $c_{[1]} \leq \cdots \leq c_{[N]}$ are the order statistics of $\{c_i, i = 1, \ldots, N\}$.*

Define

$$T_n = \sum_{k=1}^{K_N} \sum_{i \in [k]} \sum_{q=0}^{1} [\{A_i(q) - \bar{A}_{[k]}(q)\} b(U_i, q) + \bar{A}_{[k]}(q) \cdot b_i(q)],$$

where $\bar{A}_{[k]}(q)$ is the stratum-specific population mean of $A_i(q)$.

Step 1. In this step, we examine the scenario where $K_N/n \to 0$. To streamline the proof of Lemma S1, we introduce Lemma S2 and S3 below, with their proofs following immediately after the proof of Lemma S1.

Lemma S2. *Under Condition 1(i)–(ii) and Condition S1, there exists a constant L' independent of N, such that*

$$\max_{q=1,0} \max_{i \in [k]} |b_i(q)| \leq \frac{L'}{\sqrt{n}}, \quad \max_{q=1,0} \sum_{i \in [k]} \{b_i(q) - \bar{b}_{[k]}(q)\}^2 \leq L' \Pi_{[k]}.$$

Lemma S3. *If $u^{\mathrm{T}} V_R u > 0$, $K_N/n \to 0$, Condition 1(i)–(ii) and Condition S1 hold, then*

$$\frac{E(H_n' - T_n)^2}{\mathrm{var}(H_n')} \to 0.$$

By Lemma S3, H_n' and T_n are asymptotically equivalent in the mean, when the number of blocks K_N goes to infinity at a rate much slower than n. Considering the independence of the random variables in the summation of T_n, we only need to verify the Lindeberg–Feller condition to derive the asymptotic normality of T_n.

Recall that $T_n - E(T_n) = \sum_{k=1}^{K_N} \sum_{i \in [k]} \sum_{q=0}^{1} \{A_i(q) - \bar{A}_{[k]}(q)\} \cdot b(U_i, q)$. By Lemma S3, we have $\mathrm{var}(T_n) - \mathrm{var}(H_n') \to 0$. Therefore, $\mathrm{var}(T_n) \to u^T V_R u > 0$. It is clear that

$$
\begin{aligned}
&\max_{k=1,\ldots,K_N} \max_{i\in[k]} |\sum_{q=0}^{1}\{A_i(q)-\bar{A}_{[k]}(q)\}\cdot b(U_i,q)| \\
\le& 2\cdot \max_{k=1,\ldots,K_N} \max_{i\in[k]} \max_{q=1,0} |A_i(q)-\bar{A}_{[k]}(q)|\cdot|b(U_i,q)|.
\end{aligned}
$$

By Lemma S2, Hölder inequality, and Condition S1, we have

$$
\max_{k=1,\ldots,K_N} \max_{i\in[k]} \max_{q=1,0} |A_i(q)-\bar{A}_{[k]}(q)|\cdot|b(U_i,q)|
$$

can be upper bounded by

$$
\begin{aligned}
&\max_{k=1,\ldots,K_N} \max_{i\in[k]} \max_{q=1,0} |A_i(q)-\bar{A}_{[k]}(q)|\cdot\frac{L'}{\sqrt{n}} \\
\le& \max_{q=1,0}\|u_q\|_1\cdot \max_{k=1,\ldots,K_N} \max_{i\in[k]} \max_{q=1,0} \|R_i(q)-\bar{R}_{[k]}(q)\|_\infty\cdot\frac{L'}{\sqrt{n}}\to 0.
\end{aligned}
$$

Hence, $\forall\varepsilon>0$, when n is sufficient large, we have

$$
\max_{k=1,\ldots,K_N} \max_{i\in[k]} |\sum_{q=0}^{1}\{A_i(q)-\bar{A}_{[k]}(q)\}\cdot b(U_i,q)| < \varepsilon\sqrt{\mathrm{var}(T_n)}.
$$

Together with $\mathrm{var}(T_n)\to u^{\mathrm{T}}V_R u>0$, it leads to

$$
\begin{aligned}
\lim_{n\to\infty}\sum_{k=1}^{K_N}\sum_{i\in[k]} E\Big[\sum_{q=0}^{1}\{A_i(q)-\bar{A}_{[k]}(q)\}\cdot b(U_i,q)\Big]^2\cdot& \\
I\Big(|\sum_{q=0}^{1}\{A_i(q)-\bar{A}_{[k]}(q)\}\cdot b(U_i,q)|>\varepsilon\sqrt{\mathrm{var}(T_n)}\Big)&=0,
\end{aligned}
$$

where $I(\cdot)$ is the indicator function. Thus, Lindeberg's condition holds and we have

$$
\frac{T_n-ET_n}{\sqrt{\mathrm{var}(T_n)}}\xrightarrow{d}\mathcal{N}(0,1).
$$

By Lemma S3, we then have

$$\frac{H_n - EH_n}{\sqrt{\mathrm{var}(H_n)}} \xrightarrow{d} \mathcal{N}(0,1).$$

Step 2. We extend the results of Step 1 to the case of a general K_N. Let

$$\begin{aligned} B_1 &= \Big\{k : n_{[k]} \geq n^{1/4} / \max_{k=1,\ldots,K_N} \max_{i\in[k]} \max_{q=1,0} |A_i(q) - \bar{A}_{[k]}(q)|^{1/2}\Big\} \\ B_2 &= \Big\{k : n_{[k]} < n^{1/4} / \max_{k=1,\ldots,K_N} \max_{i\in[k]} \max_{q=1,0} |A_i(q) - \bar{A}_{[k]}(q)|^{1/2}\Big\}, \end{aligned}$$

which represent the index set for “large” and “small” strata, respectively. We have

$$H_n' = \sum_{k\in B_1} \sum_{i\in[k]} \sum_{q=0}^{1} A_i(q) b_{G_i}(q) + \sum_{k\in B_2} \sum_{i\in[k]} \sum_{q=0}^{1} A_i(q) b_{G_i}(q) := H_{n1}' + H_{n2}',$$

where $H_{n1}' = \sum_{k\in B_1} \sum_{i\in[k]} \sum_{q=0}^{1} A_i(q) b_{G_i}(q)$ and $H_{n2}' = \sum_{k\in B_2} \sum_{i\in[k]} \sum_{q=0}^{1} A_i(q) b_{G_i}(q)$. Note that H_{n1}' and H_{n2}' are independent. We can use the result of Step 1 if $|B_1| / \sum_{k\in B_1} n_{[k]} \to 0$, where $|B_1|$ denote the cardinality of B_1. In fact,

$$|B_1| / \sum_{k\in B_1} n_{[k]} \leq (\max_{k=1,\ldots,K_N} \max_{i\in[k]} \max_{q=1,0} n^{-1} |A_i(q) - \bar{A}_{[k]}(q)|^2)^{\frac{1}{4}} \to 0.$$

Hence, according to Step 1, we can induce that H_{n1}' is asymptotically normal if its variance converges to a finite and positive limit. Next, we consider the remaining term H_{n2}' via the following three cases:

(i) $\mathrm{var}(H_{n2}') \to 0$. This implies H_{n2}' converges to 0 in probability. Hence, $\mathrm{var}(H_{n1}') = \mathrm{var}(H_n') - \mathrm{var}(H_{n2}') \to u^T V_R u > 0$. Then, the conclusion of the theorem holds.

(ii) $\mathrm{var}(H_{n2}') \to c > 0$. It leads to $\mathrm{var}(H_{n1}') = \mathrm{var}(H_n') - \mathrm{var}(H_{n2}') \to u^{\mathrm{T}} V_R u - c$.

First, we show that $|B_2| \to \infty$. For $k \in B_2$ and $q = 0, 1$, by Condition 1(i), there is only $n_{[k]q}$ nonzero values of $b_{G_i}(q)$. We then have $|\sum_{i\in[k]}\sum_{q=0}^{1}\{A_i(q) - \bar{A}_{[k]}(q)\}b_{G_i}(q)|$ can be bounded by

$$
\begin{aligned}
&\sum_{q=0}^{1} n_{[k]q} \max_{i\in[k]} |A_i(q) - \bar{A}_{[k]}(q)| \frac{L'}{\sqrt{n}} \\
=&L' n_{[k]} (\max_{k=1,\ldots,K_N} \max_{i\in[k]} \max_{q=1,0} n^{-1}|A_i(q) - \bar{A}_{[k]}(q)|^2)^{1/2} \\
\leq&L' (\max_{k=1,\ldots,K_N} \max_{i\in[k]} \max_{q=1,0} n^{-1}|A_i(q) - \bar{A}_{[k]}(q)|^2)^{1/4} \to 0.
\end{aligned}
$$

Hence, $\mathrm{var}\{\sum_{i\in[k]}\sum_{q=0}^{1} A_i(q)b_{G_i}(q)\}$ uniformly converges to zero. If $|B_2| \not\to \infty$, then

$$
\mathrm{var}(H'_{n_2}) = \sum_{k\in B_2} \mathrm{var}\Big\{\sum_{i\in[k]}\sum_{q=0}^{1} A_i(q)b_{G_i}(q)\Big\} \to 0.
$$

This is in contradiction with $\mathrm{var}(H'_{n2}) \to c > 0$. Hence, we conclude that $|B_2| \to \infty$.

Next, we verify Lindeberg's condition for H'_{n2}. Since

$$
\max_{k\in B_2} |\sum_{i\in[k]}\sum_{q=0}^{1}\{A_i(q) - \bar{A}_{[k]}(q)\}b_{G_i}(q)| \to 0,
$$

then, $\forall \varepsilon > 0$, when n is large enough,

$$
\max_{k\in B_2} |\sum_{i\in[k]}\sum_{q=0}^{1}\{A_i(q) - \bar{A}_{[k]}(q)\}b_{G_i}(q)| < \varepsilon\sqrt{\mathrm{var}(H'_{n2})}.
$$

Hence, with $\mathrm{var}(H_{n2'}) \to c > 0$,

$$\lim_{n\to\infty} \sum_{k\in B_2} E\Big[\sum_{i\in[k]}\sum_{q=0}^{1}\{A_i(q) - \bar{A}_{[k]}(q)\}b_{G_i}(q)\Big]^2 \cdot$$
$$I(|\sum_{i\in[k]}\sum_{q=0}^{1}\{A_i(q) - \bar{A}_{[k]}(q)\}b_{G_i}(q)| > \varepsilon\sqrt{\mathrm{var}(H'_{n2})}) = 0,$$

which implies the Lindeberg's condition for H'_{n2} holds.

(iii) $\mathrm{var}(H'_{n2})$ does not converge. For any subsequence H'_{n_k} of H'_n, there exists a further subsequence $H'_{n_{k_l}}$ such that $\mathrm{var}(H'_{n_{k_l}2}) \to c' \in [0, u^{\mathrm{T}}V_R u]$. Then,

$$\mathrm{var}(H'_{n_{k_l}1}) = \mathrm{var}(H'_{n_{k_l}}) - \mathrm{var}(H'_{n_{k_l}2}) \to u^{\mathrm{T}}V_R u - c'.$$

Hence, according to previous conclusion, $H'_{n_{k_l}} \xrightarrow{d} \mathcal{N}(0, u^{\mathrm{T}}V_R u)$, which means for any subsequence H'_{n_k} of H'_n, there exists a further subsequence $H'_{n_{k_l}}$ such that $H'_{n_{k_l}} \xrightarrow{d} \mathcal{N}(0, u^{\mathrm{T}}V_R u)$. Applying Durrett (2010, Theorem 2.2.3) for $\mathrm{pr}(H'_n \le t)$, $t \in \mathbb{R}$, $H'_n \xrightarrow{d} \mathcal{N}(0, u^{\mathrm{T}}V_R u)$ and we finish the proof for Lemma S1.

