Model-robust Inference for Seamless II/III Trials with Covariate Adaptive Randomization

Kun Yi and Lucy Xia

Department of ISOM, HKUST

Supplementary Material

Notation and Definition

Throughout the supplementary material, we employ the following notation, not necessarily introduced in the text.

$\|\boldsymbol{x}\|$	For any column vector $\boldsymbol{x}$, $\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^\top \boldsymbol{x}}$.
$\|\boldsymbol{M}\|$	For a square matrix $\boldsymbol{M} \in \mathbb{R}^{q\times q}$, $\|\boldsymbol{M}\| = \sup\{\|\boldsymbol{M}\boldsymbol{x}\| : \boldsymbol{x} \in \mathbb{R}^q, \|\boldsymbol{x}\| = 1\}$.
$\boldsymbol{X}_n = o_p(\boldsymbol{1})$	For any sequence of real-valued vectors $\boldsymbol{X}_n$, if $\boldsymbol{X}_n \xrightarrow{p} \boldsymbol{0}$, then we denote $\boldsymbol{X}_n = o_p(\boldsymbol{1})$.

$\boldsymbol{X}_n = O_p(\mathbf{1})$	For any sequence of real-valued vectors $\boldsymbol{X}_n$, if $\|\boldsymbol{X}_n\|$ is bounded in probability, then we denote $\boldsymbol{X}_n = O_p(\mathbf{1})$.
$\boldsymbol{M}_n = o_p(\mathbf{1})$	For any sequence of real-valued square matrices $\boldsymbol{M}_n \in \mathbb{R}^{q\times q}$, if every element of $\boldsymbol{M}_n$ converges in probability to 0, then we denote $\boldsymbol{M}_n = o_p(\mathbf{1})$.
$\boldsymbol{M}_n = O_p(\mathbf{1})$	For any sequence of real-valued square matrices $\boldsymbol{M}_n \in \mathbb{R}^{q\times q}$, if every element of $\boldsymbol{M}_n$ is bounded in probability, then we denote $\boldsymbol{M}_n = O_p(\mathbf{1})$.
$[K]$	$\{1, \ldots, K\}$.
$[s_{\max}]$	$\{1, \ldots, s_{\max}\}$.
$\boldsymbol{Y}_i$	$(Y_i(0), \ldots, Y_i(K))^\top$, where $Y_i(k)$ is the potential outcome of the i-th patient in treatment group k.
$\boldsymbol{V}$	$(Y(0), \ldots, Y(K), \boldsymbol{X})$, where $Y(k)$ is the treatment outcome of patients in treatment group k and $\boldsymbol{X}$ are baseline covariates .

$\boldsymbol{V}_i$	$(Y_i(0), \ldots, Y_i(K), \boldsymbol{X_i})$, where $\boldsymbol{X}_i$ are baseline covariates.
S	Stratification variable with possible values $s \in [s_{\max}]$.
T_i^k	Treatment indicator (1 if i-th patient is assigned to treatment $k \in 0 \cup [K]$, 0 otherwise) .
$\boldsymbol{T}_i$	$(T_i^0, T_i^1, \ldots, T_i^K)$.
$n_k(s)$	Number of patients assigned to treatment k within stratum $s \in [s_{\max}]$.
n_k	Number of patients assigned to treatment k.
$n(s)$	Number of patients within stratum $s \in [s_{\max}]$.
$D_n^k(s)$	Imbalance measure for treatment k in stratum s.
$\boldsymbol{\theta}$	Parameter vector of dimension $K + 1 + q$.

$\boldsymbol{\psi}^k(\boldsymbol{\theta})$	Estimating equation for treatment group k, defined as $\boldsymbol{\psi}^k(Y(k), \boldsymbol{X}; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the parameter vector, $\boldsymbol{X}$ are baseline covariates and $Y(k)$ is the potential outcome under treatment k.
$\boldsymbol{\psi}_i^k(\boldsymbol{\theta})$	The estimating equation for the i-th patient under treatment k, which is defined as $\boldsymbol{\psi}^k(Y_i(k), \boldsymbol{X}_i; \boldsymbol{\theta})$.
$\psi^{k,j}(\boldsymbol{\theta})$	The j-th component (scalar) of the vector-valued estimating function $\boldsymbol{\psi}^k(\boldsymbol{\theta})$.
$\psi_i^{k,j}(\boldsymbol{\theta})$	The j-th component (scalar) of the vector-valued estimating function $\boldsymbol{\psi}_i^k(\boldsymbol{\theta})$.
$\dot{\boldsymbol{\psi}}_i^k(\boldsymbol{\theta})$	First derivative matrix with respect to $\boldsymbol{\theta}$: $\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\psi}_i^k(\boldsymbol{\theta})$.
$\ddot{\boldsymbol{\psi}}_i^{k,j}(\boldsymbol{\theta})$	Second derivative matrix for the j-th component with respect to $\boldsymbol{\theta}$: $\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \psi_i^{k,j}(\boldsymbol{\theta})$.
$\boldsymbol{e}_j$	Standard basis vector in $\mathbb{R}^{K+1}$ (1 in position j, 0 elsewhere).
$\boldsymbol{\Psi}$	$\frac{1}{K+1} \sum_{k \in 0 \cup [K]} \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\psi}^k(Y(k), \boldsymbol{X}; \boldsymbol{\theta}) \big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \right]$.

$\boldsymbol{h}(\boldsymbol{X};\boldsymbol{\theta})$	$(h^0(\boldsymbol{X};\boldsymbol{\theta}),...,h^K(\boldsymbol{X};\boldsymbol{\theta}))^\top$.
$\boldsymbol{h}(\boldsymbol{X}_i;\boldsymbol{\theta})$	$(h^0(\boldsymbol{X}_i;\boldsymbol{\theta}),...,h^K(\boldsymbol{X}_i;\boldsymbol{\theta}))^\top$.
$\widehat{\boldsymbol{\mu}}$	$(\widehat{\mu}_0,...,\widehat{\mu}_K)^\top$, where $\widehat{\mu}_k = n^{-1}\sum_{i=1}^n h^k(\boldsymbol{X}_i,\widehat{\boldsymbol{\theta}})$ for all $k=0,...,K$.
$\boldsymbol{r}(\boldsymbol{\theta})$	$(r^0(\boldsymbol{\theta}),...,r^K(\boldsymbol{\theta}))^\top$, where $r^k(\boldsymbol{\theta}) = Y(k) - h^k(\boldsymbol{X},\boldsymbol{\theta})$ for all $k=0,...,K$.
$\boldsymbol{r}_i(\boldsymbol{\theta})$	$(r_i^0(\boldsymbol{\theta}),...,r_i^K(\boldsymbol{\theta}))^\top$, where $r_i^k(\boldsymbol{\theta}) = Y_i - h^k(\boldsymbol{X}_i,\boldsymbol{\theta})$ for all $k=0,...,K$.

S1 Assumptions

Assumption 1 (Reproduced from Assumption 1 in the main text)**.**

1. *$\{(\boldsymbol{X}_i, Y_i(0),\ldots,Y_i(K)\}_{i=1}^n$ are independently identically distributed as $(\boldsymbol{X}, Y(0),\ldots,Y(K))$, with $\mathbb{E}\left[\|\boldsymbol{X}\|^2\right]<\infty$, $\max_{k\in 0\cup[K]}\mathrm{Var}[Y(k)]<\infty$.*
2. *Let π_s denote the proportion of observations in stratum s. For any $s\in[s_{\max}]$, we have $0<\pi_s<1$, and $\sum_{s\in[s_{\max}]}\pi_s=1$.*
3. *Recall that $\mathcal{F}_{i-1}=\mathcal{X}_{i,\mathrm{ex}}\otimes\mathcal{T}_{i-1}\otimes\mathcal{X}_{i-1}\otimes\mathcal{Y}_{i-1}$ and the assignment rule*

satisfies $\boldsymbol{\phi}_i = \mathbb{E}\left[\boldsymbol{T}_i|\mathcal{F}_{i-1}\right]$ *for* $i \in [n]$*. Let us define the imbalance within stratum* s *and treatment arm* k *as* $D_n^k(s)$ *with the following decomposition*

$$\begin{aligned} D_n^k(s) &:= \sum_{i=1}^{n}\left(T_i^k - \frac{1}{K+1}\right)\mathbb{I}\{S_i = s\} \\ &= \sum_{i=1}^{n} M_i^k(s) + d_n^k(s), \quad s \in [s_{\max}], \quad k \in 0 \cup [K]. \end{aligned}$$

We assume that the CAR procedure satisfies

(a) $\mathbb{E}[\{d_n^k(s)\}^2] = o(n)$*; and*

(b) $\{M_i^k(s)\}_{i=1}^n$ *is a sequence of bounded zero-mean martingale differences with respect to* $\mathcal{F}_{i-1}$*. Further with vectorized*

$$\boldsymbol{M}_i := \operatorname{vec}\Big(\big(M_i^k(s)\big)_{s\in[s_{\max}],\, k\in[K]}\Big) \in \mathbb{R}^{s_{\max}(K+1)},$$

the averaged conditional second moment converges in probability to a block covariance matrix:

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\big[\boldsymbol{M}_i\,\boldsymbol{M}_i^\top \,\big|\, \mathcal{F}_{i-1}\big] \;\xrightarrow{p}\; \boldsymbol{\Sigma}^{\mathsf{CAR}} \;\in\; \mathbb{R}^{s_{\max}(K+1)\times s_{\max}(K+1)}.$$

Entrywise, if we index $\boldsymbol{\Sigma}^{\mathsf{CAR}}$ *by* $(s, s') \in [s_{\max}] \times [s_{\max}]$ *as a block matrix:*

$$\boldsymbol{\Sigma}^{\mathsf{CAR}} = \big[\,\boldsymbol{\Sigma}^{\mathsf{CAR}}(s, s')\,\big]_{s,s'\in[s_{\max}]}, \qquad \boldsymbol{\Sigma}^{\mathsf{CAR}}(s, s') \in \mathbb{R}^{(K+1)\times(K+1)},$$

and defined the (k, k')*-th element in* $\boldsymbol{\Sigma}^{\mathsf{CAR}}(s, s')$ *as* $\boldsymbol{\Sigma}^{\mathsf{CAR}}_{kk'}(s, s')$*, then*

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[M_i^k(s)M_i^{k'}(s') \mid \mathcal{F}_{i-1}] \xrightarrow{p} \boldsymbol{\Sigma}_{kk'}^{\mathsf{CAR}}(s,s').$$

Assumption S2 (Regularity conditions for the Z-estimator)**.**

1. *Compactness of the parameter space:* $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^{K+1+q}$*, with* $\boldsymbol{\Theta}$ *compact.*

2. *Existence of a unique solution and the invertibility condition: There exists a unique* $\boldsymbol{\theta}^* \in \mathrm{int}(\boldsymbol{\Theta})$ *such that:*

$$\frac{1}{K+1}\sum_{k=0}^{K}\mathbb{E}[\boldsymbol{\psi}^k(\boldsymbol{\theta}^*)] = \mathbf{0},$$

and $\boldsymbol{\Psi} = \sum_{k=0}^{K}\mathbb{E}[\dot{\boldsymbol{\psi}}^k(\boldsymbol{\theta}^*)]/(K+1)$ *is invertible.*

3. *Bounded second moment:*

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta},k\in 0\cup[K]} \mathbb{E}\left[\|\boldsymbol{\psi}^k(\boldsymbol{\theta})\|^2\right] < \infty, \quad \sup_{k\in 0\cup[K]} \mathbb{E}\left[\|\dot{\boldsymbol{\psi}}^k(\boldsymbol{\theta}^*)\|^2\right] < \infty.$$

4. *Differentiability and bounded second derivatives: For each* $k \in 0\cup[K]$*,*

   - $\boldsymbol{\psi}^k(\boldsymbol{\theta})$ *is twice continuously differentiable for every* $(y,\boldsymbol{x})$ *in the support of* $(Y(k),\boldsymbol{X})$ *and is dominated by an integrable function* $\boldsymbol{u}(Y(k),\boldsymbol{X})$*.*

   - *Bounded second derivatives:* $\exists\, C > 0$ *and integrable* $v(Y(k),\boldsymbol{X})$ *s.t.* $\forall j \in \{1,...,K+1+q\}$*:*

$$\|\ddot{\boldsymbol{\psi}}^{k,j}(\boldsymbol{\theta})\| < v(Y(k),\boldsymbol{X}) \text{ for } \|\boldsymbol{\theta}-\boldsymbol{\theta}^*\| < C.$$

5. *Conditions for the estimating equation:* $\widehat{\boldsymbol{\theta}}$ *satisfies*

$$\sum_{k=0}^{K}\sum_{i=1}^{n} T_i^k \boldsymbol{\psi}^k(Y_i(k),\boldsymbol{X}_i;\widehat{\boldsymbol{\theta}}) = o_p(n^{-1/2}).$$

S2 Proof of Theoretical Results

S2.1 Proof of Proposition 1

Proposition 1 (Reproduced from Proposition 1 in the main text)**.** *Under Assumptions 1 and S2,*

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* = O_p(\frac{1}{\sqrt{n}}), \qquad \widehat{\boldsymbol{\mu}} - \boldsymbol{\mu} = O_p(\frac{1}{\sqrt{n}}).$$

S2.1.1 Lemmas

Lemma 1. *Under Assumption 1, for each $k \in 0 \cup [K]$ and $s \in [s_{\max}]$,*

$$\frac{n_k(s)}{n} = \frac{1}{K+1}\pi_s + o_p(1) \quad \text{and} \quad D_n^k(s) = O_p(\sqrt{n}).$$

Lemma 2. *Under Assumption 1, let $Z_i(\boldsymbol{\theta}) := Z(\boldsymbol{V}_i; \boldsymbol{\theta})$, $i = 1, \cdots, n$, be i.i.d. as $Z(\boldsymbol{\theta}) := Z(\boldsymbol{V}; \boldsymbol{\theta})$ for some measurable real-valued function Z, and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Further, for each $s \in [s_{\max}]$, let $Z^{(s)}$ denote the conditional process of $Z = \{Z(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ given $S = s$. Then, if $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\mathbb{E}\left[Z^{(s)}(\boldsymbol{\theta})\right]\right| < \infty$ and $Z^{(s)}$ is P-Glivenko-Cantelli for each $s \in [s_{\max}]$, we have*

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \frac{1}{n}\sum_{i=1}^{n} T_i^k Z_i(\boldsymbol{\theta}) - \frac{1}{K+1}\mathbb{E}[Z(\boldsymbol{\theta})] \right| \xrightarrow{p} 0, \quad \text{for all} \quad k \in 0 \cup [K].$$

Lemma 3. *For each $k \in 0 \cup [K]$ and $s \in [s_{\max}]$, (S2.7) holds.*

Lemma 4 (Consistency of $\widehat{\boldsymbol{\theta}}$)**.** *Under Assumptions 1 and S2, the Z-estimator $\widehat{\boldsymbol{\theta}}$ is consistent.*

Lemma 5. *Under Assumptions 1 and S2, we have:*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k\in 0\cup[K]} T_i^k \boldsymbol{\psi}_i^k(\boldsymbol{\theta}^*) \xrightarrow{d} N(\mathbf{0},\boldsymbol{\Lambda}),$$

where

$$\begin{aligned}\boldsymbol{\Lambda} =&\mathrm{Var}\left[\mathbb{E}\left(\frac{1}{K+1}\sum_{k\in 0\cup[K]}\boldsymbol{\psi}^k(\boldsymbol{\theta}^*)\mid S\right)\right] + \frac{1}{K+1}\sum_{k\in 0\cup[K]}\mathbb{E}\left[\mathrm{Var}\left[\boldsymbol{\psi}^k(\boldsymbol{\theta}^*)\mid S\right]\right]\\ &+\sum_{k,k'\in 0\cup[K]}\sum_{s,s'\in[s_{\max}]}\boldsymbol{\Sigma}_{kk'}^{\mathsf{CAR}}(s,s')\mathbb{E}\left[\boldsymbol{\psi}^k(\boldsymbol{\theta}^*)\mid S=s\right]\mathbb{E}\left[\boldsymbol{\psi}^{k'}(\boldsymbol{\theta}^*)\mid S=s'\right]^{\top}.\end{aligned}$$

Lemma 6. *Under Assumptions 1 and S2, the estimator $\widehat{\boldsymbol{\theta}}$ is asymptotically linear. Specifically, we have the following result:*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*) = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\Psi}^{-1}\boldsymbol{\psi}(\boldsymbol{T}_i,\boldsymbol{X}_i,Y_i;\boldsymbol{\theta}^*) + o_p(\mathbf{1}). \tag{S2.1}$$

Lemma 7. *Under Assumptions 1 and S2, we have:*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*) \xrightarrow{d} N(\mathbf{0},\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}\{\boldsymbol{\Psi}^{-1}\}^{\top}).$$

Lemma 8. *Under Assumptions 1 and S2, $\widehat{\boldsymbol{\mu}}$ is consistent.*

Lemma 9. *Under Assumptions 1 and S2, the population mean estimator $\widehat{\boldsymbol{\mu}}$ have the asymptotic linearity:*

$$\sqrt{n}(\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\boldsymbol{h}(\boldsymbol{X}_i;\boldsymbol{\theta}^*)-\boldsymbol{\mu}\} - \frac{K+1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k\in 0\cup[K]} T_i^k\{h^k(\boldsymbol{X}_{\boldsymbol{i}};\boldsymbol{\theta}^*)-Y_i\}\boldsymbol{e}_k + o_p(\mathbf{1}).$$

Lemma 10. *Under Assumptions 1 and S2, we have:*

$$\sqrt{n}(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Gamma}), \tag{S2.2}$$

where

$$\boldsymbol{\Gamma} = (K+1)\text{diag}\left\{\text{Var}\left[r^0(\boldsymbol{\theta}^*)\right], \ldots, \text{Var}\left[r^K(\boldsymbol{\theta}^*)\right]\right\} + \text{Var}\left[\boldsymbol{Y}\right] - \text{Var}\left[\boldsymbol{r}\right]$$
$$- (K+1)^2 \sum_{k,k' \in 0 \cup [K]} \sum_{s,s' \in [s_{\max}]} \{\boldsymbol{\Sigma}^{\mathsf{CR}}_{kk'}(s,s') - \boldsymbol{\Sigma}^{\mathsf{CAR}}_{kk'}(s,s')\} \boldsymbol{L}(s,k) \boldsymbol{L}(s',k')^\top,$$
$$\boldsymbol{L}(s,k) = \mathbb{E}\left[r^k(\boldsymbol{\theta}^*) \mid S = s\right] \boldsymbol{e}_k.$$

