Statistical Inference with Generalized Estimating Equations in Longitudinal Studies following Covariate-Adaptive Randomization

Ruichen Fu and Yang Liu

*Institute of Statistics and Big Data, Renmin University of China, China**

Supplementary Material

We summarize the main results in Section S1 and discuss the construction of the stratification variables used for the proposed method in Section S2. Details of the proposed adjusted tests are presented in Section S3, and the proofs of the theoretical results are given in Section S4. Additional discussions on the validity of Assumption 3.1.5 and the numerical implementation of GEE are provided in Section S5. Supplementary numerical studies for continuous and binary outcomes are presented in Section S6. Finally, the description of the covariate effects estimated from the outcome model considered in Section 6 of the main text is provided in Section S7.

S1 Summary of Main Results

In this section, we summarize the framework, assumptions, and theoretical results. We consider the following outcome model:

*Corresponding author (Email: yangliu2022@ruc.edu.cn)

$$\begin{aligned}\mathbb{E}[Y_{ij}|A_i,\boldsymbol{X}_{i,j}] =& A_i\mu(\eta^1_{i,j}) + (1-A_i)\mu(\eta^0_{i,j}),\\ \text{and} \quad \eta^a_{i,j} =& \gamma_a + \boldsymbol{X}_{i,j}^\top\boldsymbol{\beta} \quad \text{for} \quad a\in\{0,1\},\end{aligned} \tag{2.1}$$

where $\mu(\cdot)$ represents the correctly specified link function; γ_1 and γ_0 are the effects of treatment and control, respectively; and $\boldsymbol{\beta}$ represents the effect of covariates. The working model used for the adjustment can be written as

$$\begin{aligned}\mathbb{E}[Y_{i,j}|A_i,\boldsymbol{X}_{i,j}] \sim& A_i m(\eta^{*1}_{i,j}) + (1-A_i)m(\eta^{*0}_{i,j}),\\ \text{and} \quad \eta^{*a}_{i,j} =& \boldsymbol{g}_{i,j,a}^\top\boldsymbol{\theta}^*_{\mathsf{in}} \quad \text{for} \quad a\in\{0,1\},\end{aligned} \tag{2.2}$$

where $\boldsymbol{g}_{i,j,a} = (a, 1-a, \boldsymbol{X}_{i,j,\mathsf{in}}^\top)^\top$, $m(\cdot)$ is the given link function used for estimating the parameter of interest $\boldsymbol{\theta}^* = (\gamma_1^*, \gamma_0^*, \boldsymbol{\beta}_{\mathsf{in}}^{*\top})^\top$, and the $*$ on the superscript indicates that the value of the parameters may be different from the true value, since $m(\cdot)$ can be different from $\eta(\cdot)$ and the covariates $\boldsymbol{X}_{i,j,\mathsf{in}}$ may not equal to $\boldsymbol{X}_{i,j}$.

For $a\in\{0,1\}$, let $\boldsymbol{Y}_i(a) = (Y_{i,1}(a),\ldots,Y_{i,n_i}(a))^\top$, $\boldsymbol{m}_i(a) = (m_{i,1}(a),\ldots,m_{i,n_i}(a))^\top$, $m'_{i,j}(a) = \frac{d}{d\eta}m(\eta)|_{\eta=\eta^{*a}_{i,j}}$ and denote $v_{i,j}(a)$ as the model-based estimator for $Var[Y_{i,j}(a)|\boldsymbol{X}_{i,j,\mathsf{in}}]$. Furthermore, let $\widetilde{\boldsymbol{M}}_i(a) = \mathrm{diag}\{m'_{i,1}(a),\ldots,m'_{i,n_i}(a)\}$, $\boldsymbol{G}_{i,a} = (\boldsymbol{g}_{i,1,a},\ldots,\boldsymbol{g}_{i,n_i,a})^\top$, and $\boldsymbol{\Lambda}_{i,a} = \mathrm{diag}\{v_{i,1}(a),\ldots,v_{i,n_i}(a)\}$, $\boldsymbol{Y}_i = A_i\boldsymbol{Y}_i(1)+$

$(1-A_i)\boldsymbol{Y}_i(0)$, and $\boldsymbol{m}_i = A_i\boldsymbol{m}_i(1) + (1-A_i)\boldsymbol{m}_i(0)$. Then, the generalized estimation equation associated with (2.2) considers solving

$$\begin{aligned}\mathbf{0}_{p_0+2} = \Psi_{N,\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*) &= \sum_{i=1}^{N}\left(\frac{\partial \boldsymbol{m}_i}{\partial \boldsymbol{\theta}^*}\right)^{\top}\boldsymbol{V}_i^{-1}(\boldsymbol{\alpha}^*)\left(\boldsymbol{Y}_i - \boldsymbol{m}_i\right)\\ &= \sum_{i=1}^{N} I\{A_i = a\}\boldsymbol{L}_{i,a}^{\top}(\boldsymbol{\theta}^*,\boldsymbol{\alpha}^*)\boldsymbol{r}_i(a),\end{aligned} \tag{2.3}$$

where $\boldsymbol{r}_i(a) = \boldsymbol{Y}_i(a) - \boldsymbol{m}_i(a)$, $\boldsymbol{L}_{i,a}(\boldsymbol{\theta}^*,\boldsymbol{\alpha}^*) = \boldsymbol{V}_i^{-1}(\boldsymbol{\alpha}^*)\widetilde{\boldsymbol{M}}_i(a)\boldsymbol{G}_{i,a}$ for $a \in \{0,1\}$, $\boldsymbol{V}_i(\boldsymbol{\alpha}^*)$ is the working covariance matrix used to improve the estimation of $\boldsymbol{\theta}^*$, and $\boldsymbol{\alpha}^*$ is a q-vector of parameters used for estimating the covariances.

Based on (2.3), the estimate of $\boldsymbol{\theta}^*$ is denoted by $\hat{\boldsymbol{\theta}}^* = (\hat{\gamma}_1^*, \hat{\gamma}_0^*, \hat{\boldsymbol{\beta}}_{\mathsf{in}}^{*\top})^{\top}$ and the corresponding estimate for $\tau^* = \gamma_1^* - \gamma_0^*$ is $\hat{\tau}^* = \hat{\gamma}_1^* - \hat{\gamma}_0^*$. To conduct the inference of τ, we consider the hypothesis $H_0 : \tau = 0$ against $H_1 : \tau \neq 0$ with the following Wald test statistic $T_{\mathsf{Wald}} = \mathsf{se}_{\mathsf{Wald}}^{-1}(\hat{\tau}^*)\hat{\tau}^*$, where

$$N\mathsf{se}^2_{\mathsf{Wald}}(\hat{\tau}^*) = \boldsymbol{l}^{\top}\tilde{\Psi}_N^{-1}(\hat{\boldsymbol{\theta}}^*,\hat{\boldsymbol{\alpha}}^*)\hat{\Sigma}_{\Psi\Psi}(\hat{\boldsymbol{\theta}}^*,\hat{\boldsymbol{\alpha}}^*)\tilde{\Psi}_N^{-1}(\hat{\boldsymbol{\theta}}^*,\hat{\boldsymbol{\alpha}}^*)\boldsymbol{l}, \tag{S1.1}$$

$$\begin{aligned}\tilde{\Psi}_N(\hat{\boldsymbol{\theta}}^*,\hat{\boldsymbol{\alpha}}^*) = \tfrac{1}{N}\textstyle\sum_{i=1}^{N}\Big[&A_i\boldsymbol{L}_{i,1}^{\top}(\hat{\boldsymbol{\theta}}^*,\hat{\boldsymbol{\alpha}}^*)V_i(\hat{\boldsymbol{\alpha}}^*)\boldsymbol{L}_{i,1}(\hat{\boldsymbol{\theta}}^*,\hat{\boldsymbol{\alpha}}^*)\\ &+ (1-A_i)\boldsymbol{L}_{i,0}^{\top}(\hat{\boldsymbol{\theta}}^*,\hat{\boldsymbol{\alpha}}^*)V_i(\hat{\boldsymbol{\alpha}}^*)\boldsymbol{L}_{i,0}(\hat{\boldsymbol{\theta}}^*,\hat{\boldsymbol{\alpha}}^*)\Big],\end{aligned}$$

$$\begin{aligned}\hat{\Sigma}_{\Psi\Psi}(\hat{\boldsymbol{\theta}}^*,\hat{\boldsymbol{\alpha}}^*) = \tfrac{1}{N}\textstyle\sum_{i=1}^{N}\Big[&A_i\boldsymbol{L}_{i,1}^{\top}(\hat{\boldsymbol{\theta}}^*,\hat{\boldsymbol{\alpha}}^*)(\boldsymbol{Y}_i - \hat{\boldsymbol{m}}_i(1))(\boldsymbol{Y}_i - \hat{\boldsymbol{m}}_i(1))^{\top}\boldsymbol{L}_{i,1}(\hat{\boldsymbol{\theta}}^*,\hat{\boldsymbol{\alpha}}^*)\\ &+ (1-A_i)\boldsymbol{L}_{i,0}^{\top}(\hat{\boldsymbol{\theta}}^*,\hat{\boldsymbol{\alpha}}^*)(\boldsymbol{Y}_i - \hat{\boldsymbol{m}}_i(0))(\boldsymbol{Y}_i - \hat{\boldsymbol{m}}_i(0))^{\top}\boldsymbol{L}_{i,0}(\hat{\boldsymbol{\theta}}^*,\hat{\boldsymbol{\alpha}}^*)\Big],\end{aligned}$$

where $\boldsymbol{l} = (1, -1, \mathbf{0}_{p_0}^\top)^\top$, and $\hat{\boldsymbol{m}}_i(a)$ are the estimated values of $\boldsymbol{m}_i(a)$ by replacing $(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)$ with $(\hat{\boldsymbol{\theta}}^*, \hat{\boldsymbol{\alpha}}^*)$.

The assumptions used for developing our theoretical results are summarized as follows.

Assumption 3.1. 1. s_{max} and k_{max} are finite, $\{K_i, S_i\}$ are i.i.d. as $\{K, S\}$ and $\{\boldsymbol{X}_i, \boldsymbol{Y}_i(1), \boldsymbol{Y}_i(0)\}$ are i.i.d. as $\{\boldsymbol{X}^k, \boldsymbol{Y}^k(1), \boldsymbol{Y}^k(0)\}$ with finite third moments given $K_i = k$ for $1 \le k \le k_{\mathsf{max}}$ and $1 \le i \le N$.

2. Given $\boldsymbol{X}_i$ and A_i, $\boldsymbol{Y}_i$ satisfies (2.1) for $1 \le i \le N$.

3. On the support of $m(\cdot)$, $m(\cdot)$ is bounded and twice differentiable.

4. Let $\{\boldsymbol{V}^k(\boldsymbol{\alpha}), \boldsymbol{G}_1^k, \boldsymbol{G}_0^k, \boldsymbol{L}_1^k(\boldsymbol{\theta}, \boldsymbol{\alpha}), \boldsymbol{L}_0^k(\boldsymbol{\theta}, \boldsymbol{\alpha}), \widetilde{\boldsymbol{M}}^k(1), \widetilde{\boldsymbol{M}}^k(0), \boldsymbol{r}^k(1), \boldsymbol{r}^k(0)\}$ denote the population analogue of $\{\boldsymbol{V}_i(\boldsymbol{\alpha}), \boldsymbol{G}_{i,1}, \boldsymbol{G}_{i,0}, \boldsymbol{L}_{i,1}(\boldsymbol{\theta}, \boldsymbol{\alpha}), \boldsymbol{L}_{i,0}(\boldsymbol{\theta}, \boldsymbol{\alpha}), \widetilde{\boldsymbol{M}}_i(1), \widetilde{\boldsymbol{M}}_i(0), \boldsymbol{r}(1), \boldsymbol{r}(0)\}$ given $K_i = k$. Suppose $\hat{\boldsymbol{\alpha}}^* - \boldsymbol{\alpha}^* = O_p(N^{-1/2})$ for some $\boldsymbol{\alpha}^* = \boldsymbol{\alpha}(\boldsymbol{\theta}^*)$, each element of $\boldsymbol{V}^k(\boldsymbol{\alpha})$ is continuous and bounded in $\boldsymbol{\alpha}$ for $1 \le k \le k_{\mathsf{max}}$.

5. Let

$$\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\theta}) = 2^{-1} \textstyle\sum_{k=1}^{k_{\mathsf{max}}} \pi_{\cdot,k} \sum_{a \in \{0,1\}} \mathbf{E}\left[\boldsymbol{L}_a^{k\top}(\boldsymbol{\theta}, \boldsymbol{\alpha}) \boldsymbol{r}^k(a)\right],$$

$$\widetilde{\Psi}_{\boldsymbol{\alpha}}(\boldsymbol{\theta}) = 2^{-1} \textstyle\sum_{k=1}^{k_{\mathsf{max}}} \pi_{\cdot,k} \sum_{a \in \{0,1\}} \mathbf{E}[\boldsymbol{L}_a^{k\top}(\boldsymbol{\theta}, \boldsymbol{\alpha}) \boldsymbol{V}^k(\boldsymbol{\alpha}) \boldsymbol{L}_a^k(\boldsymbol{\theta}, \boldsymbol{\alpha})],$$

$$\mathcal{T}_{1,1} = 2^{-1} \textstyle\sum_{k=1}^{k_{\mathsf{max}}} \pi_{\cdot,k} \sum_{a \in \{0,1\}} \mathbf{E}\left[\boldsymbol{G}_a^{k\top} \left\{\frac{\partial}{\partial \boldsymbol{\theta}} \widetilde{\boldsymbol{M}}(a)\right\} \{\boldsymbol{V}^k(\boldsymbol{\alpha})\}^{-1} \boldsymbol{r}^k(a)\right],$$

$$\mathcal{T}_{1,2} = 2^{-1}\sum_{k=1}^{k_{\max}}\pi_{\cdot,k}\sum_{a\in\{0,1\}}\mathbf{E}\left[\boldsymbol{G}_a^{k\top}\widetilde{\boldsymbol{M}}^k(a)\frac{\partial}{\partial\boldsymbol{\theta}}\{\boldsymbol{\Lambda}^k\}^{-1/2}\left\{\boldsymbol{C}^k(\boldsymbol{\alpha})\right\}^{-1}\{\boldsymbol{\Lambda}^k\}^{-1/2}\boldsymbol{r}^k(a)\right],$$

$$\mathcal{T}_{1,3} = 2^{-1}\sum_{k=1}^{k_{\max}}\pi_{\cdot,k}\sum_{a\in\{0,1\}}\mathbf{E}\left[\boldsymbol{G}_a^{k\top}\widetilde{\boldsymbol{M}}^k(a)\{\boldsymbol{\Lambda}^k\}^{-1/2}\left\{\boldsymbol{C}^k(\boldsymbol{\alpha})\right\}^{-1}\frac{\partial}{\partial\boldsymbol{\theta}}\{\boldsymbol{\Lambda}^k\}^{-1/2}\boldsymbol{r}^k(a)\right],$$

$$\text{and}\quad \mathcal{T}_2(\boldsymbol{\alpha}) = 2^{-1}\sum_{k=1}^{k_{\max}}\pi_{\cdot,k}\sum_{a\in\{0,1\}}\mathbf{E}\left[\boldsymbol{G}_a^{k\top}\widetilde{\boldsymbol{M}}^k(a)\frac{\partial}{\partial\boldsymbol{\alpha}}\left\{\boldsymbol{V}^k(\boldsymbol{\alpha})\right\}^{-1}\boldsymbol{r}^k(a)\right].$$

Then $\tilde{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)$ is positive definite, $\mathcal{T}_1 = \mathcal{T}_{1,1} + \mathcal{T}_{1,2} + \mathcal{T}_{1,3} = \mathbf{0}$, and $\mathcal{T}_2(\boldsymbol{\alpha}) = \mathbf{0}$ for all $\boldsymbol{\alpha}$ in the neighborhood of $\boldsymbol{\alpha}^*$.

