Statistica Sinica Preprint No: SS-2025-0428	
Title	Statistical Inference with Generalized Estimating Equations in Longitudinal Studies Following Covariate-adaptive Randomization
Manuscript ID	SS-2025-0428
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0428
Complete List of Authors	Ruichen Fu and Yang Liu
Corresponding Authors	Yang Liu
E-mails	yangliu2022@ruc.edu.cn
Notice: Accepted author version.	

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**

Abstract: Longitudinal outcomes are often involved in covariate-adaptive randomized (CAR) trials. However, most existing studies disregard the longitudinal structure and focus on a cross-sectional analysis. In this article, we utilize generalized estimating equations (GEE) to conduct inference for the marginal treatment effect following CAR. In addition, we introduce an additional stratification variable to allow for the analysis with unbalanced longitudinal data. We show that the treatment estimate is consistent under the null hypothesis, even when omitted covariates are involved in the analysis. This enables us to construct tests that preserve the Type I error rate even when the working model used for analysis is misspecified. By deriving the asymptotic normality of the treatment effect estimator, we show that the robust property of GEE still holds under CAR designs, but the balance of covariates used in design and the additional stratification affects the classical Wald tests, yielding reduced Type I error. To remedy this issue, we propose an adjustment procedure and show that the

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

adjustment recovers the nominal significance level of the test for most commonly used CAR procedures. Extensive simulation studies corroborate our theoretical findings and highlight the practical advantages of the proposed methods in terms of both estimation efficiency and inferential validity.

Key words and phrases: Covariate-adaptive randomization; longitudinal data; generalized estimating equations; conservative tests; bootstrap adjustment

1. Introduction

Covariate-adaptive randomization (CAR) is widely used in clinical trials to achieve balanced treatment allocation with respect to important baseline covariates (Rosenberger et al., 2008). Typically, patients are stratified according to key prognostic factors, and the patients within each stratum are assigned with permuted-block or biased-coin designs to achieve balance (Zelen, 1974; Baldi Antognini and Zagoraiou, 2011). When the number of strata is large relative to the total sample size, it may be more efficient to achieve covariate balance based on the marginal distributions of covariates (Pocock and Simon, 1975) or through a weighted combination of multiple imbalance measures (Hu and Hu, 2012). For a comprehensive overview of recent developments in CAR, see Hu et al. (2014), Ma et al. (2024), Liu et al. (2025), and the references therein.

The broad application has spurred substantial progress in the develop-

ment of valid and efficient inference procedures for CAR designs. Within the linear model framework, Shao et al. (2010), Ma et al. (2015), Ma et al. (2020) elucidate how the covariate balance achieved by CAR can reduce the Type I error of the tests for treatment effect and further propose adjustment methods to restore nominal significance levels. Subsequent research has extended these results to generalized linear models (Li et al., 2019), survival analysis (Ye and Shao, 2020), and quantile regression (Liu et al., 2025), among others. To mitigate concerns about model misspecification, model-robust inference methods have been developed under both linear models (Bugni et al., 2018; Ma et al., 2022; Liu et al., 2022; Ye et al., 2024) and generalized linear models (Wang and Carey, 2003). However, these existing methods primarily focus on settings with a single outcome measurement per subject. In contrast, valid inference for longitudinal studies, where the responses are repeatedly measured over time, remains largely unexplored and warrants further methodological investigation.

In many therapeutic areas, including diabetes and psychiatry, longitudinal endpoints are frequently employed when disease progression is gradual and long-term follow-up is essential. Longitudinal studies enable the assessment of sustained treatment efficacy and provide clinically meaningful insights into both the timing and durability of treatment effects.

Among various analytical frameworks, the generalized estimating equation (GEE; Liang and Zeger, 1986) has been widely adopted in longitudinal trials (Wang, 2014). The advantage of GEE lies in its ability to consistently estimate marginal treatment effects even when the within-subject covariance structure is misspecified. Nevertheless, conventional GEE-based analyses typically overlook the influence of the randomization procedure. Because the covariate balance achieved by CAR can alter the distribution of test statistics used for inference (Ma et al., 2015, 2022; Li et al., 2019; Liu and Hu, 2023), understanding the impact of CAR on GEE-based inference and developing valid inferential methods for longitudinal CAR trials remain pressing research needs.

Unlike the setting with independent and identically distributed observations, inference for GEE estimators in longitudinal studies becomes considerably more complex from two main aspects. First, the GEE explicitly incorporates within-subject correlation into its estimating equations, which invalidates the desirable property of standard Z-estimation based methods used for treatment effect estimation, namely, the consistent estimation of the average treatment effect derived from the score equation (Wang et al., 2023). Moreover, when a model-based approach is adopted, the estimated treatment effect can be sensitive to the covariates used in

the working model, particularly under nonlinear models (Gail et al., 1984). It is therefore important to identify conditions under which consistent estimation of the treatment effect can still be ensured with GEE. Second, the impact of unequal numbers of repeated measurements across subjects, together with the induced correlation structure and the balancing property of CAR, on the asymptotic distribution of estimators remains poorly understood. Therefore, how to develop an appropriate variance estimate to correct the impact of CAR is not clear.

In this article, we study inference for the marginal treatment effect based on GEE and longitudinal CAR trials. To accommodate the repeated measure structure, we introduce an additional stratification variable that captures heterogeneity in the working correlation matrices derived from unbalanced longitudinal data. Within this framework, we derive the asymptotic properties of the GEE-adjusted treatment effect estimator for a class of commonly used CAR procedures. We show that the estimator remains consistent under the null hypothesis even when certain important covariates are omitted from the adjustment model. Moreover, we demonstrate that the standard GEE-based Wald test can be conservative in terms of reduced Type I errors. To address this issue, we incorporate the newly introduced stratification variable to construct an adjusted test statistic and establish

the theoretical properties of the proposed adjusted test. Simulation studies and a real clinical trial example illustrate that the proposed adjusted test effectively controls the Type I error rate while achieving higher power.

The rest of the paper is organized as follows. Section 2 introduces the proposed framework for testing the treatment effect adjusted by GEE under CAR designs. The theoretical properties of the GEE-adjusted treatment effect estimator are developed in Section 3. Based on these results, Section 4 presents our proposed adjustment to address the issue of the conservative Type I error rates under GEE and CAR. Numerical studies and a real clinical trial example are provided in Sections 5 and 6, respectively. Finally, concluding remarks are given in Section 7. Proofs and additional simulation results are presented in the supplementary material.

2. Notation and Framework

Consider a trial with N patients to compare two treatments. Let A_i denote the treatment assignment for the i-th patient, i.e., $A_i = 1$ if the i-th patient is assigned to the treatment and $A_i = 0$ otherwise. Then, $N_1 = \sum_{i=1}^{N} A_i$ and $N_0 = N - N_1$ are the numbers of patients in the treatment and control groups, respectively. In the design stage, the patients are often stratified according to their baseline factors. Suppose there are $s_{\max}$ strata

used for the design with CAR. Let S_i denote the random variable representing the stratum of the i-th patient, such that S_i takes values in $1, \ldots, s_{\mathsf{max}}$. Additional details on the construction of S_i are provided in Section S2.1 of the Supplementary Materials. Then the associated within-stratum imbalance is $D_N(s,\cdot) = \sum_{i=1}^{N}(2A_i - 1)I\{S_i = s\}$. Typically, stratified permuted block randomization (STRPB, Zelen, 1974) or minimization-based methods such as Hu and Hu (2012)'s procedure (HH) and Pocock and Simon (1975)'s procedure (PS) are used to mitigate $\{D_N(s,\cdot), 1 \le s \le s_{\mathsf{max}}\}$ based on different criteria. For a comprehensive overview of these procedures, readers are referred to Hu et al. (2014), and Rosenberger and Lachin (2015) and the references therein.