□

Proof of Lemma S2. By Condition 1(i), we have

$$\max_{i\in[k]} |b_i(q)| = \frac{1}{\sqrt{n}} \cdot \frac{nN_{[k]}}{Nn_{[k]q}} \le \frac{c_3}{c_1}\frac{1}{\sqrt{n}}.$$

Then, we have

$$\bar{b}_{[k]}(q) = \frac{n_{[k]q}}{N_{[k]}}\frac{1}{\sqrt{n}} \cdot \frac{nN_{[k]}}{Nn_{[k]q}} = \frac{\sqrt{n}}{N}.$$

Hence, $\sum_{i\in[k]}\{b_i(q)-\bar{b}_{[k]}(q)\}^2$ can be upper bounded by

$$(N_{[k]}-n_{[k]q})\frac{n}{N^2}+n_{[k]q}\frac{n}{N^2}\Big(\frac{N_{[k]}}{n_{[k]q}}-1\Big)^2=\frac{N_{[k]}}{N}\frac{n}{n_{[k]q}}\frac{N_{[k]}-n_{[k]q}}{N}\leq\frac{c_3}{c_1}\Pi_{[k]}.$$

□

Proof of Lemma S3. Since $\mathrm{var}(H'_n)\to u^{\mathrm{T}}V_Ru>0$, it suffices to show that

$$E(H'_n-T_n)^2\to 0.$$

Let $\bar{b}_{[k]}(q)$ be the stratum-specific population mean of $b_i(q)$. By the proof of Theorem 3.1 in Hájek (1961) (see 3.9–3.11), we have

$$\begin{aligned}
&E(H'_n-T_n)^2\\
=&\sum_{k=1}^{K_N}E\Big[\sum_{i\in[k]}\sum_{q=0}^{1}A_i(q)b_{G_i}(q)-\{A_i(q)-\bar{A}_{[k]}(q)\}b(U_i,q)-\bar{A}_{[k]}(q)b_i(q)\Big]^2\\
\leq&\sum_{k=1}^{K_N}2\sum_{q=0}^{1}E\Big[\sum_{i\in[k]}A_i(q)b_{G_i}(q)-\{A_i(q)-\bar{A}_{[k]}(q)\}b(U_i,q)-\bar{A}_{[m]}(q)b_i(q)\Big]^2\\
\leq&2\sum_{q=0}^{1}\sum_{k=1}^{K_N}\frac{1}{N_{[k]}-1}\sum_{i\in[k]}\{A_i(q)-\bar{A}_{[k]}(q)\}^2\\
&\times 2\sqrt{2}\times\max_{i\in[k]}|b_i(q)-\bar{b}_{[k]}(q)|\times\Big[\sum_{i\in[k]}\{b_i(q)-\bar{b}_{[k]}(q)\}^2\Big]^{1/2}.
\end{aligned}$$

By Condition 1(i), Condition S1, and Lemma S2,

$$
\begin{aligned}
& E(H'_n - T_n)^2 \\
\leq & 8\sqrt{2}L(L')^{3/2}\Big(\sum_{k=1}^{K_N}\frac{1}{\sqrt{n}}\sqrt{\Pi_{[k]}}\Big) \\
\leq & 8\sqrt{2}L(L')^{3/2}\Big\{\frac{1}{\sqrt{n}}\Big(\sum_{k=1}^{K_N}\Pi_{[k]}\sum_{k=1}^{K_N}1\Big)^{1/2}\Big\} = 8\sqrt{2}L(L')^{3/2}\Big(\frac{K_N}{n}\Big)^{1/2} \to 0.
\end{aligned}
$$

□

Taking suitable values of $R_i(1)$ and $R_i(0)$, we can prove Theorem 1.

Proof of Theorem 1. It is easy to show that $E\hat{\tau} = \tau$, $E\hat{\tau}_X = 0$, and $E\hat{\delta}_W = 0$. We define pseudo potential outcome vectors for unit $i \in [k]$ as

$$
R_i(1) = \begin{pmatrix} Y_i(1) - \bar{Y}(1) \\ X_i \\ e_{[k]1}(W_i - \bar{W}_{[k]}) \end{pmatrix}, \quad R_i(0) = \begin{pmatrix} Y_i(0) - \bar{Y}(0) \\ X_i \\ -e_{[k]0}(W_i - \bar{W}_{[k]}) \end{pmatrix}.
$$

Then, we have

$$
\begin{aligned}
\bar{R}_1 - \bar{R}_0 &= \sum_{k=1}^{K_N}\Pi_{[k]}\frac{1}{n_{[k]1}}\sum_{i\in[k]}Z_iT_iR_i(1) - \sum_{k=1}^{K_N}\Pi_{[k]}\frac{1}{n_{[k]0}}\sum_{i\in[k]}Z_i(1-T_i)R_i(0) \\
&= \sum_{k=1}^{K_N}\Pi_{[k]}\begin{pmatrix} \hat{\tau}_{[k]} \\ \hat{\tau}_{[k]X} \\ \hat{\delta}_{[k]W} \end{pmatrix} = \begin{pmatrix} \hat{\tau} - \tau \\ \hat{\tau}_X \\ \hat{\delta}_W \end{pmatrix}.
\end{aligned}
$$

By Proposition S1, $E\{\sqrt{n}(\bar{R}_1 - \bar{R}_0)\} = 0$ and $\mathrm{cov}\{\sqrt{n}(\bar{R}_1 - \bar{R}_0)\} = V$. Condition 1 implies that V has a finite limit and $\{R_i(1)\}_{i=1}^N$, $\{R_i(0)\}_{i=1}^N \in \mathcal{M}_L$. By

Lemma S1, the conclusion of Theorem 1 holds.

□

S6.2 Proof of Corollary 1

Proof. By setting $e_{[k]1} = e_1$ and $f_{[k]} = f$ for all $k = 1, \ldots, K_N$, we have $\Pi_{[k]} = \pi_{[k]}$ and then can express $V_{\tau\tau}$ as follows:

$$V_{\tau\tau} = \sum_{k=1}^{K_N} \frac{\Pi_{[k]}^2}{\pi_{[k]}} (e_{[k]1}^{-1} S_{[k]1}^2 + e_{[k]0}^{-1} S_{[k]0}^2 - f_{[k]} S_{[k]\tau}^2)$$
$$= \sum_{k=1}^{K_N} \Pi_{[k]} (e_1^{-1} S_{[k]1}^2 + e_0^{-1} S_{[k]0}^2 - f S_{[k]\tau}^2) = \sum_{k=1}^{K_N} \Pi_{[k]} V_{[k]\tau\tau}.$$

Let $S_t^2 = (N-1)^{-1} \sum_{i=1}^N \{Y_i(t) - \bar{Y}(t)\}^2$, $t = 0, 1$, and $S_\tau^2 = (N-1)^{-1} \sum_{i=1}^N (\tau_i - \tau)^2$ represent the variances of $\{Y_i(t)\}$ and $\{\tau_i\}$, respectively. Then, by Yang et al. (2023, Theorem 1),

$$V_{\tau\tau,\mathrm{C}} = e_1^{-1} S_1^2 + e_0^{-1} S_0^2 - f S_\tau^2.$$

We can decompose S_t^2 as follows:

$$S_t^2 = (N-1)^{-1} \sum_{k=1}^{K_N} \sum_{i \in [k]} \left[\{Y_i(t) - \bar{Y}_{[k]}(t)\} + \{\bar{Y}_{[k]}(t) - \bar{Y}(t)\} \right]^2$$
$$= (N-1)^{-1} \sum_{k=1}^{K_N} \left[(N_{[k]} - 1) S_{[k]t}^2 + N_{[k]} \{\bar{Y}_{[k]}(t) - \bar{Y}(t)\}^2 \right].$$

Similarly,

$$S_\tau^2 = (N-1)^{-1} \sum_{k=1}^{K_N} \left\{ (N_{[k]} - 1) S_{[k]\tau}^2 + N_{[k]} (\tau_{[k]} - \tau)^2 \right\}.$$

Hence, we can decompose $V_{\tau\tau,\mathrm{C}}$ as follows:

$$V_{\tau\tau,\mathrm{C}} = (N-1)^{-1}\sum_{k=1}^{K_N}(N_{[k]}-1)V_{[k]\tau\tau} + (N-1)^{-1}\sum_{k=1}^{K_N} N_{[k]}d_{[k]},$$

where

$$\begin{aligned} d_{[k]} &= e_1^{-1}\{\bar{Y}_{[k]}(1)-\bar{Y}(1)\}^2 + e_0^{-1}\{\bar{Y}_{[k]}(0)-\bar{Y}(0)\}^2 - f(\tau_{[k]}-\tau)^2 \\ &= [(e_0/e_1)^{1/2}\{\bar{Y}_{[k]}(1)-\bar{Y}(1)\} + (e_1/e_0)^{1/2}\{\bar{Y}_{[k]}(0)-\bar{Y}(0)\}]^2 \\ &\quad +(1-f)(\tau_{[k]}-\tau)^2 \geq 0. \end{aligned}$$

Hence, the difference of the asymptotic variances is

$$V_{\tau\tau,\mathrm{C}} - V_{\tau\tau} = (N-1)^{-1}\sum_{k=1}^{K_N} N_{[k]}d_{[k]} - (N-1)^{-1}\sum_{k=1}^{K_N}(1-\Pi_{[k]})V_{[k]\tau\tau}.$$

□

S6.3 Proof of Theorem 2

Proof. For (i), note that $\Pi_{[k]}/\pi_{[k]} = f/f_{[k]}$ and $f = \sum_{k=1}^{K_N}\Pi_{[k]}f_{[k]}$, then we have

$$\begin{aligned} V_{\tau\tau} &= f\Big[\sum_{k=1}^{K_N}\Big\{\frac{\Pi_{[k]}}{f_{[k]}}\Big(\frac{S^2_{[k]1}}{e_{[k]1}} + \frac{S^2_{[k]0}}{e_{[k]0}}\Big) - \Pi_{[k]}S^2_{[k]\tau}\Big\}\Big] \\ &= f\Big[\sum_{k=1}^{K_N}\Big\{\frac{\Pi^2_{[k]}}{f_{[k]}\Pi_{[k]}}\Big(\frac{S^2_{[k]1}}{e_{[k]1}} + \frac{S^2_{[k]0}}{e_{[k]0}}\Big) - \Pi_{[k]}S^2_{[k]\tau}\Big\}\Big] \\ &\geq f\Big\{\frac{\sum_{k=1}^{K_N}\big(\Pi_{[k]}\sqrt{S^2_{[k]1}/e_{[k]1} + S^2_{[k]0}/e_{[k]0}}\big)^2}{\sum_{k=1}^{K_N} f_{[k]}\Pi_{[k]}} - \sum_{k=1}^{K_N}\Pi_{[k]}S^2_{[k]\tau}\Big\} \\ &= \sum_{k=1}^{K_N}\big(\Pi_{[k]}\sqrt{S^2_{[k]1}/e_{[k]1} + S^2_{[k]0}/e_{[k]0}}\big)^2 - f\sum_{k=1}^{K_N}\Pi_{[k]}S^2_{[k]\tau}, \end{aligned}$$

where the inequality is due to the Cauchy–Schwarz inequality with equality holds if and only if $f_{[k]}$ is proportion to $\sqrt{S^2_{[k]1}/e_{[k]1} + S^2_{[k]0}/e_{[k]0}}$. Since $f = \sum_{k=1}^{K_N} \Pi_{[k]} f_{[k]}$, then

$$f_{[k]}/f = \sqrt{e^{-1}_{[k]1} S^2_{[k]1} + e^{-1}_{[k]0} S^2_{[k]0}} \Big/ \Big(\sum_{k'=1}^{K_N} \Pi_{[k']} \sqrt{e^{-1}_{[k']1} S^2_{[k']1} + e^{-1}_{[k']0} S^2_{[k']0}} \Big)$$

minimizes $V_{\tau\tau}$, the asymptotic variance of $\hat{\tau}$.