Assumption 3.2. Let $\mathcal{F}_{i-1} = \sigma\{\{S_j, \boldsymbol{X}_j, \boldsymbol{Y}_j(1), \boldsymbol{Y}_j(0), K_j\}_{j=1}^{i-1}, S_i\}$ denote the cumulated information before the assignment of the i-th patient. For $1 \le s \le s_{\mathsf{max}}$ and $1 \le k \le k_{\mathsf{max}}$, the joint imbalance satisfies

$$D_N(s,k) = \sum_{i=1}^{N} d_i(s,k) + \varepsilon(s,k),$$

where (i) $\mathbf{E}[\epsilon^2(s,k)] = o(N)$, and (ii) $\{d_i(s,k), 1 \le s \le s_{\mathsf{max}} \text{ and } 1 \le k \le k_{\mathsf{max}}\}_{i=1}^{N}$ is a sequence of zero-mean martingale differences with respect to $\mathcal{F}_{i-1}$ and has a covariance kernel $\gamma\{(s_1,k_1),(s_2,k_2)\} < \infty$ such that for $1 \le s_1, s_2 \le s_{\mathsf{max}}$ and $1 \le k_1, k_2 \le k_{\mathsf{max}}$

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{E}\left[d_i(s_1,k_1)d_i(s_2,k_2)\middle|\,\mathcal{F}_{i-1}\right] = \gamma\{(s_1,k_1),(s_2,k_2)\} + o_p(1).$$

Theorem 3.1. Suppose Assumptions 3.1 and 3.2 hold and the working model (2.2) is implemented with the GEE (2.3). Let

$$\boldsymbol{R}(s,k) = \mathbf{E}\left[\boldsymbol{L}_1^{k\top}(\boldsymbol{\theta}^*,\boldsymbol{\alpha}^*)\boldsymbol{r}^k(1) - \boldsymbol{L}_0^{k\top}(\boldsymbol{\theta}^*,\boldsymbol{\alpha}^*)\boldsymbol{r}^k(0)|S=s\right]$$

for $1 \le s \le s_{\mathsf{max}}$ and $1 \le k \le k_{\mathsf{max}}$. Then

$$\begin{aligned}
&\sqrt{N}\left(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\right) = 2\left\{\widetilde{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)\right\}^{-1}\sqrt{N}\Psi_{N,\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*) + o_p(1),\\
\text{and}\quad &\sqrt{N}\Psi_{N,\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*) \xrightarrow{\mathbf{P}} \mathcal{N}(\mathbf{0}_{p_0+2}, 2^{-1}\Xi_{\Psi\Psi}(\boldsymbol{\theta}^*,\boldsymbol{\alpha}^*)),\\
\text{where}\quad &\Xi_{\Psi\Psi}(\boldsymbol{\theta}^*,\boldsymbol{\alpha}^*) = \Sigma_{\Psi\Psi}(\boldsymbol{\theta}^*,\boldsymbol{\alpha}^*) - 2^{-1}\left[\Sigma_{\mathsf{CR}}(\boldsymbol{\theta}^*,\boldsymbol{\alpha}^*) - \Sigma_{\mathsf{CAR}}(\boldsymbol{\theta}^*,\boldsymbol{\alpha}^*)\right],\\
&\Sigma_{\Psi\Psi}(\boldsymbol{\theta}^*,\boldsymbol{\alpha}^*) = \sum_{k=1}^{k_{\mathsf{max}}}\pi_{\cdot,k}\sum_{a\in\{0,1\}}\mathbf{E}\left[\boldsymbol{L}_a^{k\top}(\boldsymbol{\theta}^*,\boldsymbol{\alpha}^*)\boldsymbol{r}^k(a)\boldsymbol{r}^{k\top}(a)\boldsymbol{L}_a^k(\boldsymbol{\theta}^*,\boldsymbol{\alpha}^*)\right],\\
&\Sigma_{\mathsf{CR}}(\boldsymbol{\theta}^*,\boldsymbol{\alpha}^*) = \sum_{k=1}^{k_{\mathsf{max}}}\sum_{s=1}^{s_{\mathsf{max}}}\pi_{s,k}\boldsymbol{R}(s,k)\boldsymbol{R}^\top(s,k),\\
\text{and}\quad &\Sigma_{\mathsf{CAR}}(\boldsymbol{\theta}^*,\boldsymbol{\alpha}^*) = \sum_{k_1,k_2}\sum_{s_1,s_2}\gamma\{(s_1,k_1),(s_2,k_2)\}\boldsymbol{R}(s_1,k_1)\boldsymbol{R}^\top(s_2,k_2).
\end{aligned}$$

For CR and the class of CAR procedures satisfying

$$D_N(s,\cdot) = o_p(N^{-1/2}) \quad \text{for} \quad 1 \le s \le s_{\max}, \tag{3.1}$$

the covariance $\gamma\big((s_1,k_1),(s_2,k_2)\big)$ admits closed-form expressions. We analyze their properties in the following corollary.

Corollary 3.1. 1. Under CR, Assumption 3.2 holds with

$$\gamma\{(s_1, k_1), (s_2, k_2)\} = I\{s_1 = s_2 = s, k_1 = k_2 = k\}\pi_{s,k}$$

for $1 \leq s_1, s_2 \leq s_{\mathsf{max}}$ and $1 \leq k_1, k_2 \leq k_{\mathsf{max}}$ and thus $\Sigma_{\mathsf{CR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) = \Sigma_{\mathsf{CAR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)$.

2. If the CAR procedure satisfies (3.1), then Assumption 3.2 holds with

$$\begin{aligned}\gamma\{(s_1, k_1), (s_2, k_2)\} = I\{s_1 = s_2 = s, k_1 = k_2 = k\}\pi_{s,k} \\ - I\{s_1 = s_2 = s\}\pi_{k_1|s}\pi_{k_2|s}\pi_{s,\cdot}\end{aligned}$$

for $1 \leq s_1, s_2 \leq s_{\mathsf{max}}$ and $1 \leq k_1, k_2 \leq k_{\mathsf{max}}$ and thus

$$\Sigma_{\mathsf{CR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) - \Sigma_{\mathsf{CAR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) = \sum_{s=1}^{s_{\mathsf{max}}} \pi_{s,\cdot} \sum_{k_1,k_2} \pi_{k_1|s}\pi_{k_2|s}\boldsymbol{R}(s, k_1)\boldsymbol{R}^\top(s, k_2)$$

is positive definite.

Based on the above results, we next evaluate the property of the Wald tests in the following corollary.

Corollary 3.2. Under the conditions of Theorem 3.1 and the true response model (2.1), the following statements hold.

---

1. Let $\boldsymbol{l} = (1, -1, \mathbf{0}_{p_0}^\top)^\top$, then

$$
\begin{aligned}
&\sqrt{N}\left(\hat{\tau}^* - \tau^*\right) \xrightarrow{\mathbf{D}} \mathcal{N}\left(0, 2\sigma_\tau^2(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)\right), \\
\text{and} \quad & N\mathsf{se}_{\mathsf{Wald}}^2(\hat{\tau}^*) = 2\sigma_{\mathsf{Wald}}^2(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) + o_p(1), \\
\text{where} \quad & \sigma_\tau^2(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) = \boldsymbol{l}^\top \left\{\widetilde{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)\right\}^{-1} \Xi_{\Psi\Psi}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) \left\{\left[\widetilde{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)\right]^{-1}\right\}^\top \boldsymbol{l}, \\
\text{and} \quad & \sigma_{\mathsf{Wald}}^2(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) = \boldsymbol{l}^\top \left\{\widetilde{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)\right\}^{-1} \Sigma_{\Psi\Psi}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) \left\{\left[\widetilde{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)\right]^{-1}\right\}^\top \boldsymbol{l}.
\end{aligned}
$$

   Here, $\Xi_{\Psi\Psi}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)$ and $\Sigma_{\Psi\Psi}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)$ are given in Theorem 3.1.

2. Under $H_0 : \gamma_1 = \gamma_0$, $\hat{\tau}^* = O_p(N^{-1/2})$ and T_{Wald} converges to a normal random variable with mean zero and variance

$$
\lambda_{\mathsf{Wald}}^2 = \sigma_{\mathsf{Wald}}^{-2}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)\sigma_\tau^2(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) \leq 1.
$$

   Then, T_{Wald} following CAR procedures may be conservative in terms of having reduced Type I error.

3. Under $H_1 : \gamma_1 - \gamma_0 = \delta/\sqrt{N}$ with $\delta > 0$, there exists δ^* such that $\sqrt{N}(\gamma_1^* - \gamma_0^*) = \delta^*$, and $T_{\mathsf{Wald}} - \{2\sigma_{\mathsf{Wald}}^2(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)\}^{-1/2}\delta^*$ converges to a normal random variable with mean zero and variance $\lambda_{\mathsf{Wald}}^2$. Furthermore, let c denote the significance level of the test, then the power of T_{Wald} is

$$1 - \Phi\left(\frac{z_{1-c/2}}{\lambda_{\mathsf{Wald}}} - \frac{\delta^*}{\sqrt{2\sigma_\tau^2(\boldsymbol{\theta}^*,\boldsymbol{\alpha}^*)}}\right) + \Phi\left(-\frac{z_{1-c/2}}{\lambda_{\mathsf{Wald}}} - \frac{\delta^*}{\sqrt{2\sigma_\tau^2(\boldsymbol{\theta}^*,\boldsymbol{\alpha}^*)}}\right) + o(1), \quad (3.2)$$

where $\Phi(x)$ is the cumulated distribution function for a standard normal distribution and $z_{1-c/2}$ is its $(1-c/2)$-th quantile.

Theorem 4.1. Under the conditions of Corollary 3.2, the following statements hold.

1. Under $H_0 : \gamma_1 = \gamma_0$, T_{adj} converges to a standard normal random variable. In this case, T_{adj} following CAR procedures attains the nominal level of the test.

2. Under $H_1 : \gamma_1 - \gamma_0 = \delta/\sqrt{N}$ with $\delta > 0$, there exists δ^* such that $\sqrt{N}(\gamma_1^* - \gamma_0^*) = \delta^*$ and $T_{\mathsf{adj}} - \{2\sigma_\tau^2(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)\}^{-1/2}\delta^*$ converges to a standard normal random variable. Furthermore, the power of T_{adj} is

$$1 - \Phi\left(\frac{z_{1-c/2}}{\lambda_{\mathsf{Wald}}} - \frac{\delta^*}{\sqrt{2\sigma_\tau^2(\boldsymbol{\theta}^*,\boldsymbol{\alpha}^*)}}\right) + \Phi\left(-\frac{z_{1-c/2}}{\lambda_{\mathsf{Wald}}} - \frac{\delta^*}{\sqrt{2\sigma_\tau^2(\boldsymbol{\theta}^*,\boldsymbol{\alpha}^*)}}\right) + o(1).$$

S2 Construction of the Stratification Variables

In this section, we describe the construction of the stratification variables (S_i, K_i). Here, S_i is used in the design of the CAR procedure, while K_i denotes an additional stratification variable associated with potential missingness or unbalanced observations, which may in turn affect the subsequent inference.

S2.1 The Stratification Variable Used for Design

Let $\boldsymbol{X}_{i,\mathrm{Ex}} = (X_{i,1}, \ldots, X_{i,q})^\top$ denote the q baseline covariates used at the design stage. These covariates are observed prior to randomization and may differ from those used in the analysis stage. Furthermore, some components of $\boldsymbol{X}_{i,\mathrm{Ex}}$ may be continuous. Let $\mathcal{C} = \{j : X_{i,j} \text{ is continuous},\ 1 \leq j \leq q\}$ denote the index set of continuous covariates. For each $j \in \mathcal{C}$, define a discrete function $d_j(\cdot)$ that maps $X_{i,j}$ into s_j discrete levels $\{x_j^l\}_{l=1}^{s_j}$. The resulting discretized covariate $\tilde{X}_{i,j}$ can be defined as

$$\tilde{X}_{i,j} = \begin{cases} d_j(X_{i,j}) = x_j^l, & \text{if } j \in \mathcal{C}, \\ X_{i,j}, & \text{if } j \notin \mathcal{C}. \end{cases}$$

For $j \notin \mathcal{C}$, $X_{i,j}$ is discrete and is assumed to have s_j levels. Consequently, the induced stratification scheme yields a total of $s_{\max} = \prod_{j=1}^{q} s_j$ strata.

In practice, discretization is often guided by clinically meaningful thresholds or data-driven criteria such as quantiles. For example, as noted in Section 2.2.7 of ICH (1998), diastolic blood pressure is often dichotomized at 90 mmHg based on clinical relevance. In modern oncology trials, including umbrella or basket trials (FDA, 2022), continuous or high-dimensional genomic variables are routinely discretized into biomarker-defined subgroups to facilitate treatment allocation.

When the number of baseline covariates is large, the resulting number of strata $s_{\max}$ may become substantial relative to the sample size N. In such settings, over-stratification can deteriorate the performance of the randomization procedure and is generally discouraged in practice (ICH, 1998). Accordingly, it is common to restrict stratification to a small set of key prognostic covariates, and the number of stratification factors is typically limited in practical applications (see, e.g., Section 3.2 of ICH, 1998).

Nevertheless, in many contemporary applications, a large number of strata may be unavoidable. A prominent example arises in multicenter clinical trials, where the number of participating sites can be substantial, and site indicators are routinely incorporated into the stratification scheme to account for between-center heterogeneity. In such settings, $s_{\max}$ may be large even when only a small number of additional covariates are included.

As discussed in ICH (1998), a common practical strategy is to combine similar strata, for example, by grouping geographically proximate sites or sites with similar clinical characteristics, so as to reduce the effective number of strata while preserving the main sources of heterogeneity. Alternatively, one may select only the most important covariates for stratification, thereby reducing the dimensionality of the strata, as considered in Guo et al. (2025).

S2.2 Additional Stratification Variable

We introduce K_i to account for the impact of potential imbalance on inference in longitudinal data arising from CAR experiments. K_i is defined as the stratification variable, such that, if $K_i = k$ then $n_i = n^k, \boldsymbol{C}_i(\boldsymbol{\alpha}) = \boldsymbol{C}^k(\boldsymbol{\alpha})$, for $1 \leq k \leq k_{\mathsf{max}}$. Here, k_{max} represents the number of distinct values of the pair $\{n_i, \boldsymbol{C}_i(\alpha)\}$, where n_i denotes the number of observations for subject i, and $\boldsymbol{C}_i(\boldsymbol{\alpha})$ represents the working correlation matrix, respectively. Below, we provide additional details on how K_i can be incorporated in practical applications.

Example 1 (Independent and Exchangeable Correlations)**.** The independence or exchangeable working correlation structures can be specified as

$$\boldsymbol{C}_i(\boldsymbol{\alpha}) = \boldsymbol{I}_{n_i} \quad \text{or} \quad \boldsymbol{C}_i(\boldsymbol{\alpha}) = (1-\alpha)\boldsymbol{I}_{n_i} + \alpha \mathbf{1}_{n_i}\mathbf{1}_{n_i}^\top,$$

where $\boldsymbol{I}_n$ denotes the $n \times n$ identity matrix and $\mathbf{1}_n$ denotes the n-vector of ones. In either case, $\boldsymbol{C}_i(\boldsymbol{\alpha})$ is uniquely determined by the number of observations n_i. Hence, k_{max} represents the number of distinct values taken by n_i, such that $K_i = k$ whenever $n_i = n^k$. In particular, when $n_i = n$ for all $1 \leq i \leq N$, that is, the study is balanced, $k_{\mathsf{max}} = 1$ and, K_i may be omitted.

Example 2 (AR(1) Correlation)**.** Let n denote the total number of scheduled visits. In practice, the patient may occasionally miss a scheduled visit. Hence, the number of observed measurements for the i-th patient n_i, may be less than n. For $1 \leq k \leq n_i$, let t_k denote the time of the k-th observed visit, where $1 \leq t_1 \leq \ldots \leq t_{n_i} \leq n$. Then, the AR(1) correlation matrix is specified by

$$[\boldsymbol{C}_i(\boldsymbol{\alpha})]_{k_1,k_2} = \alpha^{|t_{k_1} - t_{k_2}|}, \qquad 1 \leq k_1, k_2 \leq n_i.$$

To further illustrate the role of K_i under the AR(1) structure, consider the case where $n = 4$ and $n_i \in \{4, 3\}$. Suppose that when $n_i = 3$, the observed visit times are either $\{1, 2, 4\}$ or $\{1, 3, 4\}$. Then there are three distinct correlation patterns in total, so that $k_{\mathsf{max}} = 3$, that is, when $n_i = 4$, we have

$$\boldsymbol{C}_i(\boldsymbol{\alpha}) = \begin{pmatrix} 1 & \alpha & \alpha^2 & \alpha^3 \\ \alpha & 1 & \alpha & \alpha^2 \\ \alpha^2 & \alpha & 1 & \alpha \\ \alpha^3 & \alpha^2 & \alpha & 1 \end{pmatrix}.$$

When $n_i = 3$, for observed visit times $\{1, 2, 4\}$ and $\{1, 3, 4\}$, the corresponding correlation matrices are

$$\boldsymbol{C}_i(\boldsymbol{\alpha}) = \begin{pmatrix} 1 & \alpha & \alpha^3 \\ \alpha & 1 & \alpha^2 \\ \alpha^3 & \alpha^2 & 1 \end{pmatrix}, \quad \text{and} \quad \boldsymbol{C}_i(\boldsymbol{\alpha}) = \begin{pmatrix} 1 & \alpha^2 & \alpha^3 \\ \alpha^2 & 1 & \alpha \\ \alpha^3 & \alpha & 1 \end{pmatrix},$$

respectively.

In summary, the value of k_{max} depends on the choice of the working correlation matrix $\boldsymbol{C}_i(\boldsymbol{\alpha})$ and the distinct values of $\{n_i\}_{i=1}^N$. Under the independence and exchangeable working correlation structures, k_{max} depends only on the distinct values of $\{n_i\}_{i=1}^N$ and may remain relatively small when the number of such distinct values is limited. For other working correlation structures that depend on the observation times, such as the AR(1) correlation structure, the unbalanced visit schedules may lead to a larger value of k_{max}. In such cases, one may combine levels of K_i that contain only a small number of subjects, thereby keeping k_{max} in a reasonable scale.

In Example 2, we assume that all N subjects have the same number of scheduled visits, although the number of observed visits may vary across subjects. In some applications, however, the planned number of visits may itself differ across subjects due to differences in patients' health conditions or treatment pathways. For example, patients in better health may require fewer follow-up visits than those in poorer health. In such settings, additional modeling adjustments may be required to account for this source of heterogeneity together with potential missing visits. We leave extensions of the GEE framework to these more general settings for future research.