S2.1.2 Main proof of Proposition 1

Proof. The result follows immediately from Lemmas 7, 10. □

S2.1.3 Proof of lemmas

Proof of Lemma 1. The proof follows similarly from Proposition 1.1 in (Liu and Hu, 2023) □

Proof of Lemma 2. The proof contains three steps. In the first step, we rearrange the samples and transform the sample function into a sorted sample function. In the second step, we prove that the sample function is equivalent to the sum of the sample functions at each stratum. In the third step, we prove that the sample function of each stratum satisfies uniform convergence by using Lemma 3. Then, by combining the conclusions of the first and second steps, we prove the uniform convergence of the sample function.
Step 1: Let $(\widetilde{\boldsymbol{T}}_1, \widetilde{\boldsymbol{V}}_1, \ldots, \widetilde{\boldsymbol{T}}_1, \widetilde{\boldsymbol{V}}_n)$ be a permutation of $(\boldsymbol{T}_1, \boldsymbol{V}_1, \ldots, \boldsymbol{T}_n, \boldsymbol{V}_n)$

such that each $(\boldsymbol{T}_i, \boldsymbol{V}_i)$ is first ordered by strata (starting with those having $S_i = 1$, then $S_i = 2$, and so on) and then by treatment group (with $T_i^0 = 1$ coming first) within each stratum. By construction, the process $\left\{\frac{1}{n}\sum_{i=1}^n T_i^k Z_i(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\right\}$ is identical to $\left\{\frac{1}{n}\sum_{i=1}^n \widetilde{T}_i^k Z(\widetilde{\boldsymbol{V}}_i; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\right\}$. Thus, with probability 1, we have

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \frac{1}{n}\sum_{i=1}^n T_i^k Z_i(\boldsymbol{\theta}) - \frac{1}{K+1}\mathbb{E}[Z(\boldsymbol{\theta})] \right| = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \frac{1}{n}\sum_{i=1}^n \widetilde{T}_i^k Z(\widetilde{\boldsymbol{V}}_i; \boldsymbol{\theta}) - \frac{1}{K+1}\mathbb{E}[Z(\boldsymbol{\theta})] \right|. \quad \text{(S2.3)}$$

Step 2: For each $s \in [s_{\max}]$, let $\boldsymbol{V}_1^{(s)}, \ldots, \boldsymbol{V}_n^{(s)}$ be i.i.d. copies of $\boldsymbol{V}^{(s)}$, independent of $(\widetilde{\boldsymbol{T}}_1, \ldots, \widetilde{\boldsymbol{T}}_n)$ and $(\widetilde{\boldsymbol{V}}_1, \ldots, \widetilde{\boldsymbol{V}}_n)$. We use the notation $\boldsymbol{V}_i^{(S_i)}$ to denote $\sum_{s\in[s_{\max}]} \mathbb{I}\{S_i = s\}\boldsymbol{V}_i^{(s)}$. Let $\widetilde{S}_i$ denote the stratum variable encoded in $\widetilde{\boldsymbol{V}}_i$. Then, we have the following derivation:

$$\begin{aligned}
\mathbb{P}(\widetilde{\boldsymbol{V}}_1, \ldots, \widetilde{\boldsymbol{V}}_n | \widetilde{\boldsymbol{T}}_1, \widetilde{S}_1, \ldots, \widetilde{\boldsymbol{T}}_1, \widetilde{S}_n) &= \mathbb{P}(\widetilde{\boldsymbol{V}}_1, \ldots, \widetilde{\boldsymbol{V}}_n | \widetilde{S}_1, \ldots, \widetilde{S}_n) \\
&= \prod_{i=1}^n \mathbb{P}(\widetilde{\boldsymbol{V}}_i | \widetilde{S}_i) = \prod_{i=1}^n \mathbb{P}(\boldsymbol{V}_i^{(\widetilde{S}_i)}) \\
&= \mathbb{P}(\boldsymbol{V}_1^{(\widetilde{S}_1)}, \ldots, \boldsymbol{V}_n^{(\widetilde{S}_n)} | \widetilde{\boldsymbol{T}}_1, \widetilde{S}_1, \ldots, \widetilde{\boldsymbol{T}}_1, \widetilde{S}_n).
\end{aligned}$$

By the chain rule of joint probability, we obtain the following equality of the joint distributions:

$$\mathbb{P}(\widetilde{\boldsymbol{T}}_1, \widetilde{\boldsymbol{V}}_1, \ldots, \widetilde{\boldsymbol{T}}_1, \widetilde{\boldsymbol{V}}_n) = \mathbb{P}(\widetilde{\boldsymbol{T}}_1, \boldsymbol{V}_1^{(\widetilde{S}_1)}, \ldots, \widetilde{\boldsymbol{T}}_1, \boldsymbol{V}_n^{(\widetilde{S}_n)})$$

unconditionally. This implies that, for any $\varepsilon > 0$,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{n}\sum_{i=1}^{n}\widetilde{T}_i^k Z(\widetilde{\boldsymbol{V}}_i;\boldsymbol{\theta}) - \frac{1}{K+1}\mathbb{E}[Z(\boldsymbol{\theta})]\right| > \varepsilon\right) \\
=&\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{n}\sum_{i=1}^{n}\widetilde{T}_i^k Z(\boldsymbol{V}_i^{(\widetilde{S}_i)};\boldsymbol{\theta}) - \frac{1}{K+1}\mathbb{E}[Z(\boldsymbol{\theta})]\right| > \varepsilon\right). \qquad \text{(S2.4)}
\end{aligned}$$

Step 3: Based on our construction of $(\widetilde{\boldsymbol{T}}_i, \widetilde{\boldsymbol{V}}_i)$, we have

$$\begin{aligned}
\sum_{i=1}^{n}\widetilde{T}_i^k Z(\boldsymbol{V}_i^{(s=\widetilde{S}_i)};\boldsymbol{\theta}) &= \sum_{s\in[s_{\max}]}\sum_{i=N(s)+\sum_{j=0}^{k-1}n_j(s)+1}^{N(s)+\sum_{j=0}^{k}n_j(s)} Z(\boldsymbol{V}_i^{(s)};\boldsymbol{\theta}) \\
&= \sum_{s\in[s_{\max}]}\sum_{i=N(s)+\sum_{j=0}^{k-1}n_j(s)+1}^{N(s)+\sum_{j=0}^{k}n_j(s)} Z_i^{(s)}(\boldsymbol{\theta}), \qquad \text{(S2.5)}
\end{aligned}$$

where $N(s) = \sum_{i=1}^{n}\mathbb{I}\{S_i < s\}$, $n_k(s) = \sum_{i=1}^{n}T_i^k\mathbb{I}\{S_i = s\}$ and $Z_i^{(s)}(\boldsymbol{\theta}) = Z(\boldsymbol{V}_i^{(s)};\boldsymbol{\theta})$.

The fact that the corresponding joint distributions are identical implies:

$$\mathbb{P}(\boldsymbol{V}^{(s)}_{N(s)+\sum_{j=0}^{k-1}n_j(s)+1}, \ldots, \boldsymbol{V}^{(s)}_{N(s)+\sum_{j=0}^{k-1}n_j(s)+n_k(s)}) = \mathbb{P}(\boldsymbol{V}_1^{(s)}, \ldots, \boldsymbol{V}_{n_k(s)}^{(s)}).$$

Therefore, for any $\varepsilon > 0$,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{n}\sum_{i=N(s)+\sum_{j=0}^{k-1}n_j(s)+1}^{N(s)+\sum_{j=0}^{k}n_j(s)} Z_i^{(s)}(\boldsymbol{\theta}) - \frac{1}{K+1}\pi_s\mathbb{E}[Z^{(s)}(\boldsymbol{\theta})]\right| > \varepsilon\right) \\
=&\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{n}\sum_{i=1}^{n_k(s)} Z_i^{(s)}(\boldsymbol{\theta}) - \frac{1}{K+1}\pi_s\mathbb{E}[Z^{(s)}(\boldsymbol{\theta})]\right| > \varepsilon\right). \qquad \text{(S2.6)}
\end{aligned}$$

By Lemma 3, we know:

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{n}\sum_{i=1}^{n_k(s)} Z_i^{(s)}(\boldsymbol{\theta}) - \frac{1}{K+1}\pi_s\mathbb{E}[Z^{(s)}(\boldsymbol{\theta})]\right| \xrightarrow{p} 0. \tag{S2.7}$$

Combing (S2.3), (S2.4), (S2.5), (S2.6) and (S2.7), we finish this proof. □

Proof of Lemma 3. By Lemma 1, we know that $\frac{n_k(s)}{n} \xrightarrow{p} \frac{\pi_s}{K+1}$. Then

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{n}\sum_{i=1}^{n_k(s)} Z_i^{(s)}(\boldsymbol{\theta}) - \frac{1}{K+1}\pi_s\mathbb{E}[Z^{(s)}(\boldsymbol{\theta})]\right| \\
=&\frac{n_k(s)}{n}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{n_k(s)}\sum_{i=1}^{n_k(s)} Z_i^{(s)}(\boldsymbol{\theta}) - \frac{n\pi_s}{(K+1)n_k(s)}\mathbb{E}[Z^{(s)}(\boldsymbol{\theta})]\right| \\
\leq&\frac{n_k(s)}{n}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{n_k(s)}\sum_{i=1}^{n_k(s)} Z_i^{(s)}(\boldsymbol{\theta}) - \mathbb{E}[Z^{(s)}(\boldsymbol{\theta})]\right| \\
&+\frac{n_k(s)}{n}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\left(\frac{n\frac{1}{K+1}\pi_s}{n_k(s)}-1\right)\mathbb{E}[Z^{(s)}(\boldsymbol{\theta})]\right| \\
=&\left(\frac{\pi_s}{K+1}+o_p(1)\right)\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{n_k(s)}\sum_{i=1}^{n_k(s)} Z_i^{(s)}(\boldsymbol{\theta}) - \mathbb{E}[Z^{(s)}(\boldsymbol{\theta})]\right| \\
&+o_p(1)\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\mathbb{E}[Z^{(s)}(\boldsymbol{\theta})]|.
\end{aligned}$$

Since $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\mathbb{E}[Z^{(s)}(\boldsymbol{\theta})]| < \infty$ by assumption, it suffices to show that for each $k \in 0 \cup [K]$ and $s \in [s_{\max}]$,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{n_k(s)}\sum_{i=1}^{n_k(s)} Z_i^{(s)}(\boldsymbol{\theta}) - \mathbb{E}[Z^{(s)}(\boldsymbol{\theta})]\right| \xrightarrow{p} 0. \tag{S2.8}$$

To this end, by the Skorokhod's representation theorem, we can construct $\frac{\tilde{n}_k(s)}{n}$ such that $\frac{\tilde{n}_k(s)}{n}$ has the same distribution as $\frac{n_k(s)}{n}$, $\frac{\tilde{n}_k(s)}{n} \to \frac{\pi_s}{K+1}$ almost surely and $\frac{\tilde{n}_k(s)}{n}$ is independent of $\{Z_1^{(s)}, \ldots, Z_n^{(s)}\}$. Since $(\boldsymbol{T}_1, \ldots, \boldsymbol{T}_n)$ and $(S_1, \ldots, S_n)$ are independent of $\{Z_1^{(s)}, \ldots, Z_n^{(s)}\}$, we have, for any $\varepsilon > 0$ and $n > 0$,

$$\begin{aligned} &\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{n_k(s)}\sum_{i=1}^{n_k(s)} Z_i^{(s)}(\boldsymbol{\theta}) - \mathbb{E}[Z^{(s)}(\boldsymbol{\theta})]\right| > \varepsilon\right) \\ =&\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{n\frac{\tilde{n}_k(s)}{n}}\sum_{i=1}^{n\frac{\tilde{n}_k(s)}{n}} Z_i^{(s)}(\boldsymbol{\theta}) - \mathbb{E}[Z^{(s)}(\boldsymbol{\theta})]\right| > \varepsilon\right) \\ \le&\mathbb{E}\left[\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{n\frac{\tilde{n}_k(s)}{n}}\sum_{i=1}^{n\frac{\tilde{n}_k(s)}{n}} Z_i^{(s)}(\boldsymbol{\theta}) - \mathbb{E}[Z^{(s)}(\boldsymbol{\theta})]\right| > \varepsilon \,\middle|\, \frac{\tilde{n}_k(s)}{n}\right)\right]. \end{aligned} \tag{S2.9}$$

Recall that by assumption, $Z^{(s)}$ is a P-Glivenko–Cantelli process. This implies that for any sequence $\{n_k\}$ such that $n_k \to \infty$ as $k \to \infty$, we have

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{n_k}\sum_{i=1}^{n_k} Z_i^{(s)}(\boldsymbol{\theta}) - \mathbb{E}[Z^{(s)}(\boldsymbol{\theta})]\right| \xrightarrow{p} 0.$$

Since the almost sure convergence of $\tilde{n}_k(s)$ to infinity and the independence of $\frac{\tilde{n}_k(s)}{n}$ from the sequence $\{Z_1^{(s)}, \ldots, Z_n^{(s)}\}$, we have:

$$\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{\tilde{n}_k(s)}\sum_{i=1}^{\tilde{n}_k(s)} Z_i^{(s)}(\boldsymbol{\theta}) - \mathbb{E}[Z^{(s)}(\boldsymbol{\theta})]\right| > \varepsilon \,\middle|\, \frac{\tilde{n}_k(s)}{n}\right) \to 0 \quad a.s., \tag{S2.10}$$

then by (S2.9), (S2.10) and the dominated convergence theorem, we finish the proof of (S2.8). This concludes the proof of the entire lemma. □

Proof of Lemma 4. The proof leverages the decomposition of the estimating function:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\psi}_i(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k\in 0\cup[K]} T_i^k\boldsymbol{\psi}_i^k(\boldsymbol{\theta}). \tag{S2.11}$$

By Assumption S2.3 and S2.4, we know $\{\boldsymbol{\psi}^k(\boldsymbol{\theta}) : \boldsymbol{\theta}\in\boldsymbol{\Theta}\}$ is P-Glivenko–Cantelli class, then by Lemma 2, we establish the uniform convergence:

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\frac{1}{n}\sum_{i=1}^{n}\sum_{k\in 0\cup[K]} T_i^k\boldsymbol{\psi}_i^k(\boldsymbol{\theta}) - \frac{1}{K+1}\sum_{k\in 0\cup[K]}\mathbb{E}\left[\boldsymbol{\psi}^k(\boldsymbol{\theta})\right]\right\| \xrightarrow{p} 0.$$

The result follows since $\widehat{\boldsymbol{\theta}}$ satisfies the empirical version of the estimating equation:

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{k\in 0\cup[K]} T_i^k\boldsymbol{\psi}_i^k(\widehat{\boldsymbol{\theta}}) = o_p(n^{-1/2}),$$

combined with the uniqueness of $\boldsymbol{\theta}^*$ and Theorem 5.9 of Van der Vaart (2000). □

Proof of Lemma 5. we rewrite $n^{-1/2}\sum_{i=1}^{n}\sum_{k\in 0\cup[K]} T_i^k \boldsymbol{\psi}_i^k(\boldsymbol{\theta}^*)$ as

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k\in 0\cup[K]} T_i^k \boldsymbol{\psi}_i^k(\boldsymbol{\theta}^*) \\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k\in 0\cup[K]} (T_i^k-\frac{1}{K+1})\boldsymbol{\psi}_i^k(\boldsymbol{\theta}^*)+\frac{1}{\sqrt{n}(K+1)}\sum_{i=1}^{n}\sum_{k\in 0\cup[K]}\boldsymbol{\psi}_i^k(\boldsymbol{\theta}^*) \\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k\in 0\cup[K]}\frac{1}{K+1}\boldsymbol{\psi}_i^k(\boldsymbol{\theta}^*) \\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k\in 0\cup[K]}(T_i^k-\frac{1}{K+1})\left[\boldsymbol{\psi}_i^k(\boldsymbol{\theta}^*)-\mathbb{E}\left[\boldsymbol{\psi}_i^k(\boldsymbol{\theta}^*)\mid S_i\right]\right] \\
&+\frac{1}{\sqrt{n}}\sum_{s\in[s_{\max}]}\sum_{k\in 0\cup[K]} D_n^k(s)\mathbb{E}\left[\boldsymbol{\psi}^k(\boldsymbol{\theta}^*)\mid S=s\right] \\
:=&\mathbb{G}_{n,1}^*(\boldsymbol{\psi})+\mathbb{G}_{n,2}^*(\boldsymbol{\psi})+\mathbb{G}_{n,3}(\boldsymbol{\psi}),
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbb{G}_{n,1}^*(\boldsymbol{\psi})&=\frac{1}{\sqrt{n}(K+1)}\sum_{i=1}^{n}\sum_{k\in 0\cup[K]}\boldsymbol{\psi}_i^k(\boldsymbol{\theta}^*), \\
\mathbb{G}_{n,2}^*(\boldsymbol{\psi})&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k\in 0\cup[K]}(T_i^k-\frac{1}{K+1})\left[\boldsymbol{\psi}_i^k(\boldsymbol{\theta}^*)-\mathbb{E}\left[\boldsymbol{\psi}_i^k(\boldsymbol{\theta}^*)\mid S_i\right]\right], \\
\mathbb{G}_{n,3}^*(\boldsymbol{\psi})&=\frac{1}{\sqrt{n}}\sum_{s\in[s_{\max}]}\sum_{k\in 0\cup[K]} D_n^k(s)\mathbb{E}\left[\boldsymbol{\psi}_i^k(\boldsymbol{\theta}^*)\mid S_i=s\right].
\end{aligned}
$$

By Assumption 1.3, we can rewrite

$$
\frac{1}{\sqrt{n}}D_n^k(s)\mathbb{E}\left[\boldsymbol{\psi}^k(\boldsymbol{\theta}^*)\mid S=s\right]=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}M_i^k(s)\mathbb{E}\left[\boldsymbol{\psi}^k(\boldsymbol{\theta}^*)\mid S=s\right]+o_p(\mathbf{1}).
$$