For (ii), by Cauchy–Schwarz inequality, we have

$$\frac{S^2_{[k]1}}{e_{[k]1}} + \frac{S^2_{[k]0}}{e_{[k]0}} \geq \frac{(S_{[k]1} + S_{[k]0})^2}{e_{[k]1} + e_{[k]0}} = (S_{[k]1} + S_{[k]0})^2,$$

and the equality holds if and only if $e_{[k]1}/e_{[k]0} = \sqrt{S^2_{[k]1}/S^2_{[k]0}}$, or equivalently,

$$e_{[k]1} = \sqrt{S^2_{[k]1}} \Big/ \Big(\sqrt{S^2_{[k]1}} + \sqrt{S^2_{[k]0}} \Big).$$

□

S6.4 Proof of Theorem 3

Before proving Theorem 3, we establish a lemma on the convergence of sample covariance to the population covariance.

Lemma S4. *Under Condition 1 and the stratified randomized survey experiment or the SRSRR experiment, for $t \in \{0, 1\}$, we have*

$$\sum_{k=1}^{K_N} \Pi^2_{[k]} \pi^{-1}_{[k]} e^{-1}_{[k]t} s^2_{[k]t} - \sum_{k=1}^{K_N} \Pi^2_{[k]} \pi^{-1}_{[k]} e^{-1}_{[k]t} S^2_{[k]t} = o_p(1),$$

$$\sum_{k=1}^{K_N} \Pi_{[k]}^2 \pi_{[k]}^{-1} (e_{[k]1} e_{[k]0})^{-1} S_{[k]X|\mathcal{S}}^2 - \sum_{k=1}^{K_N} \Pi_{[k]}^2 \pi_{[k]}^{-1} (e_{[k]1} e_{[k]0})^{-1} S_{[k]X}^2 = o_p(1),$$

$$\sum_{k=1}^{K_N} \Pi_{[k]}^2 \pi_{[k]}^{-1} e_{[k]t}^{-1} s_{[k]X,t} - \sum_{k=1}^{K_N} \Pi_{[k]}^2 \pi_{[k]}^{-1} e_{[k]t}^{-1} S_{[k]X,t} = o_p(1),$$

$$\sum_{k=1}^{K_N} \Pi_{[k]}^2 \pi_{[k]}^{-1} (1 - f_{[k]}) s_{[k]W,t} - \sum_{k=1}^{K_N} \Pi_{[k]}^2 \pi_{[k]}^{-1} (1 - f_{[k]}) S_{[k]W,t} = o_p(1).$$

Proof of Lemma S4. We will only prove the first statement, as the proof for the other statements is similar. To prove the first statement, we compute the expectation and upper bound of the variance, $\mathrm{var}(\sum_{k=1}^{K_N} \Pi_{[k]}^2 \pi_{[k]}^{-1} e_{[k]t}^{-1} s_{[k]t}^2)$.

First, under the stratified randomized survey experiment, note that $s_{[k]t}^2$ is an unbiased estimator of $S_{[k]t}^2$ for $t = 0, 1$. Thus,

$$E\Big(\sum_{k=1}^{K_N} \Pi_{[k]}^2 \pi_{[k]}^{-1} e_{[k]t}^{-1} s_{[k]t}^2\Big) = \sum_{k=1}^{K_N} \Pi_{[k]}^2 \pi_{[k]}^{-1} e_{[k]t}^{-1} S_{[k]t}^2.$$

Second, we show that under the stratified randomized survey experiment,

$$\mathrm{var}\Big(\sum_{k=1}^{K_N} \Pi_{[k]}^2 \pi_{[k]}^{-1} e_{[k]t}^{-1} s_{[k]t}^2\Big) \to 0.$$

The stratum-specific sample variance $s_{[k]t}^2$ can be decomposed as

$$s_{[k]t}^2 = \frac{n_{[k]t}}{n_{[k]t} - 1} \Big[\frac{1}{n_{[k]t}} \sum_{i \in [k],\ T_i = t} Z_i \{Y_i(t) - \bar{Y}_{[k]t}\}^2 - \{\bar{Y}_{[k]|\mathcal{S}}(t) - \bar{Y}_{[k]t}\}^2 \Big],$$

where $\bar{Y}_{[k]|\mathcal{S}}(t) = n_{[k]}^{-1}\sum_{i\in[k]} Z_i Y_i(t)$. Then

$$
\begin{aligned}
&\operatorname{var}\Big(\sum_{k=1}^{K_N} \Pi_{[k]}^2 \pi_{[k]}^{-1} e_{[k]t}^{-1} s_{[k]t}^2\Big) = \sum_{k=1}^{K_N} \Pi_{[k]}^4 \pi_{[k]}^{-2} e_{[k]t}^{-2} \operatorname{var}(s_{[k]t}^2)\\
=&\sum_{k=1}^{K_N} \Pi_{[k]}^4 \pi_{[k]}^{-2} e_{[k]t}^{-2} \frac{n_{[k]t}^2}{(n_{[k]t}-1)^2} \operatorname{var}\Big[\frac{1}{n_{[k]t}} \sum_{i\in[k],\ T_i=t} Z_i\{Y_i(t)-\bar{Y}_{[k]}(t)\}^2 - \{\bar{Y}_{[k]|\mathcal{S}}(t)-\bar{Y}_{[k]}(t)\}^2\Big]\\
\le& 8\sum_{k=1}^{K_N} \Pi_{[k]}^4 \pi_{[k]}^{-2} e_{[k]t}^{-2} \Big(\operatorname{var}\Big[\frac{1}{n_{[k]t}} \sum_{i\in[k],\ T_i=t} Z_i\{Y_i(t)-\bar{Y}_{[k]}(t)\}^2\Big] + \operatorname{var}\big[\{\bar{Y}_{[k]|\mathcal{S}}(t)-\bar{Y}_{[k]}(t)\}^2\big]\Big),
\end{aligned}
$$

where the last inequality holds due to $n_{[k]t}^2/(n_{[k]t}-1)^2 \le 4$ and $\operatorname{var}(X+Y) \le 2\{\operatorname{var}(X)+\operatorname{var}(Y)\}$. We consider the two terms separately. The first term

$$
\begin{aligned}
&8\sum_{k=1}^{K_N} \Pi_{[k]}^4 \pi_{[k]}^{-2} e_{[k]t}^{-2} \operatorname{var}\Big[\frac{1}{n_{[k]t}} \sum_{i\in[k],\ T_i=t} \{Y_i(t)-\bar{Y}_{[k]}(t)\}^2\Big]\\
\le& 8\sum_{k=1}^{K_N} \Pi_{[k]}^4 \pi_{[k]}^{-2} e_{[k]t}^{-2} \Big(\frac{1}{n_{[k]t}} - \frac{1}{N_{[k]}}\Big) \frac{1}{N_{[k]}-1} \sum_{i\in[k]} \{Y_i(t)-\bar{Y}_{[k]}(t)\}^4\\
\le& 8\Big(\frac{1}{n} \max_{k=1,\dots,K_N} \max_{i\in[k]} |Y_i(t)-\bar{Y}_{[k]}(t)|^2\Big) \sum_{k=1}^{K_N} \Big[\frac{1}{N_{[k]}-1} \sum_{i\in[k]} \{Y_i(t)-\bar{Y}_{[k]}(t)\}^2\Big] \frac{\Pi_{[k]}^4 \pi_{[k]}^{-2} e_{[k]t}^{-2} n}{n_{[k]t}}\\
\le& 8\frac{c_3^2}{c_1^3} \cdot \Big\{\frac{1}{n} \max_{k=1,\dots,K_N} \max_{i\in[k]} |Y_i(t)-\bar{Y}_{[k]}(t)|^2\Big\} \cdot \Big(\sum_{k=1}^{K_N} \Pi_{[k]}^2 \pi_{[k]}^{-1} S_{[k]t}^2\Big) \to 0,
\end{aligned}
$$

where the last inequality is due to Condition 1(i) and

$$
\Pi_{[k]}^2 \pi_{[k]}^{-1} e_{[k]t}^{-1} \frac{n}{n_{[k]t}} = \Big(\frac{f}{f_{[k]} e_{[k]t}}\Big)^2 \le \frac{c_3^2}{c_1^2}.
$$

The second term is $8\sum_{k=1}^{K_N} \Pi_{[k]}^4 \pi_{[k]}^{-2} e_{[k]t}^{-2} \operatorname{var}[\{\bar{Y}_{[k]|\mathcal{S}}(t)-\bar{Y}_{[k]}(t)\}^2]$, which is bounded by

$$
\begin{aligned}
& 8\sum_{k=1}^{K_N}\Pi_{[k]}^4\pi_{[k]}^{-2}e_{[k]t}^{-2}\max_{i\in[k]}\{Y_i(t)-\bar{Y}_{[k]}(t)\}^2\mathrm{var}[\bar{Y}_{[k]|\mathcal{S}}(t)-\bar{Y}_{[k]}(t)] \\
\leq & 8\sum_{k=1}^{K_N}\Pi_{[k]}^4\pi_{[k]}^{-2}e_{[k]t}^{-2}\max_{i\in[k]}\{Y_i(t)-\bar{Y}_{[k]}(t)\}^2\Big(\frac{1}{n_{[k]t}}-\frac{1}{N_{[k]}}\Big)\frac{1}{N_{[k]}-1}\sum_{i\in[k]}\{Y_i(t)-\bar{Y}_{[k]}(t)\}^2 \\
\leq & 8\Big[\frac{1}{n}\max_{k=1,\ldots,K_N}\max_{i\in[k]}\big\{Y_i(t)-\bar{Y}_{[k]}(t)\big\}^2\Big]\sum_{k=1}^{K_N}S_{[k]t}^2\frac{\Pi_{[k]}^4\pi_{[k]}^{-2}e_{[k]t}^{-2}n}{n_{[k]t}} \\
\leq & 8\frac{c_3^2}{c_1^3}\cdot\Big\{\frac{1}{n}\max_{k=1,\ldots,K_N}\max_{i\in[k]}|Y_i(t)-\bar{Y}_{[k]}(t)|^2\Big\}\cdot\Big(\sum_{k=1}^{K_N}\Pi_{[k]}^2\pi_{[k]}^{-1}S_{[k]t}^2\Big)\to 0.
\end{aligned}
$$

Thus, we have $\mathrm{var}(\sum_{k=1}^{K_N}\Pi_{[k]}^2\pi_{[k]}^{-1}e_{[k]t}^{-1}s_{[k]t}^2)\to 0$ and the first statement holds under the stratified randomized survey experiment.

Additionally, under the SRSRR experiment, it is equivalent to conditioning on $(\mathrm{M}_S\leq a_S,\ \mathrm{M}_T\leq a_T)$. By Theorem 1, we have

$$
\mathrm{pr}(\mathrm{M}_S\leq a_S,\ \mathrm{M}_T\leq a_T)\to\mathrm{pr}(\chi_{J_1}^2\leq a_S,\ \chi_{J_2}^2\leq a_T)>0.
$$

Thus, for any $\epsilon>0$,

$$
\begin{aligned}
& \mathrm{pr}\Big(\Big|\sum_{k=1}^{K_N}\Pi_{[k]}^2\pi_{[k]}^{-1}e_{[k]t}^{-1}s_{[k]t}^2-\sum_{k=1}^{K_N}\Pi_{[k]}^2\pi_{[k]}^{-1}e_{[k]t}^{-1}S_{[k]t}^2\Big|>\epsilon\mid \mathrm{M}_S\leq a_S,\ \mathrm{M}_T\leq a_T\Big) \\
\leq\ & \mathrm{pr}\Big(\Big|\sum_{k=1}^{K_N}\Pi_{[k]}^2\pi_{[k]}^{-1}e_{[k]t}^{-1}s_{[k]t}^2-\sum_{k=1}^{K_N}\Pi_{[k]}^2\pi_{[k]}^{-1}e_{[k]t}^{-1}S_{[k]t}^2\Big|>\epsilon\Big)/\mathrm{pr}(\mathrm{M}_S\leq a_S,\ \mathrm{M}_T\leq a_T) \\
\to\ & 0.
\end{aligned}
$$

That is, the first statement holds under the SRSRR experiment.