S3 Additional Details for the Adjusted Tests

We first present a detailed description of the bootstrap procedure in Algorithm 1. Define the plug-in sample analogues of $\pi_{s,k}$, $\pi_{k|s}$, $\boldsymbol{R}(s,k)$, $\widetilde{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)$, and $\Sigma_{\mathsf{CR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)$ as $\hat{\pi}_{s,k}$, $\hat{\pi}_{k|s}$,$\hat{\boldsymbol{R}}(s,k)$, and

$$\widetilde{\Psi}_{N,\hat{\boldsymbol{\alpha}}}(\hat{\boldsymbol{\theta}}^*) = \frac{1}{N}\sum_{i=1}^{N}\sum_{a\in\{0,1\}} I\{A_i = a\}\boldsymbol{L}_{i,a}^{\top}(\hat{\boldsymbol{\theta}}^*, \hat{\boldsymbol{\alpha}}^*)V_i(\hat{\boldsymbol{\alpha}}^*)\boldsymbol{L}_{i,a}(\hat{\boldsymbol{\theta}}^*, \hat{\boldsymbol{\alpha}}^*),$$

$$\hat{\Sigma}_{\mathsf{CR}}(\hat{\boldsymbol{\theta}}^*, \hat{\boldsymbol{\alpha}}^*) = \sum_{k=1}^{k_{\mathsf{max}}}\sum_{s=1}^{s_{\mathsf{max}}} \hat{\pi}_{s,k}\hat{\boldsymbol{R}}(s,k)\hat{\boldsymbol{R}}^{\top}(s,k).$$

Algorithm 1 CAR Bootstrap procedure for Longitudinal Study

1: **Input**: Observed Stratification Data $\{(S_i, K_i)\}_{i=1}^N$; the number of bootstrap replications B; and the CAR procedure used in the original trial.
2: **for** $b = 1$ to B **do**
3: Draw a random sample of size N with replacement from the empirical distribution of $\{(S_i, K_i)\}_{i=1}^N$, denoted as $\{(S_i^{*b}, K_i^{*b})\}_{i=1}^N$.
4: Generate $\{A_i^{*b}\}_{i=1}^N$ based on the resampled covariate sequence $\{S_i^{*b}\}_{i=1}^N$ according to the original CAR procedure.
5: Evaluate the imbalance $\{D_N^{*b}(s,k) : 1 \le s \le s_{\max}, 1 \le k \le k_{\max}\}$, where

$$D_N^{*b}(s,k) = \sum_{i=1}^{N}(2A_i^{*b} - 1)I\{S_i^b = s, K_i^b = k\}.$$

6: **end for**
7: **Output**: $\{D_N^{*b}(s,k) : 1 \le s \le s_{\max}, 1 \le k \le k_{\max}\}_{b=1}^B$.

Based on Algorithm 1, we compute $\{D_N^{*b}(s,k)\}_{s,k}$ and

$$O_b^* = N^{-1/2}\sum_{k=1}^{k_{\mathsf{max}}}\sum_{s=1}^{s_{\mathsf{max}}} D_N^{*b}(s,k)\hat{\boldsymbol{R}}(s,k) \quad \text{for} \quad 1 \le b \le B.$$

Consequently, $\Sigma_{\mathsf{CAR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)$ can be estimated by the sample variance-covariance matrix of $\{O_b^*\}_{b=1}^B$, denoted by S_O^2. Furthermore, the proposed adjusted standard error can be calculated as

$$N\mathsf{se}^2_{\mathsf{adj}}(\hat{\tau}^*) = N\mathsf{se}^2_{\mathsf{Wald}}(\hat{\tau}^*) - \\ \boldsymbol{l}^\top \left\{\widetilde{\Psi}_{N,\hat{\boldsymbol{\alpha}}^*}(\hat{\boldsymbol{\theta}}^*)\right\}^{-1} 4^{-1} \left[\hat{\Sigma}_{\mathsf{CR}}(\hat{\boldsymbol{\theta}}^*, \hat{\boldsymbol{\alpha}}^*) - S_O^2\right] \left\{\widetilde{\Psi}_{N,\hat{\boldsymbol{\alpha}}^*}(\hat{\boldsymbol{\theta}}^*)\right\}^{-1} \boldsymbol{l}. \quad \text{(S3.1)}$$

S4 Proof of the Main Results

We prove our theoretical results in the following subsections.

S4.1 Proof of Theorem 3.1

We begin the proof of Theorem 3.1 starting from introducing some additional notation and some ancillary results. Let $N_{s,k} = \sum_{i=1}^N I\{S_i = s, K_i = k\}$, $N_{\cdot,k} = \sum_{i=1}^N I\{K_i = k\}$, and $N_{s,\cdot} = \sum_{i=1}^N I\{S_i = s\}$, denote the number of patients under different levels of stratifications. With a slight abuse of notation, we will use $\{\boldsymbol{X}_i^k, \boldsymbol{Y}_i^k(1), \boldsymbol{Y}_i^k(0)\}_{i=1}^{N_{\cdot,k}}$ to represent the observed data for the i-th patient, with $K_i = k$ for $1 \le i \le N_{\cdot,k}$ and $1 \le k \le k_{\mathsf{max}}$. For $a \in \{0,1\}$, let $h_a(\cdot|\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)$ be a scalar function, such that $h_a(\cdot|\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)$ is continuous and twice differentiable in $\boldsymbol{\theta}^*$ and $\boldsymbol{\alpha}^*$. Set $\boldsymbol{h} = (h_1, h_0)$, and let $\mathscr{H}_a$ denote the class of functions of h_a and $\mathscr{H} = \mathscr{H}_1 \otimes \mathscr{H}_0$.

For $\boldsymbol{W}_i = \text{vec}(\boldsymbol{X}_i, \boldsymbol{Y}_i(1), \boldsymbol{Y}_i(0))$, denote $\boldsymbol{W}_i(h) = \boldsymbol{W}_i \circ h = h(\boldsymbol{W}_i|\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)$, such that $\boldsymbol{W}_i(h_1 + h_0) = h_1(\boldsymbol{W}_i|\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) + h_0(\boldsymbol{W}_i|\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)$. Let $\mathbb{P}_N(\boldsymbol{h}) = N^{-1}\sum_{i=1}^N [A_i \boldsymbol{W}_i(h_1) + (1 - A_i)\boldsymbol{W}_i(h_0)]$, $\mathbf{P}(\boldsymbol{h}) = 2^{-1}\sum_{k=1}^{k_{\mathsf{max}}} \pi_k \mathbf{E}[\boldsymbol{W}^k(h_1) + \mathbf{W}^k(h_0)]$, and $\mathbb{G}_N(\boldsymbol{h}) = \mathbb{P}_N(\boldsymbol{h}) - \mathbf{P}(\boldsymbol{h})$. The following proposition provides the basis for our theoretical analysis.

Proposition 1. *1. The class of functions $\mathscr{H}$ is VC-subgraph class satisfying that $\sup_{h_a \in \mathscr{H}_a} \mathbf{E}[\{\boldsymbol{W}(h_a)\}^2] < \infty$, for $a \in \{0, 1\}$, then $\sup_{\boldsymbol{h} \in \mathscr{H}} |\mathbb{P}_N(\boldsymbol{h}) - \mathbf{P}(\boldsymbol{h})|$ converges to zero in probability.*

2. Suppose $\mathbf{P}(\boldsymbol{h}) = 0$, then $\mathbb{G}_N(\boldsymbol{h})$ converges to a zero-mean normal variable with variance $4^{-1}\{2\sigma_W^2(\boldsymbol{h}) - [\sigma_{\mathsf{CR}}^2(\boldsymbol{h}) - \sigma_{\mathsf{CAR}}^2(\boldsymbol{h})]\}$, where

$$\sigma_{\mathsf{W}}^2(\boldsymbol{h}) = \sum_{k=1}^{k_{\mathsf{max}}} \pi_{\cdot,k}\{\mathbf{E}[\{\boldsymbol{W}^k(h_1)\}^2] + \mathbf{E}[\{\boldsymbol{W}^k(h_0)\}^2]\},$$
$$\sigma_{\mathsf{CR}}^2(\boldsymbol{h}) = \sum_{k=1}^{k_{\mathsf{max}}} \sum_{s=1}^{s_{\mathsf{max}}} \pi_{s,k}\{\mathbf{E}[\boldsymbol{W}^k(h_1 - h_0)|S = s]^2\}$$
$$\sigma_{\mathsf{CAR}}^2(\boldsymbol{h}) = \sum_{k_1,k_2} \sum_{s_1,s_2} \gamma\{(s_1, k_1), (s_2, k_2)\}\mathbf{E}[\boldsymbol{W}^{k_1}(h_1 - h_0)|S = s_1]\mathbf{E}[\boldsymbol{W}^{k_2}(h_1 - h_0)|S = s_2].$$

Proof of Proposition 1. Regarding the first argument, it follows from Assumption 3.1.1 and Assumption 3.1.2 that

$$\sum_{i=1}^N (2A_i - 1)\mathbf{E}[\boldsymbol{W}_i(h_1 - h_0)|S_i, K_i]$$

$$= \sum_{i=1}^{N} \sum_{s,k} (2A_i - 1) I\{S_i = s, K_i = k\} \mathbf{E}[\boldsymbol{W}_i(h_1 - h_0) | S_i = s, K_i = k]$$

$$= \sum_{s,k} D_N(s,k) \mathbf{E}[\boldsymbol{W}^k(h_1 - h_0) | S = s],$$

then we can obtain the following decomposition

$$\mathbb{P}_N(\boldsymbol{h}) - \mathbf{P}(\boldsymbol{h}) = 2^{-1}\{\mathbb{P}_{N,1}(\boldsymbol{h}) + \mathbb{P}_{N,2}(\boldsymbol{h}) + \mathbb{P}_{N,3}(\boldsymbol{h}) + \mathbb{P}_{N,4}(\boldsymbol{h})\},$$

$$\text{where} \quad \mathbb{P}_{N,1}(\boldsymbol{h}) = N^{-1} \sum_{i=1}^{N} (2A_i - 1) \left\{ \boldsymbol{W}_i(h_1 - h_0) - \mathbf{E}[\boldsymbol{W}_i(h_1 - h_0) | S_i, K_i] \right\},$$

$$\mathbb{P}_{N,2}(\boldsymbol{h}) = N^{-1} \sum_{s,k=1}^{s_{\mathsf{max}},k_{\mathsf{max}}} D_N(s,k) \mathbf{E}[\boldsymbol{W}^k(h_1 - h_0) | S = s],$$

$$\mathbb{P}_{N,3}(\boldsymbol{h}) = \sum_{k=1}^{k_{\mathsf{max}}} \frac{N_{\cdot,k}}{N} \cdot \frac{1}{N_{\cdot,k}} \sum_{i=1}^{N_{\cdot,k}} \left[\boldsymbol{W}_i^k(h_1 + h_0) - \mathbf{E}\left[\boldsymbol{W}^k(h_1 + h_0)\right] \right],$$

$$\text{and} \quad \mathbb{P}_{N,4}(\boldsymbol{h}) = \sum_{k=1}^{k_{\mathsf{max}}} \left(\frac{N_{\cdot,k}}{N} - \pi_{\cdot,k} \right) \mathbf{E}\left[\boldsymbol{W}^k(h_1 + h_0)\right].$$

Notice that

$$\mathbf{E}\left[\boldsymbol{W}_i(h_1 - h_0) - \mathbf{E}[\boldsymbol{W}_i(h_1 - h_0) | S_i, K_i] | \mathcal{F}_{i-1}\right]$$

$$= \mathbf{E}\left[\mathbf{E}[\boldsymbol{W}_i(h_1 - h_0) | S_i, K_i] - \mathbf{E}[\boldsymbol{W}_i(h_1 - h_0) | S_i, K_i] | \mathcal{F}_{i-1}\right] = 0.$$

Consequently, $(2A_i - 1)\{\boldsymbol{W}_i(h_1 - h_0) - \mathbf{E}[\boldsymbol{W}_i(h_1 - h_0) | S_i, K_i]\}_{i=1}^{N}$ is a sequence of zero-mean martingale differences. Then, following a similar proof of Lemma S1.3 in the supplementary materials of Liu et al. (2025), we can show that $\sup_{\boldsymbol{h}\in\mathscr{H}} |\mathbb{P}_{N,k}(\boldsymbol{h})| = o_p(1)$ for $1 \leq k \leq 4$. This completes the proof for the first argument.

For the second argument, notice that

$$
\begin{aligned}
\mathbb{G}_N(\boldsymbol{h}) =& \frac{1}{2}\left\{\mathbb{G}_{N,1}(\boldsymbol{h}) + \mathbb{G}_{N,2}(\boldsymbol{h}) + \mathbb{G}_{N,3}(\boldsymbol{h}) + \mathbb{G}_{N,4}(\boldsymbol{h})\right\} + o_p(1)\\
\text{where} \quad \mathbb{G}_{N,j}(\boldsymbol{h}) =& \sum_{i=1}^{N} Z_{i,j}(\boldsymbol{h}) \quad \text{for } 1 \leq j \leq 4,\\
Z_{i,1}(\boldsymbol{h}) =& (2A_i - 1)\{\boldsymbol{W}_i(h_1 - h_0) - \mathbf{E}[\boldsymbol{W}_i(h_1 - h_0)|S_i, K_i]\}\\
Z_{i,2}(\boldsymbol{h}) =& \sum_{s,k} d_i(s,k)\mathbf{E}[\boldsymbol{W}_i(h_1 - h_0)|S_i = s, K_i = k]\\
Z_{i,3}(\boldsymbol{h}) =& \boldsymbol{W}_i(h_1 + h_0) - \mathbf{E}[\boldsymbol{W}_i(h_1 + h_0)|S_i = s, K_i = k]\\
\text{and} \quad Z_{i,4}(\boldsymbol{h}) =& \sum_{s,k} I\{S_{i+1} = s, K_{i+1} = k\}\\
& \times \mathbf{E}[\boldsymbol{W}_{i+1}(h_1 + h_0)|S_{i+1} = s, K_{i+1} = k].
\end{aligned}
$$

Let $\boldsymbol{Z}_i(\boldsymbol{h}) = (Z_{i,1}(\boldsymbol{h}), Z_{i,2}(\boldsymbol{h}), Z_{i,3}(\boldsymbol{h}), Z_{i,4}(\boldsymbol{h})^\top$, then under Assumption 3.1 and $\sup_{h_a\in\mathscr{H}_a} \mathbf{E}[\{\boldsymbol{W}(h_a)\}^2] < \infty$ for $a \in \{0,1\}$, $\{\boldsymbol{Z}_i(\boldsymbol{h})^\top\}_{i=1}^N$ is sequence of multivariate martingale difference with respect to $\mathcal{F}_i$ satisfying $\mathbf{E}[\|\boldsymbol{Z}_i(\boldsymbol{h})\|^2] < \infty$. We next derive the conditional variances for $\{\boldsymbol{Z}_i(\boldsymbol{h})\}_{i=1}^N$ as follows:

$$
\frac{1}{N}\sum_{i=1}^{N} \mathbf{E}[\{Z_{i,1}(\boldsymbol{h})\}^2|\mathcal{F}_{i-1}] = \sum_{s,k} \pi_{s,k}\mathbf{V}[\boldsymbol{W}^k(h_1 - h_0)|S = s] + o_p(1),
$$

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\mathbf{E}[\{Z_{i,2}(\boldsymbol{h})\}^2|\mathcal{F}_{i-1}] =& \sum_{s_1,s_2,k_1,k_2}\gamma\{(s_1,k_1),(s_2,k_2)\} \\
&\times\mathbf{E}[\boldsymbol{W}^{k_1}(h_1-h_0)|S=s_1]\mathbf{E}[\boldsymbol{W}^{k_2}(h_1-h_0)|S=s_2]+o_p(1) \\
=&\sigma^2_{\mathsf{CAR}}(\boldsymbol{h})+o_p(1), \\
\frac{1}{N}\sum_{i=1}^{N}\mathbf{E}[\{Z_{i,3}(\boldsymbol{h})\}^2|\mathcal{F}_{i-1}] =& \sum_{s,k}\pi_{s,k}\mathbf{V}[\boldsymbol{W}^{k}(h_1+h_0)|S=s]+o_p(1), \\
\frac{1}{N}\sum_{i=1}^{N}\mathbf{E}[\{Z_{i,4}(\boldsymbol{h})\}^2|\mathcal{F}_{i-1}] =& \sum_{s,k}\pi_{s,k}\mathbf{E}[\boldsymbol{W}^{k}(h_1+h_0)|S=s]^2+o_p(1).
\end{aligned}
$$

Furthermore, it is easy to check that the conditional covariances between $\{Z_{i,j}(\boldsymbol{h}), 1\le j\le 4\}$ are zero, and thus the asymptotic variance of $\sqrt{N}\mathbb{G}_N(\boldsymbol{h})$ equals to

$$
\begin{aligned}
&\frac{1}{N}\sum_{j=1}^{4}\sum_{i=1}^{N}\mathbf{E}[\{Z_{i,j}(\boldsymbol{h})\}^2|\mathcal{F}_{i-1}] \\
=&\sum_{s,k}\pi_{s,k}\left(\mathbf{V}[\boldsymbol{W}^{k}(h_1-h_0)|S=s]+\mathbf{E}[\boldsymbol{W}^{k}(h_1-h_0)|S=s]^2\right) \\
&-\left(\sum_{s,k}\pi_{s,k}\mathbf{E}[\boldsymbol{W}^{k}(h_1-h_0)|S=s]^2-\sigma^2_{\mathsf{CAR}}(\boldsymbol{h})\right) \\
&+\sum_{s,k}\pi_{s,k}\left(\mathbf{V}[\boldsymbol{W}^{k}(h_1+h_0)|S=s]+\mathbf{E}[\boldsymbol{W}^{k}(h_1+h_0)|S=s]^2\right)+o_p(1) \\
=&\sum_{s,k}\pi_{s,k}\left(\mathbf{E}[\left\{\boldsymbol{W}^{k}(h_1-h_2)\right\}^2+\left\{\boldsymbol{W}^{k}(h_1+h_2)\right\}^2]\right)
\end{aligned}
$$

$$- \left(\sum_{s,k} \pi_{s,k} \mathbf{E}[\boldsymbol{W}^k (h_1 - h_0) | S = s]^2 - \sigma^2_{\mathsf{CAR}}(\boldsymbol{h}) \right) + o_p(1)$$

$$= 2\sigma^2_{\mathsf{W}}(\boldsymbol{h}) - \left(\sigma^2_{\mathsf{CR}}(\boldsymbol{h}) - \sigma^2_{\mathsf{CAR}}(\boldsymbol{h})\right) + o_p(1).$$