Define a vector of martingale differences:

$$
\Delta G_i(\boldsymbol{\psi})=\left((\Delta G_{i,1}(\boldsymbol{\psi}))^\top,\ (\Delta G_{i,2}(\boldsymbol{\psi}))^\top,\ (\Delta G_{i,3}(\boldsymbol{\psi}))^\top,\ (\Delta G_{i,4}(\boldsymbol{\psi}))^\top\right)^\top,
$$

where

$$\Delta G_{i,1}(\boldsymbol{\psi}) = \frac{1}{K+1} \sum_{k \in 0 \cup [K]} \left[\boldsymbol{\psi}_i^k(\boldsymbol{\theta}^*) - \mathbb{E}\left[\boldsymbol{\psi}_i^k(\boldsymbol{\theta}^*) \mid S_i \right] \right],$$

$$\Delta G_{i,3}(\boldsymbol{\psi}) = \sum_{k \in 0 \cup [K]} (T_i^k - \frac{1}{K+1}) \left[\boldsymbol{\psi}_i^k(\boldsymbol{\theta}^*) - \mathbb{E}\left[\boldsymbol{\psi}_i^k(\boldsymbol{\theta}^*) \mid S_i \right] \right],$$

$$\Delta G_{i,4}(\boldsymbol{\psi}) = \sum_{k \in 0 \cup [K]} \sum_{s \in [s_{\max}]} M_i^k(s) \mathbb{E}\left[\boldsymbol{\psi}_i^k(\boldsymbol{\theta}^*) \mid S_i = s \right],$$

$$\text{and } \Delta G_{i,2}(\boldsymbol{\psi}) = \begin{cases} \frac{1}{K+1} \sum_{k \in 0 \cup [K]} \mathbb{E}\left[\boldsymbol{\psi}_{i+1}^k(\boldsymbol{\theta}^*) \mid S_{i+1} \right], & \text{for } 1 \le i \le n-1 \\ 0, & \text{for } i = n \end{cases}$$

with this construction, we have

$$\mathbb{G}_{n,1}^*(\boldsymbol{\psi}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\Delta G_{i,1}(\boldsymbol{\psi}) + \Delta G_{i,2}(\boldsymbol{\psi})) + o_p(\mathbf{1}),$$
$$\mathbb{G}_{n,2}^*(\boldsymbol{\psi}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Delta G_{i,3}(\boldsymbol{\psi}),$$
$$\mathbb{G}_{n,3}^*(\boldsymbol{\psi}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Delta G_{i,4}(\boldsymbol{\psi}) + o_p(\mathbf{1}).$$

Since Assumption 1.2 implies that $\boldsymbol{T}_i \perp \boldsymbol{V}_i$ given S_i, we have

$$\mathbb{E}[\Delta G_i(\boldsymbol{\psi}) \mid \mathcal{F}_{i-1}] = 0,$$

so $\{\Delta G_i(\boldsymbol{\psi})\}_{i=1}^n$ is a sequence of zero-mean martingale differences with respect to the filtration $\mathcal{F}_{i-1}$.

Additionally, since $\mathbb{E}\|\boldsymbol{\psi}^k(\boldsymbol{\theta}^*)\|^2 < \infty$ for all $k \in \{0, \ldots, K\}$ and Assumption 1.4, we obtain

$$\mathbb{E}\left[\sum_{i=1}^{n}\|\Delta G_i(\boldsymbol{\psi})\|\right]=O(n).$$

Therefore, the conditional Lindeberg condition (Hall and Heyde, 2014, Corollary 3.1) holds for $\{\Delta G_i(\boldsymbol{\psi})\}_{i=1}^n$. To derive the normality, it suffices to calculate the conditional covariance matrices for $\{\Delta G_i(\boldsymbol{\psi})\}_{i=1}^n$. Let $\boldsymbol{\psi}(\boldsymbol{\theta})=(K+1)^{-1}\sum_{k\in 0\cup[K]}\boldsymbol{\psi}^k(\boldsymbol{\theta})$. The following results hold:

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}E[\Delta G_{i,1}(\boldsymbol{\psi})(\Delta G_{i,1}(\boldsymbol{\psi}))^\top|\mathcal{F}_{i-1}]=&\mathbb{E}\left[\mathrm{Cov}\left[\boldsymbol{\psi}(\boldsymbol{\theta}^*),\boldsymbol{\psi}(\boldsymbol{\theta}^*)^\top\mid S\right]\right]+o_p(\mathbf{1}),\\
\frac{1}{n}\sum_{i=1}^{n}E[\Delta G_{i,2}(\boldsymbol{\psi})(\Delta G_{i,2}(\boldsymbol{\psi}))^\top|\mathcal{F}_{i-1}]=&\mathrm{Cov}\left[\mathbb{E}\left[\boldsymbol{\psi}(\boldsymbol{\theta}^*)\mid S\right],\mathbb{E}\left[\boldsymbol{\psi}(\boldsymbol{\theta}^*)^\top\mid S\right]\right]+o_p(\mathbf{1}),\\
\frac{1}{n}\sum_{i=1}^{n}E[\Delta G_{i,3}(\boldsymbol{\psi})(\Delta G_{i,3}(\boldsymbol{\psi}))^\top|\mathcal{F}_{i-1}]=&\frac{1}{K+1}\sum_{k\in 0\cup[K]}\mathbb{E}\left[\mathrm{Cov}\left[\boldsymbol{\psi}^k(\boldsymbol{\theta}^*),\boldsymbol{\psi}^k(\boldsymbol{\theta}^*)^\top\mid S\right]\right]\\
&-\mathbb{E}\left[\mathrm{Cov}\left[\boldsymbol{\psi}(\boldsymbol{\theta}^*),\boldsymbol{\psi}(\boldsymbol{\theta}^*)^\top\mid S\right]\right]+o_p(\mathbf{1}).
\end{aligned}$$

In addition, we have

$$\begin{aligned}
\mathbb{E}[\Delta G_{i,1}(\boldsymbol{\psi})(\Delta G_{i,4}(\boldsymbol{\psi}))^\top|\mathcal{F}_{i-1}]&=\mathbb{E}[\Delta G_{i,1}(\boldsymbol{\psi})(\Delta G_{i,2}(\boldsymbol{\psi}))^\top|\mathcal{F}_{i-1}]\\
&=\mathbb{E}[\Delta G_{i,3}(\boldsymbol{\psi})(\Delta G_{i,2}(\boldsymbol{\psi}))^\top|\mathcal{F}_{i-1}]\\
&=\mathbb{E}[\Delta G_{i,4}(\boldsymbol{\psi})(\Delta G_{i,2}(\boldsymbol{\psi}))^\top|\mathcal{F}_{i-1}]\\
&=\mathbb{E}[\Delta G_{i,3}(\boldsymbol{\psi})(\Delta G_{i,4}(\boldsymbol{\psi}))^\top|\mathcal{F}_{i-1}]=0.
\end{aligned}$$

Moreover,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\Delta G_{i,3}(\boldsymbol{\psi})(\Delta G_{i,1}(\boldsymbol{\psi}))^\top|\mathcal{F}_{i-1}]$$

$$=\frac{1}{n(K+1)}\sum_{s\in[s_{\max}]}\sum_{k,k'\in 0\cup[K]}\mathbb{I}(S_i=s)\mathbb{E}[(T_i^k-\frac{1}{K+1})|\mathcal{F}_{i-1}]$$

$$\times\mathbb{E}\left[\left[\boldsymbol{\psi}_i^k(\boldsymbol{\theta}^*)-\mathbb{E}\left[\boldsymbol{\psi}_i^k(\boldsymbol{\theta}^*)\mid S_i\right]\right]\times\left[\boldsymbol{\psi}_i^k(\boldsymbol{\theta}^*)-\mathbb{E}\left[\boldsymbol{\psi}_i^k(\boldsymbol{\theta}^*)\mid S_i\right]\right]^\top|S_i=s\right].$$

And,

$$D_n^k(s)-\sum_{i=1}^{n}\mathbb{I}(S_i=s)\mathbb{E}[(T_i^k-\frac{1}{K+1})|\mathcal{F}_{i-1}]=\sum_{i=1}^{n}\mathbb{I}(S_i=s)(T_i^k-\mathbb{E}[T_i^k|\mathcal{F}_{i-1}]).$$

Since $\{\mathbb{I}(S_i=s)(T_i^k-\mathbb{E}[T_i^k|\mathcal{F}_{i-1}])\}$ forms a sequence of bounded martingale differences, by the martingale convergence theorem, we have

$$\frac{1}{n}\{D_n^k(s)-\sum_{i=1}^{n}\mathbb{I}(S_i=s)\mathbb{E}[(T_i^k-\frac{1}{K+1})|\mathcal{F}_{i-1}]\}=o_p(1).$$

By Lemma 1, $n^{-1}D_n^k(s)=o_p(1)$, which implies

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}(S_i=s)\mathbb{E}[(T_i^k-\frac{1}{K+1})|\mathcal{F}_{i-1}]=o_p(1).$$

Consequently,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\Delta G_{i,3}(\boldsymbol{\psi})(\Delta G_{i,1}(\boldsymbol{\psi}))^\top|\mathcal{F}_{i-1}]=o_p(\mathbf{1}).$$

Assumptions 1.2 and 1.3 further imply that

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\Delta G_{i,4}(\boldsymbol{\psi})(\Delta G_{i,4}(\boldsymbol{\psi}))^{\top}|\mathcal{F}_{i-1}]$$
$$=\frac{1}{n}\sum_{i=1}^{n}E\left[\sum_{k\in 0\cup[K]}\sum_{s\in[s_{\max}]}M_i^k(s)\mathbb{E}\left[\boldsymbol{\psi}_i^k(\boldsymbol{\theta}^*)\mid S=s\right]\right.$$
$$\left.\times\sum_{k'\in 0\cup[K]}\sum_{s'\in[s_{\max}]}M_i^{k'}(s')\mathbb{E}\left[\boldsymbol{\psi}_i^{k'}(\boldsymbol{\theta}^*)\mid S=s'\right]^{\top}|\mathcal{F}_{i-1}\right]+o_p(\mathbf{1})$$
$$=\sum_{k,k'\in 0\cup[K]}\sum_{s,s'\in[s_{\max}]}\boldsymbol{\Sigma}_{kk'}^{\mathsf{CAR}}(s,s')\mathbb{E}\left[\boldsymbol{\psi}^k(\boldsymbol{\theta}^*)\mid S=s\right]\mathbb{E}\left[\boldsymbol{\psi}^{k'}(\boldsymbol{\theta}^*)\mid S=s'\right]^{\top}+o_p(\mathbf{1}).$$

Combining the above results, we obtain

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\sum_{j=1}^{4}\Delta G_{i,j}(\boldsymbol{\psi})\sum_{j=1}^{4}(\Delta G_{i,j}(\boldsymbol{\psi}))^{\top}\middle|\mathcal{F}_{i-1}\right]$$
$$=\frac{1}{n}\sum_{j=1}^{4}\sum_{i=1}^{n}\mathbb{E}\left[\Delta G_{i,j}(\boldsymbol{\psi})(\Delta G_{i,j}(\boldsymbol{\psi}))^{\top}\middle|\mathcal{F}_{i-1}\right]$$
$$=\mathbb{E}\left[\mathrm{Cov}\left[\boldsymbol{\psi}(\boldsymbol{\theta}^*),\boldsymbol{\psi}(\boldsymbol{\theta}^*)^{\top}\mid S\right]\right]+\mathrm{Cov}\left[\mathbb{E}\left[\boldsymbol{\psi}(\boldsymbol{\theta}^*)\mid S_i\right],\mathbb{E}\left[\boldsymbol{\psi}(\boldsymbol{\theta}^*)^{\top}\mid S\right]\right]$$
$$+\frac{1}{K+1}\sum_{k\in 0\cup[K]}\mathbb{E}\left[\mathrm{Cov}\left[\boldsymbol{\psi}^k(\boldsymbol{\theta}^*),\boldsymbol{\psi}^k(\boldsymbol{\theta}^*)^{\top}\mid S\right]\right]-\mathbb{E}\left[\mathrm{Cov}\left[\boldsymbol{\psi}(\boldsymbol{\theta}^*),\boldsymbol{\psi}(\boldsymbol{\theta}^*)^{\top}\mid S\right]\right]$$
$$+\sum_{k,k'\in 0\cup[K]}\sum_{s,s'\in[s_{\max}]}\boldsymbol{\Sigma}_{kk'}^{\mathsf{CAR}}(s,s')\mathbb{E}\left[\boldsymbol{\psi}^k(\boldsymbol{\theta}^*)\mid S=s\right]\mathbb{E}\left[\boldsymbol{\psi}^{k'}(\boldsymbol{\theta}^*)\mid S=s'\right]^{\top}+o_p(\mathbf{1})$$
$$=\mathrm{Var}\left[\mathbb{E}\left(\boldsymbol{\psi}(\boldsymbol{\theta}^*)\mid S\right)\right]+\frac{1}{K+1}\sum_{k\in 0\cup[K]}\mathbb{E}\left[\mathrm{Var}\left[\boldsymbol{\psi}^k(\boldsymbol{\theta}^*)\mid S\right]\right]$$
$$+\sum_{k,k'\in 0\cup[K]}\sum_{s,s'\in[s_{\max}]}\boldsymbol{\Sigma}_{kk'}^{\mathsf{CAR}}(s,s')\mathbb{E}\left[\boldsymbol{\psi}^k(\boldsymbol{\theta}^*)\mid S=s\right]\mathbb{E}\left[\boldsymbol{\psi}^{k'}(\boldsymbol{\theta}^*)\mid S=s'\right]^{\top}+o_p(\mathbf{1})$$
$$=\boldsymbol{\Lambda}+o_p(\mathbf{1}).$$

By the martingale central limit theorem (Hall and Heyde, 2014) and Slutsky's theorem, we finish the proof of this lemma. □

Proof of Lemma 6. By Assumption S2.5 and performing a multivariate Taylor expansion of the function $\sum_{i=1}^{n}\sum_{k\in 0\cup[K]} T_i^k \boldsymbol{\psi}_i^k(\boldsymbol{\theta})$ around $\boldsymbol{\theta}=\boldsymbol{\theta}^*$, we obtain

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\sum_{k\in 0\cup[K]} T_i^k \boldsymbol{\psi}_i^k(\widehat{\boldsymbol{\theta}}) \\
=&\frac{1}{n}\sum_{i=1}^{n}\sum_{k\in 0\cup[K]} T_i^k \boldsymbol{\psi}_i^k(\boldsymbol{\theta}^*) + \frac{1}{n}\sum_{i=1}^{n}\sum_{k\in 0\cup[K]} T_i^k \dot{\boldsymbol{\psi}}_i^k(\boldsymbol{\theta}^*)(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*) \\
&+\sum_{j=1}^{K+1+q}\frac{1}{2}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)^\top\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{k\in 0\cup[K]} T_i^k \ddot{\boldsymbol{\psi}}_i^{k,j}(\widetilde{\boldsymbol{\theta}})\right)(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)\boldsymbol{e}_j = o_p(n^{-1/2}).
\end{aligned}
$$

From Assumption S2.4, there exists a ball Q centered at $\boldsymbol{\theta}^*$ such that $\|\ddot{\boldsymbol{\psi}}_i^{k,j}(\boldsymbol{\theta})\|$ is dominated by an integrable function $v(Y_i(k),\boldsymbol{X}_i)$ for each $j\in[K+1+q]$ and $\boldsymbol{\theta}\in Q$. Since $\widetilde{\boldsymbol{\theta}}$ lies on the line segment between $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^*$ and $\mathbb{P}(\widetilde{\boldsymbol{\theta}}\in Q)\to 1$, we have

$$
\left\|\frac{1}{n}\sum_{i=1}^{n}\sum_{k\in 0\cup[K]} T_i^k \ddot{\boldsymbol{\psi}}_i^{k,j}(\widetilde{\boldsymbol{\theta}})\right\| = O_p(1), \tag{S2.12}
$$

and thus the last term in the Taylor expansion is $O_p(\mathbf{1})\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\|^2$.