□

Next, we prove Theorem 3.

Proof of Theorem 3. By Theorem 1, we have $\sqrt{n}(\hat{\tau}-\tau, \hat{\tau}_X^{\mathrm{T}}, \hat{\delta}_W^{\mathrm{T}})^{\mathrm{T}} \xrightarrow{d} N(0, V)$. Let $(A, B^{\mathrm{T}}, C^{\mathrm{T}})^{\mathrm{T}} \sim N(0, V)$ denote the limiting distribution of $\sqrt{n}(\hat{\tau}-\tau, \hat{\tau}_X^{\mathrm{T}}, \hat{\delta}_W^{\mathrm{T}})^{\mathrm{T}}$. Let $V_1 \ \dot{\sim} \ V_2$ denote two random vectors V_1 and V_2 having the same asymptotic distribution. The asymptotic distribution of $\sqrt{n}(\hat{\tau}-\tau)$ under the SRSRR experiment is the asymptotic distribution of $\sqrt{n}(\hat{\tau}-\tau)$ conditional on $(\mathrm{M}_S \le a_S,\ \mathrm{M}_T \le a_T)$, i.e.,

$$\sqrt{n}(\hat{\tau}-\tau) \mid (\mathrm{M}_S \le a_S,\ \mathrm{M}_T \le a_T).$$

Note that $\mathrm{M}_S = \hat{\delta}_W^{\mathrm{T}} \mathrm{cov}(\hat{\delta}_W)^{-1} \hat{\delta}_W$ and $\mathrm{M}_T = \hat{\tau}_X^{\mathrm{T}} \mathrm{cov}(\hat{\tau}_X \mid \mathcal{S})^{-1} \hat{\tau}_X$. By Lemma S4, we can obtain $\mathrm{cov}(\sqrt{n}\hat{\tau}_X \mid \mathcal{S}) - \mathrm{cov}(\sqrt{n}\hat{\tau}_X) = o_p(1)$. Then,

$$\begin{aligned}\sqrt{n}(\hat{\tau}-\tau) \mid \text{SRSRR} \ &\dot{\sim} \ \sqrt{n}(\hat{\tau}-\tau) \mid \sqrt{n}\hat{\tau}_X^{\mathrm{T}} V_{XX}^{-1} \sqrt{n}\hat{\tau}_X \le a_T, \sqrt{n}\hat{\delta}_W^{\mathrm{T}} V_{WW}^{-1} \sqrt{n}\hat{\delta}_W \le a_S \\ &\dot{\sim} \ A \mid B^{\mathrm{T}} V_{XX}^{-1} B \le a_T, C^{\mathrm{T}} V_{WW}^{-1} C \le a_S.\end{aligned}$$

Consider $\xi = A - V_{\tau X} V_{XX}^{-1} B - V_{\tau W} V_{WW}^{-1} C$. Accordingly,

$$\mathrm{cov}(\xi, B) = V_{\tau X} - (V_{\tau X} V_{XX}^{-1}) V_{XX} - 0 = 0,$$

$$\mathrm{cov}(\xi, C) = V_{\tau W} - 0 - (V_{\tau W} V_{WW}^{-1}) V_{WW} = 0.$$

Note that $\mathrm{cov}(B, C) = 0$. Hence, normal random variables ξ, B, C are independent. It implies that $A \mid B^{\mathrm{T}} V_{XX}^{-1} B \le a_T, C^{\mathrm{T}} V_{WW}^{-1} C \le a_S$ further equals to

$$\begin{aligned}&\xi + V_{\tau X} V_{XX}^{-1} B + V_{\tau W} V_{WW}^{-1} C \mid B^{\mathrm{T}} V_{XX}^{-1} B \le a_T, C^{\mathrm{T}} V_{WW}^{-1} C \le a_S \\ =&\xi + (V_{\tau X} V_{XX}^{-1} B \mid B^{\mathrm{T}} V_{XX}^{-1} B \le a_T) + (V_{\tau W} V_{WW}^{-1} C \mid C^{\mathrm{T}} V_{WW}^{-1} C \le a_S),\end{aligned}$$

where the above three terms are independent. The first term ξ is a normal variable with zero mean and variance $\operatorname{var}(\xi) = \operatorname{cov}(\xi, A) = V_{\tau\tau} - V_{\tau X} V_{XX}^{-1} V_{X\tau} - V_{\tau W} V_{WW}^{-1} V_{W\tau}$. Note that $V_{XX}^{-1/2} B \sim \mathcal{N}(0, I_{J_2})$, then,

$$V_{\tau X} V_{XX}^{-1} B \mid B^{\mathrm{T}} V_{XX}^{-1} B \leq a_T = V_{\tau X} V_{XX}^{-1/2} (V_{XX}^{-1/2} B) \mid (V_{XX}^{-1/2} B)^{\mathrm{T}} (V_{XX}^{-1/2} B) \leq a_T.$$

Thus, the second term follows a truncated normal distribution with scaling $V_{\tau X} V_{XX}^{-1} V_{X\tau}$ (i.e, $V_{\tau X} V_{XX}^{-1} B \mid B^{\mathrm{T}} V_{XX}^{-1} B \leq a_T \sim V_{\tau\tau}^{1/2} \sqrt{R_X^2} L_{J_2, a_T}$; see Li et al. (2018) and Yang et al. (2023)). Similarly, $V_{WW}^{-1/2} C \sim \mathcal{N}(0, I_{J_1})$,

$$V_{\tau W} V_{WW}^{-1} C \mid C^{\mathrm{T}} V_{WW}^{-1} C \leq a_S = V_{\tau W} V_{WW}^{-1/2} (V_{WW}^{-1/2} C) \mid (V_{WW}^{-1/2} C)^{\mathrm{T}} (V_{WW}^{-1/2} C) \leq a_S.$$

Thus, the third term follows a truncated normal distribution with scaling $V_{\tau W} V_{WW}^{-1} V_{W\tau}$.

□

S6.5 Proof of Corollary 2

Proof. By Theorem 3, the asymptotic variance of $\sqrt{n}(\hat{\tau} - \tau)$ under the SRSRR experiment is

$$\begin{aligned}
&\operatorname{var}\Big\{ V_{\tau\tau}^{1/2} \big(\sqrt{1 - R_W^2 - R_X^2} \cdot \varepsilon + \sqrt{R_W^2} \cdot L_{J_1, a_S} + \sqrt{R_X^2} \cdot L_{J_2, a_T} \big) \Big\} \\
=& V_{\tau\tau} \Big\{ (1 - R_W^2 - R_X^2) + R_W^2 \cdot \nu_{J_1, a_S} + R_X^2 \cdot \nu_{J_2, a_T}) \Big\} \\
=& \Big\{ 1 - (1 - \nu_{J_1, a_S}) R_W^2 - (1 - \nu_{J_2, a_T}) R_X^2 \Big\} V_{\tau\tau}.
\end{aligned}$$

By Theorem 1, the asymptotic variance of $\sqrt{n}(\hat{\tau} - \tau)$ under the stratified randomized survey experiment is $V_{\tau\tau}$. Thus, the percentage reduction in the asymptotic variance is $[100\{(1-\nu_{J_1,a_S})R_W^2+(1-\nu_{J_2,a_T})R_X^2\}]\%$.

Since the truncated normal distribution is more concentrated at zero than the normal (or truncated normal) distribution with the same or larger variance (Li et al., 2018, Lemma A3), the asymptotic distribution of $\hat{\tau}$ under the SRSRR experiment is more concentrated at τ than that under the stratified randomized survey experiment.

□

S6.6 Proof of Theorem 4

Proof. Let $V_1 \prec V_2$ denote random variable V_1 being asymptotically more concentrated at zero than V_2. By Proposition S1,

$$V_{\tau\tau}=\sum_{k=1}^{K_N}\Pi_{[k]}^2\pi_{[k]}^{-1}(e_{[k]1}^{-1}S_{[k]1}^2+e_{[k]0}^{-1}S_{[k]0}^2-f_{[k]}S_{[k]\tau}^2).$$

Recall that $\hat{V}_{\tau\tau}=\sum_{k=1}^{K_N}\Pi_{[k]}^2\pi_{[k]}^{-1}(e_{[k]1}^{-1}s_{[k]1}^2+e_{[k]0}^{-1}s_{[k]0}^2)$. By Lemma S4, we have $\hat{V}_{\tau\tau}-V_{\tau\tau}=\sum_{k=1}^{K_N}\Pi_{[k]}^2\pi_{[k]}^{-1}f_{[k]}S_{[k]\tau}^2+o_p(1)$.

Similarly, by Proposition S1 and Lemma S4, we have

$$\hat{V}_{XX}-V_{XX}=o_p(1),\quad \hat{V}_{X\tau}-V_{X\tau}=o_p(1),\quad \hat{V}_{W\tau}-V_{W\tau}=o_p(1).$$

Moreover, by definition, $\hat{V}_{\tau\tau}\hat{R}_X^2-V_{\tau\tau}R_X^2=o_p(1)$ and $\hat{V}_{\tau\tau}\hat{R}_W^2-V_{\tau\tau}R_W^2=o_p(1)$. Thus,

$$\begin{aligned}
&V_{\tau\tau}^{1/2}\Big\{\sqrt{1-R_W^2-R_X^2}\cdot\varepsilon+\sqrt{R_W^2}\cdot L_{J_1,a_S}+\sqrt{R_X^2}\cdot L_{J_2,a_T}\Big\}\\
\prec\ &\hat{V}_{\tau\tau}^{1/2}\Big\{\sqrt{1-\hat{R}_W^2-\hat{R}_X^2}\cdot\varepsilon+\sqrt{\hat{R}_W^2}\cdot L_{J_1,a_S}+\sqrt{\hat{R}_X^2}\cdot L_{J_2,a_T}\Big\}.
\end{aligned}$$

That is,

$$\hat{V}_{\tau\tau}^{1/2}\left\{\sqrt{1-\hat{R}_W^2-\hat{R}_X^2}\cdot\varepsilon+\sqrt{\hat{R}_W^2}\cdot L_{J_1,a_S}+\sqrt{\hat{R}_X^2}\cdot L_{J_2,a_T}\right\}$$

is a conservative estimator for the asymptotic distribution of $\sqrt{n}(\hat{\tau}-\tau)$ under the SRSRR experiment. By Yang et al. (2023, Lemma B21),

$$\left[\hat{\tau}-n^{-1/2}\nu_{1-\alpha/2}(\hat{V}_{\tau\tau},\hat{R}_W^2,\hat{R}_X^2),\ \hat{\tau}+n^{-1/2}\nu_{1-\alpha/2}(\hat{V}_{\tau\tau},\hat{R}_W^2,\hat{R}_X^2)\right]$$

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

S6.7 Proof of Theorem 5

Proof. We try to adjust potential outcomes to apply Theorems 1 and 3 and derive the asymptotic property. We will transform the target estimator into an easily handled one.

Let $\hat{\tau}_{\mathrm{opt}}=\hat{\tau}-\beta_{\mathrm{opt}}^{\mathrm{T}}\hat{\tau}_C-\gamma_{\mathrm{opt}}^{\mathrm{T}}\hat{\delta}_E$ be the optimal projection estimator. First, we show that $\hat{\tau}_{\mathrm{adj}}$ has the same asymptotic distribution as $\hat{\tau}_{\mathrm{opt}}$. Applying Lemma S4 to the covariates C_i and E_i, we have $\hat{\beta}-\beta_{\mathrm{opt}}=o_p(1)$ and $\hat{\gamma}-\gamma_{\mathrm{opt}}=o_p(1)$. Applying Theorem 1 to the potential outcomes $Y_i(t)$ and covariates C_i and E_i, we have

$$\hat{\tau}_C=O_p(n^{-1/2}),\quad \hat{\delta}_E=O_p(n^{-1/2}).$$

Thus, it leads to

$$\hat{\tau}_{\mathrm{adj}}-\hat{\tau}_{\mathrm{opt}}=(\beta_{\mathrm{opt}}-\hat{\beta})^{\mathrm{T}}\hat{\tau}_C+(\gamma_{\mathrm{opt}}-\hat{\gamma})^{\mathrm{T}}\hat{\delta}_E=o_p(n^{-1/2}).$$

Hence, we have

$$\sqrt{n}(\hat{\tau}_{\mathrm{adj}} - \tau) = \sqrt{n}(\hat{\tau}_{\mathrm{opt}} - \tau) + o_p(1).$$

It implies that $\hat{\tau}_{\mathrm{adj}}$ has the same asymptotic distribution as $\hat{\tau}_{\mathrm{opt}}$.