Finally, checking the conditional Lindeberg's condition (Corollary 3.1 of Hall and Heyde, 1980) for $\{\boldsymbol{Z}_i(\boldsymbol{h})\}_{i=1}^N$ yields the asymptotic normality of $\mathbb{G}_N(\boldsymbol{h})$. □

Next, we prove Theorem 3.1 by combining Liang and Zeger (1986) and the asymptotic property of M-estimation as in Theorem 5.21 of van der Vaart (1998) by utilizing Proposition 1. Write $\tilde{\boldsymbol{\alpha}} = \hat{\boldsymbol{\alpha}}(\boldsymbol{\theta}^*) = \hat{\boldsymbol{\alpha}}^*$, then $\hat{\boldsymbol{\theta}}^*$ should satisfy $\Psi_{N,\tilde{\boldsymbol{\alpha}}}(\hat{\boldsymbol{\theta}}^*) = o_p(n^{-1/2})$. Now the derivative of $\Psi_{N,\tilde{\boldsymbol{\alpha}}}(\boldsymbol{\theta}^*)$ can be written as

$$\frac{\partial}{\partial \boldsymbol{\theta}^*} \Psi_{N,\tilde{\boldsymbol{\alpha}}}(\boldsymbol{\theta}^*) = \mathcal{T}_{N,1} + \mathcal{T}_{N,2}(\tilde{\boldsymbol{\alpha}}) \frac{\partial}{\partial \boldsymbol{\theta}^*} \tilde{\boldsymbol{\alpha}} - \tilde{\Psi}_{N,\tilde{\boldsymbol{\alpha}}}(\boldsymbol{\theta}^*), \tag{S4.1}$$

$$\mathcal{T}_{N,1} = \mathcal{T}_{N,1,1} + \mathcal{T}_{N,1,2} + \mathcal{T}_{N,1,3},$$

$$\mathcal{T}_{N,1,1} = \frac{1}{N} \sum_{i=1}^N \sum_{a \in \{0,1\}} I\{A_i = a\} \boldsymbol{G}_{i,a}^\top \frac{\partial}{\partial \boldsymbol{\theta}^*} \widetilde{\boldsymbol{M}}_i(a) \boldsymbol{V}_i^{-1}(\tilde{\boldsymbol{\alpha}}) \boldsymbol{r}_i(a),$$

$$\mathcal{T}_{N,1,2} = \frac{1}{N} \sum_{i=1}^N \sum_{a \in \{0,1\}} I\{A_i = a\} \boldsymbol{G}_{i,a}^\top \widetilde{\boldsymbol{M}}_i(a) \frac{\partial}{\partial \boldsymbol{\theta}^*} \boldsymbol{\Lambda}_i^{-1/2} \left\{\boldsymbol{C}_i(\tilde{\boldsymbol{\alpha}})\right\}^{-1} \boldsymbol{\Lambda}_i^{-1/2} \boldsymbol{r}_i(a),$$

$$\mathcal{T}_{N,1,3} = \frac{1}{N} \sum_{i=1}^N \sum_{a \in \{0,1\}} I\{A_i = a\} \boldsymbol{G}_{i,a}^\top \widetilde{\boldsymbol{M}}_i(a) \boldsymbol{\Lambda}_i^{-1/2} \left\{\boldsymbol{C}_i(\tilde{\boldsymbol{\alpha}})\right\}^{-1} \frac{\partial}{\partial \boldsymbol{\theta}^*} \boldsymbol{\Lambda}_i^{-1/2} \boldsymbol{r}_i(a),$$

$$\mathcal{T}_{N,2}(\tilde{\boldsymbol{\alpha}}) = \frac{1}{N}\sum_{i=1}^{N}\sum_{a\in\{0,1\}} I\{A_i = a\}\boldsymbol{G}_{i,a}^{\top}\widetilde{\boldsymbol{M}}_i(a)\left[\frac{\partial}{\partial\tilde{\boldsymbol{\alpha}}}\boldsymbol{V}_i^{-1}(\tilde{\boldsymbol{\alpha}})\right]\boldsymbol{r}_i(a),$$

$$\text{and}\quad \tilde{\Psi}_{N,\tilde{\boldsymbol{\alpha}}}(\boldsymbol{\theta}^*) = \frac{1}{N}\sum_{i=1}^{N}\sum_{a\in\{0,1\}} I\{A_i = a\}\boldsymbol{G}_{i,a}^{\top}\widetilde{\boldsymbol{M}}_i(a)\boldsymbol{V}_i^{-1}(\tilde{\boldsymbol{\alpha}})\widetilde{\boldsymbol{M}}_i(a)\boldsymbol{G}_{i,a}.$$

Proposition 1.1 implies that

$$\mathcal{T}_{N,1} = \mathcal{T}_1 + o_p(1),\quad \mathcal{T}_{N,2} = \mathcal{T}_2 + o_p(1),\quad \text{and}\quad \tilde{\Psi}_{N,\tilde{\boldsymbol{\alpha}}}(\boldsymbol{\theta}^*) = \tilde{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*) + o_p(1),$$

as $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha}^* + o_p(1)$. Assumption 3.1.5 further yields that

$$\frac{\partial}{\partial\boldsymbol{\theta}^*}\Psi_{N,\tilde{\boldsymbol{\alpha}}}(\boldsymbol{\theta}^*) = -2^{-1}\tilde{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*) + o_p(1), \tag{S4.2}$$

and $\tilde{\Psi}_{\boldsymbol{\alpha}}(\boldsymbol{\theta}^*)$ is continuous in $\boldsymbol{\alpha}$. Since $m(\cdot)$ and $\boldsymbol{V}(\cdot)$ are continuous and twice differentiable, we shall take $\boldsymbol{W}_i(h_a) := \boldsymbol{L}_{i,a}^{\top}(\boldsymbol{\theta},\boldsymbol{\alpha})\boldsymbol{r}_i(a)$ for $a \in \{0,1\}$, and set $\boldsymbol{h} \in (h_1, h_0)$. Then following Lemma S1.3 of Liu et al. (2025), we can show that $\mathscr{H}$ formed by $\boldsymbol{h}$ is a Donsker class. By Theorem 5.21 of van der Vaart (1998) and (S4.2), we have

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\right) = 2\left\{\tilde{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)\right\}^{-1}\sqrt{N}\Psi_{N,\tilde{\boldsymbol{\alpha}}}(\boldsymbol{\theta}^*) + o_p(1). \tag{S4.3}$$

By Assumption 3.1.3, i.e., $\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^* = O_p(N^{-1/2})$, it follows the Taylor expansion of $\sqrt{N}\Psi_{N,\tilde{\boldsymbol{\alpha}}}(\boldsymbol{\theta}^*)$ that

$$
\begin{aligned}
&\sqrt{N}\Psi_{N,\tilde{\boldsymbol{\alpha}}}(\boldsymbol{\theta}^*) \\
&= \sqrt{N}\Psi_{N,\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*) + \left\{\frac{\partial}{\partial \boldsymbol{\alpha}^*}\Psi_{N,\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)\right\}\sqrt{N}\left(\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\right) + O_p(\|\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\|) \\
&= \sqrt{N}\Psi_{N,\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*) + o_p(1),
\end{aligned}
\tag{S4.4}
$$

where

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\alpha}^*}\Psi_{N,\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*) &= \frac{1}{N}\sum_{i=1}^{N}\sum_{a\in\{0,1\}} I\{A_i = a\}\boldsymbol{G}_{i,a}^{\top}\widetilde{\boldsymbol{M}}_i(a)\left[\frac{\partial}{\partial \boldsymbol{\alpha}^*}\boldsymbol{V}_i^{-1}(\boldsymbol{\alpha}^*)\right]\boldsymbol{r}_i(a) \\
&= o_p(1)
\end{aligned}
$$

follows from Proposition 1.1 and the last part of Assumption 3.2.3. Then putting (S4.3) and (S4.4) together, we have

$$
\sqrt{N}\left(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\right) = \left\{\tilde{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)/2\right\}^{-1}\sqrt{N}\Psi_{N,\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*) + o_p(1).
$$

This yields the representation given in Theorem 3.1. Finally, applying Proposition 1.2 to the $\boldsymbol{h}$ described above, we obtain

$$
\sqrt{N}\Psi_{N,\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*) \xrightarrow{\mathbf{D}} \mathcal{N}(\mathbf{0}_{p_0+2}, 2^{-1}\Xi_{\Psi\Psi}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)),
$$

$$\text{where} \quad \Xi_{\Psi\Psi}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) = \Sigma_{\Psi\Psi}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) - 2^{-1}\left[\Sigma_{\mathsf{CR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) - \Sigma_{\mathsf{CAR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)\right],$$

$$\Sigma_{\Psi\Psi}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) = \sum_{k=1}^{k_{\max}} \pi_{\cdot,k} \sum_{a\in\{0,1\}} \mathbf{E}\left[\boldsymbol{L}_a^{k\top}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)\boldsymbol{r}^k(a)\boldsymbol{r}^{k\top}(a)\boldsymbol{L}_a^k(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)\right],$$

$$\Sigma_{\mathsf{CR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) = \sum_{k=1}^{k_{\max}} \sum_{s=1}^{s_{\max}} \pi_{s,k}\boldsymbol{R}(s,k)\boldsymbol{R}^\top(s,k),$$

$$\Sigma_{\mathsf{CAR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) = \sum_{k_1,k_2} \sum_{s_1,s_2} \gamma\{(s_1,k_1),(s_2,k_2)\}\boldsymbol{R}(s_1,k_1)\boldsymbol{R}^\top(s_2,k_2),$$

$$\text{and} \quad \boldsymbol{R}(s,k) = \mathbf{E}\left[\boldsymbol{L}_1^{k\top}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)\boldsymbol{r}^k(1) - \boldsymbol{L}_0^{k\top}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)\boldsymbol{r}^k(0)|S=s\right].$$

This completes the proof of this theorem.

S4.2 Proof of Corollary 3.1

1. Under CR, it is easy to see that Assumption 3.2 holds with

$$\mathbf{E}[(2T_i-1)(2T_j-1)I\{S_i=s_1, K_i=k_1\}I\{S_j=s_2, K_j=k_2\}] = 0,$$

for any s_1, s_2 and k_1, k_2, and

$$\mathbf{E}[(2T_i-1)^2 I\{S_i=s_1, K_i=k_1\}I\{S_i=s_2, K_i=k_2\}]$$

$$= I\{s_1=s_2=s, k_1=k_2=k\}\pi_{s,k}.$$

Consequently, $\Sigma_{\mathsf{CAR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) = \Sigma_{\mathsf{CR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)$.

2. We shall consider K_i as unobserved covariates. By the proof given in the supplementary materials of Liu and Hu (2022), the CAR that satisfies (3.1) leads to

$$d_i(s,k) = (2T_i - 1)\left\{I\{K_i = k\} - \pi_{k|s}\right\} I\{S_i = s\},$$

and thus

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\mathbf{E}\left[d_i(s_1,k_1)d_i(s_2,k_2)|\mathcal{F}_{i-1}\right] \\
&= I\{s_1 = s_2 = s, k_1 = k_2 = k\}\pi_{s,\cdot}\pi_{k|s}(1-\pi_{k|s}) \\
&- I\{s_1 = s_2 = s, k_1 \neq k_2\}\pi_{s,\cdot}\pi_{k_1|s}\pi_{k_2|s} + o_p(1).
\end{aligned}$$

Furthermore, we have

$$\begin{aligned}
\Sigma_{\mathsf{CR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) - \Sigma_{\mathsf{CAR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) &= \sum_{s=1}^{s_{\mathsf{max}}}\pi_{s,\cdot}\sum_{k_1,k_2}\pi_{k_1|s}\pi_{k_2|s}\boldsymbol{R}(s,k_1)\boldsymbol{R}^\top(s,k_2) \\
&= \mathrm{V}[\mathbf{E}[\boldsymbol{R}(S,K)|S]]
\end{aligned}$$

is positive definite, unless $\boldsymbol{R}(S,K)$ is a constant of $\boldsymbol{S}$. This completes the proof for Corollary 3.1.

S4.3 Proof of Corollary 3.2

We first claim the following result that ensures the mean of T_{Wald} is zero under $H_0 : \gamma_1 = \gamma_0$.

Proposition 2. *Suppose (2.1) holds and (2.2) is used as the adjustment with GEE. Then under $H_0 : \gamma_1 = \gamma_0$, $\gamma_1^* = \gamma_0^*$ and $\hat{\tau}^*$ is a consistent estimate of zero.*

Proof of Proposition 2. Under the true outcome model (2.1) and $H_0 : \gamma_1 = \gamma_0$, then the conditional expectations of the first and second elements of $\Psi_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)$ are equal. This further yields that $\gamma_1^* = \gamma_0^*$, since $\boldsymbol{\theta}^* = (\gamma_1^*, \gamma_0^*, \boldsymbol{\beta}^{*\top})^\top$ is the root of $\Psi_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*) = \mathbf{0}_{p_0+2}$. $\square$

Next, by Proposition 2, under $H_0 : \gamma_1 = \gamma_0$, we have

$$\sqrt{N}\hat{\tau}^* \xrightarrow{\mathbf{D}} \mathcal{N}\left(0, 2\sigma_\tau^2(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)\right),$$

and thus

$$T_{\mathsf{Wald}} \xrightarrow{\mathbf{D}} \mathcal{N}(0, \lambda_{\mathsf{Wald}}^2),$$
$$\text{where} \quad \lambda_{\mathsf{Wald}}^2 = \sigma_{\mathsf{Wald}}^{-2}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)\sigma_\tau^2(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*).$$

By (S1.1), Proposition 1.1, and some calculations, we can show that

$$\tilde{\Psi}_{N,\hat{\boldsymbol{\alpha}}^*}(\hat{\boldsymbol{\theta}}^*) = 2^{-1}\widetilde{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*) + o_p(1),$$
$$\text{and} \quad \hat{\Sigma}_{\Psi\Psi}(\hat{\boldsymbol{\theta}}^*, \hat{\boldsymbol{\alpha}}^*) = 2^{-1}\Sigma_{\Psi\Psi}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) + o_p(1).$$

It thus follows that

$$\begin{aligned} N\mathsf{se}^2_{\mathsf{Wald}}(\hat{\tau}^*) &= \boldsymbol{l}^\top \left\{\widetilde{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)/2\right\}^{-1} \Sigma_{\Psi\Psi}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)/2 \left\{\widetilde{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)/2\right\}^{-1} \boldsymbol{l} + o_p(1) \\ &= 2\sigma^2_{\mathsf{Wald}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) + o_p(1). \end{aligned}$$

This completes the first part of Corollary 3.2.

Now considering the alternative hypothesis, we first illustrate the existence of δ^* under $H_1 : \gamma_1 - \gamma_0 = N^{-1/2}\delta$. Recall that the pseudo-true parameter $\boldsymbol{\theta}^* = (\gamma_1^*, \gamma_0^*, \boldsymbol{\beta}_{\mathsf{in}}^{*\top})^\top$ is defined as the unique solution to the population estimating equation $\mathbf{E}[\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\theta}^*)] = \mathbf{0}_{p_0+2}$. Consequently, $\delta^* = \gamma_1^* - \gamma_0^* = \boldsymbol{l}^\top\boldsymbol{\theta}^*$, where $\boldsymbol{l} = (1, -1, \mathbf{0}_{p_0}^\top)^\top$, is well defined as a linear transformation of $\boldsymbol{\theta}^*$.

On the other hand, to see the relationships between δ^* and δ, under the local alternative $\tau = \gamma_1 - \gamma_0 = N^{-1/2}\delta$, the pseudo-true parameter may be regarded as a function of τ, denoted by $\boldsymbol{\theta}^*(\tau)$. By Assumption 3.1.5, $\tilde{\Psi}_{\alpha}(\boldsymbol{\theta}^*(\tau))$ with respect to $\boldsymbol{\theta}^*(\tau)$ is nonsingular. Hence, by the implicit function theorem, $\boldsymbol{\theta}^*(\tau)$ is continuously differentiable in a neighborhood of

$\tau = 0$. Applying a first-order Taylor expansion to $\boldsymbol{l}^\top \boldsymbol{\theta}^*(\tau)$ around $\tau = 0$ establishes that

$$\delta^* \approx 2\delta \cdot \boldsymbol{l}^\top \left\{ \widetilde{\boldsymbol{\Psi}}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*) \Big|_{\gamma_1=\gamma_0} \right\}^{-1} \frac{\partial}{\partial \tau} \boldsymbol{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*) \Big|_{\gamma_1=\gamma_0}.$$

Lastly, by Theorem 3.1, we have

$$\sqrt{N}(\hat{\tau}^* - \tau^*) \xrightarrow{\mathbf{D}} \mathcal{N}\left(0, 2\sigma_\tau^2(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)\right),$$

where $\sigma_\tau^2(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) = \boldsymbol{l}^\top \left\{ \widetilde{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*) \right\}^{-1} \Xi_{\Psi\Psi}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) \left\{ \widetilde{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*) \right\}^{-1} \boldsymbol{l}$. Then applying the same argument used in the proof of the first part of Corollary 3.2 yields the conclusion for the second part of this corollary.