For $1 \le i \le N$, suppose the i-th patient has n_i observations, for $1 \le j \le n_i$, let $\boldsymbol{X}_{i,j} = (X_{i,j,1}, \cdots, X_{i,j,p})^\top$ be a p-vector of baseline covariates and $Y_{i,j}(a)$ represent the potential outcome of the i-th patient at the j-th observation for $a \in \{0,1\}$ and $1 \le j \le n_i$, so that the observed outcome is $Y_{i,j} = A_i Y_{i,j}(1) + (1 - A_i)Y_{i,j}(0)$. As clinical studies often involve various types of outcomes, we consider the following outcome model:

$$\mu_{i,j} = \mathbf{E}[Y_{i,j} \mid A_i, \boldsymbol{X}_{i,j}] = A_i \mu(\eta^1_{i,j}) + (1 - A_i)\mu(\eta^0_{i,j}), \qquad (2.1)$$

$$\eta^a_{i,j} = \gamma_a + \boldsymbol{X}^\top_{i,j}\boldsymbol{\beta}, \qquad a \in \{0,1\},$$

where γ_1, γ_0, and $\boldsymbol{\beta}$ represent the effects of treatment, control, and covariates, respectively. Here, $\mu(\cdot)$ denotes the unknown true link function, and $(\gamma_1, \gamma_0, \boldsymbol{\beta}^\top)^\top$ is the parameter vector associated with the true outcome model (2.1). The goal of the experiment is to conduct inference for the conditional treatment effect on the transformed scale, i.e., $\tau = \gamma_1 - \gamma_0$. When the link function is the identity, τ coincides with the average treatment effect (ATE). For a general link function $\mu(\cdot)$, however, the two quantities differ and target distinct comparisons. Whereas the ATE compares the mean potential outcomes, τ is defined on the linear predictor scale and therefore does not depend on the realized covariate values or on the specific form of $\mu(\cdot)$. For instance, τ corresponds to the conditional log-odds ratio comparing $A_i = 1$ with $A_i = 0$ for the binary outcomes with the logit link, i.e., $\mu(x) = \log\{x/(1-x)\}$ and

$$\begin{aligned}
&\log\left\{\frac{\mathbf{P}[Y_{i,j}=1|A_i=1,\boldsymbol{X}_{i,k}]}{1-\mathbf{P}[Y_{i,j}=1|A_i=1,\boldsymbol{X}_{i,k}]}\right\} - \log\left\{\frac{\mathbf{P}[Y_{i,j}=1|A_i=0,\boldsymbol{X}_{i,k}]}{1-\mathbf{P}[Y_{i,j}=1|A_i=0,\boldsymbol{X}_{i,k}]}\right\} \\
&= \log\left\{\frac{\mathbf{E}[Y_{i,j}|A_i=1,\boldsymbol{X}_{i,k}]}{1-\mathbf{E}[Y_{i,j}|A_i=1,\boldsymbol{X}_{i,k}]}\right\} - \log\left\{\frac{\mathbf{E}[Y_{i,j}|A_i=0,\boldsymbol{X}_{i,k}]}{1-\mathbf{E}[Y_{i,j}|A_i=0,\boldsymbol{X}_{i,k}]}\right\} \\
&= \eta^1_{i,j} - \eta^0_{i,j} = \gamma_1 - \gamma_0.
\end{aligned}$$

Consequently, it provides a clear and interpretable measure for comparing treatment effects across treatment arms (see p. 6 of FDA, 2023) and is a commonly adopted estimand in clinical studies (see applications such as

Claggett et al., 2022; Goligher et al., 2023; Glantz et al., 2024).

In practice, the inference for τ often involves certain model misspecification. When adjusting the estimate of treatment effect, it is common that some or even no covariates are involved in the analysis (Gail et al., 1984). Furthermore, the link function used for the analysis could be different from $\mu(\cdot)$, e.g., use logistic link function while the true link function is probit link function. We thus assume that the link function used for the analysis and the covariates used in the working model are allowed to differ from the true data generating model (2.1). Denote the p_0-covariates used in the working model as $\boldsymbol{X}_{i,j,\mathsf{in}}$ and the rest $(p - p_0)$ covariates as $\boldsymbol{X}_{i,j,-\mathsf{in}}$, such that $\boldsymbol{X}_{i,j} = (\boldsymbol{X}_{i,j,\mathsf{in}}^\top, \boldsymbol{X}_{i,j,-\mathsf{in}}^\top)^\top$ up to a permutation of the covariates without loss of generality; similarly, we rewrite $\boldsymbol{\beta} = (\boldsymbol{\beta}_{\mathsf{in}}^\top, \boldsymbol{\beta}_{-\mathsf{in}}^\top)^\top$, where $\boldsymbol{\beta}_{\mathsf{in}}$ and $\boldsymbol{\beta}_{-\mathsf{in}}$ correspond to the effects of $\boldsymbol{X}_{i,j,\mathsf{in}}$ and $\boldsymbol{X}_{i,j,-\mathsf{in}}$, respectively. For $a \in \{0, 1\}$, let $\boldsymbol{Y}_i(a) = (Y_{i,1}(a), \ldots, Y_{i,n_i}(a))^\top$ and $\boldsymbol{g}_{i,j,a} = (a, 1-a, \boldsymbol{X}_{i,j,\mathsf{in}}^\top)^\top$, then we can write $\boldsymbol{Y}_i = A_i \boldsymbol{Y}_i(1) + (1 - A_i)\boldsymbol{Y}_i(0)$ and the working model used for adjustment is

$$\mathbf{E}[\boldsymbol{Y}_i | A_i, \mathbf{X}_{i,\mathsf{in}}] \sim \boldsymbol{m}_i = A_i \boldsymbol{m}_i(1) + (1 - A_i)\boldsymbol{m}_i(0), \tag{2.2}$$

where $\mathbf{X}_{i,\mathsf{in}} = (\boldsymbol{X}_{i,1,\mathsf{in}}, \ldots, \boldsymbol{X}_{i,n_i,\mathsf{in}})$, $\boldsymbol{m}_i(a) = (m_{i,1}(a), \ldots, m_{i,n_i}(a))^\top$, $m_{i,j}(a)$ $= m(\eta)|_{\eta=\eta_{i,j}^{*a}}$, $\eta_{i,j}^{*a} = \boldsymbol{g}_{i,j,a}^\top \boldsymbol{\theta}_{\mathsf{in}}^*$, and $\boldsymbol{\theta}^* = (\gamma_1^*, \gamma_0^*, \boldsymbol{\beta}_{\mathsf{in}}^{*\top})^\top$ for $a \in \{0, 1\}$. Here,

$m(\cdot)$ is the given link function used to estimate $\boldsymbol{\theta}^*$ and may be different from $\mu(\cdot)$. The $*$ on the superscripts indicates that $\boldsymbol{\theta}^*$ and $\eta_{i,j}^{*a}$ may be different from the true values $\boldsymbol{\theta} = (\gamma_1, \gamma_0, \boldsymbol{\beta}_{\mathsf{in}}^\top)^\top$ and $\eta_{i,j}^a$, due to the previously mentioned two sorts of model misspecification. For $a \in \{0,1\}$, let $m'_{i,j}(a) = \frac{d}{d\eta} m(\eta)|_{\eta=\eta_{i,j}^{*a}}$ 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 $\widetilde{\boldsymbol{M}}_i(a) = \text{diag}\{m'_{i,1}(a), \ldots, m'_{i,n_i}(a)\}$, $\boldsymbol{G}_{i,a} = (\boldsymbol{g}_{i,1,a}, \ldots, \boldsymbol{g}_{i,n_i,a})^\top$, and $\boldsymbol{\Lambda}_{i,a} = \text{diag}\{v_{i,1}(a), \ldots, v_{i,n_i}(a)\}$, then the GEE (Liang and Zeger, 1986) used for estimating $\boldsymbol{\theta}^*$ is