Second, we show that, under the stratified randomized survey experiment, $\sqrt{n}(\hat{\tau}_{\mathrm{opt}} - \tau)$ converges in distribution to $\mathcal{N}(0, (1 - R_E^2 - R_C^2)V_{\tau\tau})$. For $i \in [k]$, considering the linear adjusted potential outcomes $Y_i^{\dagger}(1)$and $Y_i^{\dagger}(0)$ defined by

$$Y_i^{\dagger}(1) = Y_i(1) - \beta_{\mathrm{opt}}^{\mathrm{T}} C_i - e_{[k]1}\gamma_{\mathrm{opt}}^{\mathrm{T}}(E_i - E_{[k]}),$$

$$Y_i^{\dagger}(0) = Y_i(0) - \beta_{\mathrm{opt}}^{\mathrm{T}} C_i + e_{[k]0}\gamma_{\mathrm{opt}}^{\mathrm{T}}(E_i - E_{[k]}).$$

Then $\{Y_i^{\dagger}(1)\}, \{Y_i^{\dagger}(0)\} \in \mathcal{M}_L$. Define $\hat{\tau}_{[k]}^{\dagger}$, $\tau_{[k]}^{\dagger}$, $\hat{\tau}^{\dagger}$, $\tau^{\dagger}$, $(R_X^{\dagger})^2$, $(R_W^{\dagger})^2$, and $V_{\tau\tau}^{\dagger}$ similarly to $\hat{\tau}_{[k]}$, $\tau_{[k]}$, $\hat{\tau}$, τ, R_X^2, R_W^2, and $V_{\tau\tau}$ with $\{Y_i(1)\}, \{Y_i(0)\}$ replaced by $\{Y_i(1)^{\dagger}\}, \{Y_i(0)^{\dagger}\}$. Then, for $k = 1, ..., K_N$,

$$\tau_{[k]}^{\dagger} = \frac{1}{n_{[k]}} \sum_{i \in [k]} \left\{Y_i^{\dagger}(1) - Y_i^{\dagger}(0)\right\} = \tau_{[k]}, \quad \hat{\tau}_{[k]}^{\dagger} = \hat{\tau}_{[k]} - \beta_{\mathrm{opt}}^{\mathrm{T}} \hat{\tau}_{[k]C} - \gamma_{\mathrm{opt}}^{\mathrm{T}} \hat{\delta}_{[k]E},$$

$$\tau^{\dagger} = \sum_{k=1}^{K_N} \Pi_{[k]} \tau_{[k]}^{\dagger} = \sum_{k=1}^{K_N} \Pi_{[k]} \tau_{[k]} = \tau,$$

and

$$\hat{\tau}^{\dagger} = \sum_{k=1}^{K_N} \Pi_{[k]} \hat{\tau}_{[k]}^{\dagger} = \sum_{k=1}^{K_N} \Pi_{[k]} (\hat{\tau}_{[k]} - \beta_{\mathrm{opt}}^{\mathrm{T}} \hat{\tau}_{[k]C} - \gamma_{\mathrm{opt}}^{\mathrm{T}} \hat{\delta}_{[k]E}) = \hat{\tau} - \beta_{\mathrm{opt}}^{\mathrm{T}} \hat{\tau}_C - \gamma_{\mathrm{opt}}^{\mathrm{T}} \hat{\delta}_E = \hat{\tau}_{\mathrm{opt}}.$$

Applying Theorem 1 to $Y_i^{\dagger}(1), Y_i^{\dagger}(0)$, we have

$$\sqrt{n}(\hat{\tau}_{\mathrm{opt}} - \tau) = \sqrt{n}(\hat{\tau}^{\dagger} - \tau^{\dagger}) \xrightarrow{d} \mathcal{N}(0, V_{\tau\tau}^{\dagger}).$$

Moreover, we have $V^{\dagger}_{\tau\tau} = n\mathrm{var}(\hat{\tau}_{\mathrm{opt}}) = (1 - R_E^2 - R_C^2)V_{\tau\tau}$.

Third, we show that the conclusion holds under the SRSRR experiment. Applying Theorem 3 to $Y_i^{\dagger}(1)$and $Y_i^{\dagger}(0)$, we have, under the SRSRR experiment,

$$\begin{aligned}
&\sqrt{n}(\hat{\tau}_{\mathrm{opt}} - \tau)\\
=&\sqrt{n}(\hat{\tau}^{\dagger} - \tau^{\dagger})\\
\xrightarrow{d}&(V^{\dagger}_{\tau\tau})^{1/2}\Big\{\sqrt{1-(R^{\dagger}_W)^2-(R^{\dagger}_X)^2}\cdot\varepsilon + \sqrt{(R^{\dagger}_W)^2}\cdot L_{J_1,a_S} + \sqrt{(R^{\dagger}_X)^2}\cdot L_{J_2,a_T}\Big\},
\end{aligned}$$

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

By definition,

$$V^{\dagger}_{\tau\tau}(R^{\dagger}_W)^2 = n\mathrm{cov}(\hat{\tau}^{\dagger}, \hat{\delta}_W)\mathrm{cov}(\hat{\delta}_W)^{-1}\mathrm{cov}(\hat{\delta}_W, \hat{\tau}^{\dagger}).$$

Since $\mathrm{cov}(\hat{\tau}_C, \hat{\delta}_E) = 0$, then,

$$\mathrm{cov}(\hat{\tau}^{\dagger}, \hat{\delta}_W) = \mathrm{cov}(\hat{\tau} - \beta_{\mathrm{opt}}^{\mathrm{T}}\hat{\tau}_C - \gamma_{\mathrm{opt}}^{\mathrm{T}}\hat{\delta}_E, \hat{\delta}_W) = \mathrm{cov}(\hat{\tau} - \gamma_{\mathrm{opt}}^{\mathrm{T}}\hat{\delta}_E, \hat{\delta}_W) = 0,$$

where the last equation holds because of the property of linear projection and $W_i \subset E_i$. Thus, $(R^{\dagger}_W)^2 = 0$. Similarly, we can show that $(R^{\dagger}_X)^2 = 0$.

Therefore, under the SRSRR experiment, we have

$$\sqrt{n}(\hat{\tau}_{\mathrm{adj}} - \tau) \xrightarrow{d} \mathcal{N}(0, (1 - R_E^2 - R_C^2)V_{\tau\tau}).$$

Finally, we prove that $\hat{\tau}_{\mathrm{adj}}$ (or equivalently $\hat{\tau}_{\mathrm{opt}}$) has the smallest asymptotic variance among the class of linearly adjusted estimators $\{\hat{\tau} - \beta^{\mathrm{T}}\hat{\tau}_C - \gamma^{\mathrm{T}}\hat{\delta}_E,\ \beta \in \mathbb{R}^{J_4},\ \gamma \in \mathbb{R}^{J_3}\}$.

Since $\mathrm{cov}(\hat{\tau}_C, \hat{\delta}_E) = 0$ and by the property of linear projection, we have

$$\mathrm{cov}(\hat{\tau}_{\mathrm{opt}}, \hat{\tau}_C) = \mathrm{cov}(\hat{\tau} - \beta_{\mathrm{opt}}^{\mathrm{T}} \hat{\tau}_C - \gamma_{\mathrm{opt}}^{\mathrm{T}} \hat{\delta}_E, \hat{\tau}_C) = \mathrm{cov}(\hat{\tau} - \beta_{\mathrm{opt}}^{\mathrm{T}} \hat{\tau}_C, \hat{\tau}_C) = 0.$$

Similarly, $\mathrm{cov}(\hat{\tau}_{\mathrm{opt}}, \hat{\delta}_E) = 0$. Thus, $n\mathrm{var}(\hat{\tau} - \beta^{\mathrm{T}} \hat{\tau}_C - \gamma^{\mathrm{T}} \hat{\delta}_E)$ equals to

$$\begin{aligned}
& n\mathrm{var}\Big\{\hat{\tau}_{\mathrm{opt}} - (\beta - \beta_{\mathrm{opt}})^{\mathrm{T}} \hat{\tau}_C - (\gamma - \gamma_{\mathrm{opt}})^{\mathrm{T}} \hat{\delta}_E\Big\} \\
=& (1 - R_E^2 - R_C^2) V_{\tau\tau} + (\beta - \beta_{\mathrm{opt}})^{\mathrm{T}} V_{CC} (\beta - \beta_{\mathrm{opt}}) + (\gamma - \gamma_{\mathrm{opt}})^{\mathrm{T}} V_{EE} (\gamma - \gamma_{\mathrm{opt}}) \\
\geq& (1 - R_E^2 - R_C^2) V_{\tau\tau}.
\end{aligned}$$

Thus, $\hat{\tau}_{\mathrm{opt}}$ and also $\hat{\tau}_{\mathrm{adj}}$ have the smallest asymptotic variance among the class of linearly adjusted estimators $\{\hat{\tau} - \beta^{\mathrm{T}} \hat{\tau}_C - \gamma^{\mathrm{T}} \hat{\delta}_E,\ \beta \in \mathbb{R}^{J_4},\ \gamma \in \mathbb{R}^{J_3}\}$. □

S6.8 Proof of Theorem 6

Proof. By Lemma S4, we have shown that

$$\hat{V}_{\tau\tau} - V_{\tau\tau} = \sum_{k=1}^{K_N} \Pi_{[k]}^2 \pi_{[k]}^{-1} f_{[k]} S_{[k]\tau}^2 + o_p(1).$$

Moreover, applying Lemma S4 to the covariates E_i and C_i, we have

$$\hat{V}_{\tau\tau} \hat{R}_E^2 - V_{\tau\tau} R_E^2 = o_p(1), \quad \hat{V}_{\tau\tau} \hat{R}_C^2 - V_{\tau\tau} R_C^2 = o_p(1).$$

Therefore, we have $\hat{V}_{\tau\tau}(1 - \hat{R}_E^2 - \hat{R}_C^2) - V_{\tau\tau}(1 - R_E^2 - R_C^2) = \sum_{k=1}^{K_N} \Pi_{[k]}^2 \pi_{[k]}^{-1} f_{[k]} S_{[k]\tau}^2 + o_p(1)$. Accordingly, the confidence interval

$$\Big[\hat{\tau}_{\mathrm{adj}} - n^{-1/2} q_{1-\alpha/2} \hat{V}_{\tau\tau}^{1/2} \sqrt{1 - \hat{R}_E^2 - \hat{R}_C^2},\ \hat{\tau}_{\mathrm{adj}} + n^{-1/2} q_{1-\alpha/2} \hat{V}_{\tau\tau}^{1/2} \sqrt{1 - \hat{R}_E^2 - \hat{R}_C^2}\Big]$$

has an asymptotic coverage rate greater than or equal to $1-\alpha$.

For the second part of this theorem, we utilize Lemma S5 below, which has been established by Yang et al. (2023). We omit its proof here.