S4.4 Proof of Theorem 4.1

We begin the proof by restating our adjustment approach. First, we utilize

$$\hat{\Sigma}_{\mathsf{CR}}(\hat{\boldsymbol{\theta}}^*, \hat{\boldsymbol{\alpha}}^*) = \sum_{k=1}^{k_{\mathsf{max}}} \sum_{s=1}^{s_{\mathsf{max}}} \hat{\pi}_{s,k} \hat{\boldsymbol{R}}(s,k) \hat{\boldsymbol{R}}^\top(s,k)$$

for estimating $\Sigma_{\mathsf{CR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)$, where $\hat{\pi}_{s,k}$ and $\hat{\boldsymbol{R}}(s,k)$ are the empirical estimates of $\pi_{s,k}$ and $\boldsymbol{R}(s,k)$. Second, we consider estimating $\Sigma_{\mathsf{CAR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)$ with the following bootstrap based approach. Let B denote the number of simulations. For $1 \le b \le B$, generate $\{\{S_i^{*b}, K_i^{*b}\}_{i=1}^N\}_{b=1}^B$ from the empirical distribution of $\{S_i, K_i\}_{i=1}^N$. Then, generate $\{\{A_i^{*b}\}_{i=1}^N\}_{b=1}^B$ according to the same CAR procedure used as the one used for the trial and evaluate the imbalance $\{D_N^{*b}(s,k), 1 \le s \le s_{\mathsf{max}}, 1 \le k \le k_{\mathsf{max}}\}_{b=1}^B$. Next, calculate $O_b = N^{-1/2} \sum_{k=1}^{k_{\mathsf{max}}} \sum_{s=1}^{s_{\mathsf{max}}} D_N^{*b}(s,k) \hat{\boldsymbol{R}}(s,k)$ for $1 \le b \le B$ and estimate $\Sigma_{\mathsf{CAR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)$ with the sample variance of $\{O_b\}_{b=1}^B$ denoted by S_O^2. Then our proposed adjusted standard error satisfies

$$
\begin{aligned}
N\mathsf{se}^2_{\mathsf{adj}}(\hat{\tau}^*) = {} & N\mathsf{se}^2_{\mathsf{Wald}}(\hat{\tau}^*) \\
& - \boldsymbol{l}^\top \{\widetilde{\Psi}_{N,\hat{\boldsymbol{\alpha}}^*}(\hat{\boldsymbol{\theta}}^*)\}^{-1} 4^{-1} \left[\hat{\Sigma}_{\mathsf{CR}}(\hat{\boldsymbol{\theta}}^*, \hat{\boldsymbol{\alpha}}^*) - S_O^2\right] \{\widetilde{\Psi}_{N,\hat{\boldsymbol{\alpha}}^*}(\hat{\boldsymbol{\theta}}^*)\}^{-1} \boldsymbol{l},
\end{aligned}
$$

where $\widetilde{\Psi}_{N,\hat{\boldsymbol{\alpha}}^*}(\hat{\boldsymbol{\theta}}^*) = \dfrac{1}{N} \sum_{i=1}^{N} \sum_{a \in \{0,1\}} I\{A_i = a\} \hat{\boldsymbol{L}}_{i,a}^\top(\hat{\boldsymbol{\theta}}^*, \hat{\boldsymbol{\alpha}}^*) V_i(\hat{\boldsymbol{\alpha}}^*) \boldsymbol{L}_{i,a}^\top(\hat{\boldsymbol{\theta}}^*, \hat{\boldsymbol{\alpha}}^*)$.

Then the adjusted test statistics is $T_{\mathsf{adj}} = \mathsf{se}^{-1}_{\mathsf{adj}}(\hat{\tau}^*)\hat{\tau}^*$.

The proof consists of two steps. First, it follows from Proposition 1 that $\hat{\boldsymbol{R}}(s,k) = R(s,k) + o_p(1)$ and thus $\hat{\Sigma}_{\mathsf{CR}}(\hat{\boldsymbol{\theta}}^*, \hat{\boldsymbol{\alpha}}^*) = \Sigma_{\mathsf{CR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) + o_p(1)$.

For the second step, we aim to show that $S_O^2 = \Sigma_{\mathsf{CAR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) + o_p(1)$. Since $\hat{\boldsymbol{R}}(s,k)$ is consistent, it remains to show that $\{D_N^{*b}(s,k), 1 \le s \le s_{\mathsf{max}}, 1 \le k \le k_{\mathsf{max}}\}_{b=1}^B$ has the same distribution as $\{D_N(s,k), 1 \le s \le s_{\mathsf{max}}, 1 \le k \le k_{\mathsf{max}}\}$. To see this, let $\mathbf{\Pi}_{s,k}$ denote the probabilities of the joint strata for $\{\pi_{s,k}, 1 \le s \le s_{\mathsf{max}}, 1 \le k \le k_{\mathsf{max}}\}$, $\mathbf{\Pi}^*_{s,k}$ denote the empirical probabilities of $\{\hat{\pi}_{s,k}, 1 \le s \le s_{\mathsf{max}}, 1 \le k \le k_{\mathsf{max}}\}$, and $\mathbf{P}_{\mathbf{\Pi}}$ and $\mathbf{P}_{\mathbf{\Pi}^*}$ denote the associated probability measures. Let $\boldsymbol{D}_N(\boldsymbol{s}, \boldsymbol{k})$ be the vector of be the joint imbalance $\{D_N(s,k), 1 \leq s \leq s_{\mathsf{max}}, 1 \leq k \leq k_{\mathsf{max}}\}$ following the probability measure $\mathbf{P}_{\boldsymbol{\Pi}}$. In a similar manner, the vector of the joint imbalance $\{D_N^*(s,k), 1 \leq s \leq s_{\mathsf{max}}, 1 \leq k \leq k_{\mathsf{max}}\}$ evaluated from bootstrap, which is denoted by $\boldsymbol{D}_N^*(\boldsymbol{s}, \boldsymbol{k})$, can be demonstrated as $\boldsymbol{D}_N(\boldsymbol{s}, \boldsymbol{k})$ following the probability measure $\mathbf{P}_{\boldsymbol{\Pi}^*}$. By Assumption 3.2, the martingale central limit theorem yields that

$$N^{-1/2}\boldsymbol{D}_N(\boldsymbol{s}, \boldsymbol{k}) \xrightarrow[\mathbf{P}_{\boldsymbol{\Pi}}]{\mathbf{D}} \mathcal{N}\left(\mathbf{0}_{\mathsf{s}_{\mathsf{max}}\mathsf{k}_{\mathsf{max}}},\ \Xi_{\mathbf{D}}(\boldsymbol{s}, \boldsymbol{k})\right)$$

under the probability measure $\mathbf{P}_{\boldsymbol{\Pi}}$ for some covariance matrix $\Xi_{\mathbf{D}}(\boldsymbol{s}, \boldsymbol{k})$. It remains to show that

$$N^{-1/2}\boldsymbol{D}_N(\boldsymbol{s}, \boldsymbol{k}) \xrightarrow[\mathbf{P}_{\boldsymbol{\Pi}^*}]{\mathbf{D}} \mathcal{N}\left(\mathbf{0}_{\mathsf{s}_{\mathsf{max}}\mathsf{k}_{\mathsf{max}}},\ \Xi_{\mathbf{D}}(\boldsymbol{s}, \boldsymbol{k})\right), \tag{S4.5}$$

i.e., $N^{-1/2}\boldsymbol{D}_N(\boldsymbol{s}, \boldsymbol{k})$ under the probability measure $\mathbf{P}_{\boldsymbol{\Pi}^*}$ converges to the same asymptotic distribution as $N^{-1/2}\boldsymbol{D}_N(\boldsymbol{s}, \boldsymbol{k})$ under the probability measure $\mathbf{P}_{\boldsymbol{\Pi}}$. The proof of (S4.5) employs Le Cam's third lemma (Example 6.7 of van der Vaart, 1998) in a similar manner as (Lemma S3.1 of Liu and Hu, 2023) by adapting $\{D_N(s,\cdot), 1 \leq s \leq s_{\mathsf{max}}\}$ and $\{\pi_{s,\cdot}, 1 \leq s \leq s_{\mathsf{max}}\}$ to $\{D_N(s,k), 1 \leq s \leq s_{\mathsf{max}}, 1 \leq k \leq k_{\mathsf{max}}\}$ and $\{\pi_{s,k}, 1 \leq s \leq s_{\mathsf{max}}$ and $1 \leq$

$k \le k_{\mathsf{max}}\}$. Applying (S4.5) and Beran (1997) to $\{O_b\}_{b=1}^B$ yields

$$S_O^2 \xrightarrow{\mathbf{P}_{\mathbf{\Pi}^*}} \Sigma_{\mathsf{CAR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*),$$

and consequently,

$$\begin{aligned}
N\mathsf{se}_{\mathsf{adj}}^2(\hat{\tau}^*) &= \boldsymbol{l}^\top \left\{\widetilde{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)/2\right\}^{-1} 2^{-1}\Xi_{\Psi\Psi}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) \left\{\widetilde{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)/2\right\}^{-1} \boldsymbol{l} + o_p(1) \\
&= 2\boldsymbol{l}^\top \left\{\widetilde{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)\right\}^{-1} \Xi_{\Psi\Psi}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) \left\{\widetilde{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)\right\}^{-1} \boldsymbol{l} + o_p(1) \\
&= 2\sigma_\tau^2(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) + o_p(1).
\end{aligned}$$

Consequently, under $H_0 : \gamma_1 = \gamma_2$,

$$\mathsf{se}_{\mathsf{adj}}^{-1}(\hat{\tau}^*)\hat{\tau}^* \xrightarrow{\mathbf{D}} \mathcal{N}(0, 1).$$

Then, the rest of the proof follows from a similar argument as the proof of Theorem 3.1.

S5 Additional Details Regarding the Numerical Implementation of GEE and Assumption 3.1.5

We evaluate several common settings under which $\mathcal{T}_1 = 0$ or $\mathcal{T}_2(\boldsymbol{\alpha}) = 0$

in Section S5.1. Section S5.2 further examines the relative magnitudes of $\mathcal{T}_1$ and $\mathcal{T}_2(\boldsymbol{\alpha}^*)$ compared with $\widetilde{\Psi}_{\boldsymbol{\alpha}}(\boldsymbol{\theta}^*)$. Lastly, we compare the standard Fisher-scoring implementation of GEE, which ignores $\mathcal{T}_1$ and $\mathcal{T}_2(\boldsymbol{\alpha}^*)$, with the Newton–Raphson algorithm that incorporates these terms into the gradient. Overall, the numerical results indicate that ignoring $\mathcal{T}_1$ and $\mathcal{T}_2(\boldsymbol{\alpha}^*)$ has limited impact on the implementation of GEE for estimating τ^*.

S5.1 Common Settings

The common settings under which $\mathcal{T}_1 = \mathbf{0}$ or $\mathcal{T}_2(\boldsymbol{\alpha}) = \mathbf{0}$ are listed in the following proposition.

Proposition 3. *1. If the working model is correctly specified, i.e.,* $\mu(\cdot) = m(\cdot)$ *and* $\mathbf{X}_{i,\mathsf{in}} = \mathbf{X}_{i,j}$*, then* $\mathcal{T}_1 = \mathbf{0}$ *and* $\mathcal{T}_2(\boldsymbol{\alpha}^*) = \mathbf{0}$.

2. If the identity matrix is assumed for the working correlation, then $\mathcal{T}_2(\boldsymbol{\alpha}^*) = \mathbf{0}$*; in addition, if* $m(\cdot)$ *considers a canonical link for the exponential family, then* $\mathcal{T}_1 = \mathbf{0}$ *and* $\mathcal{T}_2(\boldsymbol{\alpha}^*) = \mathbf{0}$.

3. If the Gaussian-identity link function is used , i.e., the dispersion parameter ϕ_i *is not considered or does not depend on* $\boldsymbol{\theta}^*$*, then* $\mathcal{T}_1 = \mathbf{0}$.

4. If $m(\cdot)$ *considers the a canonical link for the exponential family and the* *covariates used for adjustment are subject-invariant, i.e.,* $\boldsymbol{X}_{i,j} = \boldsymbol{X}_i$ *and* $\eta^a_{i,j} = \eta^a_i$*, then* $\mathcal{T}_1 = \mathbf{0}$.

We note that a correctly specified working model or a canonical link with working independence correlation guarantee $\mathcal{T}_1 = \mathbf{0}$ and $\mathcal{T}_2(\boldsymbol{\alpha}^*) = \mathbf{0}$. On the other hand, the identity link function or the canonical link with subject-invariant covariates imply $\mathcal{T}_1 = \mathbf{0}$. We evaluate the relative magnitude of $\mathcal{T}_2(\boldsymbol{\alpha}^*)$ in Section S5.2.

Proof of Proposition 3. We derive for each of the four cases considered in Proposition 3 as given below.

1. Under the correctly specified working model, i.e., $\mu(\cdot) = m(\cdot)$ and $\mathbf{X}_{i,\mathsf{in}} = \mathbf{X}_i$, we have

$$\begin{aligned}\mathbf{E}[r_{i,j}(a)|\mathbf{X}_i] &= \mathbf{E}[Y_{i,j}(a) - m_{i,j}(a)|\mathbf{X}_i] \\ &= \mathbf{E}[Y_{i,j}(a) - \mu_{i,j}(a)|\mathbf{X}_i] = 0.\end{aligned}$$

This implies $\mathcal{T}_1 = \mathbf{0}$ and $\mathcal{T}_2(\boldsymbol{\alpha}^{\boldsymbol{*}}) = \mathbf{0}$.

2. Under the independent working correlation, i.e., $\boldsymbol{C}_i(\boldsymbol{\alpha}^*) = \boldsymbol{I}_{n_i}$, it is straightforward to see that $\frac{\partial}{\partial \boldsymbol{\alpha}}\boldsymbol{C}^k_i(\boldsymbol{\alpha}^*) = \mathbf{0}$, $\boldsymbol{V}_i(\boldsymbol{\alpha}^*) = \boldsymbol{\Lambda}^{1/2}_{i,a}\boldsymbol{C}^k_i(\boldsymbol{\alpha}^*)\boldsymbol{\Lambda}^{1/2}_{i,a} = \boldsymbol{\Lambda}_{i,a}$, and $\mathcal{T}_2(\boldsymbol{\alpha}^*) = \mathbf{0}$.

Next, suppose the working model utilizes the canonical link for the exponential family, it is easy to check that $\phi_i \cdot m'_{i,j}(a) = v_{i,j}(a)$ and thus $\phi_i \widetilde{\boldsymbol{M}}_i(a) = \boldsymbol{\Lambda}_{i,a}$ and $\widetilde{\boldsymbol{M}}_i(a)\{\boldsymbol{V}_i(\boldsymbol{\alpha}^*)\}^{-1} = \phi_i^{-1}\boldsymbol{\Lambda}_{i,a}\boldsymbol{\Lambda}_{i,a}^{-1} = \phi_i^{-1}\boldsymbol{I}_{n_i}$, where ϕ_i represent the dispersion parameter (McCullagh and J. A., 2019). This suggests that the score function becomes

$$\Psi_{N,\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*) = \frac{1}{N}\sum_{i=1}^{N}\left[A_i\boldsymbol{G}_{i,1}^\top\boldsymbol{r}_i(1) + (1-A_i)\boldsymbol{G}_{i,0}^\top\boldsymbol{r}_i(0)\right]\phi_i^{-1}.$$

The derivative of $\Psi_{N,\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)$ can be calculated as

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\theta}^*}\Psi_{N,\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*) &= -\frac{1}{N}\sum_{i=1}^{N}\left[A_i\boldsymbol{G}_{i,1}^\top\widetilde{\boldsymbol{M}}_i(1)\boldsymbol{G}_{i,1} + (1-A_i)\boldsymbol{G}_{i,0}^\top\widetilde{\boldsymbol{M}}_i(0)\boldsymbol{G}_{i,0}\right]\phi_i^{-1} \\ &= -\tilde{\Psi}_{N,\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*),\end{aligned}$$

which implies that $\mathcal{T}_{N,1} = \mathbf{0}$ and $\mathcal{T}_{N,2}(\boldsymbol{\alpha}^*) = \mathbf{0}$.

3. Now consider the identity link function $m(\eta) = \eta$. Since $m'(\eta) = 1$, and $m''(\eta) = 0$, it can be shown that $\widetilde{\boldsymbol{M}}_i(a) = \boldsymbol{I}_{n_i}$, and $\frac{\partial}{\partial\boldsymbol{\theta}^*}\widetilde{\boldsymbol{M}}_i(a) = \mathbf{0}$. Furthermore, identity link is a special case of canonical link, so that $\phi_i\widetilde{\boldsymbol{M}}_i(a) = \phi_i\boldsymbol{I}_{n_i} = \boldsymbol{\Lambda}_{i,a}$. Hence, $\frac{\partial}{\partial\boldsymbol{\theta}^*}\boldsymbol{\Lambda}_{i,a} = \mathbf{0}$. These results imply $\mathcal{T}_{N,1,1} = \mathcal{T}_{N,1,2} = \mathcal{T}_{N,1,3} = \mathbf{0}$.