By Lemma 2 and Assumption S2.2, we have

$$
\frac{1}{n}\sum_{i=1}^{n}\sum_{k\in 0\cup[K]} T_i^k \dot{\boldsymbol{\psi}}_i^k(\boldsymbol{\theta}^*) \xrightarrow{p} \frac{1}{K+1}\sum_{k\in 0\cup[K]}\mathbb{E}\left[\dot{\boldsymbol{\psi}}^k(\boldsymbol{\theta}^*)\right] = \boldsymbol{\Psi}.
$$

Combining the above, we get the equation

$$
\frac{1}{n}\sum_{i=1}^{n}\sum_{k\in 0\cup[K]} T_i^k \boldsymbol{\psi}_i^k(\boldsymbol{\theta}^*) = -\boldsymbol{\Psi}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*) - O_p(\mathbf{1})\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\|^2 + o_p(1)(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*). \tag{S2.13}
$$

Multiplying both sides of (S2.13) by $\sqrt{n}$ and by Lemma 5, we have:

$$-\sqrt{n}\boldsymbol{\Psi}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)-\sqrt{n}\ o_p(\mathbf{1})\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\|+\sqrt{n}\ o_p(1)(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)=O_p(\mathbf{1}).$$

Since $\boldsymbol{\Psi}$ is invertible, we get

$$\sqrt{n}\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\|\le\|\boldsymbol{\Psi}^{-1}\|\sqrt{n}\|\boldsymbol{\Psi}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)\|=O_p(1)+o_p(1)\sqrt{n}\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\|,$$

which implies $\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)=O_p(\mathbf{1})$, then the asymptotic linearity of $\widehat{\boldsymbol{\theta}}$ follows from (S2.13). □

Proof of Lemma 7. The result follows immediately from Lemmas 5, 6 and Slutsky's theorem. □

Proof of Lemma 8. Note that for all $k\in 0\cup[K]$, the definition of $\boldsymbol{\psi}^k(Y_i(k),\boldsymbol{X}_i;\boldsymbol{\theta})$ implies that $\mathbb{E}[Y(k)]=\mathbb{E}[h^k(\boldsymbol{X};\boldsymbol{\theta}^*)]$; and thus

$$\frac{1}{n}\sum_{i=1}^{n}h^k(\boldsymbol{X}_i,\boldsymbol{\theta}^*)=\mathbb{E}[Y(k)]+o_p(1). \tag{S2.14}$$

Next, by the multivariate Taylor's expansion of $\sum_{i=1}^n\boldsymbol{h}(\boldsymbol{X}_i;\widehat{\boldsymbol{\theta}})$ around the point $\boldsymbol{\theta}=\boldsymbol{\theta}^*$, there exists a $\widetilde{\boldsymbol{\theta}}$ (which depends on $\widehat{\boldsymbol{\theta}}$) on the line segment between $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^*$ such that

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\left\{\boldsymbol{h}(\boldsymbol{X}_i;\widehat{\boldsymbol{\theta}})-\boldsymbol{h}(\boldsymbol{X}_i;\boldsymbol{\theta}^*)\right\}\\&=\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{h}(\boldsymbol{X}_i;\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right\}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})\\&+\sum_{k\in 0\cup[K]}\frac{1}{2}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)^\top\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}h^k(\boldsymbol{X}_i;\widetilde{\boldsymbol{\theta}})\right)(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)\boldsymbol{e}_k.\end{aligned}$$

From Assumption S2.4 and (S2.12), we know $\left\|\sum_{i=1}^{n} \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} h^k(\boldsymbol{X}_i;\widetilde{\boldsymbol{\theta}})/n\right\| = O_p(1)$, and thus the last term in the Taylor expansion is $O_p(\mathbf{1})\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\|^2$, By Lemma 4, we have

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\boldsymbol{h}(\boldsymbol{X}_i;\widehat{\boldsymbol{\theta}}) - \boldsymbol{h}(\boldsymbol{X}_i;\boldsymbol{\theta}^*)\right\} = o_p(\mathbf{1}).$$

Combined with (S2.14), we complete the proof of this lemma. □

Proof of Lemma 9. Denote $\boldsymbol{C} = \big((K+1)\boldsymbol{I}_{(K+1)\times(K+1)}, \mathbf{0}_{(K+1)\times q}\big)$, from the definition of $\boldsymbol{\Psi}$, it follows that

$$\boldsymbol{\Psi} = \frac{1}{K+1}\sum_{k\in 0\cup[K]}\mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{\psi}^k(\boldsymbol{\theta})\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right] = \frac{1}{K+1}\begin{bmatrix} \mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\theta}}h^0(\boldsymbol{X};\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right] \\ \mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\theta}}h^1(\boldsymbol{X};\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right] \\ \vdots \\ \mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\theta}}h^K(\boldsymbol{X};\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right] \\ \sum_{k\in 0\cup[K]}\mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{\psi}_1^k(\boldsymbol{X},Y;\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right] \end{bmatrix}$$

$$= \frac{1}{K+1}\begin{bmatrix} \mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{h}(\boldsymbol{X};\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right] \\ \sum_{k\in 0\cup[K]}\mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{\xi}^k(\boldsymbol{X},Y;\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right] \end{bmatrix}.$$

Moreover, since $\boldsymbol{\Psi}\boldsymbol{\Psi}^{-1} = \boldsymbol{I}_{(K+1+q)\times(K+1+q)}$, we have

$$\boldsymbol{\Psi}\boldsymbol{\Psi}^{-1} = \frac{1}{K+1}\begin{bmatrix} \mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{h}(\boldsymbol{X};\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right] \\ \sum_{k\in 0\cup[K]}\mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{\xi}^k(\boldsymbol{X},Y;\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right] \end{bmatrix} \times \boldsymbol{\Psi}^{-1} = \boldsymbol{I}_{(K+1+q)\times(K+1+q)}.$$

Then,

$$\mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{h}(\boldsymbol{X};\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right] \times \boldsymbol{\Psi}^{-1} = (K+1)\boldsymbol{C}.$$

By the multivariate Taylor's expansion and Lemma 6, we have:

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\boldsymbol{h}(\boldsymbol{X}_i;\widehat{\boldsymbol{\theta}})-\boldsymbol{h}(\boldsymbol{X}_i;\boldsymbol{\theta}^*)\right\}\\
&=\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{h}(\boldsymbol{X}_i;\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right\}\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})+\sqrt{n}\left\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right\|^2 O_p(\mathbf{1})\\
&=-\frac{1}{\sqrt{n}}\left\{\mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{h}(\boldsymbol{X};\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right]+o_p(\mathbf{1})\right\}\sum_{i=1}^{n}\boldsymbol{\Psi}^{-1}\boldsymbol{\psi}(\boldsymbol{T}_i,\boldsymbol{X}_i,Y_i;\boldsymbol{\theta}^*)+o_p(\mathbf{1})\\
&=-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{C}\boldsymbol{\psi}(\boldsymbol{T}_i,\boldsymbol{X}_i,Y_i;\boldsymbol{\theta}^*)+o_p(\mathbf{1})\\
&=-\frac{K+1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k\in 0\cup[K]}T_i^k\{h^k(\boldsymbol{X}_{\boldsymbol{i}};\boldsymbol{\theta}^*)-Y_i\}\boldsymbol{e}_k+o_p(\mathbf{1}).
\end{aligned}
$$

Combined with (S2.14), we finish the proof of this lemma. □

Proof of Lemma 10. By simple calculations, we have the variance-covariance matrix of imbalance measure under complete randomization, denoted as

$$
\boldsymbol{\Sigma}^{\mathsf{CR}}=\begin{bmatrix}\boldsymbol{\Sigma}_D(1) & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Sigma}_D(2) & \cdots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma}_D(s_{\max})\end{bmatrix},
$$

where,

$$
\boldsymbol{\Sigma}_D(s)=\pi_s\left[\operatorname{diag}\left(\frac{\mathbf{1}_K}{K+1}\right)-\frac{1}{(K+1)^2}\mathbf{1}_K\mathbf{1}_K^\top\right].
$$

By Lemma 9, we have:

$$
\begin{aligned}
\sqrt{n}(\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\boldsymbol{h}(\boldsymbol{X}_i;\boldsymbol{\theta}^*)-\boldsymbol{\mu}\} - \frac{K+1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k\in 0\cup[K]} T_i^k r_i^k(\boldsymbol{\theta}^*)\boldsymbol{e}_k + o_p(\mathbf{1}) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\boldsymbol{h}(\boldsymbol{X}_i;\boldsymbol{\theta}^*)-\boldsymbol{\mu}-\sum_{k\in 0\cup[K]} r_i^k(\boldsymbol{\theta}^*)\boldsymbol{e}_k\} \\
&- \frac{K+1}{\sqrt{n}}\sum_{s\in[s_{\max}]}\sum_{k\in 0\cup[K]} D_n^k(s)\mathbb{E}\left[r_i^k(\boldsymbol{\theta}^*)\boldsymbol{e}_k|S=s\right] \\
&- \frac{K+1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k\in 0\cup[K]}(T_i^k-\frac{1}{K+1})\left[r_i^k(\boldsymbol{\theta}^*)\boldsymbol{e}_k-\mathbb{E}\left[r_i^k(\boldsymbol{\theta}^*)\boldsymbol{e}_k\mid S_i\right]\right]+o_p(\mathbf{1}) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\boldsymbol{Y}_i-\boldsymbol{\mu}\} - \frac{K+1}{\sqrt{n}}\sum_{s\in[s_{\max}]}\sum_{k\in 0\cup[K]} D_n^k(s)\mathbb{E}\left[r_i^k(\boldsymbol{\theta}^*)|S=s\right]\boldsymbol{e}_k \\
&- \frac{K+1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k\in 0\cup[K]}(T_i^k-\frac{1}{K+1})\left[r_i^k(\boldsymbol{\theta}^*)-\mathbb{E}\left[r_i^k(\boldsymbol{\theta}^*)\mid S_i\right]\right]\boldsymbol{e}_k+o_p(\mathbf{1}).
\end{aligned}
$$

Following a similar derivation of Lemma 5, we know

$$
\sqrt{n}(\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu}) \xrightarrow{d} N(\mathbf{0},\boldsymbol{\Gamma}),
$$

$$
\begin{aligned}
\boldsymbol{\Gamma} &= \mathrm{Var}\left[\boldsymbol{Y}\right] + (K+1)\sum_{k\in 0\cup[K]}\mathbb{E}\left[\mathrm{Var}\left[r^k(\boldsymbol{\theta}^*)\mid S\right]\right]\boldsymbol{e}_k\boldsymbol{e}_k^\top - \mathbb{E}\left[\mathrm{Var}\left[\boldsymbol{r}\mid S\right]\right] \\
&+ (K+1)^2\sum_{k,k'\in 0\cup[K]}\sum_{s,s'\in[s_{\max}]}\boldsymbol{\Sigma}^{\mathsf{CAR}}_{kk'}(s,s')\boldsymbol{e}_k E\left[r^k(\boldsymbol{\theta}^*)\mid S=s\right]\mathbb{E}\left[r^{k'}(\boldsymbol{\theta}^*)\mid S=s'\right]\boldsymbol{e}_{k'}^\top \\
&= \mathrm{Var}\left[\boldsymbol{Y}\right]-\mathrm{Var}\left[\boldsymbol{r}\right]+\{\mathrm{Var}\left[\boldsymbol{r}\right]-\mathbb{E}\left[\mathrm{Var}\left[\boldsymbol{r}\mid S\right]\right]+(K+1)\sum_{k\in 0\cup[K]}\mathbb{E}\left[\mathrm{Var}\left[r^k(\boldsymbol{\theta}^*)\mid S\right]\right]\boldsymbol{e}_k\boldsymbol{e}_k^\top \\
&+ (K+1)^2\sum_{k,k'\in 0\cup[K]}\sum_{s,s'\in[s_{\max}]}\boldsymbol{\Sigma}^{\mathsf{CAR}}_{kk'}(s,s')\boldsymbol{e}_k E\left[r^k(\boldsymbol{\theta}^*)\mid S=s\right]\mathbb{E}\left[r^{k'}(\boldsymbol{\theta}^*)\mid S=s'\right]\boldsymbol{e}_{k'}^\top\} \\
&= \mathrm{Var}\left[\boldsymbol{Y}\right]-\mathrm{Var}\left[\boldsymbol{r}\right]+(K+1)\sum_{k\in 0\cup[K]}\mathbb{E}\left[(r^k(\boldsymbol{\theta}^*))^2\right]\boldsymbol{e}_k\boldsymbol{e}_k^\top-(K+1)\sum_{k\in 0\cup[K]}\mathbb{E}\left[\mathbb{E}\left[r^k(\boldsymbol{\theta}^*)\mid S\right]^2\right]\boldsymbol{e}_k\boldsymbol{e}_k^\top \\
&+ \mathrm{Var}\left[\mathbb{E}\left[\boldsymbol{r}\mid S\right]\right]+(K+1)^2\sum_{k,k'\in 0\cup[K]}\sum_{s,s'\in[s_{\max}]}\boldsymbol{\Sigma}^{\mathsf{CAR}}_{kk'}(s,s')\mathbb{E}\left[r^k(\boldsymbol{\theta}^*)\mid S=s\right]\mathbb{E}\left[r^{k'}(\boldsymbol{\theta}^*)\mid S=s'\right]\boldsymbol{e}_k\boldsymbol{e}_{k'}^\top \\
&= (K+1)\mathrm{diag}\left\{\mathrm{Var}\left[r^0(\boldsymbol{\theta}^*)\right],\ldots,\mathrm{Var}\left[r^K(\boldsymbol{\theta}^*)\right]\right\}+\mathrm{Var}\left[\boldsymbol{Y}\right]-\mathrm{Var}\left[\boldsymbol{r}\right] \\
&- (K+1)^2\sum_{k,k'\in 0\cup[K]}\sum_{s,s'\in[s_{\max}]}\{\boldsymbol{\Sigma}^{\mathsf{CR}}_{kk'}(s,s')-\boldsymbol{\Sigma}^{\mathsf{CAR}}_{kk'}(s,s')\}\boldsymbol{L}(s,k)\boldsymbol{L}(s',k')^\top.
\end{aligned}
$$

□

S2.2 Proof of Theorem 2 and Corollary 3

Theorem 2 (Reproduced from Theorem 2 in the main text)**.** *Under Assumptions 1 and S2, $\widehat{\boldsymbol{\mu}}$ in (2.2) satisfies:*

$$\sqrt{n}(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Gamma}), \tag{S2.15}$$

with $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_{\mathsf{base}} - \boldsymbol{\Gamma}_{\mathsf{adj}} - \boldsymbol{\Gamma}_{\mathsf{CAR,adj}}$. Here,

- $\boldsymbol{\Gamma}_{\mathsf{base}} = (K+1)\mathrm{diag}\left\{\mathrm{Var}\left[Y(0)\right], \ldots, \mathrm{Var}\left[Y(K)\right]\right\}$ *represents the baseline covariance matrix that does not account for covariate adjustment or the CAR procedures.*

- *With $r^k(\boldsymbol{\theta}^*) = Y(k) - h^k(\boldsymbol{X}, \boldsymbol{\theta}^*)$ and $\boldsymbol{r} = (r^0(\boldsymbol{\theta}^*), \ldots, r^K(\boldsymbol{\theta}^*))^\top$,*

$$\boldsymbol{\Gamma}_{\mathsf{adj}} = (K+1)\mathrm{diag}\left\{\mathrm{Var}\left[Y(0)\right] - \mathrm{Var}\left[r^0(\boldsymbol{\theta}^*)\right], \ldots, \mathrm{Var}\left[Y(K)\right] - \mathrm{Var}\left[r^K(\boldsymbol{\theta}^*)\right]\right\}$$
$$+ \mathrm{Var}\left[\boldsymbol{r}\right] - \mathrm{Var}\left[\boldsymbol{Y}\right]$$

captures the contribution from covariate adjustment in Z-estimation.

- *With $\boldsymbol{L}(s,k) = \mathbb{E}\left[r^k(\boldsymbol{\theta}^*) \mid S = s\right] \boldsymbol{e}_k$,*

$$\boldsymbol{\Gamma}_{\mathsf{CAR,adj}} = (K+1)^2 \sum_{k,k' \in 0 \cup [K]} \sum_{s,s' \in [s_{\max}]} \{\boldsymbol{\Sigma}^{\mathsf{CR}}_{kk'}(s,s') - \boldsymbol{\Sigma}^{\mathsf{CAR}}_{kk'}(s,s')\} \boldsymbol{L}(s,k) \boldsymbol{L}(s',k')^\top$$

depicts the interplay between covariate adjustment in Z-estimation and the CAR procedures.

S2.2.1 Corollary

Corollary 3 (Reproduced from Corollary 3 in the main text). *1. Under CR, Assumption 1 holds with $\mathbf{\Sigma}^{\mathsf{CR}}_{kk'}(s,s') = \mathbf{\Sigma}^{\mathsf{CAR}}_{kk'}(s,s')$ for $s,s' \in [s_{\max}]$ and $k,k' \in 0 \cup [K]$, and $\mathbf{\Gamma}_{\mathsf{CAR,adj}} = 0$.*

2. When $q = 0$ in the Z-estimation, $\mathbf{\Gamma}_{\mathsf{adj}} = 0$.

3. Under HH or STRPB, $\mathbf{\Gamma}_{\mathsf{CAR,adj}}$ is positive definite because $\mathbf{\Sigma}^{\mathsf{CAR}}_{kk'}(s,s') = 0$ for all $k,k' \in 0 \cup [K]$ and $s,s' \in [s_{\max}]$.

S2.2.2 Main proof of Theorem 2

Proof. By Lemma 10, we decompose $\mathbf{\Gamma}$ into three parts:

$$
\begin{aligned}
\mathbf{\Gamma} &= (K+1)\mathrm{diag}\left\{\mathrm{Var}\left[r^0(\boldsymbol{\theta}^*)\right], \ldots, \mathrm{Var}\left[r^K(\boldsymbol{\theta}^*)\right]\right\} + \mathrm{Var}\left[\boldsymbol{Y}\right] - \mathrm{Var}\left[\boldsymbol{r}\right] \\
&- (K+1)^2 \sum_{k,k' \in 0\cup[K]} \sum_{s,s' \in [s_{\max}]} \{\mathbf{\Sigma}^{\mathsf{CR}}_{kk'}(s,s') - \mathbf{\Sigma}^{\mathsf{CAR}}_{kk'}(s,s')\}\boldsymbol{L}(s,k)\boldsymbol{L}(s',k')^\top \\
&= (K+1)\mathrm{diag}\left\{\mathrm{Var}\left[Y(0)\right], \ldots, \mathrm{Var}\left[Y(K)\right]\right\} + \mathrm{Var}\left[\boldsymbol{Y}\right] - \mathrm{Var}\left[\boldsymbol{r}\right] \\
&+ (K+1)\mathrm{diag}\left\{\mathrm{Var}\left[r^0(\boldsymbol{\theta}^*)\right] - \mathrm{Var}\left[Y(0)\right], \ldots, \mathrm{Var}\left[r^K(\boldsymbol{\theta}^*)\right] - \mathrm{Var}\left[Y(K)\right]\right\} \\
&- (K+1)^2 \sum_{k,k' \in 0\cup[K]} \sum_{s,s' \in [s_{\max}]} \{\mathbf{\Sigma}^{k,k'}_{\mathsf{CR}}(s,s') - \mathbf{\Sigma}^{\mathsf{CAR}}_{kk'}(s,s')\}\boldsymbol{L}(s,k)\boldsymbol{L}(s',k')^\top \\
&:= \mathbf{\Gamma}_{\mathsf{base}} - \mathbf{\Gamma}_{\mathsf{adj}} - \mathbf{\Gamma}_{\mathsf{CAR,adj}}.
\end{aligned}
$$

Furthermore, to facilitate estimation, we expand $\boldsymbol{\Gamma}_{\mathsf{adj}}$ as follows:

$$\begin{aligned}\boldsymbol{\Gamma}_{\mathsf{adj}} &= (K+1)\cdot \operatorname{diag}\left\{\operatorname{Var}[Y(0)] - \operatorname{Var}[r^0(\boldsymbol{\theta}^*)], \ldots, \operatorname{Var}[Y(K)] - \operatorname{Var}[r^K(\boldsymbol{\theta}^*)]\right\} \\ &\quad + \operatorname{Var}[\boldsymbol{r}] - \operatorname{Var}[\boldsymbol{Y}] \\ &= (K+1)\cdot \operatorname{diag}\left\{\operatorname{Var}[Y(0)] - \operatorname{Var}[r^0(\boldsymbol{\theta}^*)], \ldots, \operatorname{Var}[Y(K)] - \operatorname{Var}[r^K(\boldsymbol{\theta}^*)]\right\} \\ &\quad + \operatorname{Var}[\boldsymbol{h}(\boldsymbol{X};\boldsymbol{\theta}^*)] - \operatorname{Cov}[\boldsymbol{h}(\boldsymbol{X};\boldsymbol{\theta}^*), \boldsymbol{Y}] - \operatorname{Cov}[\boldsymbol{h}(\boldsymbol{X};\boldsymbol{\theta}^*), \boldsymbol{Y}]^\top,\end{aligned}$$

where the second to fourth lines are obtained by substituting $\boldsymbol{r} = \boldsymbol{h}(\boldsymbol{X};\boldsymbol{\theta}^*)$ and expanding the expression using the relationship between covariance and variance, thereby yielding a form more amenable to estimation. □

S2.2.3 Proof of Corollary 3

Proof. For the first claim, since the complete randomization procedure satisfies Assumption 1, the result follows directly from Theorem 2.