Lemma S5 (Yang et al. (2023), Lemma B15)**.** *For any positive integers K_1, K_2 and constants a_1, a_2, suppose that $\varepsilon, L_{K_1,a_1}, L_{K_2,a_2}$ are mutually independent. Then, for any nonnegative constants $b_0 \le \bar{b}_0$, $b_1 \le \bar{b}_1$, $b_2 \le \bar{b}_2$, and any constant $c > 0$,*

$$\mathrm{pr}\Big(|b_0\varepsilon + b_1 L_{K_1,a_1} + b_2 L_{K_2,a_2}| \le c\Big) \le \mathrm{pr}\Big(|\bar{b}_0\varepsilon + \bar{b}_1 L_{K_1,a_1} + \bar{b}_2 L_{K_2,a_2}| \le c\Big).$$

Since $W_i \subset E_i$ and $X_i \subset C_i$, then, $R_E^2 \ge R_W^2$ and $R_C^2 \ge R_X^2$. Taking $b_0 = V_{\tau\tau}^{1/2}\sqrt{1-R_E^2-R_C^2}$, $\bar{b}_0 = V_{\tau\tau}^{1/2}\sqrt{1-R_W^2-R_X^2}$, $b_1 = b_2 = 0$, $\bar{b}_1 = V_{\tau\tau}^{1/2}\sqrt{R_W^2}$ and $\bar{b}_2 = V_{\tau\tau}^{1/2}\sqrt{R_X^2}$, the asymptotic distribution of $\hat{\tau}_{\mathrm{adj}}$ is more concentrated around τ than $\hat{\tau}$ under the SRSRR experiment. Moreover, by the convergence of $\hat{V}_{\tau\tau}$, $\hat{V}_{\tau\tau}\hat{R}_E^2$, $\hat{V}_{\tau\tau}\hat{R}_C^2$, $\hat{V}_{\tau\tau}\hat{R}_W^2$, and $\hat{V}_{\tau\tau}\hat{R}_X^2$, the proposed confidence interval in Theorem 6 is asymptotically shorter than, at least as short as, the confidence interval based on $\hat{\tau}$ in Theorem 4.

$\square$

S6.9 Proof of Theorem S1

Proof. By Theorem 3, under the SRSRR experiment, we have

$$\mathrm{var}(\sqrt{n}\hat{\tau}) = V_{\tau\tau} - (1-\nu_{J_1,a_S})V_{\tau W}V_{WW}^{-1}V_{W\tau} - (1-\nu_{J_2,a_T})V_{\tau X}V_{XX}^{-1}V_{X\tau}.$$

To prove (i), by Proposition S1, we have

$$\frac{\partial}{\partial(1/f_{[k]})}V_{\tau\tau} = \Pi_{[k]}f(e_{[k]1}^{-1}S_{[k]1}^2 + e_{[k]0}^{-1}S_{[k]0}^2),$$

$$\frac{\partial}{\partial(1/f_{[k]})}V_{\tau W} = \Pi_{[k]}fS_{[k]\tau W}, \quad \frac{\partial}{\partial(1/f_{[k]})}V_{WW} = \Pi_{[k]}fS_{[k]W}^2,$$

$$\frac{\partial}{\partial(1/f_{[k]})}V_{\tau X} = \Pi_{[k]}f(e_{[k]1}^{-1}S_{[k]1,X}+e_{[k]0}^{-1}S_{[k]0,X}), \quad \frac{\partial}{\partial(1/f_{[k]})}V_{XX} = \Pi_{[k]}f(e_{[k]1}e_{[k]0})^{-1}S_{[k]X}^2.$$

Since $f_{[k]}$ satisfies $\sum_{k=1}^{K_N}\Pi_{[k]}f_{[k]} = f$, then, considering the Lagrange function:

$$\mathrm{var}(\sqrt{n}\hat{\tau}) + \lambda\Big(\sum_{k=1}^{K_N}\Pi_{[k]}f_{[k]} - f\Big).$$

For the optimal $\{f_{[k]}\}_{k=1}^{K_N}$, we have

$$\begin{aligned}
0 =&\frac{\partial}{\partial(1/f_{[k]})}\Big\{\mathrm{var}(\sqrt{n}\hat{\tau}) + \lambda\Big(\sum_{k=1}^{K_N}\Pi_{[k]}f_{[k]} - f\Big)\Big\}\\
=&\Pi_{[k]}f(e_{[k]1}^{-1}S_{[k]1}^2 + e_{[k]0}^{-1}S_{[k]0}^2) - 2(1-\nu_{J_1,a_S})\Pi_{[k]}fS_{[k]\tau W}V_{WW}^{-1}V_{W\tau}\\
&+(1-\nu_{J_1,a_S})\Pi_{[k]}fV_{\tau W}V_{WW}^{-1}S_{[k]W}^2V_{WW}^{-1}V_{W\tau}\\
&-2(1-\nu_{J_2,a_T})\Pi_{[k]}f(e_{[k]1}^{-1}S_{[k]1,X} + e_{[k]0}^{-1}S_{[k]0,X})V_{XX}^{-1}V_{X\tau}\\
&+(1-\nu_{J_2,a_T})\Pi_{[k]}f(e_{[k]1}e_{[k]0})^{-1}V_{\tau X}V_{XX}^{-1}S_{[k]X}^2V_{XX}^{-1}V_{X\tau} - \lambda\Pi_{[k]}f_{[k]}^2.
\end{aligned}$$

Hence,

$$\Big(\frac{|\lambda|}{f}\Big)^{1/2} = \frac{A_k}{f_{[k]}} = \frac{\Pi_{[k]}A_k}{\Pi_{[k]}f_{[k]}} = \frac{\sum_{k=1}^{K_N}\Pi_{[k]}A_k}{\sum_{k=1}^{K_N}\Pi_{[k]}f_{[k]}} = \frac{\sum_{k=1}^{K_N}\Pi_{[k]}A_k}{f}.$$

To prove (ii) of the theorem, we have

$$\frac{\partial}{\partial(e_{[k]t})}V_{\tau\tau} = -\Pi_{[k]}^2\pi_{[k]}^{-1}e_{[k]t}^{-2}S_{[k]t}^2,$$

$$\frac{\partial}{\partial(e_{[k]t})}V_{\tau X} = -\Pi_{[k]}^2\pi_{[k]}^{-1}e_{[k]t}^{-2}S_{[k]t,X}, \quad \frac{\partial}{\partial(e_{[k]t})}V_{XX} = -\Pi_{[k]}^2\pi_{[k]}^{-1}e_{[k]t}^{-2}S_{[k]X}^2, \quad t = 0, 1.$$

Since $e_{[k]1}, e_{[k]0}$ satisfies $e_{[k]1} + e_{[k]0} = 1$, considering the Lagrange function

$$\mathrm{var}(\sqrt{n}\hat{\tau}) + \sum_{k=1}^{K_N} \lambda_k(e_{[k]1} + e_{[k]0} - 1).$$

For the optimal $\{e_{[k]t}\}_{k=1}^{K_N}$, $t = 0, 1$, we have

$$\begin{aligned}
0 =& \frac{\partial}{\partial e_{[k]t}}\Big\{\mathrm{var}(\sqrt{n}\hat{\tau}) + \sum_{k=1}^{K_N} \lambda_k(e_{[k]1} + e_{[k]0} - 1)\Big\} \\
=& \frac{\Pi_{[k]}^2}{\pi_{[k]}}\Big(-\frac{S_{[k]t}^2}{e_{[k]t}^2}\Big) + 2(1-\nu_{J_2,a_T})\frac{\Pi_{[k]}^2}{\pi_{[k]}}\frac{S_{[k]t,X}}{e_{[k]t}^2}V_{XX}^{-1}V_{X\tau} \\
&- (1-\nu_{J_2,a_T})V_{\tau X}V_{XX}^{-1}\frac{\Pi_{[k]}^2}{\pi_{[k]}}\frac{S_{[k]t}^2}{e_{[k]t}^2}V_{XX}^{-1}V_{X\tau} + \lambda_k.
\end{aligned}$$

Hence, we have

$$\begin{aligned}
\frac{e_{[k]1}}{e_{[k]0}} =& \Big\{\Big|\frac{S_{[k]1}^2 + (1-\nu_{J_2,a_T})V_{\tau X}V_{XX}^{-1}S_{[k]X}^2V_{XX}^{-1}V_{X\tau} - 2(1-\nu_{J_2,a_T})S_{[k]1,X}V_{XX}^{-1}V_{X\tau}}{S_{[k]0}^2 + (1-\nu_{J_2,a_T})V_{\tau X}V_{XX}^{-1}S_{[k]X}^2V_{XX}^{-1}V_{X\tau} - 2(1-\nu_{J_2,a_T})S_{[k]0,X}V_{XX}^{-1}V_{X\tau}}\Big|\Big\}^{1/2} \\
=& \frac{a_{[k]1}}{a_{[k]0}}.
\end{aligned}$$

Thus,

$$e_{[k]1} = (a_{[k]1})/(a_{[k]1} + a_{[k]0}).$$

□

S6.10 Proof of Proposition S1

Proof. First, we show that $\hat{\tau}$ is an unbiased estimator of τ. Since $E\hat{\tau}_{[k]} = \tau_{[k]}$, then

$$E\hat{\tau} = \sum_{k=1}^{K_N} \Pi_{[k]}E\hat{\tau}_{[k]} = \sum_{k=1}^{K_N} \Pi_{[k]}\tau_{[k]} = \tau.$$

Next, we compute the covariance of $\hat{\tau}$. Simple calculation gives

$$\mathrm{cov}(\bar{W}_{\mathcal{S}} - \bar{W}) = \mathrm{cov}\Big(\sum_{k=1}^{K_N} \Pi_{[k]} \bar{W}_{[k]\mathcal{S}}\Big) = \mathrm{cov}\Big\{\sum_{k=1}^{K_N} \Pi_{[k]} \frac{1}{n_{[k]}} \sum_{i\in[k]} Z_i(W_i - \bar{W}_{[k]})\Big\}$$

$$= \sum_{k=1}^{K_N} \frac{\Pi_{[k]}^2}{n_{[k]}^2} \mathrm{cov}\Big\{\sum_{i\in[k]} Z_i(W_i - \bar{W}_{[k]})\Big\}$$

$$+ \sum_{k_1\neq k_2} \frac{\Pi_{[k_1]}}{n_{[k_1]}} \frac{\Pi_{[k_2]}}{n_{[k_2]}} \mathrm{cov}\Big\{\sum_{i\in[k_1]} Z_i(W_i - \bar{W}_{[k_1]}), \sum_{j\in[k_2]} Z_j(W_j - \bar{W}_{[k_2]})\Big\}.$$

For $1 \le k \le K_N$, $\mathrm{cov}\{\sum_{i\in[k]} Z_i(W_i - \bar{W}_{[k]})\}$ can be further expressed as

$$E\Big[\Big\{\sum_{i\in[k]} Z_i(W_i - \bar{W}_{[k]})\Big\}\Big\{\sum_{i\in[k]} Z_i(W_i - \bar{W}_{[k]})^{\mathrm{T}}\Big\}\Big]$$

$$= \sum_{i\in[k]} E\big\{Z_i(W_i - \bar{W}_{[k]})(W_i - \bar{W}_{[k]})^{\mathrm{T}}\big\} + \sum_{i\neq j; i,j\in[k]} E\big\{Z_iZ_j(W_i - \bar{W}_{[k]})(W_j - \bar{W}_{[k]})^{\mathrm{T}}\big\}$$

$$= \sum_{i\in[k]} \frac{n_{[k]}}{N_{[k]}}(W_i - \bar{W}_{[k]})(W_i - \bar{W}_{[k]})^{\mathrm{T}} + \sum_{i\neq j; i,j\in[k]} \frac{n_{[k]}}{N_{[k]}} \frac{n_{[k]}-1}{N_{[k]}-1}(W_i - \bar{W}_{[k]})(W_j - \bar{W}_{[k]})^{\mathrm{T}}$$

$$= \sum_{i\in[k]} \Big(\frac{n_{[k]}}{N_{[k]}} - \frac{n_{[k]}}{N_{[k]}} \frac{n_{[k]}-1}{N_{[k]}-1}\Big)(W_i - \bar{W}_{[k]})(W_i - \bar{W}_{[k]})^{\mathrm{T}} = n_{[k]}\Big(1 - \frac{n_{[k]}}{N_{[k]}}\Big) S_{[k]W}^2,$$

where the second to last equality is due to

$$0 = \sum_{i,j\in[k]} (W_i - \bar{W}_{[k]})(W_j - \bar{W}_{[k]})^{\mathrm{T}}$$

$$= \sum_{i\in[k]} (W_i - \bar{W}_{[k]})(W_i - \bar{W}_{[k]})^{\mathrm{T}} + \sum_{i\neq j; i,j\in[k]} (W_i - \bar{W}_{[k]})(W_j - \bar{W}_{[k]})^{\mathrm{T}}.$$