4. If $m(\cdot)$ corresponds to the canonical link for the exponential family, then $\phi_i\widetilde{\boldsymbol{M}}_i(a) = \boldsymbol{\Lambda}_{i,a}$. Suppose further that the covariates used for adjustment are subject-invariant, i.e., $\boldsymbol{X}_{i,j} = \boldsymbol{X}_i$. Then the linear predictor is constant over repeated measurements within subject, and $v_{i,j}(\cdot)$ is a function of $\boldsymbol{X}_{i,j}$ only. Denote $v_i(\boldsymbol{\theta}^*)$ as the working variance for the i-th subject, then we have $\boldsymbol{\Lambda}_{i,a} = v_i(\boldsymbol{\theta}^*)\boldsymbol{I}_{n_i}$ and

$$\begin{aligned}\widetilde{\boldsymbol{M}}_i(a)\,[\boldsymbol{V}_i(\boldsymbol{\alpha}^*)]^{-1} &= \phi_i^{-1} v_i(\boldsymbol{\theta}^*)\boldsymbol{I}_{n_i}[v_i^{1/2}(\boldsymbol{\theta}^*)\boldsymbol{C}_i(\boldsymbol{\alpha}^*)v^{1/2}(\boldsymbol{\theta}^*)]^{-1} \\ &= \phi_i^{-1}[\boldsymbol{C}_i(\boldsymbol{\alpha}^*)]^{-1}.\end{aligned}$$

Since $\frac{\partial}{\partial \boldsymbol{\theta}}\widetilde{\boldsymbol{M}}_i(a)\,[\boldsymbol{V}_i(\boldsymbol{\alpha}^*)]^{-1} = \mathbf{0}$, then we have $\frac{\partial}{\partial \boldsymbol{\theta}^*}\Psi_{N,\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*) = \mathcal{T}_{N,2}(\boldsymbol{\alpha}) - \tilde{\Psi}_{N,\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)$ with $\mathcal{T}_{N,1} = \mathbf{0}$. $\square$

S5.2 The Evaluation of the Relative Magnitude of $\mathcal{T}_1$ and $\mathcal{T}_2$

We complement the common settings stated in Proposition 3 with the numerical evaluations of $\mathcal{T}_1$ and $\mathcal{T}_2(\boldsymbol{\alpha}^*)$ in settings where they are not necessarily zeros. Since $\frac{\partial}{\partial \boldsymbol{\theta}^*}\boldsymbol{\alpha}$ is bounded, the contribution of $\mathcal{T}_2(\boldsymbol{\alpha}^*)\frac{\partial}{\partial \boldsymbol{\theta}^*}\boldsymbol{\alpha}^*$ to the Hessian matrix $\frac{\partial}{\partial \boldsymbol{\theta}^*}\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\theta}^*)$ can be majorly due to $\mathcal{T}_2(\boldsymbol{\alpha}^*)$. For a matrix $\boldsymbol{A}$, we consider the Frobenius norm, i.e., $\|A\|_F = \sqrt{\mathrm{tr}(A^\top A)}$, to characterize the magnitude of the matrix. Then the relative magnitude of $\mathcal{T}_1$ and $\mathcal{T}_2(\boldsymbol{\alpha}^*)$ is characterized by the ratios, $\{\|\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)\|_F\}^{-1}\|\mathcal{T}_1\|_F$ and $\{\|\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)\|_F\}^{-1}\|\mathcal{T}_2(\boldsymbol{\alpha}^*)\|_F$, respectively.

In each scenario, $\boldsymbol{\theta}^*$ and $\boldsymbol{\alpha}^*$ are obtained from the `geese()` function in

R package `geepack`, and the expectations defining $\mathcal{T}_1$, $\mathcal{T}_2(\boldsymbol{\alpha}^*)$, and $\widetilde{\boldsymbol{\Psi}}$ are computed via 10,000 repetitions of Monte Carlo simulations.

Scenario 1: Gaussian Identity Link

Under the Gaussian identity link function, $\mathcal{T}_1 = \mathbf{0}$, we evaluate the magnitude of $\mathcal{T}_2(\boldsymbol{\alpha}^*)$ under the same setting as we considered in Section 5 of our main text. We briefly summarize the setting as follows. Suppose $N = 500$, K_i are *i.i.d.* multinomial random variables taking values $3, 4, 5$ with probabilities 0.2, 0.3, and 0.5, respectively. Suppose $\boldsymbol{X}_{i,j}$ are identical across j and $\boldsymbol{X}_{i,j} = \boldsymbol{X}_i = (X_{i,1}, X_{i,2}, X_{i,3})^\top$, and $X_{i,1}$, $X_{i,2}$, $X_{i,3}$ are i.i.d., $2Bernoulli(1/2) - 1$, $N(0,1)$, and $N(0,1)$, respectively. Consider the following outcome generating model:

$$\begin{aligned} Y_{i,j} = A_{i,j}\gamma_1 + (1 - A_{i,j})\gamma_0 + X_{i,1}\beta_1 + X_{i,2}\beta_2 + X_{i,3}\beta_3 \\ + \exp\{X_{i,1}\beta_1 + X_{i,2}\beta_2\} + \epsilon_{i,j}, \end{aligned} \tag{S5.1}$$

where $\gamma_1 = \gamma_0 = 0.5$, $\beta_1 = \beta_2 = \beta_3 = 1$, and $\{\epsilon_{i,j}, 1 \le j \le n_i\}_{i=1}^N$ are generated with $\sigma^2 = 1.5^2$ and AR(1) correlation matrices with $\alpha = 0.7$.

Since the consistency of $\hat{\boldsymbol{\theta}}^*$ and $\hat{\boldsymbol{\alpha}}^*$ does not depend on the design, we assume A_i are generated according to complete randomization. For the inference stage, we apply (2.2) with identity link with M_1, adjusting no covariates; M_2, adjusting $X_{i,1}$ and $X_{i,2}$; and M_3, adjusting $X_{i,1}$, $X_{i,2}$ and $X_{i,3}$, with the linear form of the predictor, and compare the exchangeable (EXCH) and the AR(1) working correlation matrices for the implementation of GEE. We report $\|\mathcal{T}_2(\boldsymbol{\alpha}^*)\|_F$ and $\{\|\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)\|_F\}^{-1}\|\mathcal{T}_2(\boldsymbol{\alpha}^*)\|_F$ under exchangeable and AR(1) working correlations in Table S1.

Table S1: Numerical evaluation of $\|\mathcal{T}_2(\boldsymbol{\alpha}^*)\|_F$ and $\{\|\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)\|_F\}^{-1}\|\mathcal{T}_2(\boldsymbol{\alpha}^*)\|_F$ for continuous outcomes.

Corr	Model	$\|\mathcal{T}_2(\boldsymbol{\alpha}^*)\|_F$ (10^{-2})	$\|\mathcal{T}_2(\boldsymbol{\alpha}^*)\|_F/\|\widetilde{\boldsymbol{\Psi}}\|_F$ (%)
EXCH	M_1	0.81	1.67
	M_2	0.20	4.84
	M_3	0.20	3.97
AR(1)	M_1	0.11	2.23
	M_2	0.17	3.99
	M_3	0.19	3.69

From Table S1, it can be seen that the ratio $\{\|\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)\|_F\}^{-1}\|\mathcal{T}_2(\boldsymbol{\alpha}^*)\|_F$ is uniformly below 5% across all combinations of working models and non-independent working correlation structures. This suggests that $\mathcal{T}_2(\boldsymbol{\alpha}^*)$ has relative small impact compared to $\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)$.

Scenario 2: Binary Outcomes with Subject-Varying Covariates

In this section, we examine a binary-outcome setting in which neither $\mathcal{T}_1$ nor $\mathcal{T}_2(\boldsymbol{\alpha}^*)$ is zero according to conditions stated in Proposition 3. We consider the following logistic-link data-generating model:

$$\begin{aligned}\eta_{i,j} &= (1-A_i)\gamma_0 + A_i\gamma_1 + X_{i,j,1}\beta_1 + X_{i,j,2}\beta_2 + X_{i,j,3}\beta_3 \\ &\quad + \exp\{X_{i,j,1}\beta_1 + X_{i,j,2}\beta_2\}, \\ Y_{i,j} &= I(\epsilon_{i,j} \le \eta_{i,j}),\end{aligned}$$

where $\gamma_0 = 0.5$, $\gamma_1 = \tau + \gamma_0$, $\tau = 0$ and $\beta_1 = \beta_2 = \beta_3 = 1$. To avoid settings in which $\mathcal{T}_1 = \mathbf{0}$, we generate covariates independently across subjects i and the observations j for $1 \le j \le n_i$. Specifically, we let $X_{i,j,1}$ be i.i.d. Uniform$(0,1)$, $X_{i,j,2} = 2Z_{i,j} - 1$ with $Z_{i,j}$ being Bernoulli$(1/2)$, and $X_{i,j,3}$ be i.i.d.$\mathcal{N}(0,1)$. Furthermore, the random errors $\{\epsilon_{i,j} : 1 \le j \le n_i\}_{i=1}^N$ are generated from a multivariate logistic distribution with AR(1) correlation matrix and correlation parameter $\alpha = 0.7$, using the Gaussian copula method of Emrich and Piedmonte (1991a).

We implement the same working models M_1–M_3 considered in Scenario 1, using a logistic link and the corresponding linear working predictors and A_i is generated according to complete randomization. Because the data-generating model includes the nonlinear term $\exp\{X_{i,j,1}\beta_1 + X_{i,j,2}\beta_2\}$, whereas the working models M_1–M_3 do not, the working models are misspecified. The values of $\|\mathcal{T}_1\|_F$, $\|\mathcal{T}_2(\boldsymbol{\alpha}^*)\|_F$, $\|\mathcal{T}_1\|_F/\|\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)\|_F$, and

$\|\mathcal{T}_2(\boldsymbol{\alpha}^*)\|_F/\|\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)\|_F$ are evaluated in Table S2.

Table S2: Numerical evaluation of $\|\mathcal{T}_1\|_F$ and $\|\mathcal{T}_2(\boldsymbol{\alpha}^*)\|_F$ relative to $\|\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)\|_F$ for binary outcomes.

Corr	Model	$\|\mathcal{T}_1\|_F$ (10^{-18})	$\|\mathcal{T}_2(\boldsymbol{\alpha}^*)\|_F$ (10^{-2})	$\|\mathcal{T}_1\|_F/\|\widetilde{\boldsymbol{\Psi}}\|_F$ (%, 10^{-16})	$\|\mathcal{T}_2(\boldsymbol{\alpha}^*)\|_F/\|\widetilde{\boldsymbol{\Psi}}\|_F$ (%)
	M_1	19.70	0.13	17.14	1.17
EXCH	M_2	1.46	0.45	0.97	3.01
	M_3	2.83	0.52	1.82	3.37
	M_1	8.74	1.18	7.45	1.01
AR(1)	M_2	2.64	7.45	1.80	5.06
	M_3	2.25	7.91	1.49	5.25

Table S2 suggests that both $\mathcal{T}_1$ and $\mathcal{T}_2(\boldsymbol{\alpha}^*)$ are small compared to $\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)$ in the binary-outcome setting. First, although the simulation uses measurement-level covariates together with non-independent working correlation structures, a setting in which $\mathcal{T}_1$ is not expected to be zero, the Frobenius norm of $\mathcal{T}_1$ remains on the order of 10^{-18} across all scenarios, and its relative magnitude is on the order of 10^{-16}%. Thus, $\mathcal{T}_1$ is numerically indistinguishable from zero in this setting. Second, $\mathcal{T}_2(\boldsymbol{\alpha}^*)$ is nonzero but remains small in absolute magnitude. Its Frobenius norm is on the order of 10^{-2}, ranging from 0.13×10^{-2} to 7.91×10^{-2} across different settings. The magnitude of $\mathcal{T}_2(\boldsymbol{\alpha}^*)$ is larger under the AR(1) working correlation than under the exchangeable working correlation, particularly for the enriched working models M_2 and M_3. Finally, after normalization by $\|\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)\|_F$, the relative magnitude of $\mathcal{T}_2(\boldsymbol{\alpha}^*)$ ranges from 1.01% to 5.25% and remains below 6% in all cases, consistent with the findings in Scenario 1.

S5.3 Impact on the Estimation of τ^*

In this section, we examine the impact of the additional derivative terms $\mathcal{T}_1$ and $\mathcal{T}_2(\boldsymbol{\alpha}^*)$ on the estimation of τ^*. By (S4.1) and following Liang and Zeger (1986), for a given current estimate $\tilde{\boldsymbol{\alpha}}$ of the working correlation parameter, the Newton–Raphson update for solving $\Psi_{N,\tilde{\boldsymbol{\alpha}}}(\boldsymbol{\theta}^*) = \mathbf{0}$ is given by

$$\hat{\boldsymbol{\theta}}_{\mathrm{NR}}^{*(j+1)} = \hat{\boldsymbol{\theta}}_{\mathrm{NR}}^{*(j)} - \left[\frac{\partial}{\partial \boldsymbol{\theta}^*} \Psi_{N,\tilde{\boldsymbol{\alpha}}}(\boldsymbol{\theta}^*) \bigg|_{\boldsymbol{\theta}^* = \hat{\boldsymbol{\theta}}_{\mathrm{NR}}^{*(j)}} \right]^{-1} \Psi_{N,\tilde{\boldsymbol{\alpha}}} \left(\hat{\boldsymbol{\theta}}_{\mathrm{NR}}^{*(j)} \right),$$

where $\frac{\partial}{\partial \boldsymbol{\theta}^*} \Psi_{N,\tilde{\boldsymbol{\alpha}}}(\boldsymbol{\theta}^*) = \mathcal{T}_{N,1} + \mathcal{T}_{N,2}(\tilde{\boldsymbol{\alpha}}) \frac{\partial \tilde{\boldsymbol{\alpha}}}{\partial \boldsymbol{\theta}^*} - \widetilde{\boldsymbol{\Psi}}_{N,\tilde{\boldsymbol{\alpha}}}(\boldsymbol{\theta}^*)$. On the other hand, the Fisher scoring update used in the standard GEE algorithm considers

$$\hat{\boldsymbol{\theta}}_{\mathrm{FS}}^{*(j+1)} = \hat{\boldsymbol{\theta}}_{\mathrm{FS}}^{*(j)} + \left[\widetilde{\boldsymbol{\Psi}}_{N,\tilde{\boldsymbol{\alpha}}}(\boldsymbol{\theta}^*) \bigg|_{\boldsymbol{\theta}^* = \hat{\boldsymbol{\theta}}_{\mathrm{FS}}^{*(j)}} \right]^{-1} \Psi_{N,\tilde{\boldsymbol{\alpha}}} \left(\hat{\boldsymbol{\theta}}_{\mathrm{FS}}^{*(j)} \right).$$

Thus, the difference between the Newton–Raphson and Fisher Scoring is due to the additional terms $\mathcal{T}_{N,1}$ and $\mathcal{T}_{N,2}(\tilde{\boldsymbol{\alpha}}) \frac{\partial}{\partial \boldsymbol{\theta}^*} \tilde{\boldsymbol{\alpha}}$ in the derivative of the estimating equation. In our numerical implementation, the standard GEE estimator is computed using the `geese()` function in the `geepack` package in R software.

To compare Newton-Raphson with Fisher Scoring, we assume the covariates and the outcomes follow the same setting as described in Scenario 2 in Section S5.2 with $\tau = 0$ ($\gamma_1 = \gamma_0 = 0.5$) and $\tau = 0.2$ ($\gamma_1 = 0.7, \gamma_0 = 0.5$), the four randomization procedures (CR, HH, SPB, PS), three working models (M_1, M_2, M_3), and 10,000 replications per configuration. For trial design and stratification, we use only the baseline covariates, $X_{i,1,1}$ and $X_{i,1,2}$ ($1 \leq i \leq N$), consistent with standard practice in longitudinal clinical trials. And $X_{i,1,2}$ is discretized to $\tilde{X}_{i,1,2}$ such that $\tilde{X}_{i,1,2} = 1$ if $X_{i,1,2} \geq 0$, and $\tilde{X}_{i,1,2} = 0$ if $X_{i,1,2} < 0$, yielding a total of $s_{\mathsf{max}} = 4$ strata. We report mean and standard deviation of $\hat{\tau}^*$ from each method, along with their differences in Table S3.

Table S3 shows that the Newton–Raphson update based on the exact Hessian and the Fisher scoring update implemented by `geese()` produce virtually identical performance on the estimation of τ^*. Across all randomization procedures, working models, and values of τ, the means and $\sqrt{N}$-scaled standard deviations of $\hat{\tau}^*_{\mathrm{FS}}$ and $\hat{\tau}^*_{\mathrm{NR}}$ are quite close. The differences in the sample means are only on the order of 10^{-8}, while the differences in the $\sqrt{N}$-scaled standard deviations are on the order of 10^{-7}. These discrepancies are therefore almost negligible. This confirms that, although $\mathcal{T}_1$ and $\mathcal{T}_2(\boldsymbol{\alpha}^*)$ are used in the exact Hessian and hence may affect the iterative path, their contribution has essentially limited effect on the numerical

Table S3: Comparison of $\hat{\tau}^*$ between Newton–Raphson (NR, exact Hessian) and Fisher scoring (FS, `geese()`).

τ	Assignment	Model	$\hat{\tau}^*_{\text{FS}}$	$\hat{\tau}^*_{\text{NR}}$	Difference	
			mean ($\sqrt{N}$sd)	mean ($\sqrt{N}$sd)	mean (10^{-8})	$\sqrt{N}$sd (10^{-7})
0.0	CR	M_1	0.001 (2.521)	0.001 (2.521)	0.580	2.503
		M_2	0.001 (2.485)	0.001 (2.485)	0.754	3.997
		M_3	0.000 (2.471)	0.000 (2.471)	5.303	-11.452
	HH	M_1	0.000 (2.433)	0.000 (2.433)	0.075	0.052
		M_2	0.000 (2.481)	0.000 (2.481)	0.147	0.155
		M_3	0.000 (2.474)	0.000 (2.474)	-2.610	-2.842
	SPB	M_1	-0.000 (2.421)	-0.000 (2.421)	-0.891	2.988
		M_2	-0.000 (2.467)	-0.000 (2.467)	-2.564	1.261
		M_3	-0.000 (2.462)	-0.000 (2.462)	-0.338	3.271
	PS	M_1	0.000 (2.419)	0.000 (2.419)	0.094	-0.046
		M_2	0.000 (2.465)	0.000 (2.465)	0.164	0.070
		M_3	0.000 (2.447)	0.000 (2.447)	0.172	0.157
0.2	CR	M_1	0.100 (2.557)	0.100 (2.557)	2.719	1.091
		M_2	0.102 (2.522)	0.102 (2.522)	9.042	5.063
		M_3	0.102 (2.507)	0.102 (2.507)	19.229	5.648
	HH	M_1	0.100 (2.474)	0.100 (2.474)	-1.014	2.686
		M_2	0.102 (2.522)	0.102 (2.522)	1.594	7.766
		M_3	0.103 (2.513)	0.103 (2.513)	12.659	-1.188
	SPB	M_1	0.099 (2.452)	0.099 (2.452)	0.558	4.340
		M_2	0.101 (2.496)	0.101 (2.496)	4.606	11.852
		M_3	0.101 (2.492)	0.101 (2.492)	13.901	20.145
	PS	M_1	0.099 (2.457)	0.099 (2.457)	-1.250	0.173
		M_2	0.101 (2.502)	0.101 (2.502)	3.473	0.755
		M_3	0.102 (2.482)	0.102 (2.482)	7.619	0.514

solution of $\boldsymbol{\Psi}_N(\boldsymbol{\theta}, \hat{\boldsymbol{\alpha}}) = \mathbf{0}$.