For the second claim, when $q = 0$, i.e., when no covariate is included, we have $h^k(\boldsymbol{X}, \boldsymbol{\theta}^*) = h^k(\boldsymbol{\theta}^*)$, which is a constant. Consequently, $\operatorname{Var}[r^k(\boldsymbol{\theta}^*)] = \operatorname{Var}[Y(k)]$ and $\operatorname{Var}[\boldsymbol{r}] = \operatorname{Var}[\boldsymbol{Y}]$.

For the third claim, it is straightforward to verify that $D_n^k(s) = O_p(1)$ for $k \in 0 \cup [K]$ and $s \in [s_{\max}]$ under the STRPB procedure. Under the HH procedure, where $w_s > 0$ as noted in (Hu et al., 2023), and by Theorem 3.2 in (Hu et al., 2023), we also have $D_n^k(s) = O_p(1)$ for $k \in 0 \cup [K]$ and $s \in [s_{\max}]$. Consequently, $\boldsymbol{\Sigma}_{kk'}^{\mathsf{CAR}}(s, s') = 0$ for all $k, k' \in 0 \cup [K]$ and for all

$s, s' \in [s_{\max}]$ under both the STRPB and HH procedures.

□

S2.3 Proof of Theorem 4

Theorem 4 (Reproduced from Theorem 4 in the main text)**.** *Consider the parameter of interest* $\boldsymbol{\delta} = (g(\mu_1) - g(\mu_0), \ldots, g(\mu_K) - g(\mu_0))^\top$*, and denote by* $\widehat{\boldsymbol{\delta}}$ *its estimator obtained by replacing* $\boldsymbol{\mu}$ *with* $\widehat{\boldsymbol{\mu}}$*. Then under Assumptions 1 and S2,*

$$\sqrt{n}(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{G}\boldsymbol{\Gamma}\boldsymbol{G}^\top), \tag{S2.16}$$

where $\boldsymbol{G} = \frac{\partial \boldsymbol{\delta}}{\partial \boldsymbol{\mu}}$ *is the* $K \times (K+1)$ *Jacobian matrix of* $\boldsymbol{\delta}$ *evaluated at* $\boldsymbol{\mu}$*, and* $\boldsymbol{\Gamma}$ *is defined in Theorem 2.*

S2.3.1 Main proof of Theorem 4

Proof. The result follows immediately from Theorem 2 and the delta method (Casella and Berger, 2002, Theorem 5.5.28). □

S2.4 Proof of Proposition 5

Proposition 5 (Reproduced from Proposition 5 in the main text)**.** *Under Assumptions 1 and S2,* $\widehat{\boldsymbol{\Gamma}}_{\mathsf{conv}} \xrightarrow{p} \boldsymbol{\Gamma}_{\mathsf{conv}}$ *and* $\widehat{\boldsymbol{\Gamma}} \xrightarrow{p} \boldsymbol{\Gamma}$.

S2.4.1 Lemmas

Lemma 11. *Under Assumptions 1 and S2, the estimator* $\widehat{\boldsymbol{\Gamma}} := \widehat{\boldsymbol{\Gamma}}_{\mathsf{base}} - \widehat{\boldsymbol{\Gamma}}_{\mathsf{adj}} - \widehat{\boldsymbol{\Gamma}}_{\mathsf{CAR,adj}}$ *are consistent, where:*

$$\widehat{\boldsymbol{\Gamma}}_{\mathsf{base}} := \operatorname{diag}\left\{\frac{1}{\widehat{\pi}(0)}\widehat{\operatorname{Var}}\left[Y(0)\right], \cdots, \frac{1}{\widehat{\pi}(K)}\left[Y(K)\right]\right\},$$

$$\widehat{\boldsymbol{\Gamma}}_{\mathsf{adj}} := \widehat{\operatorname{Var}}\left[\boldsymbol{h}(\boldsymbol{X};\boldsymbol{\theta}^*)\right] - \widehat{\operatorname{Cov}}\left[\boldsymbol{h}(\boldsymbol{X};\boldsymbol{\theta}^*), \boldsymbol{Y}\right] - \widehat{\operatorname{Cov}}\left[\boldsymbol{h}(\boldsymbol{X};\boldsymbol{\theta}^*), \boldsymbol{Y}\right]^\top$$
$$+ \operatorname{diag}\left\{\frac{1}{\widehat{\pi}(0)}\{\widehat{\operatorname{Var}}\left[Y(0)\right] - \widehat{\operatorname{Var}}\left[r^0(\boldsymbol{\theta}^*)\right]\}, \cdots, \frac{1}{\widehat{\pi}(K)}\{\widehat{\operatorname{Var}}\left[Y(K)\right] - \widehat{\operatorname{Var}}\left[r^K(\boldsymbol{\theta}^*)\right]\}\right\},$$

$$\widehat{\boldsymbol{\Gamma}}_{\mathsf{CAR,adj}} := \widehat{\boldsymbol{\Gamma}}_{\mathsf{CR}} - \widehat{\boldsymbol{\Gamma}}_{\mathsf{CAR}}, \text{ with}$$

$$\widehat{\boldsymbol{\Gamma}}_{\mathsf{CR}} := \sum_{k,k' \in 0\cup[K]} \sum_{s,s' \in [s_{\max}]} \frac{1}{\widehat{\pi}(k)} \frac{1}{\widehat{\pi}(k')} \widehat{\boldsymbol{\Sigma}}^{\mathsf{CR}}_{kk'}(s,s') \widehat{\boldsymbol{L}}(s,k) \widehat{\boldsymbol{L}}(s',k')^\top,$$

$$\widehat{\boldsymbol{\Gamma}}_{\mathsf{CAR}} := \sum_{k,k' \in 0\cup[K]} \sum_{s,s' \in [s_{\max}]} \frac{1}{\widehat{\pi}(k)} \frac{1}{\widehat{\pi}(k')} \widehat{\boldsymbol{\Sigma}}^{\mathrm{CAR}}_{kk'}(s,s') \widehat{\boldsymbol{L}}(s,k) \widehat{\boldsymbol{L}}(s',k')^\top.$$

$$\widehat{\operatorname{Var}}\left[Y(k)\right] := \frac{1}{n_k}\sum_{i=1}^n T_i^k [Y_i - \overline{Y}(k)]^2, \quad \overline{Y}(k) := \frac{1}{n_k}\sum_{i=1}^n T_i^k Y_i,$$

$$\widehat{\operatorname{Cov}}\left[\boldsymbol{h}(\boldsymbol{X};\boldsymbol{\theta}^*), \boldsymbol{Y}\right] := \begin{bmatrix} \frac{1}{n_0}\sum_{i=1}^n T_i^0 \{Y_i - \overline{Y}(0)\} \boldsymbol{h}(\boldsymbol{X}_i;\widehat{\boldsymbol{\theta}})\}^\top \\ \frac{1}{n_1}\sum_{i=1}^n T_i^1 \{Y_i - \overline{Y}(1)\} \boldsymbol{h}(\boldsymbol{X}_i;\widehat{\boldsymbol{\theta}})^\top \\ \vdots \\ \frac{1}{n_K}\sum_{i=1}^n T_i^K \{Y_i - \overline{Y}(K)\} \boldsymbol{h}(\boldsymbol{X}_i;\widehat{\boldsymbol{\theta}})^\top \end{bmatrix},$$

$$\widehat{\operatorname{Var}}\left[r^k(\boldsymbol{\theta}^*)\right] := \frac{1}{\widehat{\pi}(k)}\sum_{i=1}^n T_i^k (r_i^k(\widehat{\boldsymbol{\theta}}))^2,$$

$$\widehat{\pi}_s := \frac{n(s)}{n}, \quad \widehat{\pi}(k) := \frac{1}{n}\sum_{i=1}^n T_i^k, \quad \widehat{\boldsymbol{L}}(s,k) := \frac{1}{n_k(s)}\sum_{i=1}^n T_i^k r_i^k(\widehat{\boldsymbol{\theta}}) \mathbb{I}(S_i = s) \boldsymbol{e}_k.$$

First, by substituting $\widehat{\pi}_s$ *and* $\widehat{\pi}(k)$*, we can compute* $\widehat{\boldsymbol{\Sigma}}^{\mathsf{CR}}_{kk'}(s,s')$*. Next, we estimate* $\widehat{\boldsymbol{\Gamma}}_{\mathsf{CAR}}$ *using a bootstrap-based approach. Let* B *denote the number of bootstrap replicates, for each* $1 \le b \le B$*, generate* $\{\{S_{i,b}\}_{i=1}^n\}_{b=1}^B$ *from the empirical distribution of* $\{S_i\}_{i=1}^n$*. Then, generate* $\{\{\boldsymbol{T}^*_{i,b}\}_{i=1}^n\}_{b=1}^B$ *according to the same CAR procedure as used in the trial, and evaluate the imbalance* $\{D^{*k}_{n,b}(s), 1 \le s \le s_{\max}, 0 \le k \le K\}_{b=1}^B$*. Subsequently, each element of* $\widehat{\boldsymbol{\Sigma}}^{\mathsf{CAR}}$ *can be estimated as*

$$\widehat{\boldsymbol{\Sigma}}^{\mathsf{CAR}}_{kk'}(s,s') = \frac{1}{n}\left\{\frac{1}{B}\sum_{b=1}^{B} D^{*k}_{n,b}(s) D^{*k'}_{n,b}(s') - \left(\frac{1}{B}\sum_{b=1}^{B} D^{*k}_{n,b}(s)\right)\left(\frac{1}{B}\sum_{b=1}^{B} D^{*k'}_{n,b}(s')\right)\right\}.$$

By plugging in $\widehat{\boldsymbol{L}}(s,k)$ *and* $\widehat{\boldsymbol{\Sigma}}^{\mathsf{CAR}}_{kk'}(s,s')$*, we obtain* $\widehat{\boldsymbol{\Gamma}}_{\mathsf{CAR}}$*. Furthermore, when the dimension of* $\widehat{\boldsymbol{\Sigma}}^{\mathsf{CAR}}$ *is large and memory efficiency is a concern, we can instead compute*

$$\boldsymbol{O}_b = \sum_{k=0}^{K}\sum_{s=1}^{s_{\max}} D^{*k}_{n,b}(s)\,\widehat{\boldsymbol{L}}(s,k), \quad 1 \le b \le B,$$

and directly estimate $\widehat{\boldsymbol{\Gamma}}_{\mathrm{CAR}}$ *using the sample covariance of* $\{\boldsymbol{O}_b\}_{b=1}^B$*, that is,*

$$\widehat{\boldsymbol{\Gamma}}_{\mathrm{CAR}} = \frac{1}{nB}\sum_{b=1}^{B} \boldsymbol{O}_b \boldsymbol{O}_b^\top - \frac{1}{n}\left(\frac{1}{B}\sum_{b=1}^{B}\boldsymbol{O}_b\right)\left(\frac{1}{B}\sum_{b=1}^{B}\boldsymbol{O}_b\right)^\top.$$

It is straightforward to verify that both methods yield equivalent estimates of $\widehat{\boldsymbol{\Gamma}}_{\mathrm{CAR}}$*.*

Lemma 12. *Under Assumptions 1 and S2,* $\widehat{\mathrm{Cov}}\left[\boldsymbol{h}(\boldsymbol{X};\boldsymbol{\theta}^*), \boldsymbol{Y}\right]$ *is consistent.*

Lemma 13. *Under Assumption 1,* $\widehat{\boldsymbol{\Sigma}}^{\mathsf{CAR}}$ *is consistent.*

Lemma 14. *For any* $\tilde{\boldsymbol{\pi}} \in \tilde{\boldsymbol{\Pi}}^0(\boldsymbol{\kappa})$*, where the latter is defined in the proof of Lemma 13,* (S2.18) *holds.*

S2.4.2 Main proof of Proposition 5

Proof. The result follows immediately from Lemma 11. □

S2.4.3 Proof of lemmas

Proof of Lemma 11. By Lemma 12, we know

$$\widehat{\mathrm{Cov}}\left[\boldsymbol{h}(\boldsymbol{X};\boldsymbol{\theta}^*), \boldsymbol{Y}\right] \xrightarrow{p} \mathrm{Cov}\left[\boldsymbol{h}(\boldsymbol{X};\boldsymbol{\theta}^*), \boldsymbol{Y}\right].$$

Similarly, we can prove the consistency of $\widehat{\boldsymbol{\Gamma}}_{\mathsf{base}}$, $\widehat{\boldsymbol{\Gamma}}_{\mathsf{adj}}$ and $r_s^k(\widehat{\boldsymbol{\theta}})$. By Lemma 13 and Slutsky's theorem, we can prove the consistency of $\widehat{\boldsymbol{\Gamma}}_{\mathsf{CAR}}$ and $\widehat{\boldsymbol{\Gamma}}_{\mathsf{CR}}$, then we finish the proof of this theorem. □

Proof of Lemma 12. By the multivariate Taylor's expansion of $\sum_{i=1}^n T_i^k Y_i \boldsymbol{h}(\boldsymbol{X}_i;\widehat{\boldsymbol{\theta}})$ around the point $\boldsymbol{\theta} = \boldsymbol{\theta}^*$, there exists a $\widetilde{\boldsymbol{\theta}}$ (which depends on $\widehat{\boldsymbol{\theta}}$) on the line segment between $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^*$ such that

$$\frac{1}{n_k}\sum_{i=1}^n T_i^k Y_i \left\{\boldsymbol{h}(\boldsymbol{X}_i;\widehat{\boldsymbol{\theta}}) - \boldsymbol{h}(\boldsymbol{X}_i;\boldsymbol{\theta}^*)\right\} = \left\{\frac{1}{n_k}\sum_{i=1}^n T_i^k Y_i \frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{h}(\boldsymbol{X}_i;\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right\}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$$
$$+ \sum_{k' \in 0 \cup [K]} \frac{1}{2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)^\top \left(\frac{1}{n_k}\sum_{i=1}^n T_i^k Y_i \frac{\partial^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top} h^{k'}(\boldsymbol{X}_i;\widetilde{\boldsymbol{\theta}})\right)(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)\boldsymbol{e}_{k'}.$$

From Assumptions 1.1, S2 and (S2.12), we know $\left\|\sum_{i=1}^{n} T_i^k Y_i \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} h^{k'}(\boldsymbol{X}_i;\widetilde{\boldsymbol{\theta}})/n_k\right\| = O_p(1)$, and thus the last term in the Taylor expansion is $O_p(\mathbf{1})\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2$. By Lemma 4, we have

$$\frac{1}{n_k}\sum_{i=1}^{n} T_i^k Y_i \left\{\boldsymbol{h}(\boldsymbol{X}_i;\widehat{\boldsymbol{\theta}}) - \boldsymbol{h}(\boldsymbol{X}_i;\boldsymbol{\theta}^*)\right\} = o_p(\mathbf{1}).$$

Then, by Assumption 1.1 and Lemma 8, we have

$$\begin{aligned}\frac{1}{n_k}\sum_{i=1}^{n} T_i^k Y_i \{\boldsymbol{h}(\boldsymbol{X}_i;\widehat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\mu}}\} &= \frac{1}{n_k}\sum_{i=1}^{n} T_i^k Y_i \{\boldsymbol{h}(\boldsymbol{X}_i;\boldsymbol{\theta}^*) - \boldsymbol{\mu}\} + o_p(\mathbf{1}) \\ &= \mathbb{E}\left[Y(k)(\boldsymbol{h}(\boldsymbol{X}_i;\boldsymbol{\theta}^*) - \boldsymbol{\mu})\right] + o_p(\mathbf{1}),\end{aligned}$$

so we finish the proof of this lemma. □

Proof of Lemma 13. Let $\boldsymbol{D}_n^{\mathcal{K}}(\mathcal{S}) \in \mathbb{R}^{(K+1)\times s_{\max}}$ denote the vector of observed within-stratum imbalances $\{D_n^k(s), k \in \{0\}\cup[K], s \in [s_{\max}]\}$. Denote the vector of true stratum probabilities $\{\pi_s, s \in [s_{\max}]\}$ by $\boldsymbol{\pi}^0$, and the vector of corresponding empirical probabilities $\{\pi_s^n = n^{-1}n(s), s \in [s_{\max}]\}$ by $\boldsymbol{\pi}^n$. Let $\mathcal{P}_{\boldsymbol{\pi}^0}$ and $\mathcal{P}_{\boldsymbol{\pi}^n}$ denote the probability measures associated with the parameters $\boldsymbol{\pi}^0$ and $\boldsymbol{\pi}^n$, respectively. Under Assumption 1, there exists a covariance matrix $\Xi_D(\boldsymbol{\pi}^0)$ (which may depend on $\boldsymbol{\pi}^0$) such that

$$\frac{\boldsymbol{D}_n^{\mathcal{K}}(\mathcal{S})}{\sqrt{n}} \xrightarrow{d}_{\mathcal{P}_{\boldsymbol{\pi}^0}} \mathcal{N}\Big(\mathbf{0}_{|\mathcal{V}|}, \Xi_D(\boldsymbol{\pi}^0)\Big), \tag{S2.17}$$

where $|\mathcal{V}| = (K+1)\times s_{\max}$ and the subscript $\mathcal{P}_{\boldsymbol{\pi}^0}$ indicates that convergence holds under that probability measure. Given (S2.17), we can employ the bootstrap framework based on local asymptotic normality (Beran, 1997) to establish the desired result.