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

where $\boldsymbol{r}_i(a) = \boldsymbol{Y}_i(a) - \boldsymbol{m}_i(a)$, $\boldsymbol{L}_{i,a}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}) = \boldsymbol{V}_{i,a}^{-1}(\boldsymbol{\alpha})\widetilde{\boldsymbol{M}}_i(a)\boldsymbol{G}_{i,a}$, $\boldsymbol{V}_{i,a}(\boldsymbol{\alpha}) = \boldsymbol{\Lambda}_{i,a}^{1/2}\boldsymbol{C}_i(\boldsymbol{\alpha})\boldsymbol{\Lambda}_{i,a}^{1/2}$ for $a \in \{0,1\}$, and $\boldsymbol{V}_i(\boldsymbol{\alpha}) = A_i\boldsymbol{V}_{i,1}(\boldsymbol{\alpha}) + (1-A_i)\boldsymbol{V}_{i,0}(\boldsymbol{\alpha})$. Here, $\boldsymbol{C}_i(\boldsymbol{\alpha})$ is the working correlation matrix used to enhance the estimation of $\boldsymbol{\theta}^*$ and $\boldsymbol{\alpha}$ is a q-vector of additional parameters used for estimating the correlation matrix. For instance, the independent, exchangeable, and AR(1) working correlation matrices can be written as $\boldsymbol{C}^k(0) = \boldsymbol{I}_{n^k}$, $\boldsymbol{C}^k(\alpha) = (1-\alpha)\boldsymbol{I}_{n^k} + \alpha\mathbf{1}_{n^k}\mathbf{1}_{n^k}^\top$, and $[\boldsymbol{C}^k(\alpha)]_{j_1,j_2} = \alpha^{|j_2-j_1|\dagger}$, where j_1, j_2 represent two observation times. We denote the estimate of $\boldsymbol{\theta}^*$ from (2.3) as

$^\dagger[\boldsymbol{B}]_{j,j}$ represents the (j,j)-th element of the matrix $\boldsymbol{B}$ and $\boldsymbol{I}_p$ is the $p \times p$ identity matrix.

$\hat{\boldsymbol{\theta}}^* = (\hat{\gamma}_1^*, \hat{\gamma}_0^*, \hat{\boldsymbol{\beta}}_{\mathsf{in}}^{*\top})^\top$ and the corresponding estimate for τ^* as $\hat{\tau}^* = \hat{\gamma}_1^* - \hat{\gamma}_0^*$.

To address the issue of unbalanced data in longitudinal studies, we introduce an additional stratification variable, denoted by K_i, defined according to the values of $\{n_i, \boldsymbol{C}_i(\boldsymbol{\alpha})\}$. Specifically, if $K_i = k$ for $1 \leq k \leq k_{\mathsf{max}}$ and $1 \leq i \leq N$, then subject i satisfies $n_i = n^k, \boldsymbol{C}_i(\boldsymbol{\alpha}) = \boldsymbol{C}^k(\boldsymbol{\alpha})$. Under this formulation, balanced longitudinal studies, where $n_i = n$ for all $1 \leq i \leq N$, arise as a special case, since $\boldsymbol{C}_i(\boldsymbol{\alpha}) = \boldsymbol{C}(\boldsymbol{\alpha})$ and $K_i = 1$ with probability one. For unbalanced longitudinal studies, suppose, for example, that $\{n_i\}_{i=1}^N$ takes values in $\{3, 5\}$. Under the independence or exchangeable working correlation structures, we then have $k_{\mathsf{max}} = 2$, with $n^1 = 3$ and $n^2 = 5$. For other working correlation structures, the value of k_{max} may additionally depend on the form of $\boldsymbol{C}^k(\boldsymbol{\alpha})$. For a detailed discussion of the use of K_i and the impact of potential missing or irregular visits, see Section S2.2 of the Supplementary Materials. We note that the proposed formulation is restricted to settings where k_{max} remains relatively small compared with N. Nevertheless, it provides an initial framework for addressing the challenges arising from unbalanced numbers of observations in longitudinal studies under CAR. We assume (S_i, K_i) are i.i.d. as (S, K), the joint probability mass of (S, K) is $\pi_{s,k} = \mathbf{P}\{S_i = s, K_i = k\}$; the marginal probabilities with respect to S and K are $\pi_{s,\cdot} = \mathbf{P}\{S_i = s\}$ and $\pi_{\cdot,k} = \mathbf{P}\{K_i = k\}$ for

$1 \leq s \leq s_{\mathsf{max}}$ and $1 \leq k \leq k_{\mathsf{max}}$, respectively; and the conditional probability of $K = k$ given $S = s$ is denoted by $\pi_{k|s} = \pi_{s,\cdot}^{-1}\pi_{s,k}$ for $1 \leq s \leq s_{\mathsf{max}}$ and $1 \leq k \leq k_{\mathsf{max}}$.

For the inference of τ, we consider the hypothesis $H_0 : \tau = 0$ against $H_1 : \tau \neq 0$ with the Wald test statistic $T_{\mathsf{Wald}} = \mathsf{se}_{\mathsf{Wald}}^{-1}(\hat{\tau}^*)\hat{\tau}^*$, where $\mathsf{se}_{\mathsf{Wald}}(\hat{\tau}^*)$ is the model-based sandwich standard error and its explicit form is given in the supplementary materials. It is worth noting that the unadjusted test statistics or the omitted covariates analysis (Gail et al., 1984, 1988) can be considered as a special case of T_{Wald} with (2.2) by setting $\boldsymbol{X}_{i,j} = \boldsymbol{X}_{i,j,-\mathsf{in}}$, $\boldsymbol{\beta} = \boldsymbol{\beta}_{-\mathsf{in}}$ and $\boldsymbol{\theta}^* = (\gamma_1^*, \gamma_0^*)^\top$. Unlike the balanced longitudinal study, the property of T_{Wald} depends on the joint imbalance with respect to $\{S_i, K_i\}_{i=1}^N$, denoted by $D_N(s,k) = \sum_{i=1}^N d_i(s,k)$ and $d_i(s,k) = (2A_i - 1)I\{S_i = s, K_i = k\}$ for $1 \leq s \leq s_{\mathsf{max}}$ and $1 \leq k \leq k_{\mathsf{max}}$, as will be shown in Section 3. Furthermore, T_{Wald} may be subject to the misspecified link function as well as the issue of omitted covariates. We demonstrate the issue of the inference with T_{Wald} in the following section.

3. Theoretical Results

We begin with introducing the assumptions used for the development of our theoretical results.

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* $1 \le k \le k_{\max}$ *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* $\{\boldsymbol{V}_{i,a}(\boldsymbol{\alpha}), \boldsymbol{G}_{i,1}, \boldsymbol{G}_{i,0}, \boldsymbol{L}_{i,1}(\boldsymbol{\theta}, \boldsymbol{\alpha}), \boldsymbol{L}_{i,0}(\boldsymbol{\theta}, \boldsymbol{\alpha}), \widetilde{\boldsymbol{M}}_i(1), \widetilde{\boldsymbol{M}}_i(0), \boldsymbol{r}_i(1), \boldsymbol{r}_i(0)\}$ *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* $1 \le k \le k_{\max}$.

5. *Let*

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

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

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

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

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

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

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* $1 \le s \le s_{\max}$ *and* $1 \le k \le k_{\max}$, *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)* $\{d_i(s,k), 1 \le s \le s_{\max}$ *and* $1 \le k \le k_{\max}\}_{i=1}^{N}$ *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* $1 \le s_1, s_2 \le s_{\max}$ *and* $1 \le k_1, k_2 \le k_{\max}$

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

Assumptions 3.1 and 3.2 are introduced to describe the data-generating process and to analyze the impact of CAR. Specifically, Assumption 3.1 allows us to evaluate the properties of the GEE estimator within a Z-estimation framework. This assumption can be viewed as an extension of the usual i.i.d. framework by incorporating an additional stratification

variable K, which accounts for the unbalanced number of observations in longitudinal studies.

In addition, Assumption 3.1.5 ensures the local identifiability of the pseudo-true parameter $\boldsymbol{\theta}^*$ under model misspecification. Under misspecification, $\frac{\partial}{\partial \boldsymbol{\theta}}\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\theta}) = -\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*) + \mathcal{T}_1 + \mathcal{T}_2(\boldsymbol{\alpha})\frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{\theta}}$, where $\mathcal{T}_1$ and $\mathcal{T}_2(\boldsymbol{\alpha})$ are the additional terms arising from model misspecification. When $\mathcal{T}_1 = \mathbf{0}$ and $\mathcal{T}_2(\boldsymbol{\alpha}) = \mathbf{0}$, the positive definiteness of $\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)$ guarantees that (2.3) admits a unique local root and ensures the $\sqrt{N}$-consistency of $\hat{\boldsymbol{\theta}}^*$.

Under misspecification, however, $\mathcal{T}_1$ and $\mathcal{T}_2(\boldsymbol{\alpha})$ may not vanish exactly. Nevertheless, the Fisher-scoring used in standard GEE typically ignores these terms (Liang and Zeger, 1986). To assess the validity of this approximation, Section S5 of the Supplementary Materials first presents several common settings under which $\mathcal{T}_1 = \mathbf{0}$ or $\mathcal{T}_2(\boldsymbol{\alpha}) = \mathbf{0}$ holds, and then evaluates the relative magnitudes of $\mathcal{T}_1$ and $\mathcal{T}_2(\boldsymbol{\alpha})$ compared with $\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)$ under misspecified models. We further compare the estimation of τ obtained from the standard GEE with that obtained from the algorithm incorporating $\mathcal{T}_1$ and $\mathcal{T}_2(\boldsymbol{\alpha})$. The numerical results indicate that $\mathcal{T}_1$ and $\mathcal{T}_2(\boldsymbol{\alpha})$ are typically small relative to $\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)$ and have limited impact on the estimation of τ^*. Therefore, following Liang and Zeger (1986), we retain the approximation $\mathcal{T}_1 = \mathbf{0}$ and $\mathcal{T}_2(\boldsymbol{\alpha}) = \mathbf{0}$ in our theoretical development.

Since the properties of $\hat{\tau}^*$ also depend on the distribution $D_N(s,k)$ under a given CAR procedure, Assumption 3.2 characterizes the behavior of $\{D_N(s,k) : 1 \le s \le s_{\max}, 1 \le k \le k_{\max}\}$ for a broad class of CAR designs. Because $\{K_i\}_{i=1}^N$ are not used in the design stage, they can be treated as unobserved covariates, as discussed in Liu and Hu (2022). Consequently, Assumption 3.2 holds for randomization procedures such as complete randomization (CR), stratified permuted-block randomization (STRPB), Pocock and Simon's procedure (PS), and Hu and Hu's procedure (HH). Importantly, this assumption enables us to propose a valid variance estimator for $\hat{\tau}^*$ using the bootstrap method (Liu and Hu, 2023), as detailed in Section 4. We begin by studying the properties of $\hat{\boldsymbol{\theta}}^*$ in the following theorem.

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*

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

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

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

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

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

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

The asymptotic distribution given in Theorem 3.1 demonstrates the impact of CAR on $\hat{\boldsymbol{\theta}}^*$. In particular, the difference $\Sigma_{\mathsf{CR}}(\boldsymbol{\theta}^*,\boldsymbol{\alpha}^*)-\Sigma_{\mathsf{CAR}}(\boldsymbol{\theta}^*,\boldsymbol{\alpha}^*)$ describes the impact of CAR on the asymptotic variance of $\hat{\boldsymbol{\theta}}^*$. 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 $\gamma\{(s_1,k_1),(s_2,k_2)\}$ admits closed-form expressions. We summarize 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 $1\le s_1,s_2\le s_{\mathsf{max}}$ *and* $1\le k_1,k_2\le k_{\mathsf{max}}$ *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

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

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

$$\Sigma_{\mathsf{CR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) - \Sigma_{\mathsf{CAR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) = \mathrm{V}[\mathbf{E}[\boldsymbol{R}(S, K)|S]]$$

is positive definite, unless $\mathbf{E}[\boldsymbol{R}(S, K)|S]$ *is a constant of* S.

Corollary 3.1, together with Theorem 3.1, leads to the following threefold conclusions. First, Corollary 3.1.1 indicates that for CR, $\Sigma_{\mathsf{CR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) = \Sigma_{\mathsf{CAR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)$, which implies $\Xi_{\Psi\Psi}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) = \Sigma_{\Psi\Psi}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)$, i.e., the asymptotic variance of $\hat{\boldsymbol{\theta}}^*$ following CR is

$$2\left\{\widetilde{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)\right\}^{-1} \Sigma_{\Psi\Psi}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) \left\{\left[\widetilde{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)\right]^{-1}\right\}^{\top}.$$

Second, for CAR procedures satisfying condition (3.1), Corollary 3.1.2 suggests that

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

which is positive definite. This demonstrates the variance reduction of $\hat{\boldsymbol{\theta}}^*$ attributable to the improved covariate balance under the procedures satisfying (3.1). Note that this condition holds for STRPB, HH, and stratified biased coin randomization (Baldi Antognini and Zagoraiou, 2011), thereby providing closed-form asymptotic variances of $\hat{\boldsymbol{\theta}}^*$ under these designs. Lastly, PS is commonly used in practice. However, it is well known

that $\gamma\big((s_1,k_1),(s_2,k_2)\big)$ does not admit a closed form (Pocock and Simon, 1975; Hu et al., 2023). When the biased-coin probability used in PS is sufficiently large, PS typically achieves finer balance than CR (Zhao et al., 2024). Therefore, PS may also yield a positive definite value of $\Sigma_{\Psi\Psi}(\boldsymbol{\theta}^*,\boldsymbol{\alpha}^*)-\Xi_{\Psi\Psi}(\boldsymbol{\theta}^*,\boldsymbol{\alpha}^*)$.

It is worth noting that the performance of $\hat{\boldsymbol{\theta}}^*$ also depends on the covariate adjustment with the working model (2.2). First, the influence of covariate balance on the asymptotic variance of $\hat{\boldsymbol{\theta}}^*$ is alleviated when the covariates used in the design—particularly the most prognostic baseline covariates—are also included in the working model. For instance, when $\boldsymbol{R}(s,k)=\mathbf{0}_{p_0+2}$ for all $1\le s\le s_{\mathsf{max}}$ and $1\le k\le k_{\mathsf{max}}$, we have $\Xi_{\Psi\Psi}(\boldsymbol{\theta}^*,\boldsymbol{\alpha}^*)=\Sigma_{\Psi\Psi}(\boldsymbol{\theta}^*,\boldsymbol{\alpha}^*)$, implying that $\hat{\boldsymbol{\theta}}^*$ has the same asymptotic variance under all randomization procedures. This condition holds when the working model is fully correctly specified, i.e., the link function satisfies $m(\cdot)=\mu(\cdot)$ and all baseline covariates are adjusted i.e., $\boldsymbol{X}_{i,j}=\boldsymbol{X}_{i,j,\mathsf{in}}$. Moreover, adjusting for more influential baseline covariates typically improves its approximation to the underlying data generating model, thereby enhancing efficiency. In particular, when the working model is correctly specified, $\hat{\boldsymbol{\theta}}^*$ obtains the optimal efficiency (Tsiatis, 2006).