We also observe that, when $\tau = 0.2$, the estimated pseudo-true treatment contrast $\hat{\tau}^*$ is approximately 0.10 across all settings, which is substantially smaller than $\tau = 0.2$. This difference is not caused by the numerical algorithm; rather, it reflects the pseudo-true parameter targeted by the misspecified logistic working model. This is consistent with the bias phenomenon for misspecified GLM discussed by Gail et al. (1984).

S5.4 Summary

Overall, the results in Section S5.1–Section S5.3 suggest that the additional derivative terms $\mathcal{T}_1$ and $\mathcal{T}_2(\boldsymbol{\alpha}^*)$ have limited practical impact on the implementation of GEE for estimating τ^*. First, Proposition 3 identifies several common settings in which these terms are zero. In particular, both terms are zero when the working mean model is correctly specified, and $\mathcal{T}_2(\boldsymbol{\alpha}^*)$ equals zero under working independence. Moreover, under canonical links with working independence, both $\mathcal{T}_1$ and $\mathcal{T}_2(\boldsymbol{\alpha}^*)$ are zero. The term $\mathcal{T}_1$ also equals zero under the Gaussian identity-link working model and under canonical-link models with subject-invariant covariates. These results show that the additional terms are absent in many standard GEE settings.

Second, in settings where these terms are not necessarily zero, their numerical magnitudes remain small relative to $\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)$. For continuous outcomes under the Gaussian identity link, $\mathcal{T}_1 = \mathbf{0}$ analytically, and the relative magnitude of $\mathcal{T}_2(\boldsymbol{\alpha}^*)$ is uniformly below 5% across all working models and working correlation structures. For binary outcomes with subject-varying covariates and misspecified logistic working models, $\mathcal{T}_1$ is not expected to be zero, but its Frobenius norm is on the order of 10^{-18}, making it numerically indistinguishable from zero. The term $\mathcal{T}_2(\boldsymbol{\alpha}^*)$ is nonzero but remains moderate, with relative magnitude ranging from 1.01% to 5.25% and staying below 6% in all cases.

Finally, the direct comparison between Newton–Raphson, which incorporates the exact Hessian including $\mathcal{T}_1$ and $\mathcal{T}_2(\boldsymbol{\alpha}^*)$, and the standard Fisher scoring algorithm implemented by `geese()`, shows virtually identical estimates of τ^*. Across all randomization procedures, working models, and values of τ, the differences in the sample means of $\hat{\tau}^*$ are on the order of 10^{-8}, and the differences in the $\sqrt{N}$-scaled standard deviations are on the order of 10^{-7}. These findings indicate that, although $\mathcal{T}_1$ and $\mathcal{T}_2(\boldsymbol{\alpha}^*)$ appear in the exact derivative of the estimating equation, ignoring them in the Fisher scoring implementation has negligible numerical effect on the resulting estimator in the scenarios considered.

Taken together, these theoretical and numerical results justify the use of the standard Fisher scoring implementation of GEE for computing $\hat{\tau}^*$, while treating the additional derivative terms as having negligible impact on the numerical solution in the settings considered.

S6 Additional Simulation Results

In this section, we present two additional numerical studies to further demonstrate the properties of proposed procedure. In particular, we consider continuous outcome in Section S6.1 and binary outcome in Section S6.2.

For both of the two studies, we assume the sample size $N = 500$ and the following covariates settings. Suppose K_i are *i.i.d.* multinomial random variables taking values $3, 4, 5$ with probabilities 0.2, 0.3, and 0.5, respectively. Suppose there are three baseline covariates such that $\boldsymbol{X}_{i,j}$ are identical across j and $\boldsymbol{X}_{i,j} = \boldsymbol{X}_i = (X_{i,1}, X_{i,2}, X_{i,3})^\top$, and $X_{i,1}$, $X_{i,2}$, $X_{i,3}$ are i.i.d., $2Bernoulli(1/2) - 1$, $N(0, 1)$, and $N(0, 1)$, respectively. For the design stage, $X_{i,1}$ and $X_{i,2}$ are used and $X_{i,2}$ is discretized to $\tilde{X}_{i,2}$ such that $\tilde{X}_{i,2} = 1$ if $X_{i,2} \geq 0$, and $\tilde{X}_{i,2} = 0$ if $X_{i,2} < 0$, yielding a total of $s_{\mathsf{max}} = 4$ strata. Based on $\{X_{i,1}, \tilde{X}_{i,2}\}_{i=1}^{500}$, we consider the following four designs: (1) complete randomization (CR), (2) stratified permuted-block randomization (STRPB), (3) Pocock and Simon's procedure (PS), and (4) Hu and Hu's procedure (HH). For the inference stage, we apply (2.2) with M_1, adjusting no covariates; M_2, adjusting the covariates used in CAR, i.e., $X_{i,1}$ and $X_{i,2}$; and M_3, adjusting $X_{i,1}$, $X_{i,2}$ and $X_{i,3}$. We consider the identity link for the continuous outcome evaluated in Section S6.1, and the logistic link for the binary outcome evaluated in Section S6.2. We also compare the independent (IND), the exchangeable (EXCH) and the AR(1) working correlation matrices for the implementation of GEE. The nominal level of the test is

$c = 0.05$ and all simulation results are evaluated with 10,000 repetitions.

S6.1 Study 1: Continuous Outcome

We consider the following outcome generating model:

$$Y_{i,j} = A_i\gamma_1 + (1 - A_i)\gamma_0 + X_{i,1}\beta_1 + X_{i,2}\beta_2 + X_{i,3}\beta_3 + \epsilon_{i,j}, \tag{S6.1}$$

where $\gamma_0 = 0.5$, $\gamma_1 = \gamma_0 + \tau$, $\beta_1 = \beta_2 = \beta_3 = 1$ for $1 \leq i \leq 500$. Here, $\{\epsilon_{i,j}, 1 \leq j \leq n_i\}_{i=1}^N$ are independent and follow a multivariate normal distribution with mean $\mathbf{0}_{n_i}$ and variance-covariance matrix $\sigma^2\boldsymbol{C}_i(\boldsymbol{\alpha})$, $\sigma^2 = 1.5^2$ and $\boldsymbol{C}_i(\boldsymbol{\alpha})$ is the $n_i \times n_i$ AR(1) correlation matrix with the parameter $\alpha = 0.7$. We report the bias, standard deviation of $\hat{\tau}^*$ when $\tau = 0$ as well as the Type I error rates of T_{Wald} in Table S4. We compare the performance of T_{Wald} and the proposed T_{adj} on controlling the Type I errors in Table S5. The power curves of the two tests with the exchangeable working correlation matrix are evaluated in Figures S1–S3.

Table S4 yields three main conclusions. First, the estimator $\hat{\tau}^*$ is consistent even when some baseline covariates are omitted from the working model under $H_0 : \tau = 0$. The biases of $\hat{\tau}^*$ under models M_1, M_2, and M_3 are all close to zero across different designs and working correlation structures used in GEE. Second, when the covariates used in the CAR procedure

Table S4: The bias (10^{-1}) and standard deviation (s.d.) of $\hat{\tau}^*$, and the Type I error rates (%) of T_{Wald} under $\tau = 0$ for Study 1 (Cor: correlation).

Design	Cor	M_1	M_2	M_3
		$\hat{\tau}^*$:bias($\sqrt{N}$sd) Type I	$\hat{\tau}^*$:bias($\sqrt{N}$sd) Type I	$\hat{\tau}^*$:bias($\sqrt{N}$sd) Type I
CR	IND	0.03(4.27) 4.84	0.02(3.16) 4.82	0.01(2.42) 4.95
	EXCH	0.02(4.22) 5.02	0.02(3.13) 4.90	0.01(2.41) 4.93
	AR(1)	0.02(4.19) 5.04	0.02(3.09) 4.84	0.01(2.36) 4.95
PS	IND	0.00(3.48) 1.63	-0.02(3.20) 5.53	-0.00(2.43) 5.10
	EXCH	-0.01(3.42) 1.52	-0.02(3.16) 5.30	-0.01(2.42) 4.70
	AR(1)	-0.00(3.38) 1.38	-0.02(3.12) 5.18	-0.00(2.36) 4.80
STRPB	IND	-0.00(3.41) 1.42	-0.01(3.16) 4.95	-0.01(2.43) 5.01
	EXCH	-0.01(3.35) 1.37	-0.01(3.13) 4.83	-0.01(2.42) 4.98
	AR(1)	-0.01(3.34) 1.30	-0.01(3.10) 4.94	-0.01(2.36) 5.02
HH	IND	0.01(3.39) 1.33	0.01(3.15) 5.00	0.01(2.45) 5.07
	EXCH	0.01(3.33) 1.25	0.02(3.12) 4.87	0.02(2.44) 5.13
	AR(1)	0.01(3.32) 1.29	0.02(3.09) 4.74	0.01(2.39) 5.29

Table S5: Type I error rates (%) and Powers of T_{Wald} and T_{adj} under $\tau = 0$ and $\tau = 0.4$ for Study 1 (Cor: correlation).

Design	τ	Cor for GEE	M_1		M_2		M_3	
			T_{Wald}	T_{adj}	T_{Wald}	T_{adj}	T_{Wald}	T_{adj}
CR	0	IND	4.84	-	4.82	-	4.95	-
		EXCH	5.02	-	4.90	-	4.93	-
		AR(1)	5.04	-	4.84	-	4.95	-
	0.4	IND	55.68	-	80.85	-	95.87	-
		EXCH	56.52	-	81.77	-	96.10	-
		AR(1)	56.82	-	82.30	-	96.66	-
PS	0	IND	1.63	5.05	5.05	5.10	5.10	5.08
		EXCH	1.52	4.85	5.30	5.20	4.70	4.72
		AR(1)	1.38	5.25	5.17	5.17	4.80	4.67
	0.4	IND	56.35	75.85	80.65	80.70	96.20	96.15
		EXCH	56.55	73.85	80.55	80.50	95.35	95.30
		AR(1)	59.25	78.30	82.85	82.80	97.30	97.30
STRPB	0	IND	1.42	4.73	4.95	4.95	5.01	5.01
		EXCH	1.37	4.71	4.83	4.82	4.98	5.00
		AR(1)	1.30	4.79	4.94	4.92	5.02	5.01
	0.4	IND	56.27	74.15	80.56	80.54	95.73	95.73
		EXCH	57.77	75.27	80.98	80.95	95.95	95.94
		AR(1)	57.78	75.95	81.98	81.93	96.46	96.44
HH	0	IND	1.33	4.88	5.00	5.00	5.07	5.07
		EXCH	1.25	4.97	4.87	4.85	5.13	5.13
		AR(1)	1.29	4.87	4.74	4.72	5.29	5.29
	0.4	IND	56.07	72.98	79.99	80.00	95.41	95.41
		EXCH	57.17	74.58	80.68	80.65	95.61	95.61
		AR(1)	57.67	75.73	81.76	81.76	96.17	96.16

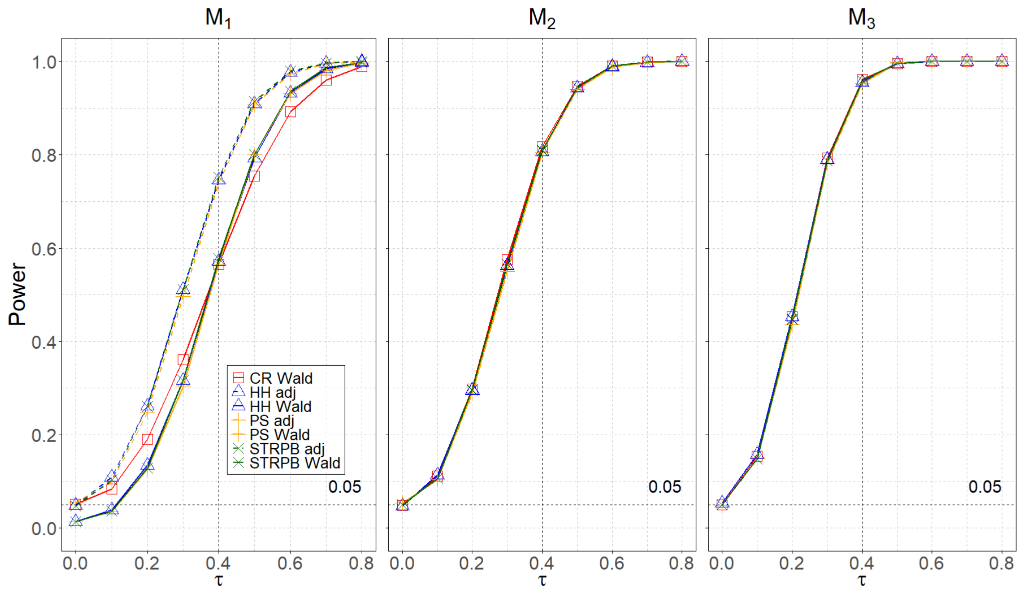

Figure S1: Power curves of T_{Wald} and T_{adj} under the exchangeable working correlation matrix for Study 1.

are not properly adjusted in the working model, the unadjusted Wald test T_{Wald} can exhibit conservative Type I error rates. Specifically, under M_1, the Type I errors of T_{Wald} for the PS, STRPB, and HH designs fall below the nominal level of 5%, whereas they are close to 5% under the working model that adjusts the covariates used in design, i.e., M_2 and M_3. Third, as more influential covariates are included or the working correlation structure is correctly specified as the true data generating process, the efficiency for estimating τ^* improves. In particular, the standard deviations of $\hat{\tau}^*$ under the AR(1) correlation structure are moderately smaller than those obtained using exchangeable or independent working correlations, demonstrating the benefit of GEE-based covariate adjustment for enhancing the efficiency for estimating the treatment effect.

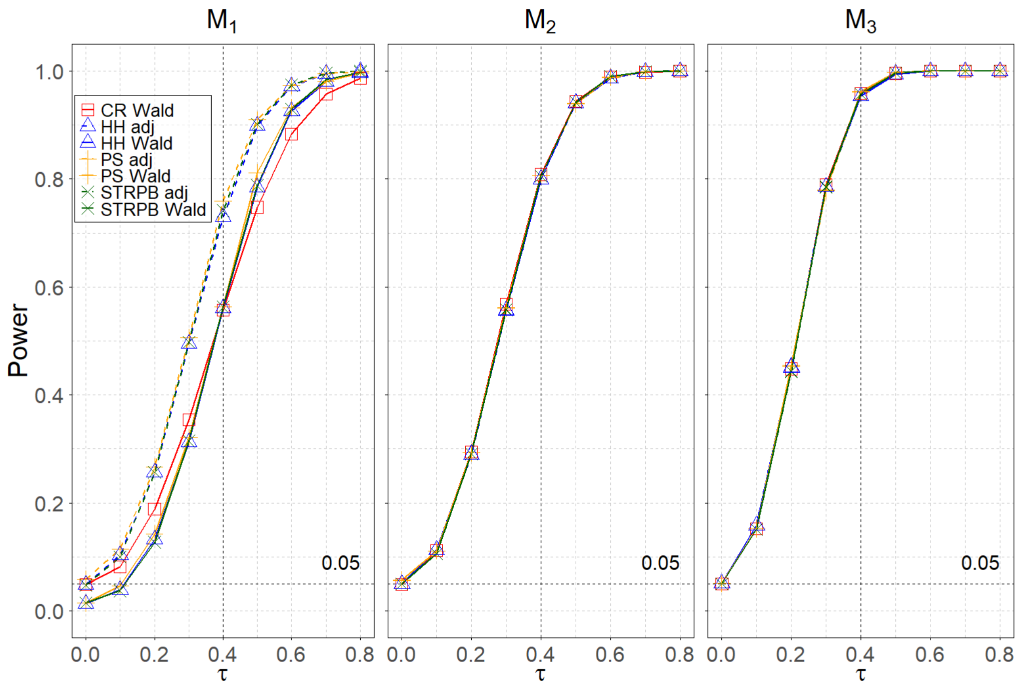

Figure S2: Power curves of T_{Wald} and T_{adj} under the independence working correlation martix for Study 1.

Table S5 demonstrates the desirable properties of T_{adj} compared with T_{Wald}. Since T_{Wald} under M_1 does not include the covariates used in CAR, its Type I error rates are below the nominal level of 5% under the CAR procedures. It is worth noting that T_{Wald} adjusted with M_2 and M_3 can be regarded as model-based adjustments that incorporate the covariates used in the design. In this study, the outcome model (S6.1) is linear and correctly specified by the working model M_3. Consequently, the model-based adjustment performs well: the Type I errors of T_{Wald} with M_2 and M_3 attain the nominal 5% level across the PS, STRPB, and HH. In contrast, T_{adj} explicitly adjusts for the impact of CAR on subsequent inference and maintains nominal Type I error rates regardless of the working model used and the CAR procedure used for design. Consequently, T_{adj} may achieve higher empirical power after correcting the conservativeness of T_{Wald}.