The proof proceeds as follows. To apply the local asymptotic normality framework, define the local neighborhood

$$\widetilde{\boldsymbol{\Pi}}^0(\boldsymbol{\kappa}) = \left\{\tilde{\boldsymbol{\pi}} : \tilde{\boldsymbol{\pi}} = \boldsymbol{\pi}^0 + n^{-1/2}\boldsymbol{\kappa}\right\},$$

where $\boldsymbol{\kappa} \in \mathbb{R}^{s_{\max}}$ is a local parameter. For any $\tilde{\boldsymbol{\pi}} \in \widetilde{\boldsymbol{\Pi}}^0(\boldsymbol{\kappa})$, we use Le Cam's third lemma (see (Van der Vaart, 2000, Example 6.7, p. 90)) to show that under $\mathcal{P}_{\tilde{\boldsymbol{\pi}}}$ the distribution of $n^{-1/2}\boldsymbol{D}_n^{\mathcal{K}}(\mathcal{S})$ remains as in (S2.17), i.e.,

$$\frac{\boldsymbol{D}_n^{\mathcal{K}}(\mathcal{S})}{\sqrt{n}} \xrightarrow{d}_{\mathcal{P}_{\tilde{\boldsymbol{\pi}}}} \mathcal{N}\Big(\mathbf{0}_{|\mathcal{V}|}, \, \Xi_D(\boldsymbol{\pi}^0)\Big). \tag{S2.18}$$

Consequently, the sample covariance matrix $\widehat{\Xi}_D(\tilde{\boldsymbol{\pi}})$ computed from B random draws $\{\boldsymbol{D}_{n,b}^{\mathcal{K}}(\mathcal{S})\}_{b=1}^B$ of $\boldsymbol{D}_n^{\mathcal{K}}(\mathcal{S})$ under $\mathcal{P}_{\tilde{\boldsymbol{\pi}}}$ converges to $\Xi_D(\boldsymbol{\pi}^0)$ by (S2.18) and the law of large numbers. Because (S2.18) holds for every $\boldsymbol{\kappa} \in \mathbb{R}^{|\mathcal{S}|}$, it holds in particular for the $\boldsymbol{\kappa}$ that satisfies $n^{1/2}(\boldsymbol{\pi}^n - \boldsymbol{\pi}^0) = \boldsymbol{\kappa} + o_p(1)$. By the central limit theorem, $n^{1/2}(\boldsymbol{\pi}^n - \boldsymbol{\pi}^0) = O_p(1)$, so $\boldsymbol{\pi}^n \in \widetilde{\boldsymbol{\Pi}}^0(\boldsymbol{\kappa})$ for some $\boldsymbol{\kappa} \in \mathbb{R}^{|\mathcal{S}|}$. Hence, by the extended continuous mapping theorem (Van der Vaart and Wellner, 1996, Theorem 1.11.1, p. 67), $\widehat{\Xi}_D(\boldsymbol{\pi}^n)$ converges to $\Xi_D(\boldsymbol{\pi}^0)$ under $\mathcal{P}_{\boldsymbol{\pi}^n}$. Noting that the bootstrap samples $\{\boldsymbol{D}_{n,b}^{*\mathcal{K}}(\mathcal{S})\}_{b=1}^B$ in Lemma 11 are precisely $\{\boldsymbol{D}_{n,b}^{\mathcal{K}}(\mathcal{S})\}_{b=1}^B$ under $\mathcal{P}_{\boldsymbol{\pi}^n}$, we obtain

$$\widehat{\mathbf{\Sigma}}_{kk'}^{\mathsf{CAR}}(s,s') = \mathbf{\Sigma}_{kk'}^{\mathsf{CAR}}(s,s') + o_p(1).$$

□

Proof of Lemma 14. For notational simplicity, write the probability mass function of $\tilde{\boldsymbol{x}}_i$ under parameter $\boldsymbol{\pi}$ as

$$\pi(\tilde{\boldsymbol{x}}_i) = \prod_{s\in\mathcal{S}} [\pi_s]^{\mathbb{I}\{\tilde{\boldsymbol{x}}_i = \boldsymbol{x}(s)\}},$$

and let $\tilde{\pi}(\tilde{\boldsymbol{x}}_i)$ and $\pi^0(\tilde{\boldsymbol{x}}_i)$ denote the analogous functions with parameters $\tilde{\boldsymbol{\pi}}$ and $\boldsymbol{\pi}^0$, respectively. To apply Le Cam's third lemma, consider the log-likelihood ratio

$$l_n(\tilde{\boldsymbol{\pi}}, \boldsymbol{\pi}^0) = \sum_{i=1}^n \log\big\{\tilde{\pi}(\tilde{\boldsymbol{x}}_i)/\pi^0(\tilde{\boldsymbol{x}}_i)\big\}.$$

Let κ_s be the sth component of $\boldsymbol{\kappa}$. Expanding $\log(1+x)$ via Taylor series gives

$$\begin{aligned}
l_n(\tilde{\boldsymbol{\pi}}, \boldsymbol{\pi}^0) &= \sum_{i=1}^n \log\Big\{ \prod_{s\in[s_{\max}]} \Big(\frac{\tilde{\pi}_s}{\pi_s^0}\Big)^{\mathbb{I}\{\tilde{\boldsymbol{x}}_i = \boldsymbol{x}(s)\}} \Big\} \\
&= \sum_{s\in[s_{\max}]} \sum_{i=1}^n \mathbb{I}\{\tilde{\boldsymbol{x}}_i = \boldsymbol{x}(s)\} \log\Big\{1 + \frac{\kappa_s}{\sqrt{n}\,\pi_s^0}\Big\} \\
&= \sum_{i=1}^n \sum_{s\in[s_{\max}]} \mathbb{I}\{\tilde{\boldsymbol{x}}_i = \boldsymbol{x}(s)\} \frac{\kappa_s}{\sqrt{n}\,\pi_s^0} - \frac{1}{2n} \sum_{s\in[s_{\max}]} \sum_{i=1}^n \mathbb{I}\{\tilde{\boldsymbol{x}}_i = \boldsymbol{x}(s)\} \Big(\frac{\kappa_s}{\pi_s^0}\Big)^2 + O_p(n^{-1/2}) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{s\in[s_{\max}]} \big(\mathbb{I}\{\tilde{\boldsymbol{x}}_i = \boldsymbol{x}(s)\} - \pi_s^0\big) \frac{\kappa_s}{\pi_s^0} - \frac{1}{2} \sum_{s\in[s_{\max}]} \frac{\kappa_s^2}{\pi_s^0} + o_p(1),
\end{aligned}$$

provided $\sum_{s\in\mathcal{S}} \kappa_s = \sqrt{n} \sum_{s\in\mathcal{S}} (\pi_s - \pi_s^0) = 0$.

We now establish joint normality of $\big(n^{-1/2}\boldsymbol{D}_n^{\mathcal{K}}(\mathcal{S})^\top,\ l_n(\tilde{\boldsymbol{\pi}},\boldsymbol{\pi}^0)\big)^\top$. Define

$$\Delta l_i(\tilde{\boldsymbol{\pi}},\boldsymbol{\pi}) = \mathbb{E}[l_n(\tilde{\boldsymbol{\pi}},\boldsymbol{\pi}) \mid \mathcal{F}_i] - \mathbb{E}[l_n(\tilde{\boldsymbol{\pi}},\boldsymbol{\pi}) \mid \mathcal{F}_{i-1}] = \sum_{s\in\mathcal{S}} \big(\mathbb{I}\{\tilde{\boldsymbol{x}}_{i+1} = \boldsymbol{x}(s)\} - \pi_s^0\big)\frac{\kappa_s}{\pi_s^0},$$

and let $\widetilde{M}_i = \big(M_i^{\mathcal{K}}(\mathcal{S})^\top,\ \Delta l_i(\tilde{\boldsymbol{\pi}},\boldsymbol{\pi})\big)^\top$, where $M_i^{\mathcal{K}}(\mathcal{S})$ is the $|\mathcal{V}|$-vector of martingale differences $\{M_i^k(s), k\in\{0\}\cup[K], s\in[s_{\max}]\}$. The last component of $\widetilde{M}_i$ is bounded, so $\widetilde{M}_i$ is a sequence of $(|\mathcal{V}|+1)$-dimensional zero-mean martingale differences satisfying the conditional Lindeberg condition. Moreover, under Assumption 1 we have

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}\big[M_i^{\mathcal{K}}(\mathcal{S}) M_i^{\mathcal{K}}(\mathcal{S})^\top \mid \mathcal{F}_{i-1}\big] \xrightarrow{p}_{\mathcal{P}_{\boldsymbol{\pi}^0}} \boldsymbol{\Xi}_D(\boldsymbol{\pi}^0),$$

$$\begin{aligned}
&\sum_{i=1}^n \mathbb{E}\big[M_i^{\mathcal{K}}(\mathcal{S})\,\Delta l_i(\tilde{\boldsymbol{\pi}},\boldsymbol{\pi}) \mid \mathcal{F}_{i-1}\big] \\
&\qquad = \sum_{i=1}^n \mathbb{E}\big[M_i^{\mathcal{K}}(\mathcal{S}) \mid \mathcal{F}_{i-1}\big]\,\mathbb{E}\Big[\sum_{s\in\mathcal{S}}\big(\mathbb{I}\{\tilde{\boldsymbol{x}}_{i+1} = \boldsymbol{x}(s)\} - \pi_s^0\big) \,\Big|\, \mathcal{F}_{i-1}\Big]\frac{\kappa_s}{\pi_s^0} = \mathbf{0}_{|\mathcal{V}|},
\end{aligned}$$

and

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \mathbb{E}\big[\Delta l_i(\tilde{\boldsymbol{\pi}},\boldsymbol{\pi})^2 \mid \mathcal{F}_{i-1}\big] &= \sum_{s\in\mathcal{S}}\Big(\frac{\kappa_s}{\pi_s^0}\Big)^2 \pi_s^0(1-\pi_s^0) - \sum_{s\neq s'}\kappa_s\kappa_{s'} \\
&= \sum_{s\in\mathcal{S}}\frac{\kappa_s^2}{\pi_s^0} - \Big(\sum_{s\in\mathcal{S}}\kappa_s\Big)^2 = \sum_{s\in\mathcal{S}}\frac{\kappa_s^2}{\pi_s^0}.
\end{aligned}$$

Applying the martingale central limit theorem and the Cramér–Wold device yields

$$\begin{pmatrix} n^{-1/2}\boldsymbol{D}_n^{\mathcal{K}}(\mathcal{S}) \\ l_n(\tilde{\boldsymbol{\pi}},\boldsymbol{\pi}^0) \end{pmatrix} \xrightarrow{d}_{\mathcal{P}_{\boldsymbol{\pi}^0}} \mathcal{N}\left(\begin{pmatrix} \mathbf{0}_{|\mathcal{V}|} \\ -\dfrac{\nu^2}{2} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Xi}_D(\boldsymbol{\pi}^0) & \mathbf{0}_{|\mathcal{V}|}^\top \\ \mathbf{0}_{|\mathcal{V}|} & \nu^2 \end{pmatrix}\right),$$

where $\nu^2 = \sum_{s\in\mathcal{S}}(\pi_s^0)^{-1}\kappa_s^2$. Le Cam's third lemma therefore implies that (S2.18) holds. □

S2.5 Proof of Theorem 6 and Theorem 7

Theorem 6 (Reproduced from Theorem 6 in the main text)**.** *Under Assumptions 1 and S2, we consider the individual hypothesis test $H_0 : \delta_k = 0$ versus $H_1 : \delta_k > 0$. Define the individual conventional and model-robust Wald test statistics for $k \in [K]$ as*

$$W_k^{\mathsf{conv}} = \frac{\sqrt{n}\,\widehat{\delta}_k}{\sqrt{[\widehat{\boldsymbol{G}}\,\widehat{\boldsymbol{\Gamma}}_{\mathsf{conv}}\,\widehat{\boldsymbol{G}}^\top]_{(k,k)}}}, \quad \text{and} \quad W_k^{\mathsf{robust}} = \frac{\sqrt{n}\,\widehat{\delta}_k}{\sqrt{[\widehat{\boldsymbol{G}}\,\widehat{\boldsymbol{\Gamma}}\,\widehat{\boldsymbol{G}}^\top]_{(k,k)}}}.$$

W_k^{robust} always produces a valid type I error by Theorem 4 and Proposition 1. Because $\boldsymbol{\Gamma}_{\mathsf{conv}} - \boldsymbol{\Gamma}$ is positive definite under HH and STRPB, the conventional statistic W_k^{conv} exhibits a reduced type I error under these procedures.

Theorem 7 (Reproduced from Theorem 7 in the main text)**.** *Under Assumptions 1 and S2, we aim to conduct the following hypothesis test*

$$H_0^1 : \delta_1 = \delta_2 = \cdots = \delta_K = 0,$$

$$H_1^1 : \exists k \in [K], \delta_k > 0.$$

In particular, $(W_1^{\mathsf{robust}}, \ldots, W_K^{\mathsf{robust}})^\top \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{R})$ with

$$\boldsymbol{R}_{(k,k')} = \frac{[\boldsymbol{G}\boldsymbol{\Gamma}\boldsymbol{G}^\top]_{(k,k')}}{\sqrt{[\boldsymbol{G}\boldsymbol{\Gamma}\boldsymbol{G}^\top]_{(k,k)}[\boldsymbol{G}\boldsymbol{\Gamma}\boldsymbol{G}^\top]_{(k',k')}}}, \quad k, k' \in [K].$$

S2.5.1 Main proof of Theorem 6

Proof. By Theorem 4 and Proposition 5, the robust statistic W_k^{robust} satisfies $W_k^{\mathsf{robust}} \xrightarrow{d} N(0,1)$, and therefore controls the type I error at the nominal level. In contrast, Corollary 3 implies that $\boldsymbol{\Gamma}_{\mathsf{conv}} - \boldsymbol{\Gamma}$ is positive definite under the HH and STRPB procedures. Consequently, the conventional statistic W_k^{conv} converges in distribution to $N(0,\varrho)$ for some $0 < \varrho < 1$, leading to a reduced type I error under these randomization schemes. □

S2.5.2 Main proof of Theorem 7

Proof. For any fixed vector $\boldsymbol{a} \in \mathbb{R}^K$, consider the linear combination

$$\boldsymbol{a}^\top (W_1^{\mathsf{robust}}, \dots, W_K^{\mathsf{robust}})^\top,$$

since $\widehat{\boldsymbol{\delta}}$ is asymptotically normal and $\widehat{\boldsymbol{G}}\,\widehat{\boldsymbol{\Gamma}}\,\widehat{\boldsymbol{G}}^\top$ is consistent, Slutsky's theorem implies that this linear combination converges in distribution to a univariate normal random variable. Because this holds for every $\boldsymbol{a}$, the Cramér–Wold theorem ensures that $(W_1^{\mathsf{robust}}, \dots, W_K^{\mathsf{robust}})^\top$ is jointly asymptotically multivariate normal. Moreover, a direct computation verifies that its limiting covariance matrix is

$$\boldsymbol{R}_{(k,k')} = \frac{[\boldsymbol{G}\boldsymbol{\Gamma}\boldsymbol{G}^\top]_{(k,k')}}{\sqrt{[\boldsymbol{G}\boldsymbol{\Gamma}\boldsymbol{G}^\top]_{(k,k)}[\boldsymbol{G}\boldsymbol{\Gamma}\boldsymbol{G}^\top]_{(k',k')}}},$$

which matches the correlation matrix stated in the theorem. □

S3 Additional Numerical Results

S3.1 Example 1

In this section, Tables S2, S3, S4 and Figures S1, S2 present additional simulation results from Example 1. More specifically, Table S2 reports the results of conventional Wald tests across different estimands and randomization procedures; Tables S3 and S4 compare the conventional Wald test with the proposed model-robust version for the LOR and ATE metrics, respectively; Figures S1 and S2 display the statistical power of the conventional and model-robust tests under different randomization schemes as $\iota_{\max}$ increases for LOR and ATE, respectively.

Table S2: Results from conventional Wald tests for Example 1, the logistic model: test statistics, SD ($\times\sqrt{n}$), Type I error rates (in %) for Stage 1, Stage 2, and the overall Type I error, evaluated across different working models and randomization schemes: $(\iota_1, \iota_2) = (0, 0)$.