For $k_1 \neq k_2$, because the sampling is independent across strata, we have

$$\mathrm{cov}\Big\{\sum_{i\in[k_1]} Z_i(W_i - \bar{W}_{[k_1]}), \sum_{j\in[k_2]} Z_j(W_j - \bar{W}_{[k_2]})\Big\} = 0.$$

Hence,

$$\mathrm{cov}(\hat{\delta}_W) = \mathrm{cov}(\bar{W}_{\mathcal{S}} - \bar{W}) = \sum_{k=1}^{K_N} \Pi_{[k]}^2 \Big(\frac{1}{n_{[k]}} - \frac{1}{N_{[k]}}\Big) S_{[k]W}^2.$$

Moreover, $\mathrm{cov}(\hat{\tau}_X \mid \mathcal{S}) = \mathrm{cov}\{\sum_{k=1}^{K_N} \Pi_{[k]}(\bar{X}_{[k]1} - \bar{X}_{[k]0} \mid \mathcal{S})\}$, which further equals to

$$\begin{aligned}
&\sum_{k=1}^{K_N} \Pi_{[k]}^2 \mathrm{cov}\Big\{\frac{1}{n_{[k]1}} \sum_{i\in[k], i\in\mathcal{S}} X_i T_i - \frac{1}{n_{[k]0}} \sum_{i\in[k], i\in\mathcal{S}} X_i(1-T_i) \mid \mathcal{S}\Big\} \\
=&\sum_{k=1}^{K_N} \Pi_{[k]}^2 \mathrm{cov}\Big\{\sum_{i\in[k], i\in\mathcal{S}} \Big(\frac{1}{n_{[k]1}} + \frac{1}{n_{[k]0}}\Big) T_i(X_i - \bar{X}_{[k]}) \mid \mathcal{S}\Big\} \\
=&\sum_{k=1}^{K_N} \Pi_{[k]}^2 \frac{n_{[k]}^2}{n_{[k]1}^2 n_{[k]0}^2} \mathrm{cov}\Big\{\sum_{i\in[k], i\in\mathcal{S}} T_i(X_i - \bar{X}_{[k]}) \mid \mathcal{S}\Big\} \\
=&\sum_{k=1}^{K_N} \Pi_{[k]}^2 \frac{n_{[k]}^2}{n_{[k]1}^2 n_{[k]0}^2} n_{[k]1} \frac{n_{[k]} - n_{[k]1}}{n_{[k]}} S_{[k]X|\mathcal{S}}^2 = \sum_{k=1}^{K_N} \Pi_{[k]}^2 \frac{n_{[k]}}{n_{[k]1} n_{[k]0}} S_{[k]X|\mathcal{S}}^2.
\end{aligned}$$

Thus, we have

$$\mathrm{cov}(\hat{\tau}_X) = \sum_{k=1}^{K_N} \Pi_{[k]}^2 \frac{n_{[k]}}{n_{[k]1} n_{[k]0}} S_{[k]X}^2.$$

Similarly, we can compute the covariances between $\hat{\tau} - \tau$, $\hat{\tau}_X$, and $\hat{\delta}_W$, and obtain

$$\mathrm{cov}(\sqrt{n}(\hat{\tau} - \tau, \hat{\tau}_X^{\mathrm{T}}, \hat{\delta}_W^{\mathrm{T}})^{\mathrm{T}}) = \sum_{k=1}^{K_N} \Pi_{[k]}^2 \pi_{[k]}^{-1}\cdot$$

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

□

S6.11 Proof of Theorem S2

Proof. By Theorem 5, under the SRSRR experiment (or the stratified randomized survey experiment),

$$\mathrm{var}(\sqrt{n}\hat{\tau}_{\mathrm{adj}}) = V_{\tau\tau} - V_{\tau E}V_{EE}^{-1}V_{E\tau} - V_{\tau C}V_{CC}^{-1}V_{C\tau}.$$

For (i), by Proposition S1, we have

$$\frac{\partial}{\partial(1/f_{[k]})}V_{\tau\tau} = \Pi_{[k]}f(e_{[k]1}^{-1}S_{[k]1}^2 + e_{[k]0}^{-1}S_{[k]0}^2)$$

$$\frac{\partial}{\partial(1/f_{[k]})}V_{\tau E} = \Pi_{[k]}fS_{[k]\tau E}, \quad \frac{\partial}{\partial(1/f_{[k]})}V_{EE} = \Pi_{[k]}fS_{[k]E}^2.$$

$$\frac{\partial}{\partial(1/f_{[k]})}V_{\tau C} = \Pi_{[k]}f(e_{[k]1}^{-1}S_{[k]1,C}+e_{[k]0}^{-1}S_{[k]0,C}), \quad \frac{\partial}{\partial(1/f_{[k]})}V_{CC} = \Pi_{[k]}f(e_{[k]1}e_{[k]0})^{-1}S_{[k]C}^2.$$

Since $\sum_{k=1}^{K_N}\Pi_{[k]}f_{[k]} = f$, we consider the Lagrange function:

$$\mathrm{var}(\sqrt{n}\hat{\tau}_{\mathrm{adj}}) + \lambda\Big(\sum_{k=1}^{K_N}\Pi_{[k]}f_{[k]} - f\Big).$$

For the optimal $\{f_{[k]}\}_{k=1}^{K_N}$, we have

$$\begin{aligned}
0 =&\frac{\partial}{\partial(1/f_{[k]})}\Big\{\mathrm{var}(\sqrt{n}\hat{\tau}_{\mathrm{adj}}) + \lambda(\sum_{k=1}^{K_N}\Pi_{[k]}f_{[k]} - f)\Big\}\\
=&\Pi_{[k]}f(e_{[k]1}^{-1}S_{[k]1}^2 + e_{[k]0}^{-1}S_{[k]0}^2) - 2\Pi_{[k]}fS_{[k]\tau E}V_{EE}^{-1}V_{E\tau} + \Pi_{[k]}fV_{\tau E}V_{EE}^{-1}S_{[k]E}^2V_{EE}^{-1}V_{E\tau}\\
&- 2\Pi_{[k]}f(e_{[k]1}^{-1}S_{[k]1,C} + e_{[k]0}^{-1}S_{[k]0,C})V_{CC}^{-1}V_{C\tau}\\
&+ \Pi_{[k]}f(e_{[k]1}e_{[k]0})^{-1}V_{\tau C}V_{CC}^{-1}S_{[k]C}^2V_{CC}^{-1}V_{C\tau} - \lambda\Pi_{[k]}f_{[k]}^2.
\end{aligned}$$

Hence, it leads to

$$\Big(\frac{|\lambda|}{f}\Big)^{1/2} = \frac{B_k}{f_{[k]}} = \frac{\Pi_{[k]}B_k}{\Pi_{[k]}f_{[k]}} = \frac{\sum_{k=1}^{K_N}\Pi_{[k]}B_k}{\sum_{k=1}^{K_N}\Pi_{[k]}f_{[k]}} = \frac{\sum_{k=1}^{K_N}\Pi_{[k]}B_k}{f}.$$

For (ii), we have

$$\frac{\partial}{\partial(e_{[k]t})}V_{\tau\tau} = -\Pi_{[k]}^2\pi_{[k]}^{-1}e_{[k]t}^{-2}S_{[k]t}^2$$

$$\frac{\partial}{\partial(e_{[k]t})}V_{\tau C} = -\Pi_{[k]}^2\pi_{[k]}^{-1}e_{[k]t}^{-2}S_{[k]t,C}, \quad \frac{\partial}{\partial(e_{[k]t})}V_{CC} = -\Pi_{[k]}^2\pi_{[k]}^{-1}e_{[k]t}^{-2}S_{[k]C}^2.$$

Since $e_{[k]1} + e_{[k]0} = 1$, we consider the Lagrange function

$$\mathrm{var}(\sqrt{n}\hat{\tau}_{\mathrm{adj}}) + \sum_{k=1}^{K_N}\lambda_k(e_{[k]1} + e_{[k]0} - 1).$$

For the optimal $\{e_{[k]t}\}_{k=1}^{K_N}$, $t = 0, 1$, we have

$$\begin{aligned}
0 =& \frac{\partial}{\partial e_{[k]t}}\Big\{\mathrm{var}(\sqrt{n}\hat{\tau}_{\mathrm{adj}}) + \sum_{k=1}^{K_N}\lambda_k(e_{[k]1} + e_{[k]0} - 1)\Big\} \\
=& \frac{\Pi_{[k]}^2}{\pi_{[k]}}(-\frac{S_{[k]t}^2}{e_{[k]t}^2}) + 2\frac{\Pi_{[k]}^2}{\pi_{[k]}}\frac{S_{[k]t,C}}{e_{[k]t}^2}V_{CC}^{-1}V_{C\tau} - V_{\tau C}V_{CC}^{-1}\frac{\Pi_{[k]}^2}{\pi_{[k]}}\frac{S_{[k]t}^2}{e_{[k]t}^2}V_{CC}^{-1}V_{C\tau} + \lambda_k.
\end{aligned}$$

We have

$$\frac{e_{[k]1}}{e_{[k]0}} = \Big(\Big|\frac{S_{[k]1}^2 + V_{\tau C}V_{CC}^{-1}S_{[k]C}^2V_{CC}^{-1}V_{C\tau} - 2S_{[k]1,C}V_{CC}^{-1}V_{C\tau}}{S_{[k]0}^2 + V_{\tau C}V_{CC}^{-1}S_{[k]C}^2V_{CC}^{-1}V_{C\tau} - 2S_{[k]0,C}V_{CC}^{-1}V_{C\tau}}\Big|\Big)^{1/2} = \frac{b_{[k]1}}{b_{[k]0}}.$$

Thus,

$$e_{[k]1} = b_{[k]1}/(b_{[k]1} + b_{[k]0}).$$

$\square$

S6.12 Proof of Theorem S3

Proof. Recall that

$$\mathcal{M} = \{(Z, T_{\mathcal{S}}) : \text{ satisfying } M_S \le a_S, M_T \le a_T \text{ under SRSE}\}$$

is all possible sampling and treatment assignment vectors under $\widetilde{\text{SRSRR}}$. Define the acceptable treatment assignment vector when the sampling vector takes value z as

$$\mathcal{M}_2(z) = \{t : (z,t) \in \mathcal{M}\},$$

and the set of acceptable sampling under $\widetilde{\text{SRSRR}}$ as

$$\mathcal{M}_1' = \Big\{z : (z,t) \in \mathcal{M} \text{ for some } t \in \{0,1\}^n \text{ and } \sum_{i \in [k] \cap \mathcal{S}} t_i = n_{[k]1}, k = 1, \dots, K_N\Big\}.$$

Furthermore, if we only consider the sampling stage, we can define $\mathcal{M}_1$ as the sampling indicators set such that the corresponding $M_S \le a_S$.

When $\mathcal{M} = \emptyset$, Theorem S3 holds. Below we consider only the case when $\mathcal{M} \neq \emptyset$. We first consider the difference between $\widetilde{\text{SRSRR}}$ and SRSRR for sampling and assignment vectors in $\mathcal{M}$. Applying the property of total variation distance for discrete measures, we need to bound $|\text{pr}(Z = z, T_{\mathcal{S}} = t) \mid \widetilde{\text{SRSRR}}) - \text{pr}(Z = z, T_{\mathcal{S}} = t) \mid \text{SRSRR})|$ under $(z,t) \in \mathcal{M}$ and $(z,t) \notin \mathcal{M}$, respectively.