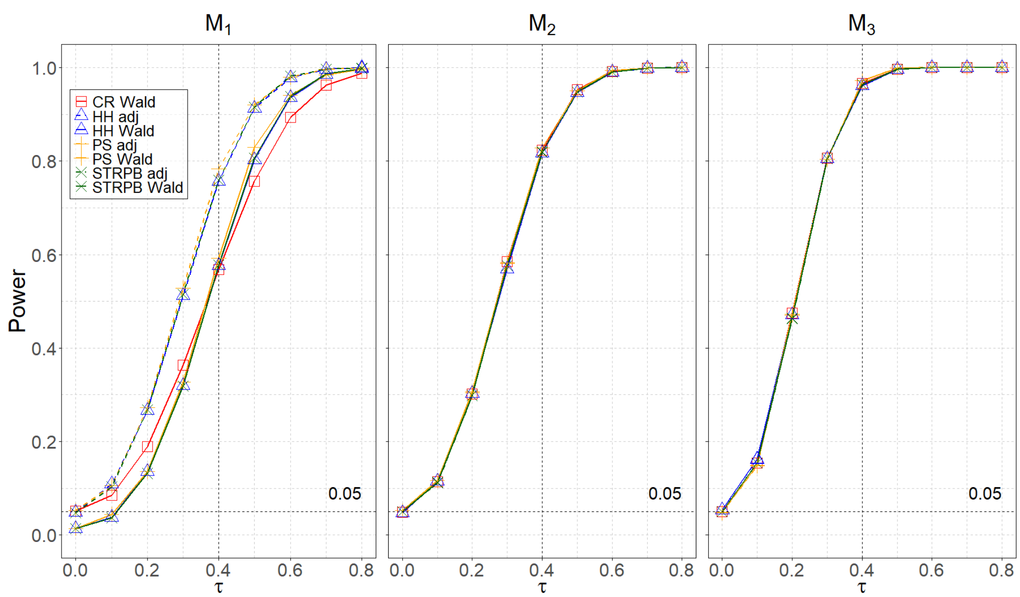

Figure S3: Power curves of T_{Wald} and T_{adj} under the AR(1) working correlation matrix for Study 1.

The results in Table S5 and Figures S1-S3 highlight the advantages of covariate adjustment using GEE. Comparing the powers of T_{adj} across models from M_1 to M_3, we observe that as more influential covariates are incorporated, the power of the tests increases, illustrating the benefit of adjusting for prognostic covariates. Furthermore, incorporating correlation information through GEE further improves efficiency and test power. For example, the power of the adjusted tests increases by approximately 2% when using the AR(1) working correlation structure under M_1 and M_2, and by about 1% under M_3, compared with using an independent working correlation.

S6.2 Study 2: Binary Outcome

Consider the following outcome model:

$$\eta_i = (1 - A_i)\gamma_0 + A_i\gamma_1 + X_{i,1}\beta_1 + X_{i,2}\beta_2 + X_{i,3}\beta_3,$$

$$\text{and} \quad Y_{i,j} = I(\epsilon_{i,j} \leq \eta_i),$$

where $\gamma_0 = 0.5$, $\gamma_1 = \gamma_0 + \tau$, $\beta_1 = \beta_2 = \beta_3 = 1$ and $\{\epsilon_{i,j}, 1 \leq j \leq n_i\}_{i=1}^N$ follows a multivariate distribution with logistic margins and exchangeable correlation matrices with $\alpha = 0.7$ based on the Gaussian copula (Emrich and Piedmonte, 1991b). We report the bias, standard deviation of $\hat{\tau}^*$ when $\tau = 0$ together with the Type I error rates of T_{Wald} in Table S6. We compare the performance of T_{Wald} and the proposed T_{adj} on controlling Type I errors in Table S7. The power curves of the two tests with the AR(1) working correlation matrix are evaluated in Figure S4.

Table S6: The bias(10^{-1}) and standard deviation (s.d.) of $\hat{\tau}^*$, and the Type I error rates (%) of T_{Wald} under $\tau = 0$ (Cor: correlation).

Design	Cor for GEE	M_1		M_2		M_3	
		$\hat{\tau}^*$:bias($\sqrt{N}$sd)	Type I	$\hat{\tau}^*$:bias($\sqrt{N}$sd)	Type I	$\hat{\tau}^*$:bias($\sqrt{N}$sd)	Type I
CR	IND	-0.00(3.74)	5.32	-0.03(4.05)	5.21	-0.02(4.35)	5.20
	EXCH	-0.02(3.69)	5.32	-0.03(3.99)	5.10	-0.01(4.30)	5.17
	AR(1)	-0.02(3.79)	5.30	-0.03(4.12)	5.07	-0.01(4.40)	5.34
PS	IND	-0.01(3.16)	2.09	-0.01(4.03)	5.20	0.03(4.34)	5.46
	EXCH	-0.01(3.14)	2.18	-0.01(4.02)	5.45	0.02(4.36)	5.53
	AR(1)	0.00(3.17)	2.18	-0.00(4.04)	4.80	0.03(4.36)	5.10
SPB	IND	0.00(3.17)	2.25	0.00(4.06)	5.13	-0.01(4.32)	4.99
	EXCH	0.00(3.12)	2.24	-0.00(4.02)	5.25	-0.01(4.28)	4.90
	AR(1)	0.00(3.23)	2.27	-0.00(4.12)	5.47	-0.01(4.37)	4.82
HH	IND	0.01(3.14)	2.10	0.01(4.01)	5.00	0.01(4.31)	4.94
	EXCH	0.01(3.08)	2.06	0.01(4.00)	4.82	0.01(4.27)	5.03
	AR(1)	0.02(3.19)	2.06	0.02(4.06)	4.74	0.01(4.36)	4.98

Table S6 delivers the following three folds of conclusions. First, the GEE estimator $\hat{\tau}^*$ remains consistent under the null hypothesis $H_0 : \tau = 0$, even when certain baseline covariates are excluded from the working logistic model. Across different CAR designs and GEE working correlation structures, the biases of $\hat{\tau}^*$ under models M_1, M_2, and M_3 are consistently close to zero. Second, when the working model fails to account for the covariates incorporated in the CAR procedure, the unadjusted Wald test T_{Wald} tends to yield reduced Type I errors. In particular, under model M_1, the Type I error rates of T_{Wald} for the PS, STRPB, and HH fall below the nominal level 5%, while under M_2 and M_3, they are close to nominal level. Third, incorporating more prognostic covariates or specifying the working correlation structure in accordance with the true data-generating process enhances the efficiency of $\hat{\tau}^*$ estimation. Notably, the standard deviations

Table S7: Type I error rates (%) and Powers of T_{Wald} and T_{adj} under $\tau = 0$ and $\tau = 0.7$ (Cor: correlation).

Design	τ	Cor for GEE	M_1 T_{Wald}	M_1 T_{adj}	M_2 T_{Wald}	M_2 T_{adj}	M_3 T_{Wald}	M_3 T_{adj}
CR	0	IND	5.32	-	5.21	-	5.20	-
		EXCH	5.32	-	5.10	-	5.17	-
		AR(1)	5.30	-	5.07	-	5.34	-
	0.7	IND	81.22	-	92.65	-	97.44	-
		EXCH	82.09	-	92.99	-	97.66	-
		AR(1)	65.22	-	78.88	-	94.10	-
PS	0	IND	2.09	5.40	5.19	5.06	5.46	5.45
		EXCH	2.18	5.28	5.45	5.33	5.53	5.45
		AR(1)	2.18	4.90	4.80	4.73	5.10	5.05
	0.7	IND	86.00	91.50	93.20	93.20	97.65	97.65
		EXCH	86.65	92.40	93.80	93.80	97.85	97.85
		AR(1)	66.55	77.45	78.60	78.55	94.50	94.40
STRPB	0	IND	3.62	4.93	4.89	4.89	4.98	4.99
		EXCH	3.75	5.07	4.89	4.87	5.12	5.10
		AR(1)	3.84	5.17	5.08	5.07	5.09	5.08
	0.7	IND	84.85	90.74	92.80	92.79	97.48	97.49
		EXCH	85.93	91.60	93.33	93.33	97.63	97.62
		AR(1)	67.66	77.76	78.85	78.72	94.27	94.21
HH	0	IND	2.87	4.79	4.45	4.41	4.75	4.75
		EXCH	3.06	4.78	4.62	4.63	4.81	4.80
		AR(1)	3.00	4.87	4.70	4.67	4.65	4.62
	0.7	IND	85.09	90.85	92.94	92.90	97.52	97.51
		EXCH	85.89	91.84	93.45	93.43	97.66	97.63
		AR(1)	68.12	77.90	79.45	79.31	94.35	94.31

of $\hat{\tau}^*$ obtained under the exchangeable structure are moderately smaller than those under independence or AR(1) structures.

Table S7 demonstrates the superior performance of T_{adj} compared with T_{Wald}. Similar to Table S5 in Study 1, T_{Wald} with M_1 has reduced Type I errors. For HH, the Type I error rates of T_{Wald} with M_1 are 2.87%, 3.06%, and 3.00%, all below the nominal 5% level, under the independence, exchangeable, and AR(1) working correlation structures, respectively. In contrast, across various CAR designs and working correlation structures under models from M_1 to M_3, the Type I error rates of T_{adj} remain consistently close to the nominal 5% level. Consequently, T_{adj} attains higher empirical power by correcting the conservativeness in T_{Wald}.

In addition, the power of the test increases as more covariates are included in the adjustment. From Table S7, it can be seen that the powers of T_{adj} under STRPB using AR(1) working correlation structure together with M_1, M_2 and M_3 are 77.76%, 78.72% and 94.21%, respectively, exhibiting a clear increasing trend. A similar pattern can be observed from Figure S4, where, for the same treatment effect $\tau = 0.4$, the power of T_{adj} increases progressively from model M_1 to M_2 and M_3.

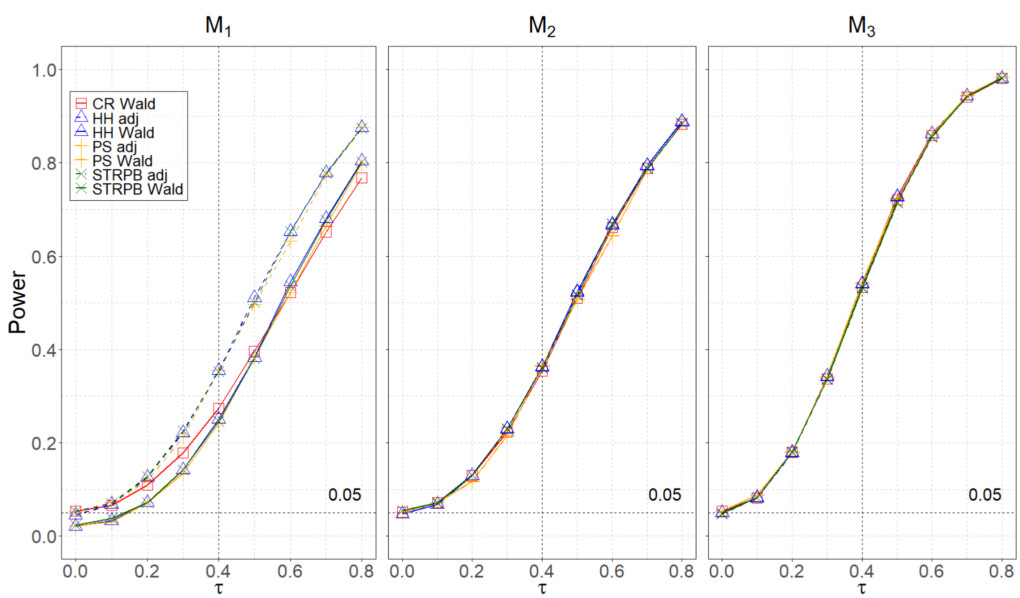

Figure S4: Power curves of T_{Wald} and T_{adj} under the AR(1) working correlation for Study 2.

Lastly, Study 2 suggests that, even when correlation exists, adjusting for correlation using GEE does not necessarily improve the efficiency of treatment effect estimation. In this study, the true correlation structure is exchangeable. As shown in Table S6, the standard deviations of $\hat{\tau}^*$ estimated with the independent working correlation can be smaller than those obtained with the AR(1) working correlation. For instance, in Table S6, under HH and M_3, the asymptotic standard deviation of $\hat{\tau}^*$ with the exchangeable working correlation matrix is 4.27, whereas those with independence and AR(1) structures are 4.31 and 4.36, respectively. However, when the true correlation structure is correctly specified as the working correlation in GEE, the resulting standard deviation of $\hat{\tau}^*$ remains lower than that under the independent working correlation. Meanwhile, the smaller standard deviation of $\hat{\tau}^*$ under the independence structure is accompanied by increased power. For instance, under HH, the power of T_{adj} with M_2 and the independence working correlation structure is 92.90%, which is higher than that under the AR(1) structure (79.31%) but lower than that under the exchangeable structure (93.43%). Among the three, the exchangeable working correlation yields the highest power. For readers interested in this phenomenon, we refer to Wang and Carey (2003) for more technical and numerical details. These findings suggest that the use of GEE for covariate adjustment may require model selection for the working correlation structure prior to conducting inference on the treatment effect. We leave this as an open problem for future research.

In conclusion, the two studies given in this section demonstrate the desirable performance of the proposed T_{adj} for covariate adjustment with GEE for longitudinal CAR trials. The results indicate that T_{adj} can effectively maintain the nominal Type I error rate and achieve higher power across various CAR designs and working correlation structures, further confirming its robustness and reliability when applied to different outcome types.

S7 Supplement to Section 6

The effect of Cocaine and PreHAMD are fitted according to cubic basis spline function. We present the effects of Cocaine and PreHAMD in Figure S5. As can be seen from the figure, the effects of both cocaine use and the level of preHAMD tend to be quadratic or cubic.

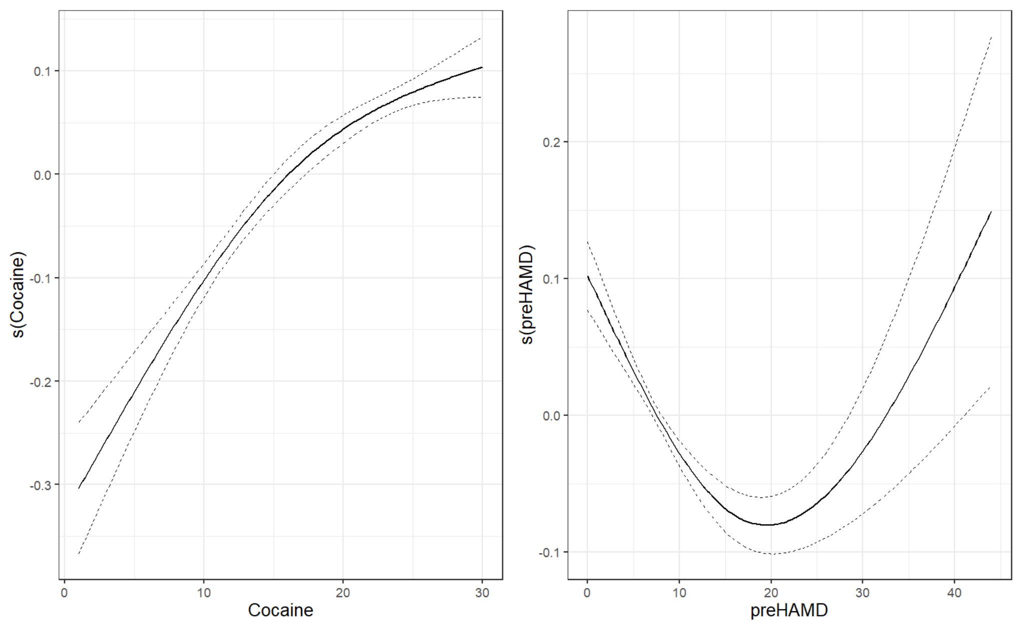

Figure S5: Estimated smooth effects of the covariates under the additive model for the cocaine dependence data.

Bibliography

Beran, R. (1997). Diagnosing bootstrap success. *Annals of the Institute of Statistical Mathematics 49*, 1–24.

Emrich, L. J. and M. R. Piedmonte (1991a). A method for generating high-dimensional multivariate binary variates. *The American Statistician 45*(4), 302–304.

Emrich, L. J. and M. R. Piedmonte (1991b). A method for generating high-dimensional multivariate binary variates. *The American Statistician 45*(4), 302–304.

FDA (2022). Master protocols: Efficient clinical trial design strategies to expedite development of oncology drugs and biologics: Guidance for industry. Guidance, Oncology Center of Excellence (OCE), CDER, and CBER, FDA.

Gail, M. H., S. Wieand, and S. Piantadosi (1984). Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika 71*(3), 431–444.

Guo, Z., Y. Liu, and L. Xia (2025). Covariate selection for optimizing balance with an innovative adaptive randomization approach. *Statistical Methods in Medical Research*, 09622802241313283.

Hall, P. and C. C. Heyde (1980). *Martingale limit theory and its application* (1st ed.). New York: Academic press.

ICH (1998). Ich harmonised tripartite guideline e9: Statistical principles for clinical trials. Guideline, International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH).

Liang, K.-Y. and S. L. Zeger (1986). Longitudinal data analysis using generalized linear models. *Biometrika 73*(1), 13–22.

Liu, Y. and F. Hu (2022). Balancing unobserved covariates with covariate-adaptive randomized experiments. *Journal of the American Statistical Association 117*(538), 875–886.

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

Liu, Y., L. Xia, and F. Hu (2025). Testing heterogeneous treatment effect with quantile regression under covariate-adaptive randomization. *Journal of Econometrics 249*, 105808.

McCullagh, P. and N. J. A. (2019). *Generalized linear models* (2nd ed.). New York: Routledge.

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge, UK: Cambridge University Press.

Wang, Y.-G. and V. Carey (2003). Working correlation structure misspecification, estimation and covariate design: implications for generalised estimating equations performance. *Biometrika 90*(1), 29–41.