Based on Theorem 3.1, we evaluate the property of T_{Wald} 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*

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

and $N\mathsf{se}^2_{\mathsf{Wald}}(\hat{\tau}^*) = 2\sigma^2_{\mathsf{Wald}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) + o_p(1)$,

where $\sigma_\tau^2(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) = \boldsymbol{l}^\top \left\{\widetilde{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)\right\}^{-1} \Xi_{\Psi\Psi}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) \left\{\left[\widetilde{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)\right]^{-1}\right\}^\top \boldsymbol{l}$,

and $\sigma^2_{\mathsf{Wald}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) = \boldsymbol{l}^\top \left\{\widetilde{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)\right\}^{-1} \Sigma_{\Psi\Psi}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) \left\{\left[\widetilde{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)\right]^{-1}\right\}^\top \boldsymbol{l}$.

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*

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

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* $T_{\mathsf{Wald}} - \{2\sigma^2_{\mathsf{Wald}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)\}^{-1/2}\delta^*$ *converges to a normal random variable with mean zero and variance* $\lambda^2_{\mathsf{Wald}}$. *Further-*

more, 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.

Corollary 3.2 evaluates the properties of T_{Wald} and leads to several key conclusions. First, under CR, we have $\sigma^2_{\mathsf{Wald}}(\boldsymbol{\theta}^*,\boldsymbol{\alpha}^*) = \sigma^2_\tau(\boldsymbol{\theta}^*,\boldsymbol{\alpha}^*)$ and $\lambda^2_{\mathsf{Wald}} = 1$, indicating that T_{Wald} attains the nominal Type I error and that the power expression in (3.2) holds with $\lambda^2_{\mathsf{Wald}} = 1$. Second, Corollary 3.2.1 shows that the covariate balance induced by CAR may lead to $\lambda^2_{\mathsf{Wald}} < 1$, resulting in T_{Wald} being conservative with reduced Type I error and lower power. It is worth noting that the conservativeness of T_{Wald} under CAR may be mitigated when $\boldsymbol{R}(s,k) = \mathbf{0}_{p_0+2}$ for all $1 \le s \le s_{\mathsf{max}}$ and $1 \le k \le k_{\mathsf{max}}$, which holds under the fully correctly specified working model. In such cases, we again obtain $\lambda^2_{\mathsf{Wald}} = 1$, and the Type I error rate of T_{Wald} remains valid.

In practice, GEE may often involve model misspecification, such as an incorrect link function or omitted covariates. Under such conditions, $\boldsymbol{R}(s,k) = \mathbf{0}_{p_0+2}$ may not hold, and consequently, T_{Wald} may exhibit reduced Type I error and insufficient power under CAR.

4. A Robust Adjusted Test

Since model-based adjustments with T_{Wald} may result in reduced Type I error rates and lower power under model misspecification, it is desirable to develop robust adjusted tests that are valid even when the working model is misspecified. A key component of such robust inference is a reliable estimator for the standard error of $\hat{\tau}^*$ that does not depend on the correct specification of the model.

By Theorem 3.1 and Corollary 3.2.1, the difference between the model-based variance $N\mathsf{se}^2_{\mathsf{Wald}}(\hat{\tau}^*)$ and the asymptotic variance of $\sqrt{N}\hat{\tau}^*$ is

$$\boldsymbol{l}^\top \left\{\widetilde{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)\right\}^{-1} [\Sigma_{\mathsf{CR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*) - \Sigma_{\mathsf{CAR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)] \left\{\widetilde{\Psi}_{\boldsymbol{\alpha}^*}(\boldsymbol{\theta}^*)\right\}^{-1} \boldsymbol{l}. \tag{4.1}$$

This allows us to estimate the variance of $\sqrt{N}\hat{\tau}^*$ by combining the model-based estimate $N\mathsf{se}^2_{\mathsf{Wald}}(\hat{\tau}^*)$ with the correction term given in (4.1).

We describe the proposed adjusted estimator as follows. 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

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

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

To estimate $\Sigma_{\mathsf{CAR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)$, we adopt a bootstrap procedure introduced in Liu and Hu (2023). Let B denote the number of bootstrap replications. For each $1 \leq b \leq B$, we compute the bootstrapped imbalance $\{D_N^{*b}(s,k)\}_{s,k}$ according to Algorithm 1 detailed in Section S3 of the Supplementary Materials. Define $O_b^* = N^{-1/2} \sum_{k=1}^{k_{\max}} \sum_{s=1}^{s_{\max}} D_N^{*b}(s,k) \hat{\boldsymbol{R}}(s,k)$ for $1 \leq b \leq B$, then we estimate $\Sigma_{\mathsf{CAR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)$ by the sample covariance matrix of $\{O_b^*\}_{b=1}^B$, denoted by S_O^2. The proposed adjusted standard error is

$$N\mathsf{se}^2_{\mathsf{adj}}(\hat{\tau}^*) = N\mathsf{se}^2_{\mathsf{Wald}}(\hat{\tau}^*) - \boldsymbol{l}^\top \left\{ \widetilde{\Psi}_{N,\hat{\boldsymbol{\alpha}}^*}(\hat{\boldsymbol{\theta}}^*) \right\}^{-1} 4^{-1} \left[\hat{\Sigma}_{\mathsf{CR}}(\hat{\boldsymbol{\theta}}^*, \hat{\boldsymbol{\alpha}}^*) - S_O^2 \right] \left\{ \widetilde{\Psi}_{N,\hat{\boldsymbol{\alpha}}^*}(\hat{\boldsymbol{\theta}}^*) \right\}^{-1} \boldsymbol{l}. \quad (4.2)$$

And thus, the corresponding adjusted test statistic is $T_{\mathsf{adj}} = \mathsf{se}^{-1}_{\mathsf{adj}}(\hat{\tau}^*)\hat{\tau}^*$.

Note that for CAR procedures satisfying the strong balance condition in (3.1), $\Sigma_{\mathsf{CAR}}(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)$ can be directly estimated based on Corollary 3.2. Specifically, we define the plug-in estimator

$$\Sigma_{\mathsf{CAR},N}(\hat{\boldsymbol{\theta}}^*, \hat{\boldsymbol{\alpha}}^*) = \sum_{s=1}^{s_{\max}} \sum_{k_1,k_2} \left\{ I\{k_1 = k_2 = k\} \hat{\pi}_{s,k} - \hat{\pi}_{k_1|s} \hat{\pi}_{k_2|s} \hat{\pi}_{s,\cdot} \right\} \times \hat{\boldsymbol{R}}(s,k_1) \hat{\boldsymbol{R}}^\top(s,k_2).$$

The corresponding adjusted standard error for $\hat{\tau}^*$ is

$$N\mathsf{se}^2_{\mathsf{adj}}(\hat{\tau}^*) = N\mathsf{se}^2_{\mathsf{Wald}}(\hat{\tau}^*) - 4^{-1}\boldsymbol{l}^\top \left\{\widetilde{\Psi}_{N,\hat{\boldsymbol{\alpha}}^*}(\hat{\boldsymbol{\theta}}^*)\right\}^{-1} \times \left[\hat{\Sigma}_{\mathsf{CR}}(\hat{\boldsymbol{\theta}}^*, \hat{\boldsymbol{\alpha}}^*) - \Sigma_{\mathsf{CAR},N}(\hat{\boldsymbol{\theta}}^*, \hat{\boldsymbol{\alpha}}^*)\right] \left\{\widetilde{\Psi}_{N,\hat{\boldsymbol{\alpha}}^*}(\hat{\boldsymbol{\theta}}^*)\right\}^{-1} \boldsymbol{l}. \quad (4.3)$$

We note that the adjustment formula in (4.2) remains valid for CAR procedures satisfying (3.1). However, the direct estimator (4.3) is computationally more efficient, as it avoids the resampling process. The properties of the proposed adjusted test statistic are given in the following theorem.

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*

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

Theorem 4.1 establishes the validity of T_{adj} for CAR procedures satisfying Assumption 3.2. Specifically, it shows that $N\mathsf{se}^2_{\mathsf{adj}}(\hat{\tau}^*)$ consistently

estimates the asymptotic variance of $\hat{\tau}^*$, thereby ensuring that T_{adj} controls the Type I error at the nominal level and generally achieves higher power compared to the model-based test statistic T_{Wald}.

Theorem 4.1 also yields two important implications. First, when the working model is correctly specified, the asymptotic variance $\sigma_\tau^2(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)$ remains invariant across different randomization procedures. As a result, the power of T_{adj} is also invariant across different randomization procedures. Second, under model misspecification, the covariates not included in the working model may still affect $\hat{\tau}^*$ through the covariate balance induced by CAR. As shown in Liu and Hu (2022), CAR procedures can affect the balance of covariates not used in randomization, which in turn may influence the performance of T_{adj} (Liu and Hu, 2023). In such cases, designs that achieve finer covariate balance may yield smaller values of $\sigma_\tau^2(\boldsymbol{\theta}^*, \boldsymbol{\alpha}^*)$, and thus lead to higher power for the adjusted test.

5. Numerical Studies

Consider a trial with 500 patients. 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. We consider the following outcome generating model:

$$Y_{i,j} = A_{i,j}\gamma_1 + (1-A_{i,j})\gamma_0 + X_{i,1}\beta_1 + X_{i,2}\beta_2 + X_{i,3}\beta_3 + \exp\{X_{i,1}\beta_1 + X_{i,2}\beta_2\} + \epsilon_{i,j}, \quad (5.1)$$

where $\gamma_0 = 0.5$, $\gamma_1 = \gamma_0 + \tau$, $\beta_1 = \beta_2 = \beta_3 = 1$, 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 $\alpha = 0.7$.

In 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 $X_{i,2} \ge 0$, 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 identity link 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 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. Table 1 reports the bias, the standard deviation of $\hat{\tau}^*$, and the Type I error rates of the Wald test, T_{Wald}, under the null

hypothesis $\tau = 0$. The performance of both T_{Wald} and the proposed T_{adj} are compared in Table 2. The power of the two tests is further evaluated in Figure 1.

Table 1: 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 $\hat{\tau}^*$:bias($\sqrt{N}$sd) Type I	M_2 $\hat{\tau}^*$:bias($\sqrt{N}$sd) Type I	M_3 $\hat{\tau}^*$:bias($\sqrt{N}$sd) Type I
CR	IND	-0.03(12.10) 4.91	0.02(7.49) 4.65	0.01(7.21) 4.79
	EXCH	-0.03(11.88) 5.04	0.02(7.38) 4.74	0.01(7.11) 4.78
	AR(1)	-0.03(11.84) 5.15	0.02(7.33) 4.99	0.01(7.05) 4.75
PS	IND	0.19(9.17) 0.80	0.05(7.42) 4.40	0.05(7.08) 4.55
	EXCH	0.10(9.01) 0.70	0.02(7.33) 4.45	0.01(7.00) 4.40
	AR(1)	0.12(8.95) 0.60	0.03(7.27) 4.40	0.03(6.94) 4.15
STRPB	IND	-0.06(8.80) 0.49	-0.07(7.01) 3.44	-0.07(6.68) 3.16
	EXCH	-0.06(8.52) 0.47	-0.07(6.87) 3.53	-0.07(6.54) 3.16
	AR(1)	-0.06(8.49) 0.47	-0.07(6.83) 3.53	-0.07(6.49) 3.10
HH	IND	0.03(8.78) 0.37	0.04(6.97) 3.30	0.03(6.71) 3.15
	EXCH	0.02(8.52) 0.35	0.03(6.86) 3.14	0.03(6.59) 3.03
	AR(1)	0.05(8.49) 0.35	0.05(6.82) 3.21	0.04(6.55) 3.12

Our results lead to the following conclusions. First, the model-based adjustment with T_{Wald} may fail to preserve the Type I error rates under model misspecification. As shown in Table 1, the Type I error of T_{Wald} adjusted with M_2 and M_3 under HH and AR(1) are about 3.21% and 3.12%, respectively. This indicates that even when all of the covariates used in design are adjusted, the Type I errors of T_{Wald} adjusted with M_2 and M_3 under PS, STRPB and HH may still be lower than the nominal level 5%.

On the other hand, the proposed T_{adj} appropriately controls the Type I error for all working models and CAR procedures. Since $\mathsf{se}_{\mathsf{adj}}(\hat{\tau}^*)$ does

not rely on whether the working model $m_i(a)$ correctly reflects the outcome generating process, T_{adj} appropriately controls the Type I error rates under various settings of CAR designs and working correlation matrices used for GEE. As shown in Table 2, the Type I error rates (when $\tau = 0$) of T_{adj} under PS, STRPB, HH and the working models M_1, M_2, M_3, are much closer to 5% than the values under T_{Wald}.

Table 2: Type I error rates (%) and Powers of T_{Wald} and T_{adj} under $\tau = 0$ and $\tau = 0.8$ (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	4.91	-	4.65	-	4.79	-
		EXCH	5.04	-	4.74	-	4.78	-
		AR(1)	5.15	-	4.99	-	4.75	-
	0.8	IND	34.43	-	71.45	-	74.58	-
		EXCH	34.89	-	72.26	-	75.74	-
		AR(1)	34.94	-	72.58	-	75.72	-
PS	0	IND	0.80	4.95	4.40	5.05	4.55	5.35
		EXCH	0.70	5.10	4.45	5.30	4.40	5.10
		AR(1)	0.60	5.15	4.40	5.15	4.15	4.90
	0.8	IND	31.20	59.40	72.20	73.80	75.00	76.50
		EXCH	31.65	60.35	72.40	74.20	75.70	76.90
		AR(1)	31.40	60.35	72.90	74.85	76.35	77.50
STRPB	0	IND	0.49	4.75	3.44	5.22	3.16	4.96
		EXCH	0.47	4.84	3.53	5.29	3.16	5.09
		AR(1)	0.47	4.85	3.53	5.31	3.10	5.05
	0.8	IND	28.77	58.73	72.25	77.04	75.96	80.00
		EXCH	29.53	60.32	73.17	77.74	76.80	80.93
		AR(1)	29.67	60.61	73.52	77.82	77.24	80.82
HH	0	IND	0.37	4.68	3.30	4.96	3.15	4.91
		EXCH	0.40	4.60	3.22	4.81	3.06	4.85
		AR(1)	0.31	4.80	3.21	5.13	3.14	5.15
	0.8	IND	29.29	58.67	72.40	76.73	76.21	80.31
		EXCH	29.59	60.21	73.43	77.62	76.84	81.00
		AR(1)	29.74	60.41	73.77	78.02	77.31	81.62

As T_{adj} recovers the nominal test level for CAR procedures, its improve-

ment in power relative to T_{Wald} becomes evident. In Table 2, the powers of T_{adj} and T_{Wald} under HH and AR(1) working correlation structure and model M_2 are 73.77% and 78.02%, respectively; under model M_3, the corresponding powers are 77.31% and 81.62%, indicating an approximate 4% gain in power.