Metric	**Procedure**		**Stage 1**		**Stage 2**		**Overall**
			Mean (SD)	**Type I**	**Mean (SD)**	**Type I**	**Type I**
logRR	CR	$\mathcal{A}_0$	0.0664 (2.9292)	5.15	-0.0004 (2.6664)	5.02	5.11
		$\mathcal{A}_1$	0.0631 (2.8222)	5.13	-0.0004 (2.5550)	5.05	4.85
		$\mathcal{A}_2$	0.0511 (2.2592)	5.49	-0.0273 (0.0908)	5.22	5.14
	STRPB	$\mathcal{A}_0$	0.0502 (2.4190)	1.76	0.0009 (2.1713)	2.41	1.53
		$\mathcal{A}_1$	0.0502 (2.4189)	2.23	0.0009 (2.1702)	2.84	2.01
		$\mathcal{A}_2$	0.0460 (2.2209)	4.84	0.0091 (0.0884)	4.70	4.62
	HH	$\mathcal{A}_0$	0.0532 (2.3964)	1.62	0.0004 (2.2353)	2.67	1.60
		$\mathcal{A}_1$	0.0531 (2.3946)	2.17	0.0004 (2.2336)	3.21	2.15
		$\mathcal{A}_2$	0.0491 (2.1902)	5.00	0.0021 (0.0912)	5.23	4.96
	PS	$\mathcal{A}_0$	0.0529 (2.4672)	1.76	0.0009 (2.2249)	2.23	1.61
		$\mathcal{A}_1$	0.0528 (2.4663)	2.37	0.0008 (2.2233)	2.88	2.07
		$\mathcal{A}_2$	0.0477 (2.2283)	4.87	0.0107 (0.0908)	4.97	4.95
LOR	CR	$\mathcal{A}_0$	0.1040 (4.6484)	5.23	-0.0007 (4.1738)	5.05	5.26
		$\mathcal{A}_1$	0.0989 (4.4781)	5.13	-0.0006 (3.9998)	5.11	4.99
		$\mathcal{A}_2$	0.0801 (3.5755)	5.51	-0.0426 (0.1422)	5.25	5.30
	STRPB	$\mathcal{A}_0$	0.0786 (3.8335)	1.78	0.0013 (3.4012)	2.43	1.56
		$\mathcal{A}_1$	0.0785 (3.8333)	2.27	0.0013 (3.3995)	2.88	2.05
		$\mathcal{A}_2$	0.0720 (3.5149)	4.88	0.0141 (0.1385)	4.77	4.70
	HH	$\mathcal{A}_0$	0.0833 (3.7977)	1.65	0.0006 (3.4996)	2.69	1.61
		$\mathcal{A}_1$	0.0832 (3.7949)	2.22	0.0006 (3.4969)	3.23	2.20
		$\mathcal{A}_2$	0.0770 (3.4677)	5.05	0.0021 (0.1427)	5.28	5.07
	PS	$\mathcal{A}_0$	0.0828 (3.9043)	1.82	0.0013 (3.4826)	2.28	1.62
		$\mathcal{A}_1$	0.0826 (3.9028)	2.42	0.0013 (3.4801)	2.89	2.12
		$\mathcal{A}_2$	0.0747 (3.5217)	4.93	0.0171 (0.1420)	5.04	5.02
ATE	CR	$\mathcal{A}_0$	0.0239 (1.0774)	5.33	-0.0002 (0.9599)	5.10	5.38
		$\mathcal{A}_1$	0.0227 (1.0379)	5.28	-0.0002 (0.9201)	5.14	5.21
		$\mathcal{A}_2$	0.0184 (0.8285)	5.60	-0.0097 (0.0327)	5.28	5.38
	STRPB	$\mathcal{A}_0$	0.0180 (0.8888)	1.81	0.0003 (0.7831)	2.46	1.58
		$\mathcal{A}_1$	0.0180 (0.8887)	2.33	0.0003 (0.7827)	2.88	2.12
		$\mathcal{A}_2$	0.0165 (0.8145)	4.92	0.0032 (0.0319)	4.81	4.76
	HH	$\mathcal{A}_0$	0.0191 (0.8805)	1.67	0.0001 (0.8054)	2.74	1.71
		$\mathcal{A}_1$	0.0191 (0.8799)	2.27	0.0001 (0.8048)	3.29	2.27
		$\mathcal{A}_2$	0.0177 (0.8039)	5.17	0.0003 (0.0328)	5.35	5.18
	PS	$\mathcal{A}_0$	0.0190 (0.9043)	1.89	0.0003 (0.8013)	2.38	1.64
		$\mathcal{A}_1$	0.0190 (0.9039)	2.51	0.0003 (0.8008)	2.92	2.19
		$\mathcal{A}_2$	0.0172 (0.8154)	4.99	0.0040 (0.0327)	5.06	5.07

Table S3: LOR results for Example 1, the logistic model: Type I error rates and power (in %) for Stage 1, Stage 2, and the combined analysis, evaluated across different working models and randomization schemes.

	Procedure	Test	Type I			Power		
			Stage 1	Stage 2	All	Stage 1	Stage 2	All
			$(\iota_1, \iota_2) = (0, 0)$			$(\iota_1, \iota_2) = (0.3, 0.4)$		
$\mathcal{A}_0$	CR		5.23	5.05	5.26	21.77	29.72	38.84
	STRPB	conv	1.78	2.43	1.56	14.76	26.48	31.89
		robust	4.89	4.92	5.04	28.42	39.29	51.01
	HH	conv	1.65	2.69	1.61	15.90	26.56	32.22
		robust	5.08	5.47	5.26	29.69	39.63	51.79
	PS	conv	1.82	2.28	1.62	15.46	26.69	32.59
		robust	4.90	4.90	4.86	28.50	39.31	51.07
$\mathcal{A}_1$	CR		5.13	5.11	4.99	23.76	31.94	41.80
	STRPB	conv	2.27	2.88	2.05	17.69	29.34	36.12
		robust	4.90	5.02	5.04	28.45	39.35	51.10
	HH	conv	2.22	3.23	2.20	18.57	29.58	36.74
		robust	5.11	5.45	5.17	29.67	39.63	51.77
	PS	conv	2.42	2.89	2.12	18.25	29.80	37.02
		robust	4.93	4.88	4.93	28.58	39.39	51.02
$\mathcal{A}_2$	CR		5.51	5.25	5.30	33.18	43.80	57.91
	STRPB	conv	4.88	4.77	4.70	31.59	44.55	56.97
		robust	4.95	4.78	4.75	31.95	44.66	57.17
	HH	conv	5.05	5.28	5.07	32.37	44.15	58.00
		robust	5.21	5.30	5.16	32.64	44.18	58.29
	PS	conv	4.93	5.04	5.02	31.92	43.98	58.11
		robust	4.98	5.05	5.04	32.07	43.98	58.19

Table S4: ATE results for Example 1, the logistic model: Type I error rates and power (in %) for Stage 1, Stage 2, and the combined analysis, evaluated across different working models and randomization schemes.

	Procedure	**Test**	**Type I**			**Power**		
			Stage 1	**Stage 2**	**All**	**Stage 1**	**Stage 2**	**All**
			$(\iota_1, \iota_2) = (0, 0)$			$(\iota_1, \iota_2) = (0.3, 0.4)$		
$\mathcal{A}_0$	CR		5.33	5.10	5.38	22.27	30.02	39.39
	STRPB	conv	1.81	2.46	1.58	15.10	26.70	32.25
		robust	4.96	4.95	5.10	28.65	39.40	51.37
	HH	conv	1.67	2.74	1.71	16.25	26.87	32.69
		robust	5.16	5.52	5.33	29.89	39.83	52.16
	PS	conv	1.89	2.38	1.64	15.87	27.01	32.97
		robust	4.97	4.98	4.99	28.73	39.48	51.38
$\mathcal{A}_1$	CR		5.28	5.14	5.21	24.18	32.18	42.24
	STRPB	conv	2.33	2.88	2.12	18.00	29.58	36.57
		robust	4.99	5.03	5.11	28.71	39.56	51.45
	HH	conv	2.27	3.29	2.27	18.87	29.78	37.16
		robust	5.17	5.51	5.28	30.04	39.79	52.05
	PS	conv	2.51	2.92	2.19	18.64	30.06	37.36
		robust	5.05	4.99	5.08	28.86	39.59	51.34
$\mathcal{A}_2$	CR		5.60	5.28	5.38	33.43	43.99	58.16
	STRPB	conv	4.92	4.81	4.76	31.83	44.69	57.19
		robust	5.00	4.82	4.79	32.22	44.79	57.39
	HH	conv	5.17	5.35	5.18	32.59	44.21	58.29
		robust	5.28	5.37	5.25	32.87	44.31	58.56
	PS	conv	4.99	5.06	5.07	32.20	44.08	58.29
		robust	5.03	5.06	5.07	32.35	44.11	58.37

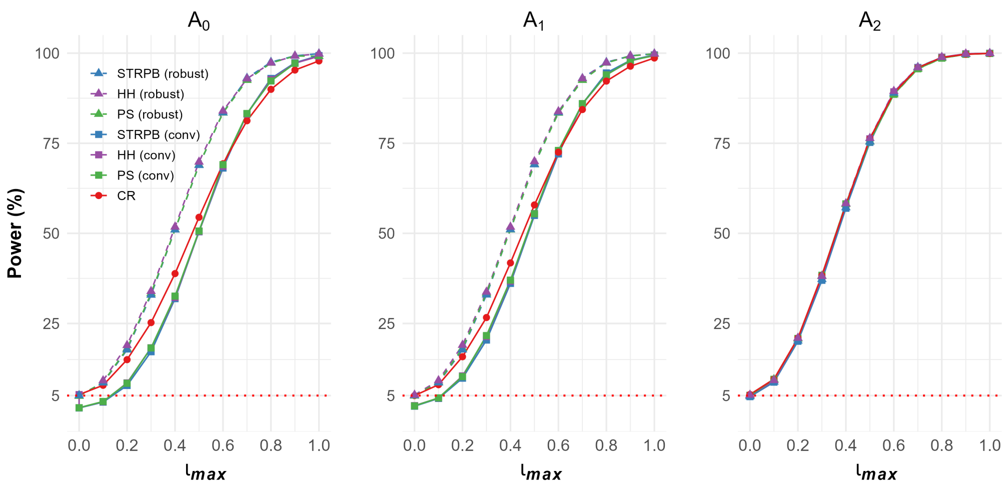

Figure S1: LOR for Example 1, power comparison of conventional and model-robust tests under $(\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2)$ across $\iota_{\max}$; the red dashed line denotes $\alpha = 0.05$ under the null hypothesis.

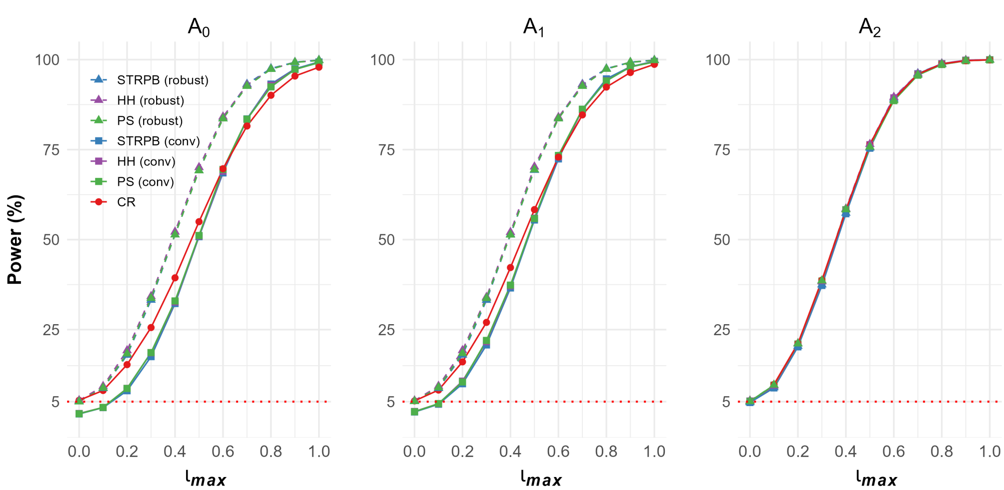

Figure S2: ATE for Example 1, power comparison of conventional and model-robust tests under $(\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2)$ across $\iota_{\max}$; the red dashed line denotes $\alpha = 0.05$ under the null hypothesis.

S3.2 Example 2

In this section, Tables S5, S6, S7 and Figures S3, S4, S5 present additional simulation results from Example 2. More specifically, Table S5 reports the results of conventional Wald tests across different estimands and randomization procedures; Tables S6 and S7 compare the conventional Wald test with the proposed model-robust version for the LOR and ATE metrics, respectively; Figures S3, S4 and S5 display the statistical power of the conventional and model-robust tests under different randomization schemes as $\iota_{\max}$ increases for logRR, LOR and ATE, respectively.

Table S5: Results from conventional Wald tests for Example 2, the probit model: test statistics, SD ($\times\sqrt{n}$), Type I error rates (in %) for Stage 1, Stage 2, and the overall Type I error, evaluated across different working models and randomization schemes: $(\iota_1, \iota_2) = (0, 0)$.

Metric	Procedure		Stage 1		Stage 2		Overall
			Mean (SD)	**Type I**	**Mean (SD)**	**Type I**	**Type I**
RR	CR	$\mathcal{A}_0$	0.0433 (1.9857)	4.92	-0.0001 (1.8169)	5.16	5.13
		$\mathcal{A}_1$	0.0394 (1.8350)	5.30	-0.0006 (1.6594)	5.16	5.54
		$\mathcal{A}_2$	0.0384 (1.7727)	5.29	-0.0005 (1.6108)	5.31	5.55
	STRPB	$\mathcal{A}_0$	0.0375 (1.7606)	2.59	-0.0002 (1.5993)	3.16	2.47
		$\mathcal{A}_1$	0.0364 (1.7021)	3.75	-0.0000 (1.5422)	4.15	3.85
		$\mathcal{A}_2$	0.0351 (1.6381)	3.51	-0.0001 (1.4839)	4.10	3.62
	HH	$\mathcal{A}_0$	0.0368 (1.7593)	2.68	-0.0015 (1.5987)	3.18	2.58
		$\mathcal{A}_1$	0.0358 (1.6961)	3.67	-0.0013 (1.5368)	3.81	3.64
		$\mathcal{A}_2$	0.0351 (1.6419)	3.79	-0.0011 (1.4861)	3.90	3.76
	PS	$\mathcal{A}_0$	0.0389 (1.8800)	3.63	0.0014 (1.6782)	3.78	3.72
		$\mathcal{A}_1$	0.0380 (1.8354)	5.21	0.0014 (1.6172)	4.97	4.84
		$\mathcal{A}_2$	0.0370 (1.7732)	5.06	0.0013 (1.5671)	4.96	4.89
LOR	CR	$\mathcal{A}_0$	0.0976 (4.5811)	4.85	-0.0003 (4.0916)	5.24	5.18
		$\mathcal{A}_1$	0.0889 (4.2311)	5.18	-0.0015 (3.7427)	5.14	5.51
		$\mathcal{A}_2$	0.0867 (4.0834)	5.20	-0.0013 (3.6324)	5.29	5.51
	STRPB	$\mathcal{A}_0$	0.0848 (4.0654)	2.58	-0.0006 (3.6036)	3.18	2.50
		$\mathcal{A}_1$	0.0823 (3.9315)	3.59	-0.0001 (3.4761)	4.19	3.80
		$\mathcal{A}_2$	0.0794 (3.7797)	3.47	-0.0002 (3.3440)	4.14	3.61
	HH	$\mathcal{A}_0$	0.0833 (4.0511)	2.63	-0.0033 (3.6028)	3.18	2.59
		$\mathcal{A}_1$	0.0811 (3.9080)	3.61	-0.0030 (3.4645)	3.81	3.62
		$\mathcal{A}_2$	0.0794 (3.7791)	3.74	-0.0024 (3.3487)	3.91	3.76
	PS	$\mathcal{A}_0$	0.0876 (4.3333)	3.60	0.0033 (3.7798)	3.80	3.73
		$\mathcal{A}_1$	0.0858 (4.2351)	5.10	0.0032 (3.6474)	4.99	4.78
		$\mathcal{A}_2$	0.0835 (4.0882)	4.96	0.0030 (3.5322)	5.02	4.87
ATE	CR	$\mathcal{A}_0$	0.0239 (1.1167)	5.12	-0.0001 (1.0058)	5.30	5.48
		$\mathcal{A}_1$	0.0218 (1.0321)	5.37	-0.0004 (0.9200)	5.23	5.70
		$\mathcal{A}_2$	0.0213 (0.9967)	5.45	-0.0003 (0.8930)	5.38	5.70
	STRPB	$\mathcal{A}_0$	0.0208 (0.9917)	2.77	-0.0001 (0.8860)	3.19	2.74
		$\mathcal{A}_1$	0.0202 (0.9592)	3.81	-0.0000 (0.8547)	4.27	4.00
		$\mathcal{A}_2$	0.0195 (0.9227)	3.60	-0.0000 (0.8224)	4.17	3.76
	HH	$\mathcal{A}_0$	0.0204 (0.9890)	2.89	-0.0008 (0.8859)	3.21	2.73
		$\mathcal{A}_1$	0.0199 (0.9541)	3.77	-0.0007 (0.8519)	3.84	3.80
		$\mathcal{A}_2$	0.0195 (0.9231)	3.84	-0.0006 (0.8236)	3.95	3.85
	PS	$\mathcal{A}_0$	0.0215 (1.0569)	3.95	0.0008 (0.9292)	3.87	3.91
		$\mathcal{A}_1$	0.0211 (1.0329)	5.17	0.0008 (0.8966)	5.08	4.96
		$\mathcal{A}_2$	0.0205 (0.9976)	5.11	0.0007 (0.8685)	5.05	5.05

Table S6: LOR results for Example 2, the probit model: Type I error rates and power (in %) for Stage 1, Stage 2, and the combined analysis, evaluated across different working models and randomization schemes.

	Procedure	Test	Type I			Power		
			Stage 1	Stage 2	All	Stage 1	Stage 2	All
			$(\iota_1, \iota_2) = (0, 0)$			$(\iota_1, \iota_2) = (0.2, 0.3)$		
$\mathcal{A}_0$	CR		5.18	5.24	4.85	26.47	37.46	48.59
	STRPB	conv	2.58	3.18	2.50	23.72	36.22	46.83
		robust	4.71	4.84	5.06	31.45	43.57	56.82
	HH	conv	2.63	3.18	2.59	22.75	35.96	46.77
		robust	4.54	4.91	4.85	30.77	42.56	55.88
	PS	conv	3.60	3.80	3.73	25.00	37.13	48.16
		robust	4.79	5.25	4.92	28.70	41.42	53.21
$\mathcal{A}_1$	CR		5.51	5.14	5.18	29.87	41.52	54.06
	STRPB	conv	3.59	4.19	3.80	28.62	41.83	53.51
		robust	4.99	5.02	5.11	33.11	45.55	58.63
	HH	conv	3.61	3.81	3.62	27.42	41.34	53.57
		robust	4.89	4.85	4.84	32.07	44.99	58.41
	PS	conv	5.10	4.99	4.78	30.31	43.09	55.41
		robust	5.02	5.01	4.80	30.00	43.07	55.32
$\mathcal{A}_2$	CR		5.51	5.29	5.20	31.84	43.54	56.22
	STRPB	conv	3.47	4.14	3.61	29.86	44.29	56.11
		robust	4.81	5.06	4.93	35.04	48.36	61.68
	HH	conv	3.74	3.91	3.76	29.27	43.78	56.44
		robust	4.90	4.91	4.96	34.38	47.66	61.40
	PS	conv	5.08	4.97	4.94	31.78	45.16	57.69
		robust	4.96	5.02	4.87	31.50	45.29	57.55

Table S7: ATE results for Example 2, the probit model: Type I error rates and power (in %) for Stage 1, Stage 2, and the combined analysis, evaluated across different working models and randomization schemes.