For convenience, we denote the combinatorial number

$$\binom{N_{[k]}}{n_{[k]1}, n_{[k]0}} = \frac{N_{[k]}!}{(N_{[k]} - n_{[k]})! n_{[k]1}! n_{[k]0}!}.$$

For any $(z,t) \in \mathcal{M}$, we can express the conditional probability under $\widetilde{\text{SRSRR}}$

$$\text{pr}(Z = z, T_{\mathcal{S}} = t \mid \widetilde{\text{SRSRR}}) = \frac{1}{|\mathcal{M}|} = \frac{1}{\prod_{k=1}^{K_N} \binom{N_{[k]}}{n_{[k]1}, n_{[k]0}} \cdot \text{pr}(M_T \le a_T, M_S \le a_S)},$$

and the conditional probability under SRSRR

$$\begin{aligned}
&\operatorname{pr}(Z=z, T_{\mathcal{S}}=t \mid \mathrm{SRSRR}) \\
&= \operatorname{pr}(Z=z \mid \mathrm{SRSRR}) \cdot \operatorname{pr}(T_{\mathcal{S}}=t \mid Z=z, \mathrm{SRSRR}) = \frac{1}{|\mathcal{M}_1|} \cdot \frac{1}{|\mathcal{M}_2(z)|} \\
&= \frac{1}{\prod_{k=1}^{K_N} \binom{N_{[k]}}{n_{[k]}} \cdot \operatorname{pr}(M_S \le a_S)} \cdot \frac{1}{\prod_{k=1}^{K_N} \binom{n_{[k]}}{n_{[k]1}} \cdot \operatorname{pr}(M_T \le a_T \mid Z=z)} \\
&= \frac{1}{\prod_{k=1}^{K_N} \binom{N_{[k]}}{n_{[k]1}, n_{[k]0}} \cdot \operatorname{pr}(M_S \le a_S) \cdot \operatorname{pr}(M_T \le a_T \mid Z=z)}.
\end{aligned}$$

Therefore, for any $(z,t) \in \mathcal{M}$, the difference of conditional probability under the two designs can be expressed as

$$\begin{aligned}
&\operatorname{pr}(Z=z, T_{\mathcal{S}}=t \mid \widetilde{\mathrm{SRSRR}}) - \operatorname{pr}(Z=z, T_{\mathcal{S}}=t \mid \mathrm{SRSRR}) \\
&= \frac{\operatorname{pr}(M_T \le a_T \mid Z=z) - \operatorname{pr}(M_T \le a_T \mid M_S \le a_S)}{\prod_{k=1}^{K_N} \binom{N_{[k]}}{n_{[k]1}, n_{[k]0}} \cdot \operatorname{pr}(M_T \le a_T, M_S \le a_S) \cdot \operatorname{pr}(M_T \le a_T \mid Z=z)}.
\end{aligned}$$

Combine the result of all $(z,t) \in \mathcal{M}$, then we have

$$\begin{aligned}
&\sum_{(z,t)\in\mathcal{M}} \left| \operatorname{pr}(Z=z, T_{\mathcal{S}}=t \mid \widetilde{\mathrm{SRSRR}}) - \operatorname{pr}(Z=z, T_{\mathcal{S}}=t \mid \mathrm{SRSRR}) \right| \\
&= \sum_{z\in\mathcal{M}_1'} |\mathcal{M}_2(z)| \cdot \frac{|\operatorname{pr}(M_T \le a_T \mid Z=z) - \operatorname{pr}(M_T \le a_T \mid M_S \le a_S)|}{\prod_{k=1}^{K_N} \binom{N_{[k]}}{n_{[k]1}, n_{[k]0}} \cdot \operatorname{pr}(M_T \le a_T, M_S \le a_S) \cdot \operatorname{pr}(M_T \le a_T \mid Z=z)} \\
&= \sum_{z\in\mathcal{M}_1'} \prod_{k=1}^{K_N} \binom{n_{[k]}}{n_{[k]1}} \cdot \operatorname{pr}(M_T \le a_T \mid Z=z) \\
&\quad \cdot \frac{|\operatorname{pr}(M_T \le a_T \mid Z=z) - \operatorname{pr}(M_T \le a_T \mid M_S \le a_S)|}{\prod_{k=1}^{K_N} \binom{N_{[k]}}{n_{[k]1}, n_{[k]0}} \cdot \operatorname{pr}(M_T \le a_T, M_S \le a_S) \cdot \operatorname{pr}(M_T \le a_T \mid Z=z)}
\end{aligned}$$

$$= \frac{1}{\mathrm{pr}(M_T \le a_T, M_S \le a_S)} \cdot \frac{1}{\prod_{k=1}^{K_N} \binom{N_{[k]}}{n_{[k]}}} \sum_{z \in \mathcal{M}_1'} |\mathrm{pr}(M_T \le a_T \mid Z = z)$$

$$- \mathrm{pr}(M_T \le a_T \mid M_S \le a_S)|.$$

Because $\mathcal{M}_1' \subset \{z \in \{0,1\}^N : \sum_{i \in [k]} z_i = n_{[k]}\}$, we have

$$\sum_{(z,t) \in \mathcal{M}} \left| \mathrm{pr}(Z = z, T_{\mathcal{S}} = t \mid \widetilde{\mathrm{SRSRR}}) - \mathrm{pr}(Z = z, T_{\mathcal{S}} = t \mid \mathrm{SRSRR}) \right|$$

$$\le \frac{1}{\mathrm{pr}(M_T \le a_T, M_S \le a_S)} \cdot \frac{1}{\prod_{k=1}^{K_N} \binom{N_{[k]}}{n_{[k]}}}$$

$$\sum_{z \in \{0,1\}^N : \sum_{i \in [k]} z_i = n_{[k]}} |\mathrm{pr}(M_T \le a_T \mid Z = z) - \mathrm{pr}(M_T \le a_T \mid M_S \le a_S)|$$

$$= \frac{1}{\mathrm{pr}(M_T \le a_T, M_S \le a_S)} \cdot E \left| \mathrm{pr}(M_T \le a_T \mid Z) - \mathrm{pr}(M_T \le a_T \mid M_S \le a_S) \right|.$$

For the difference between $\widetilde{\mathrm{SRSRR}}$ and SRSRR for sampling and assignment vectors not in $\mathcal{M}$, we have

$$\sum_{(z,t) \notin \mathcal{M}} \left| \mathrm{pr}(Z = z, T_{\mathcal{S}} = t \mid \widetilde{\mathrm{SRSRR}}) - \mathrm{pr}(Z = z, T_{\mathcal{S}} = t \mid \mathrm{SRSRR}) \right|$$

$$= \sum_{(z,t) \notin \mathcal{M}} \mathrm{pr}(Z = z, T_{\mathcal{S}} = t \mid \mathrm{SRSRR}) = 1 - \sum_{(z,t) \in \mathcal{M}} \mathrm{pr}(Z = z, T_{\mathcal{S}} = t \mid \mathrm{SRSRR})$$

$$= \sum_{(z,t) \in \mathcal{M}} \mathrm{pr}(Z = z, T_{\mathcal{S}} = t \mid \widetilde{\mathrm{SRSRR}}) - \sum_{(z,t) \in \mathcal{M}} \mathrm{pr}(Z = z, T_{\mathcal{S}} = t \mid \mathrm{SRSRR})$$

$$\le \sum_{(z,t) \in \mathcal{M}} \left| \mathrm{pr}(Z = z, T_{\mathcal{S}} = t \mid \widetilde{\mathrm{SRSRR}}) - \mathrm{pr}(Z = z, T_{\mathcal{S}} = t \mid \mathrm{SRSRR}) \right|.$$

Combining the result for $(z,t) \in \mathcal{M}$ and $(z,t) \notin \mathcal{M}$, we can bound the total variation distance between the two designs as follows:

$$
\begin{aligned}
& \mathrm{d_{TV}}(\widetilde{\mathrm{SRSRR}}, \mathrm{SRSRR}) \\
&= \frac{1}{2} \sum_{(z,t)\in\mathcal{M}} \Big| \mathrm{pr}(Z = z, T_{\mathcal{S}} = t \mid \widetilde{\mathrm{SRSRR}}) - \mathrm{pr}(Z = z, T_{\mathcal{S}} = t \mid \mathrm{SRSRR}) \Big| \\
&\quad + \frac{1}{2} \sum_{(z,t)\notin\mathcal{M}} \Big| \mathrm{pr}(Z = z, T_{\mathcal{S}} = t \mid \widetilde{\mathrm{SRSRR}}) - \mathrm{pr}(Z = z, T_{\mathcal{S}} = t \mid \mathrm{SRSRR}) \Big| \\
&\leq \sum_{(z,t)\in\mathcal{M}} \Big| \mathrm{pr}(Z = z, T_{\mathcal{S}} = t \mid \widetilde{\mathrm{SRSRR}}) - \mathrm{pr}(Z = z, T_{\mathcal{S}} = t \mid \mathrm{SRSRR}) \Big| \\
&\leq \frac{E|\mathrm{pr}(M_T \leq a_T \mid Z) - \mathrm{pr}(M_T \leq a_T \mid M_S \leq a_S)|}{\mathrm{pr}(M_T \leq a_T, M_S \leq a_S)}.
\end{aligned}
$$

Therefore, Theorem S3 holds. □

S6.13 Proof of Corollary S1

Proof. By Theorem 1, as $n \to \infty$, we have $\mathrm{pr}(M_S \leq a_S, M_T \leq a_T) \to \mathrm{pr}(\chi^2_{J_1} \leq a_S)\mathrm{pr}(\chi^2_{J_2} \leq a_T) > 0$ and $\mathrm{pr}(M_S \leq a_S) \to \mathrm{pr}(\chi^2_{J_1} \leq a_S)$. These imply that $\mathrm{pr}(M_T \leq a_T \mid M_S \leq a_S) \to \mathrm{pr}(\chi^2_{J_2} \leq a_T)$, and $\mathcal{M}$ is not an empty set when N is sufficient large. Consequently, $1(\mathcal{M} = \emptyset) \to 0$ as $n \to \infty$.

When Z is drawn from SRSE, given $\mathcal{S}$, applying CLT with $f = 1$ and considering the whole population as $\mathcal{S}$, we have $\mathrm{pr}(M_T \leq a_T \mid Z) \xrightarrow{p} \mathrm{pr}(\chi^2_{J_2} \leq a_T)$ as $n \to \infty$. Then, we have $|\mathrm{pr}(M_T \leq a_T \mid Z) - \mathrm{pr}(M_T \leq a_T \mid M_S \leq a_S)| \xrightarrow{p} 0$. Because $|\mathrm{pr}(M_T \leq a_T \mid Z) - \mathrm{pr}(M_T \leq a_T \mid M_S \leq a_S)|$ is upper bounded by 2, by Lebesgue's dominated convergence theorem, as $n \to \infty$, we have

$$E|\mathrm{pr}(M_T \leq a_T \mid Z) - \mathrm{pr}(M_T \leq a_T \mid M_S \leq a_S)| \to 0.$$

By Theorem S3, as $n \to \infty$,

$$\begin{aligned} \mathrm{d_{TV}}(\widetilde{\mathrm{SRSRR}}, \mathrm{SRSRR}) &\leq 1(\mathcal{M} = \emptyset) + \frac{E\,|\mathrm{pr}(M_T \leq a_T \mid Z) - \mathrm{pr}(M_T \leq a_T \mid M_S \leq a_S)|}{\mathrm{pr}(M_T \leq a_T, M_S \leq a_S)} \\ &\to 0. \end{aligned}$$

Therefore, Corollary S1 holds. □

Bibliography

Durrett, R. (2010) *Probability: Theory and Examples*. Cambridge: Cambridge University Press, fourth edn.

Hájek, J. (1961) Some extensions of the Wald–Wolfowitz–Noether theorem. *The Annals of Mathematical Statistics*, **32**, 506–523.

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

Liu, H., Ren, J. and Yang, Y. (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**, 136–150.

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