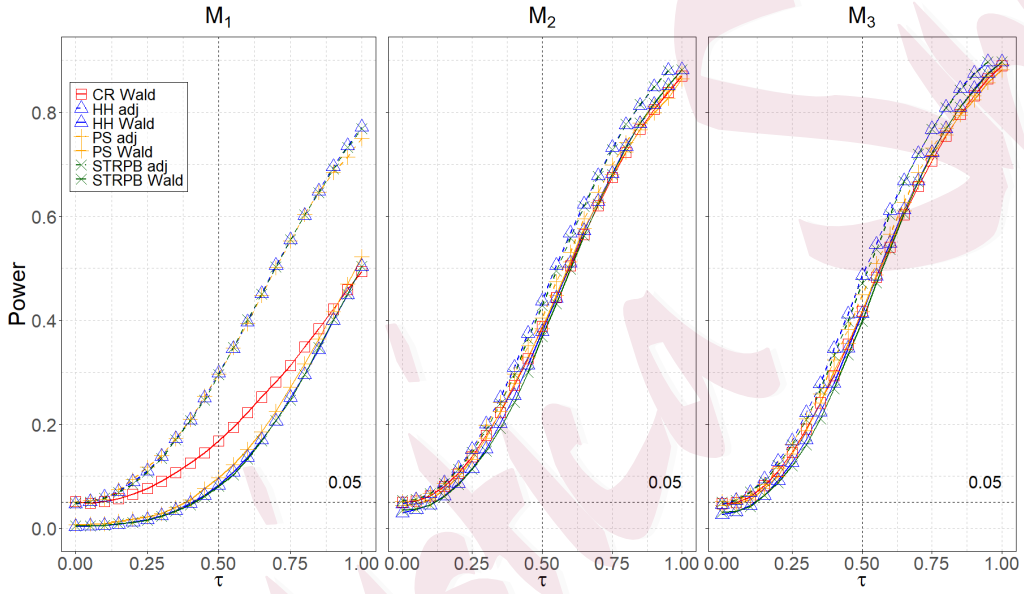

Figure 1: Power curves of T_{Wald} and T_{adj} under the exchangeable working correlation matrix.

Furthermore, the power of T_{adj} enhanced by covariate balancing is also demonstrated in Table 2. For instance, the powers ($\tau = 0.8$) of T_{adj} under M_3 are approximately 77% for PS and 81% for HH, whereas the corresponding power of T_{Wald} under CR is about 75%. This phenomenon of

the enhanced power aligns with Theorem 4.1, which indicates that designs achieving finer covariate balance yield higher statistical power. A similar pattern is illustrated in Figure 1.

Lastly, the results in Table 2 and Figure 1 highlight the advantages of the 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 T_{adj} using AR(1) and M_3 under HH is 81.62%, but the value under the independent correlation matrix is 80.31%.

In Section S2 of the Supplementary Material, we present additional simulation studies to evaluate the performance of T_{adj} and T_{Wald} for continuous and binary outcomes when the outcome models are correctly specified. The results further confirm the superiority of the proposed T_{adj} in maintaining the nominal Type I error rate and highlight the benefits of using GEE to adjust covariates and within-subject correlations, thereby enhancing the power for detecting treatment effects.

6. Empirical Application to Real Clinical Trial

The study by Elkashef et al. (2006) is used to illustrate the application of the proposed procedure for the analysis of longitudinal CAR trials by using GEE. This trial was conducted to test the effect of selegiline transdermal system (STS) on the weekly mean proportion of cocaine non-use days. The trial involves the following covariates: cocaine use (Cocaine), defined as the number of self-reported days of use during the past 30 days (range: 1–30); the Hamilton Depression Rating Scale score at screening (preHAMD); and gender (Gender), coded as 1 for male and 2 for female. We assume the responses are generated from the following fitted outcome model.

$$\begin{aligned} Y_{i,j} =& 0.770 + \tau A_i + f_1(\mathsf{Cocaine}_i) + f_2(\mathsf{preHAMD}_i) \\ &+ 0.017\mathsf{Gender}_i + \epsilon_{i,j}, \end{aligned}$$

where $f_1(\mathsf{Cocaine}_i)$ and $f_2(\mathsf{preHAMD}_i)$ are fitted according to cubic B-spline regression, and $\{\epsilon_{i,j}, 1 \le j \le n_i\}_{i=1}^{N}$ are random errors with $\sigma^2 = 0.105$, the estimated AR(1) correlation matrices, and $\alpha = 0.758$.

To perform CAR, Cocaine (0:$<$ 6; 1:otherwise) and preHAMD (1:$<$ 16, 0:otherwise) are discretized and used in design. We consider CR, STRPB, PS and HH for the design stage. For the inference stage, we compare GEE with independent (IND), exchangeable (EXCH), and AR(1) working

correlation matrices, and apply (2.2) with the identity link function with M_1, no covariates are used; M_2, adjusting preHAMD and Cocaine; and M_3, adjusting preHAMD, Cocaine, and Gender. Table 3 evaluates the Type I error ($\tau = 0.00$) and the power ($\tau = 0.07$) of T_{Wald} and T_{adj} under different design and GEE analysis schemes.

Table 3: Type I error rates ($\tau = 0$) and power ($\tau = 0.07$) for the CSP-1019 study.

Design	Cor for GEE	$\tau = 0.00$						$\tau = 0.07$					
		M_1		M_2		M_3		M_1		M_2		M_3	
		T_{Wald}	T_{adj}	T_{Wald}	T_{adj}	T_{Wald}	T_{adj}	T_{Wald}	T_{adj}	T_{Wald}	T_{adj}	T_{Wald}	T_{adj}
CR	IND	5.41	-	5.34	-	5.37	-	55.97	-	62.57	-	62.64	-
	EXCH	5.23	-	5.09	-	5.17	-	56.45	-	63.47	-	63.39	-
	AR(1)	5.32	-	5.05	-	5.15	-	58.38	-	65.45	-	65.54	-
PS	IND	3.65	4.68	5.15	5.13	5.15	5.15	57.00	61.65	62.75	62.75	62.75	62.80
	EXCH	3.65	4.83	5.08	5.10	5.12	5.15	57.80	62.15	63.50	63.50	63.95	63.90
	AR(1)	3.65	5.02	5.20	5.23	5.13	5.15	60.40	65.75	67.85	67.75	67.50	67.55
STRPB	IND	3.55	5.00	4.87	4.93	4.99	5.03	56.96	61.69	63.62	63.71	63.70	63.82
	EXCH	3.71	5.05	5.12	5.14	5.14	5.25	57.48	61.99	63.91	63.98	64.03	64.17
	AR(1)	3.84	5.13	5.10	5.18	5.05	5.12	60.19	64.94	66.86	67.06	67.03	67.16
HH	IND	3.73	4.96	5.18	5.21	5.14	5.17	56.88	61.43	63.54	63.70	63.53	63.68
	EXCH	3.82	5.10	5.28	5.32	5.30	5.32	57.42	61.65	63.47	63.59	63.73	63.81
	AR(1)	3.62	4.66	4.90	4.90	5.08	5.12	59.33	64.39	66.52	66.64	66.54	66.68

Table 3 demonstrates the superior performance of the proposed adjusted test using GEE for CAR designs. Compared with T_{Wald}, the adjusted test T_{adj} provides more reliable control of the Type I error rate across working models. For instance, under model M_1, the Type I errors of T_{Wald} with PS, STRPB, and HH designs are approximately 3%, whereas those of T_{adj} are much closer to the nominal 5% level. Under working models M_2 and M_3, both T_{Wald} and T_{adj} achieve Type I error rates close to 5%, resulting in comparable power. This may be due to the small effect of Gender compared

to other values. Regarding the performance of GEE, we observe that tests assuming an AR(1) correlation structure exhibit higher power than those based on the other working correlation matrices. Moreover, the power of the tests under AR(1) with STRPB, PS, and HH designs is higher than that under CR, indicating the advantage of using appropriate combinations of design and analysis methods to improve inference for treatment effects.

7. Conclusions

In this paper, we study covariate adjustment using GEE for the inference on treatment effects in longitudinal trials conducted under CAR. We investigate the properties of the GEE-adjusted treatment effect estimator by introducing an additional stratification variable, which facilitates the analysis of unbalanced longitudinal data. We show that the standard Wald test based on GEE can exhibit reduced Type I error rates because it ignores the impact of the CAR design. Building on the derived theoretical results, we propose an adjustment method applicable to a broad class of CAR procedures. The proposed adjustment does not rely on model assumptions and can maintain the nominal level of the test even when the working model is misspecified. Numerical studies are conducted to illustrate the advantages of the proposed method over conventional model-based approaches.

Our results can be extended in several directions. First, the proposed adjustment can be readily generalized to linear mixed-effects models and generalized linear mixed-effects models, as their asymptotic properties can be derived in a manner similar to those of GEE. Second, beyond longitudinal studies, GEE is also widely used in cluster randomized controlled trials (Offorha et al., 2023), and CAR has likewise been adopted in such trials to improve covariate balance (Ivers et al., 2012). Extending the proposed procedure to this setting would therefore be of practical interest. Third, since longitudinal trials often involve complex temporal patterns of treatment effects, semiparametric methods are frequently employed for analysis. It would be valuable to investigate how CAR influences such semiparametric analyses and to develop valid and efficient inferential procedures for longitudinal CAR trials with complex temporal structures.

Supplementary Materials

Supplementary Materials contain the proofs of theoretical results and the additional simulation results.

Acknowledgements

The authors would like to thank the editor and AE. and the two reviewers for the insightful comments that lead to a much improved version of the paper. Liu's research was supported by the National Natural Science Foundation of China grant no. 12301324, and the Fundametal Research Funds for the Central Universities and the Research Funds of University of China [grant no. 202630333]. This research was also supported by Public Computing Cloud, Renmin University of China.

References

Baldi Antognini, A. and M. Zagoraiou (2011). The covariate-adaptive biased coin design for balancing clinical trials in the presence of prognostic factors. *Biometrika 98*(3), 519–535.

Bugni, F. A., I. A. Canay, and A. M. Shaikh (2018). Inference under covariate-adaptive randomization. *Journal of the American Statistical Association 113*(524), 1784–1796.

Claggett, B. L., Z. R. McCaw, L. Tian, J. J. V. McMurray, P. S. Jhund, H. Uno, M. A. Pfeffer, S. D. Solomon, and L.-J. Wei (2022). Quantifying treatment effects in trials with multiple event-time outcomes. *NEJM Evidence 1*(10), EVIDoa2200047.

Elkashef, A., P. J. Fudala, L. Gorgon, S.-H. Li, R. Kahn, N. Chiang, F. Vocci, J. Collins, K. Jones, K. Boardman, et al. (2006). Double-blind, placebo-controlled trial of selegiline transdermal system (sts) for the treatment of cocaine dependence. *Drug and Alcohol*

Dependence 85(3), 191–197.

FDA (2023). Adjusting for covariates in randomized clinical trials for drugs and biological products guidance for industry. Technical report, U.S. Department of Health and Human Services, Food and Drug Administration (FDA).

Gail, M., W.-Y. Tan, and S. Piantadosi (1988). Tests for no treatment effect in randomized clinical trials. *Biometrika 75*(1), 57–64.

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.

Glantz, S. A., N. Nguyen, and A. L. O. da Silva (2024). Population-based disease odds for e-cigarettes and dual use versus cigarettes. *NEJM Evidence 3*(3), EVIDoa2300229.

Goligher, E. C., P. R. Lawler, T. P. Jensen, V. Talisa, L. R. Berry, E. Lorenzi, B. J. McVerry, C.-C. H. Chang, E. Leifer, C. Bradbury, et al. (2023). Heterogeneous treatment effects of therapeutic-dose heparin in patients hospitalized for covid-19. *Jama 329*(13), 1066–1077.

Hu, F., Y. Hu, Z. Ma, and W. F. Rosenberger (2014). Adaptive randomization for balancing over covariates. *Wiley Interdisciplinary Reviews: Computational Statistics 6*(4), 288–303.

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

Hu, Y. and F. Hu (2012). Asymptotic properties of covariate-adaptive randomization. *The*

Annals of Statistics 40(3), 1794–1815.

Ivers, N. M., I. J. Halperin, J. Barnsley, J. M. Grimshaw, B. R. Shah, K. Tu, R. Upshur, and M. Zwarenstein (2012). Allocation techniques for balance at baseline in cluster randomized trials: a methodological review. *Trials 13*(1), 120.

Li, X., J. Zhou, and F. Hu (2019). Testing hypotheses under adaptive randomization with continuous covariates in clinical trials. *Statistical Methods in Medical Research 28*(6), 1609–1621.

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

Liu, H., F. Tu, and W. Ma (2022, 06). Lasso-adjusted treatment effect estimation under covariate-adaptive randomization. *Biometrika 110*(2), 431–447.

Liu, X., F. Hu, and W. Ma (2025). The properties of covariate-adaptive randomization procedures with possibly unequal allocation ratio. *The Annals of Applied Statistics 19*(2), 907–925.

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 regres-

sion under covariate-adaptive randomization. *Journal of Econometrics 249*, 105808.

Ma, W., F. Hu, and L. Zhang (2015). Testing hypotheses of covariate-adaptive randomized clinical trials. *Journal of the American Statistical Association 110*(510), 669–680.

Ma, W., P. Li, L.-X. Zhang, and F. Hu (2024). A new and unified family of covariate adaptive randomization procedures and their properties. *Journal of the American Statistical Association 119*(545), 151–162.

Ma, W., Y. Qin, Y. Li, and F. Hu (2020). Statistical inference for covariate-adaptive randomization procedures. *Journal of the American Statistical Association 115*(531), 1488–1497.

Ma, W., F. Tu, and H. Liu (2022). Regression analysis for covariate-adaptive randomization: a robust and efficient inference perspective. *Statistics in Medicine 41*(29), 5645–5661.

Offorha, B. C., S. J. Walters, and R. M. Jacques (2023). Analysing cluster randomised controlled trials using glmm, gee1, gee2, and qif: results from four case studies. *BMC Medical Research Methodology 23*(1), 293.

Pocock, S. J. and R. Simon (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics 31*(1), 103–115.

Rosenberger, W. F. and J. M. Lachin (2015). *Randomization in clinical trials: theory and practice* (2nd ed.). Hoboken, N.J: John Wiley & Sons.

Rosenberger, W. F., O. Sverdlov, et al. (2008). Handling covariates in the design of clinical trials. *Statistical Science 23*(3), 404–419.

REFERENCES

Shao, J., X. Yu, and B. Zhong (2010). A theory for testing hypotheses under covariate-adaptive randomization. *Biometrika 97*(2), 347–360.

Tsiatis, A. A. (2006). *Semiparametric theory and missing data*, Volume 4. New York: Springer.

Wang, B., R. Susukida, R. Mojtabai, M. Amin-Esmaeili, and M. Rosenblum (2023). Model-robust inference for clinical trials that improve precision by stratified randomization and covariate adjustment. *Journal of the American Statistical Association 118*(542), 1152–1163.

Wang, M. (2014). Generalized estimating equations in longitudinal data analysis: a review and recent developments. *Advances in Statistics 2014*(1), 303728.

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.

Ye, T. and J. Shao (2020). Robust tests for treatment effect in survival analysis under covariate-adaptive randomization. *Journal of the Royal Statistical Society Series B: Statistical Methodology 82*(5), 1301–1323.

Ye, T., J. Shao, and Y. Yi (2024). Covariate-adjusted log-rank test: guaranteed efficiency gain and universal applicability. *Biometrika 111*(2), 691–705.

Zelen, M. (1974). The randomization and stratification of patients to clinical trials. *Journal of Chronic Diseases 27*(7), 365–375.

REFERENCES

Zhao, Z., Y. Song, W. Jiang, and D. Tu (2024). Consistent covariances estimation for stratum imbalances under minimization method for covariate-adaptive randomization. *Scandinavian Journal of Statistics 51*(2), 861–890.

Ruichen Fu

Institute of Statistics and Big Data,

Renmin University of China, Beijing, China

E-mail: fu_ruichen@ruc.edu.cn

Yang Liu

Institute of Statistics and Big Data,

Renmin University of China, Beijing, China

E-mail: yangliu2022@ruc.edu.cn