	Procedure	Test	Type I			Power		
			Stage 1	**Stage 2**	**All**	**Stage 1**	**Stage 2**	**All**
			$(\iota_1, \iota_2) = (0, 0)$			$(\iota_1, \iota_2) = (0.2, 0.3)$		
$\mathcal{A}_0$	CR		5.48	5.30	5.12	49.36	37.70	27.22
	STRPB	conv	2.77	3.19	2.74	24.51	36.59	47.51
		robust	4.94	4.89	5.20	32.29	43.74	57.49
	HH	conv	2.89	3.21	2.73	23.57	36.17	47.47
		robust	4.79	4.97	5.03	31.52	42.76	56.47
	PS	conv	3.95	3.87	3.91	25.87	37.44	48.89
		robust	4.97	5.32	5.10	29.55	41.62	53.84
$\mathcal{A}_1$	CR		5.70	5.23	5.37	54.84	41.72	30.75
	STRPB	conv	3.81	4.27	4.00	29.52	41.99	54.10
		robust	5.19	5.07	5.26	33.95	45.73	59.35
	HH	conv	3.77	3.84	3.80	28.12	41.56	54.21
		robust	5.03	4.87	5.02	32.78	45.12	59.05
	PS	conv	5.28	5.07	4.93	31.06	43.24	56.12
		robust	5.17	5.08	4.96	30.75	43.26	55.97
$\mathcal{A}_2$	CR		5.70	5.38	5.45	56.79	43.71	32.65
	STRPB	conv	3.60	4.17	3.76	30.60	44.51	56.84
		robust	4.95	5.08	5.09	35.86	48.52	62.07
	HH	conv	3.84	3.95	3.85	30.17	43.98	56.94
		robust	5.12	4.94	5.09	35.12	47.87	62.07
	PS	conv	5.23	5.01	5.11	32.64	45.42	58.35
		robust	5.11	5.05	5.05	32.24	45.48	58.04

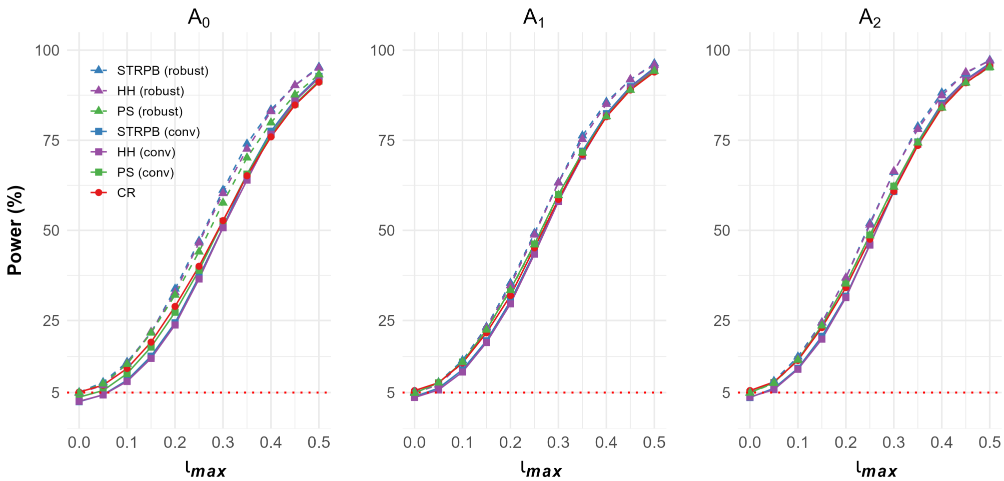

Figure S3: logRR for Example 2, power comparison of conventional and model-robust tests under $(\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2)$ across $\iota_{\max}$; the red dashed line denotes $\alpha = 0.05$ under the null hypothesis.

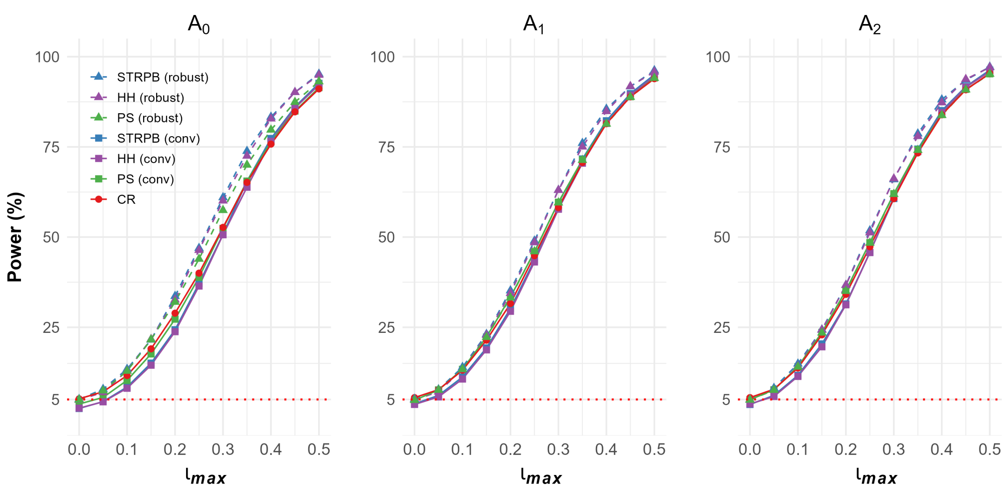

Figure S4: LOR for Example 2, power comparison of conventional and model-robust tests under $(\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2)$ across $\iota_{\max}$; the red dashed line denotes $\alpha = 0.05$ under the null hypothesis.

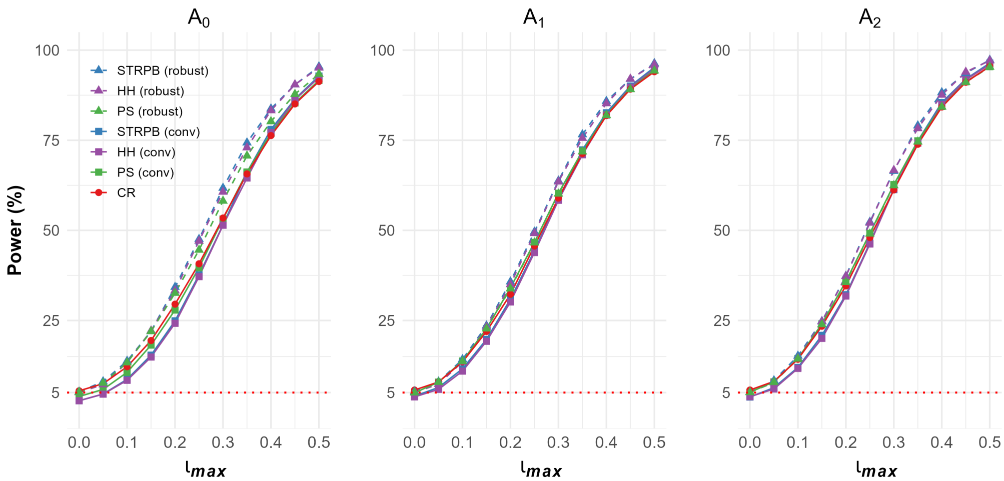

Figure S5: ATE for Example 2, power comparison of conventional and model-robust tests under $(\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2)$ across $\iota_{\max}$; the red dashed line denotes $\alpha = 0.05$ under the null hypothesis.

S3.3 Hypothetical Trial Example

In this section, Tables S8, S9 and S10 present additional results from the hypothetical trial example. More specifically, Table S8 reports the results of conventional Wald tests across different estimands and randomization procedures; Tables S9 and S10 compare the conventional Wald test with the proposed model-robust version for the LOR and ATE metrics.

Table S8: Results from conventional Wald tests for the hypothetical trial: test statistics, SD ($\times\sqrt{n}$), Type I error rates (in %) for Stage 1, Stage 2, and the overall Type I error, evaluated across different working models and randomization schemes: $(\iota_1, \iota_2) = (0, 0)$.

Metric	Procedure		Stage 1		Stage 2		Overall
			Mean (SD)	Type I	Mean (SD)	Type I	Type I
logRR	CR	$\mathcal{A}_0$	0.1040 (4.6883)	4.90	-0.0013 (4.2830)	4.86	4.76
		$\mathcal{A}_1$	0.1003 (4.6080)	4.73	0.0019 (4.1792)	5.12	4.87
		$\mathcal{A}_2$	0.0976 (4.4921)	4.74	0.0019 (4.0624)	5.18	5.09
	STRPB	$\mathcal{A}_0$	0.0960 (4.4737)	3.99	-0.0025 (4.0903)	3.81	3.33
		$\mathcal{A}_1$	0.0960 (4.4548)	4.36	-0.0026 (4.0635)	4.11	3.76
		$\mathcal{A}_2$	0.1017 (4.4499)	5.05	-0.0013 (4.0527)	4.79	4.81
	HH	$\mathcal{A}_0$	0.0988 (4.5224)	3.89	-0.0001 (4.1135)	4.30	4.11
		$\mathcal{A}_1$	0.0984 (4.4997)	4.07	-0.0000 (4.0993)	4.74	4.54
		$\mathcal{A}_2$	0.1004 (4.4611)	4.83	0.0004 (4.0596)	4.64	4.90
	PS	$\mathcal{A}_0$	0.1005 (4.5501)	4.23	0.0008 (4.1135)	4.10	3.98
		$\mathcal{A}_1$	0.1002 (4.5288)	4.69	0.0008 (4.0896)	4.46	4.43
		$\mathcal{A}_2$	0.0957 (4.4654)	4.67	0.0025 (4.0555)	4.96	4.68
LOR	CR	$\mathcal{A}_0$	0.1273 (5.7768)	4.97	-0.0015 (5.2392)	4.98	4.86
		$\mathcal{A}_1$	0.1227 (5.6778)	4.82	0.0024 (5.1142)	5.17	5.01
		$\mathcal{A}_2$	0.1194 (5.5356)	4.86	0.0023 (4.9712)	5.21	5.24
	STRPB	$\mathcal{A}_0$	0.1175 (5.5109)	4.10	-0.0030 (5.0007)	3.89	3.50
		$\mathcal{A}_1$	0.1174 (5.4879)	4.46	-0.0032 (4.9681)	4.16	3.86
		$\mathcal{A}_2$	0.1244 (5.4822)	5.19	-0.0016 (4.9581)	4.83	4.92
	HH	$\mathcal{A}_0$	0.1208 (5.5749)	4.00	-0.0002 (5.0325)	4.36	4.18
		$\mathcal{A}_1$	0.1204 (5.5461)	4.19	0.0000 (5.0150)	4.81	4.62
		$\mathcal{A}_2$	0.1229 (5.4960)	4.97	0.0004 (4.9639)	4.68	5.00
	PS	$\mathcal{A}_0$	0.1230 (5.6039)	4.34	0.0010 (5.0305)	4.16	4.08
		$\mathcal{A}_1$	0.1225 (5.5769)	4.81	0.0009 (5.0016)	4.56	4.56
		$\mathcal{A}_2$	0.1171 (5.5024)	4.79	0.0031 (4.9601)	5.02	4.90
ATE	CR	$\mathcal{A}_0$	0.0189 (0.8828)	5.12	-0.0002 (0.7813)	5.04	5.08
		$\mathcal{A}_1$	0.0182 (0.8679)	4.98	0.0004 (0.7637)	5.27	5.24
		$\mathcal{A}_2$	0.0177 (0.8467)	4.99	0.0004 (0.7424)	5.34	5.41
	STRPB	$\mathcal{A}_0$	0.0174 (0.8418)	4.23	-0.0004 (0.7445)	4.06	3.68
		$\mathcal{A}_1$	0.0174 (0.8385)	4.58	-0.0005 (0.7397)	4.21	4.00
		$\mathcal{A}_2$	0.0185 (0.8377)	5.31	-0.0002 (0.7397)	4.94	5.21
	HH	$\mathcal{A}_0$	0.0179 (0.8535)	4.09	-0.0000 (0.7509)	4.49	4.36
		$\mathcal{A}_1$	0.0178 (0.8488)	4.31	0.0000 (0.7482)	5.00	4.82
		$\mathcal{A}_2$	0.0183 (0.8398)	5.09	0.0000 (0.7393)	4.81	5.18
	PS	$\mathcal{A}_0$	0.0182 (0.8554)	4.45	0.0002 (0.7496)	4.34	4.27
		$\mathcal{A}_1$	0.0182 (0.8509)	4.90	0.0002 (0.7455)	4.68	4.72
		$\mathcal{A}_2$	0.0174 (0.8413)	4.97	0.0005 (0.7394)	5.10	5.09

Table S9: LOR results for the hypothetical trial: Type I error rates and power (in 10^{-2}) for Stage 1, Stage 2, and the combined analysis, across different working models and randomization schemes.

	Procedure	Test	Type I			Power		
			Stage 1	Stage 2	All	Stage 1	Stage 2	All
			$(\iota_1, \iota_2) = (0, 0)$			$(\iota_1, \iota_2) = (0.2, 0.4)$		
$\mathcal{A}_0$	CR		4.97	4.98	4.86	28.42	42.17	52.52
	STRPB	conv	4.10	3.89	3.50	26.52	42.17	51.90
		robust	5.11	5.07	5.06	30.28	45.16	56.43
	HH	conv	4.00	4.36	4.18	26.75	41.50	51.47
		robust	4.83	4.90	4.82	30.71	44.70	55.85
	PS	conv	4.34	4.16	4.08	27.47	42.07	52.00
		robust	5.37	4.91	5.07	30.68	44.95	55.84
$\mathcal{A}_1$	CR		4.82	5.17	5.01	28.99	44.14	54.23
	STRPB	conv	4.46	4.16	3.86	28.10	43.58	54.18
		robust	5.08	4.55	4.55	30.75	45.66	57.09
	HH	conv	4.19	4.81	4.62	28.38	43.16	53.92
		robust	4.82	5.35	5.30	31.10	45.44	56.62
	PS	conv	4.81	4.56	4.56	29.30	43.61	54.07
		robust	5.38	5.02	5.18	31.12	45.31	56.66
$\mathcal{A}_2$	CR		4.86	5.21	5.24	30.53	45.68	56.85
	STRPB	conv	5.19	4.83	4.92	30.64	45.72	56.97
		robust	5.17	4.60	4.58	30.90	45.77	57.12
	HH	conv	4.97	4.68	5.00	30.77	45.39	56.33
		robust	4.90	5.38	5.33	31.11	45.45	56.50
	PS	conv	4.79	5.02	4.90	31.33	45.39	56.97
		robust	5.35	5.03	5.21	31.52	45.41	57.04

Table S10: ATE results for the hypothetical trial: Type I error rates and power (in %) for Stage 1, Stage 2, and the combined analysis, evaluated across different working models and randomization schemes.

	Procedure	Test	Type I			Power		
			Stage 1	Stage 2	All	Stage 1	Stage 2	All
			$(\iota_1, \iota_2) = (0, 0)$			$(\iota_1, \iota_2) = (0.2, 0.4)$		
$\mathcal{A}_0$	CR		5.12	5.04	5.08	28.71	42.51	53.10
	STRPB	conv	4.23	4.06	3.68	26.83	42.50	52.36
		robust	5.28	5.21	5.28	30.69	45.49	56.92
	HH	conv	4.09	4.49	4.36	26.99	41.96	51.99
		robust	4.96	4.96	5.10	31.11	45.04	56.40
	PS	conv	4.45	4.34	4.27	27.65	42.52	52.36
		robust	5.52	4.99	5.34	30.97	45.25	56.30
$\mathcal{A}_1$	CR		4.98	5.27	5.24	29.22	44.51	54.77
	STRPB	conv	4.58	4.21	4.00	28.45	43.93	54.64
		robust	5.21	4.70	4.77	31.13	46.01	57.46
	HH	conv	4.31	5.00	4.82	28.82	43.49	54.36
		robust	4.98	5.48	5.45	31.49	45.77	56.98
	PS	conv	4.90	4.68	4.72	29.55	43.94	54.48
		robust	5.52	5.13	5.43	31.43	45.65	57.19
$\mathcal{A}_2$	CR		4.99	5.34	5.41	30.87	46.05	57.29
	STRPB	conv	5.31	4.94	5.21	31.06	46.03	57.42
		robust	5.35	4.76	4.73	31.29	46.05	57.58
	HH	conv	5.09	4.81	5.18	31.28	45.78	56.94
		robust	5.06	5.50	5.56	31.55	45.82	57.05
	PS	conv	4.97	5.10	5.09	31.73	45.71	57.38
		robust	5.44	5.14	5.41	31.87	45.73	57.47

S4 An additional example of estimand: the average treatment effect (ATE)

The ATE captures absolute risk differences and is widely used for policy and clinical decision making. For arm k,

$$\mathrm{ATE}_k = \mu_k - \mu_0, \qquad \boldsymbol{\delta} = \big(\mu_1 - \mu_0, \ldots, \mu_K - \mu_0\big)^\top.$$

Estimation can proceed from generalized linear working models (including logistic regression for binary outcomes), with inference carried out analogously to logRR and LOR.

Lemma 14. *For the ATE, $g(x) = x$ and*

$$\boldsymbol{G}_{\mathsf{ATE}} = \big(-\mathbf{1}_K, \ \ \mathrm{diag}\{1, \ldots, 1\}\big), \qquad \widehat{\boldsymbol{G}}_{\mathsf{ATE}} = \boldsymbol{G}_{\mathsf{ATE}}\big|_{\mu=\widehat{\mu}}.$$

Hence Theorem 4 implies

$$\begin{pmatrix} \widehat{\mu}_1 - \widehat{\mu}_0 \\ \vdots \\ \widehat{\mu}_K - \widehat{\mu}_0 \end{pmatrix} - \begin{pmatrix} \mu_1 - \mu_0 \\ \vdots \\ \mu_K - \mu_0 \end{pmatrix} \xrightarrow{d} N\big(0, \ \boldsymbol{G}_{\mathsf{ATE}} \, \boldsymbol{\Gamma} \, \boldsymbol{G}_{\mathsf{ATE}}^\top\big) \, .$$

References

Beran, R. (1997). Bootstrap tests and confidence regions. *Annals of Statistics 25*(1), 454 – 489.

Casella, G. and R. L. Berger (2002). *Statistical Inference*. Duxbury.

Hall, P. and C. C. Heyde (2014). *Martingale limit theory and its application*. New York: Academic press.

Hu, F., X. Ye, and L.-X. Zhang (2023). Multi-arm covariate-adaptive randomization. *Science China Mathematics 66*(1), 163–190.

Liu, Y. and F. Hu (2023). The impacts of unobserved covariates on covariate-adaptive randomized experiments. *The Annals of Statistics 51*(5), 1895–1920.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, Volume 3. Cambridge university press.

Van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer.