Supplementary Materials for “Valid test for multi-arm trials with generalized linear models under covariate-adaptive randomization

Guannan Zhai [1], Feifang Hu [1]

[1] *Department of Statistics, The George Washington University, U.S.A.*

Supplementary Material

This document contains supplementary materials for the paper “Valid test for multi-arm trials with generalized linear models under covariate-adaptive randomization”. The materials are organized into four sections. Section S1 introduces the notation and imbalance measures under multi-arm CAR. Section S2 presents ancillary results, including technical lemmas and intermediate results used to establish the main theorems. Section S3 provides detailed proofs of the main theorems in Section 3 of the paper. Section S4 studies the working variance term in the asymptotic theory for different GLMs.

S1 Notation and imbalance measures under multi-arm CAR

This section collects the imbalance measures used in Section 3 of the main paper. The notation follows Hu et al. (2023).

Consider a clinical trial with R discrete stratification covariates. The rth covariate has m_r levels. This induces $M = \sum_{r=1}^{R} m_r$ marginal categories and $S = \prod_{r=1}^{R} m_r$ strata. We use (r, m_r^*) to denote the marginal category where the rth covariate takes value $m_r^* \in \{1, \ldots, m_r\}$. We use $s = (m_1^*, \ldots, m_R^*)$ to denote the stratum where $Z_1 = m_1^*, \ldots, Z_R = m_R^*$.

Let n_k be the total number of patients assigned to arm k ($k = 0, \ldots, K$). Let $n_k(r, m_r^*)$ be the number of patients in marginal category (r, m_r^*) assigned to arm k. Let $n_k(s)$ be the number of patients in stratum s assigned to arm k. We define three imbalance vectors.

Overall imbalance: $\boldsymbol{D}_N = (D_0, \ldots, D_K)$, where $D_k = n_k - \bar{N}$ and $\bar{N} = \frac{1}{K+1}\sum_{k'=0}^{K} n_{k'} = \frac{N}{K+1}$.

Marginal imbalance: $\boldsymbol{D}_N(r, m_r^*) = \big(D_0(r, m_r^*), \ldots, D_K(r, m_r^*)\big)$, where $D_k(r, m_r^*) = n_k(r, m_r^*) - \bar{N}(r, m_r^*)$ and $\bar{N}(r, m_r^*) = \frac{1}{K+1}\sum_{k'=0}^{K} n_{k'}(r, m_r^*)$.

Within-stratum imbalance: $\boldsymbol{D}_N(s) = \big(D_0(s), \ldots, D_K(s)\big)$, where $D_k(s) = n_k(s) - \bar{N}(s)$ and $\bar{N}(s) = \frac{1}{K+1}\sum_{k'=0}^{K} n_{k'}(s)$.

Remark 1. The multi-arm CAR procedures considered in this paper can be divided into two categories in terms of within-stratum imbalance. Under stratified randomization-type procedures (when $w_s > 0$), the within-stratum, marginal, and overall imbalances are all bounded in probability, that is, $O_p(1)$ (Hu et al., 2023). Under Pocock and Simon's minimization (when $w_o = w_s = 0$), the procedure controls marginal and overall imbalances, while the within-stratum imbalances $D_k(s)$ typically grow at rate $\sqrt{N}$ and admit an asymptotic normal distribution (Hu and Zhang, 2020; Hu et al., 2023).

S2 Ancillary Results

Lemma 1. *Suppose that the overall imbalance is bounded in probability, $\boldsymbol{D}_N = O_p(\boldsymbol{1})$. Then, for every $k = 0, \ldots, K$,*

$$\frac{n_k}{N} \xrightarrow{P} \frac{1}{K+1}.$$

Proof. Since

$$\frac{n_k}{N} = \frac{1}{K+1} + \frac{D_k}{N},$$

the result follows from $D_k = O_p(1)$. □

Lemma 2. *Suppose that the equal-allocation covariate-adaptive randomization procedure satisfies Condition 1 and is invariant under permutations of the $K+1$ treatment labels. Suppose further that the number of strata is finite, that*

$$\mathbb{E}(Y^2) < \infty,$$

and that

$$D_a(s) = o_p(N), \qquad a = 0, \ldots, K,$$

for every stratum s.

Under the global null hypothesis

$$H_0 : \delta_1 = \cdots = \delta_K = 0,$$

let

$$\nu_i = \mathbb{E}(Y_i \mid \boldsymbol{Z}_i) = h\left(\delta_0 + \boldsymbol{\beta}^T \boldsymbol{Z}_i\right),$$

and define

$$\boldsymbol{R}_N = \frac{1}{\sqrt{N}} \begin{pmatrix} \sum_{i=1}^{N} (T_{i1} - T_{i0})(Y_i - \nu_i) \\ \vdots \\ \sum_{i=1}^{N} (T_{iK} - T_{i0})(Y_i - \nu_i) \end{pmatrix},$$

$$\boldsymbol{H}_N = \frac{1}{\sqrt{N}} \begin{pmatrix} \sum_{i=1}^{N} (T_{i1} - T_{i0})\nu_i \\ \vdots \\ \sum_{i=1}^{N} (T_{iK} - T_{i0})\nu_i \end{pmatrix}.$$

Assume that the components of $\boldsymbol{H}_N$ converge jointly to a multivariate normal distribution and that Condition 2 holds for each treatment-control contrast. Then

$$\left(\sqrt{2(K+1)}\,\boldsymbol{R}_N,\ \sqrt{2(K+1)}\,\boldsymbol{H}_N\right) \xrightarrow{D} (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2),$$

where $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ are independent, with

$$\boldsymbol{\xi}_1 \sim \mathcal{N}\left(\boldsymbol{0},\, \mathbb{E}\{\mathrm{Var}(Y \mid \boldsymbol{Z}_i)\} \left[\mathrm{diag}\{2\boldsymbol{1}_K\} + 2\boldsymbol{1}_K \boldsymbol{1}_K^T\right]\right),$$

$$\boldsymbol{\xi}_2 \sim \mathcal{N}\left(\boldsymbol{0},\, 2(K+1)\sigma_h^2 \boldsymbol{V}\right),$$

where

$$\boldsymbol{V} = \frac{1}{2}\left(\boldsymbol{I}_K + \boldsymbol{1}_K \boldsymbol{1}_K^T\right).$$

Thus, the diagonal entries of $\boldsymbol{V}$ are one and its off-diagonal entries are $1/2$.

Proof. Let

$$\mathcal{Z} = \{\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_N\}, \qquad \mathcal{T} = \{\boldsymbol{T}_1, \ldots, \boldsymbol{T}_N\},$$

and write

$$\varepsilon_i = Y_i - \nu_i.$$

Under the global null hypothesis and the randomization assumption,

$$\mathbb{E}(\varepsilon_i \mid \mathcal{Z}, \mathcal{T}) = 0.$$

Moreover, conditional on $(\mathcal{Z}, \mathcal{T})$, the residuals $\varepsilon_1, \ldots, \varepsilon_N$ are independent.

We first derive the limiting covariance matrix of $\boldsymbol{R}_N$. For the kth component of $\boldsymbol{R}_N$,

$$\begin{aligned}\operatorname{Var}(R_{N,k} \mid \mathcal{Z}, \mathcal{T}) &= \frac{1}{N}\sum_{i=1}^{N}(T_{ik} - T_{i0})^2 \operatorname{Var}(Y_i \mid \boldsymbol{Z}_i) \\ &= \frac{1}{N}\sum_{i=1}^{N}(T_{ik} + T_{i0}) \operatorname{Var}(Y_i \mid \boldsymbol{Z}_i).\end{aligned}$$

To evaluate the limit, let

$$\tau_s^2 = \operatorname{Var}(Y_i \mid \boldsymbol{Z}_i = s)$$

and let $n(s)$ denote the number of subjects in stratum s. By the definition of the within-stratum imbalance,

$$n_a(s) = \frac{n(s)}{K+1} + D_a(s).$$

Therefore, for every $a = 0, \ldots, K$,

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N} T_{ia} \operatorname{Var}(Y_i \mid \boldsymbol{Z}_i) &= \frac{1}{N}\sum_{s} n_a(s)\tau_s^2 \\ &= \frac{1}{K+1}\frac{1}{N}\sum_{s} n(s)\tau_s^2 + \frac{1}{N}\sum_{s} D_a(s)\tau_s^2.\end{aligned}$$

Because the number of strata is finite and $D_a(s) = o_p(N)$ for every a and s,

$$\frac{1}{N}\sum_{s} D_a(s)\tau_s^2 = o_p(1).$$

Furthermore, by the law of large numbers,

$$\frac{1}{N}\sum_{s} n(s)\tau_s^2 \xrightarrow{P} \mathbb{E}\{\operatorname{Var}(Y \mid \mathcal{Z})\}.$$

Consequently,

$$\frac{1}{N}\sum_{i=1}^{N} T_{ia} \operatorname{Var}(Y_i \mid \boldsymbol{Z}_i) \xrightarrow{P} \frac{1}{K+1}\mathbb{E}\{\operatorname{Var}(Y \mid \mathcal{Z})\}.$$

It follows that

$$\operatorname{Var}\left(R_{N,k} \mid \mathcal{Z}, \mathcal{T}\right) \xrightarrow{P} \frac{2}{K+1}\mathbb{E}\{\operatorname{Var}(Y \mid \mathcal{Z})\}.$$

For any $j \neq k$,

$$\begin{aligned}
&\operatorname{Cov}\left(R_{N,j}, R_{N,k} \mid \mathcal{Z}, \mathcal{T}\right) \\
&= \frac{1}{N}\sum_{i=1}^{N}(T_{ij} - T_{i0})(T_{ik} - T_{i0}) \operatorname{Var}(Y_i \mid \boldsymbol{Z}_i).
\end{aligned}$$

Because each subject is assigned to exactly one treatment arm,

$$(T_{ij} - T_{i0})(T_{ik} - T_{i0}) = T_{i0}, \qquad j \neq k.$$

Therefore,

$$\begin{aligned}
&\operatorname{Cov}\left(R_{N,j}, R_{N,k} \mid \mathcal{Z}, \mathcal{T}\right) \\
&= \frac{1}{N}\sum_{i=1}^{N} T_{i0} \operatorname{Var}(Y_i \mid \boldsymbol{Z}_i) \xrightarrow{P} \frac{1}{K+1}\mathbb{E}\{\operatorname{Var}(Y \mid \mathcal{Z})\}.
\end{aligned}$$

Since the number of strata is finite and $\mathbb{E}(Y^2) < \infty$, the Lindeberg condition holds for the triangular array of residuals $\{\varepsilon_i\}$ conditional on $(\mathcal{Z}, \mathcal{T})$.

Hence, by the conditional multivariate central limit theorem,

$$\boldsymbol{R}_N \xrightarrow{D} \boldsymbol{R} \sim \mathcal{N}\left(\boldsymbol{0},\ \frac{2\mathbb{E}\{\operatorname{Var}(Y \mid \mathcal{Z})\}}{K+1}\boldsymbol{V}\right).$$

We next derive the covariance matrix of the limiting distribution of $\boldsymbol{H}_N$. By the assumed joint convergence, write

$$\boldsymbol{H}_N \xrightarrow{D} \boldsymbol{H} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_H).$$

By Condition 2, for every $k = 1, \ldots, K$,

$$H_{N,k} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(T_{ik} - T_{i0})\nu_i \xrightarrow{D} \mathcal{N}(0, \sigma_h^2).$$

Therefore,

$$\mathrm{Var}(H_k) = \sigma_h^2, \qquad k = 1, \ldots, K.$$

For any $j \neq k$, the continuous mapping theorem gives

$$H_{N,j} - H_{N,k} \xrightarrow{D} H_j - H_k.$$

Moreover,

$$H_{N,j} - H_{N,k} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (T_{ij} - T_{ik}) \nu_i.$$

Because the equal-allocation CAR procedure is invariant under permutations of the $K + 1$ treatment labels, this treatment-treatment contrast has the same limiting distribution as any treatment-control contrast. Hence,

$$H_j - H_k \sim \mathcal{N}(0, \sigma_h^2).$$

It follows that

$$\begin{aligned}
\sigma_h^2 &= \mathrm{Var}(H_j - H_k) \\
&= \mathrm{Var}(H_j) + \mathrm{Var}(H_k) - 2\,\mathrm{Cov}(H_j, H_k) \\
&= 2\sigma_h^2 - 2\,\mathrm{Cov}(H_j, H_k).
\end{aligned}$$

Therefore,

$$\mathrm{Cov}(H_j, H_k) = \frac{\sigma_h^2}{2}, \qquad j \neq k.$$

Consequently,

$$\boldsymbol{\Sigma}_H = \sigma_h^2 \boldsymbol{V},$$

and hence

$$\boldsymbol{H}_N \xrightarrow{D} \boldsymbol{H} \sim \mathcal{N}\left(\mathbf{0},\, \sigma_h^2 \boldsymbol{V}\right).$$

It remains to establish the asymptotic independence of $\boldsymbol{R}_N$ and $\boldsymbol{H}_N$. Notice that $\boldsymbol{H}_N$ is measurable with respect to $(\mathcal{Z}, \mathcal{T})$. By the conditional multivariate central limit theorem, for every bounded continuous function f,

$$\mathbb{E}\left\{f(\boldsymbol{R}_N) \mid \mathcal{Z}, \mathcal{T}\right\} \xrightarrow{P} \mathbb{E}\{f(\boldsymbol{R})\},$$

where the distribution of $\boldsymbol{R}$ does not depend on $(\mathcal{Z}, \mathcal{T})$.

Therefore, for any bounded continuous functions f and g,

$$\begin{aligned}
&\mathbb{E}\left\{f(\boldsymbol{R}_N) g(\boldsymbol{H}_N)\right\} \\
&\quad = \mathbb{E}\left[g(\boldsymbol{H}_N) \mathbb{E}\left\{f(\boldsymbol{R}_N) \mid \mathcal{Z}, \mathcal{T}\right\}\right] \\
&\quad \longrightarrow \mathbb{E}\{f(\boldsymbol{R})\}\mathbb{E}\{g(\boldsymbol{H})\}.
\end{aligned}$$

Thus,

$$(\boldsymbol{R}_N, \boldsymbol{H}_N) \xrightarrow{D} (\boldsymbol{R}, \boldsymbol{H}),$$

where $\boldsymbol{R}$ and $\boldsymbol{H}$ are independent.

Finally, multiplying both components by $\sqrt{2(K+1)}$ gives

$$\left(\sqrt{2(K+1)}\, \boldsymbol{R}_N,\ \sqrt{2(K+1)}\, \boldsymbol{H}_N\right) \xrightarrow{D} (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2),$$

where

$$\boldsymbol{\xi}_1 = \sqrt{2(K+1)}\, \boldsymbol{R}, \qquad \boldsymbol{\xi}_2 = \sqrt{2(K+1)}\, \boldsymbol{H}.$$

Since

$$4\boldsymbol{V} = \operatorname{diag}\{2\mathbf{1}_K\} + 2\mathbf{1}_K \mathbf{1}_K^T,$$

we obtain

$$\begin{aligned}\operatorname{Var}(\boldsymbol{\xi}_1) &= 4\mathbb{E}\{\operatorname{Var}(Y \mid \mathcal{Z})\}\boldsymbol{V} \\ &= \mathbb{E}\{\operatorname{Var}(Y \mid \mathcal{Z})\}\left[\operatorname{diag}\{2\mathbf{1}_K\} + 2\mathbf{1}_K\mathbf{1}_K^T\right], \\ \operatorname{Var}(\boldsymbol{\xi}_2) &= 2(K+1)\sigma_h^2\boldsymbol{V}.\end{aligned}$$

This proves the asserted joint convergence. □

Lemma 3. *Suppose that the assumptions of Lemma 2 hold. Under the global null hypothesis*

$$H_0 : \boldsymbol{\delta} = \mathbf{0},$$

let

$$\psi_k = \frac{1}{n_k}\sum_{i=1}^N T_{ik}Y_i, \qquad n_k = \sum_{i=1}^N T_{ik}, \qquad k = 0, \dots, K.$$

Then

$$\psi_k - \mathbb{E}(Y) = O_p(N^{-1/2}), \qquad k = 0, \dots, K.$$

In particular,

$$\frac{1}{n_k}\sum_{i=1}^N T_{ik}Y_i \xrightarrow{P} \mathbb{E}(Y), \qquad k = 0, \dots, K.$$

Proof. Let

$$m = \mathbb{E}(Y)$$

and define the arm-specific outcome totals

$$Q_{N,k} = \sum_{i=1}^N T_{ik}Y_i, \qquad k = 0, \dots, K.$$

For every $k = 1, \dots, K$,

$$Q_{N,k} - Q_{N,0} = \sum_{i=1}^N (T_{ik} - T_{i0})Y_i.$$

Using the decomposition

$$Y_i = (Y_i - \nu_i) + \nu_i, \qquad \nu_i = \mathbb{E}(Y_i \mid \boldsymbol{Z}_i),$$

we obtain

$$\begin{aligned}\frac{Q_{N,k} - Q_{N,0}}{\sqrt{N}} &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (T_{ik} - T_{i0})(Y_i - \nu_i) \\ &\quad + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (T_{ik} - T_{i0})\nu_i.\end{aligned}$$

By Lemma 2, both terms on the right-hand side are $O_p(1)$. Therefore,

$$Q_{N,k} - Q_{N,0} = O_p(\sqrt{N}), \qquad k = 1, \ldots, K.$$

Since each subject is assigned to exactly one of the $K+1$ arms,

$$\sum_{k=0}^{K} Q_{N,k} = \sum_{i=1}^{N} Y_i.$$

Moreover,

$$\sum_{k=0}^{K} Q_{N,k} = (K+1)Q_{N,0} + \sum_{k=1}^{K} (Q_{N,k} - Q_{N,0}).$$

Consequently,

$$Q_{N,0} = \frac{1}{K+1} \left\{ \sum_{i=1}^{N} Y_i - \sum_{k=1}^{K} (Q_{N,k} - Q_{N,0}) \right\}.$$

Under the global null hypothesis, the ordinary central limit theorem gives

$$\sum_{i=1}^{N} Y_i = Nm + O_p(\sqrt{N}).$$

Together with

$$Q_{N,k} - Q_{N,0} = O_p(\sqrt{N}),$$

this yields

$$Q_{N,0} = \frac{Nm}{K+1} + O_p(\sqrt{N}).$$

For every $k = 1, \dots, K$,

$$Q_{N,k} = Q_{N,0} + (Q_{N,k} - Q_{N,0}) = \frac{Nm}{K+1} + O_p(\sqrt{N}).$$

Thus, for every $k = 0, \dots, K$,

$$Q_{N,k} = \frac{Nm}{K+1} + O_p(\sqrt{N}).$$

By Condition 1,

$$n_k = \frac{N}{K+1} + O_p(1), \qquad k = 0, \dots, K.$$

It follows that

$$\begin{aligned}
\psi_k - m &= \frac{Q_{N,k} - mn_k}{n_k} \\
&= \frac{\{Q_{N,k} - Nm/(K+1)\} - m\{n_k - N/(K+1)\}}{n_k} \\
&= O_p(N^{-1/2}).
\end{aligned}$$

Therefore,

$$\psi_k \xrightarrow{P} m = \mathbb{E}(Y), \qquad k = 0, \dots, K.$$

$\square$

S3 Proof of the Main Theorems

Proof of Theorem 4: Let

$$m = \mathbb{E}(Y), \qquad g = h^{-1}, \qquad \eta_0 = h^{-1}(m),$$

and write

$$h_0' = h'(\eta_0), \qquad \gamma_0' = \gamma'(\eta_0), \qquad A = \mathbb{E}\{\mathrm{Var}(Y \mid \mathcal{Z})\}.$$

For $k = 0, \ldots, K$, define

$$\psi_k = \frac{1}{n_k} \sum_{i=1}^{N} T_{ik} Y_i.$$

Also write

$$\widehat{\boldsymbol{\theta}} = (\widehat{\delta}_0, \widehat{\delta}_1, \ldots, \widehat{\delta}_K)^T, \qquad \widehat{\boldsymbol{\delta}} = (\widehat{\delta}_1, \ldots, \widehat{\delta}_K)^T.$$

Under the treatment-only working model, the maximum likelihood estimators satisfy

$$\widehat{\delta}_0 = g(\psi_0), \qquad \widehat{\delta}_k = g(\psi_k) - g(\psi_0), \quad k = 1, \ldots, K.$$

Recall that the overall imbalance is defined by

$$D_k = n_k - \frac{N}{K+1}.$$

Condition 1 implies $D_k = O_p(1)$. Moreover, since

$$\frac{n_k}{N} \xrightarrow{P} \frac{1}{K+1}$$

and $\mathbb{E}|Y| < \infty$, we have

$$|\psi_k| \leq \frac{N}{n_k} \frac{1}{N} \sum_{i=1}^{N} |Y_i| = O_p(1).$$

It follows that

$$\begin{aligned}
\psi_k &= \frac{K+1}{N} \sum_{i=1}^{N} T_{ik} Y_i + \frac{N - (K+1) n_k}{N} \psi_k \\
&= \frac{K+1}{N} \sum_{i=1}^{N} T_{ik} Y_i - \frac{(K+1) D_k}{N} \psi_k \\
&= \frac{K+1}{N} \sum_{i=1}^{N} T_{ik} Y_i + o_p(N^{-1/2}).
\end{aligned}$$

By Lemma 2, for every $k = 1, \ldots, K$,

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} (T_{ik} - T_{i0}) Y_i = O_p(1).$$

Therefore,

$$\frac{K+1}{N}\sum_{i=1}^{N}(T_{ik}-T_{i0})Y_i = O_p(N^{-1/2}).$$

Furthermore,

$$\frac{1}{K+1}\sum_{a=0}^{K}\left\{\frac{K+1}{N}\sum_{i=1}^{N}T_{ia}Y_i\right\} = \frac{1}{N}\sum_{i=1}^{N}Y_i = m + O_p(N^{-1/2}).$$

Consequently,

$$\psi_k - m = O_p(N^{-1/2}), \qquad k = 0, \dots, K.$$

Applying a componentwise Taylor expansion of g at m gives

$$\begin{aligned}
\widehat{\delta}_k &= g(\psi_k) - g(\psi_0) \\
&= g'(m)(\psi_k - \psi_0) + o_p(N^{-1/2}) \\
&= g'(m)\frac{K+1}{N}\sum_{i=1}^{N}(T_{ik}-T_{i0})Y_i + o_p(N^{-1/2}).
\end{aligned}$$

Since $g = h^{-1}$, the inverse function theorem gives

$$g'(m) = \frac{1}{h'(\eta_0)} = \frac{1}{h'_0}.$$

Therefore,

$$\widehat{\delta}_k = \frac{K+1}{Nh'_0}\sum_{i=1}^{N}(T_{ik}-T_{i0})Y_i + o_p(N^{-1/2}), \qquad k = 1, \dots, K. \tag{S3.1}$$

We next derive the probability limit of the model-based standard errors. Let

$$\boldsymbol{x}_i = (1, T_{i1}, \dots, T_{iK})^T$$

and

$$\widehat{\eta}_i = \widehat{\delta}_0 + \sum_{k=1}^{K}\widehat{\delta}_k T_{ik}.$$

For the general link satisfying $\theta = \gamma(\eta)$, the Fisher information matrix under the working model is

$$\mathcal{I}(\widehat{\boldsymbol{\theta}}) = \frac{1}{\phi}\sum_{i=1}^{N}h'(\widehat{\eta}_i)\gamma'(\widehat{\eta}_i)\boldsymbol{x}_i\boldsymbol{x}_i^T.$$

Because

$$\psi_k \xrightarrow{P} m \qquad \text{and} \qquad \frac{n_k}{N} \xrightarrow{P} \frac{1}{K+1},$$

we obtain

$$\frac{1}{N}\mathcal{I}(\widehat{\boldsymbol{\theta}}) \xrightarrow{P} \frac{h_0'\gamma_0'}{\phi}\begin{bmatrix} 1 & \dfrac{1}{K+1}\mathbf{1}_K^T \\ \dfrac{1}{K+1}\mathbf{1}_K & \dfrac{1}{K+1}\boldsymbol{I}_K \end{bmatrix}.$$

The inverse of the block matrix above is

$$\begin{bmatrix} K+1 & -(K+1)\mathbf{1}_K^T \\ -(K+1)\mathbf{1}_K & (K+1)\left(\boldsymbol{I}_K + \mathbf{1}_K\mathbf{1}_K^T\right) \end{bmatrix}.$$

Since

$$\boldsymbol{V} = \frac{1}{2}\left(\boldsymbol{I}_K + \mathbf{1}_K\mathbf{1}_K^T\right),$$

the treatment-effect block of the inverse Fisher information satisfies

$$N\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\delta}}) \xrightarrow{P} \frac{2(K+1)\phi}{h_0'\gamma_0'}\boldsymbol{V}. \tag{S3.2}$$

In particular, for every $k = 1, \ldots, K$,

$$\sqrt{N}\,\widehat{\mathrm{se}}(\widehat{\delta}_k) \xrightarrow{P} \left\{\frac{2(K+1)\phi}{h_0'\gamma_0'}\right\}^{1/2}. \tag{S3.3}$$

We now derive the actual asymptotic covariance matrix of $\widehat{\boldsymbol{\delta}}$. Define

$$\boldsymbol{C}_N = \frac{1}{\sqrt{N}}\begin{pmatrix} \sum_{i=1}^N (T_{i1} - T_{i0})Y_i \\ \vdots \\ \sum_{i=1}^N (T_{iK} - T_{i0})Y_i \end{pmatrix}.$$

By the decomposition

$$Y_i = (Y_i - \nu_i) + \nu_i$$

and Lemma 2,

$$\boldsymbol{C}_N \xrightarrow{D} \mathcal{N}\left(\mathbf{0}, \left\{\frac{2A}{K+1} + \sigma_h^2\right\}\boldsymbol{V}\right). \tag{S3.4}$$

Combining (S3.1) and (S3.4) gives

$$\sqrt{N}\,\widehat{\boldsymbol{\delta}} \xrightarrow{D} \mathcal{N}\left(\mathbf{0}, \frac{2(K+1)}{(h_0')^2}\left\{A + \frac{(K+1)\sigma_h^2}{2}\right\}\boldsymbol{V}\right). \tag{S3.5}$$

Finally, let

$$\boldsymbol{S} = (S_1, \ldots, S_K)^T, \qquad S_k = \frac{\widehat{\delta}_k}{\widehat{\mathrm{se}}(\widehat{\delta}_k)}.$$

It follows from (S3.3), (S3.5), and Slutsky's theorem that

$$\boldsymbol{S} \xrightarrow{D} \mathcal{N}\left(\mathbf{0}, \frac{A + (K+1)\sigma_h^2/2}{\phi h_0'/\gamma_0'}\boldsymbol{V}\right).$$

Equivalently,

$$\boldsymbol{S} \xrightarrow{D} \mathcal{N}\left(\mathbf{0}, \frac{\mathbb{E}\{\mathrm{Var}(Y \mid \mathcal{Z})\} + (K+1)\sigma_h^2/2}{\phi h'[h^{-1}\{\mathbb{E}(Y)\}]/\gamma'[h^{-1}\{\mathbb{E}(Y)\}]}\boldsymbol{V}\right),$$

where

$$v_{jk} = \begin{cases} 1, & j = k, \\ 1/2, & j \neq k. \end{cases}$$

This proves Theorem 4. Theorems 1–3 follow under their corresponding assumptions as special cases obtained by setting $\gamma'(\eta_0) = 1$ and/or $\sigma_h^2 = 0$. $\square$

S4 Additional Working Variance Analysis in GLMs

For a general link, recall that the natural parameter satisfies $\theta = \gamma(\eta)$ and

$$h(\eta) = b'\{\gamma(\eta)\}.$$

Consequently,

$$h'(\eta) = b''\{\gamma(\eta)\}\gamma'(\eta).$$

Let

$$m = \mathbb{E}(Y), \qquad \eta_0 = h^{-1}(m).$$

The model-based working variance term appearing in Theorems 3 and 4 is therefore

$$W(m) = \frac{\phi h'(\eta_0)}{\gamma'(\eta_0)} = \phi b''\{\gamma(\eta_0)\}. \tag{S4.6}$$

Thus, $W(m)$ is the variance function of the working response distribution evaluated at the marginal mean m. For a canonical link, $\gamma'(\eta) = 1$, and (S4.6) reduces to

$$W(m) = \phi h'(\eta_0).$$

Let

$$A = \mathbb{E}\{\mathrm{Var}(Y \mid \boldsymbol{Z})\}.$$

For designs with bounded within-stratum imbalances, Theorem 3 shows that the marginal asymptotic variance of each standard Wald statistic is

$$\frac{A}{W(m)}.$$

For designs satisfying Conditions 1–2, Theorem 4 gives the marginal asymptotic variance

$$\frac{A + (K+1)\sigma_h^2/2}{W(m)}.$$

Thus, for the latter designs, the additional imbalance variance must also be included when determining whether the test is conservative or anti-conservative.

In the calculations below, let

$$\mu_Z = \mathbb{E}(Y \mid \boldsymbol{Z}), \qquad m = \mathbb{E}(\mu_Z) = \mathbb{E}(Y).$$

S4.1 Bernoulli Regression with Probit Link

For a Bernoulli response with a probit link,

$$h(\eta) = \Phi(\eta), \qquad h'(\eta) = \varphi(\eta),$$

where Φ and φ are the standard normal distribution and density functions, respectively. The Bernoulli natural parameter is

$$\gamma(\eta) = \log\left\{\frac{\Phi(\eta)}{1-\Phi(\eta)}\right\},$$

so that

$$\gamma'(\eta) = \frac{\varphi(\eta)}{\Phi(\eta)\{1-\Phi(\eta)\}}.$$

Because $\phi = 1$ for a Bernoulli response,

$$\begin{aligned} W(m) &= \frac{h'(\eta_0)}{\gamma'(\eta_0)} \\ &= \Phi(\eta_0)\{1-\Phi(\eta_0)\} \\ &= m(1-m), \qquad \eta_0 = \Phi^{-1}(m). \end{aligned}$$

Moreover,

$$\begin{aligned} \mathbb{E}\{\mathrm{Var}(Y \mid \boldsymbol{Z})\} &= \mathbb{E}\{\mu_Z(1-\mu_Z)\} \\ &= m(1-m) - \mathrm{Var}(\mu_Z) \\ &\le W(m). \end{aligned}$$

The inequality is strict when $\mathrm{Var}(\mu_Z) > 0$. Therefore, for designs with bounded within-stratum imbalances, the standard model-based Wald test is asymptotically conservative when prognostic covariates are omitted from the working model.

Under Conditions 1–2,

$$A + \frac{(K+1)\sigma_h^2}{2} - W(m) = -\,\mathrm{Var}(\mu_Z) + \frac{(K+1)\sigma_h^2}{2}.$$

Thus, the direction of the size distortion depends on the relative magnitudes of the omitted-covariate and imbalance-variance terms.

S4.2 Bernoulli Regression with Complementary Log-Log Link

For a Bernoulli response with a complementary log-log link,

$$h(\eta) = 1 - \exp\{-\exp(\eta)\},$$

and

$$h'(\eta) = \exp(\eta)\exp\{-\exp(\eta)\}.$$

The Bernoulli natural parameter is

$$\gamma(\eta) = \log\left\{\frac{h(\eta)}{1-h(\eta)}\right\},$$

which gives

$$\gamma'(\eta) = \frac{h'(\eta)}{h(\eta)\{1-h(\eta)\}}.$$

Consequently,

$$\begin{aligned} W(m) &= \frac{h'(\eta_0)}{\gamma'(\eta_0)} \\ &= h(\eta_0)\{1-h(\eta_0)\} \\ &= m(1-m), \end{aligned}$$

where

$$\eta_0 = \log\{-\log(1-m)\}.$$

As in the probit case,

$$\mathbb{E}\{\mathrm{Var}(Y \mid \boldsymbol{Z})\} = m(1-m) - \mathrm{Var}(\mu_Z) \le W(m).$$

Thus, for designs with bounded within-stratum imbalances, the standard model-based Wald test is asymptotically conservative when $\mathrm{Var}(\mu_Z) > 0$.

Under Conditions 1–2,

$$A + \frac{(K+1)\sigma_h^2}{2} - W(m) = -\operatorname{Var}(\mu_Z) + \frac{(K+1)\sigma_h^2}{2},$$

so the direction of the size distortion depends on the additional imbalance variance.

S4.3 Gamma Regression with Log Link

For a Gamma response, use the usual parameterization

$$\operatorname{Var}(Y \mid \boldsymbol{Z}) = \phi\mu_Z^2.$$

Under the log link,

$$h(\eta) = \exp(\eta), \qquad h'(\eta) = \exp(\eta).$$

The Gamma natural parameter is

$$\gamma(\eta) = -\exp(-\eta), \qquad \gamma'(\eta) = \exp(-\eta).$$

It follows that

$$\begin{aligned} W(m) &= \frac{\phi h'(\eta_0)}{\gamma'(\eta_0)} \\ &= \phi \exp(2\eta_0) \\ &= \phi m^2, \qquad \eta_0 = \log(m). \end{aligned}$$

On the other hand,

$$\begin{aligned} \mathbb{E}\{\operatorname{Var}(Y \mid \boldsymbol{Z})\} &= \phi\mathbb{E}(\mu_Z^2) \\ &= \phi\{m^2 + \operatorname{Var}(\mu_Z)\} \\ &\geq W(m). \end{aligned}$$

The inequality is strict when $\operatorname{Var}(\mu_Z) > 0$. Therefore, for designs with bounded within-stratum imbalances, the standard model-based Wald test is asymptotically anti-conservative in the presence of prognostic heterogeneity.

Under Conditions 1–2,

$$A + \frac{(K+1)\sigma_h^2}{2} - W(m) = \phi \operatorname{Var}(\mu_Z) + \frac{(K+1)\sigma_h^2}{2} \geq 0.$$

S4.4 Binary Indicator Obtained from an Ordinal Endpoint

The proportional-odds cumulative logit model for a full ordinal response contains multiple cutpoints and does not have the one-parameter GLM form assumed in Theorems 1–4. Therefore, those theorems cannot be applied directly to the full ordinal likelihood.

For a fixed threshold j, one may instead define the binary response

$$Y_i^{(j)} = I(Y_i \leq j)$$

and analyze this binary endpoint separately. Let

$$p_{j,Z} = P(Y \leq j \mid \boldsymbol{Z}), \qquad m_j = \mathbb{E}(p_{j,Z}).$$

For the separate Bernoulli logit working model,

$$W_j(m_j) = m_j(1 - m_j),$$

whereas

$$\begin{aligned}
\mathbb{E}\{\operatorname{Var}(Y^{(j)} \mid \boldsymbol{Z})\} &= \mathbb{E}\{p_{j,Z}(1 - p_{j,Z})\} \\
&= m_j(1 - m_j) - \operatorname{Var}(p_{j,Z}) \\
&\leq W_j(m_j).
\end{aligned}$$

For designs with bounded within-stratum imbalances, this establishes conservativeness only for a separately analyzed binary threshold.

Under Conditions 1–2, the sign is instead determined by

$$-\operatorname{Var}(p_{j,Z}) + \frac{(K+1)\sigma_{h,j}^2}{2},$$

where $\sigma_{h,j}^2$ denotes the Condition 2 imbalance variance for the binary response $Y^{(j)}$.

These calculations do not establish the corresponding result for a proportional-odds model fitted jointly across all thresholds.

S4.5 Negative Binomial Regression with Log Link

Consider the negative binomial parameterization

$$\operatorname{Var}(Y \mid \boldsymbol{Z}) = \mu_Z + \alpha\mu_Z^2, \qquad \alpha > 0,$$

where α is treated as fixed. Under the log link,

$$h(\eta) = \exp(\eta), \qquad h'(\eta) = \exp(\eta).$$

With $\phi = 1$, the corresponding natural parameter is

$$\gamma(\eta) = \log\left\{\frac{\alpha\exp(\eta)}{1+\alpha\exp(\eta)}\right\},$$

and hence

$$\gamma'(\eta) = \frac{1}{1+\alpha\exp(\eta)}.$$

Therefore,

$$\begin{aligned} W(m) &= \frac{h'(\eta_0)}{\gamma'(\eta_0)} \\ &= m(1+\alpha m) \\ &= m + \alpha m^2. \end{aligned}$$

The true average conditional variance is

$$\begin{aligned}\mathbb{E}\{\mathrm{Var}(Y \mid \boldsymbol{Z})\} &= \mathbb{E}(\mu_Z) + \alpha\mathbb{E}(\mu_Z^2) \\ &= m + \alpha\mathbb{E}(\mu_Z^2).\end{aligned}$$

Thus,

$$\begin{aligned}\mathbb{E}\{\mathrm{Var}(Y \mid \boldsymbol{Z})\} - W(m) &= \alpha\left\{\mathbb{E}(\mu_Z^2) - m^2\right\} \\ &= \alpha\,\mathrm{Var}(\mu_Z) \\ &\geq 0.\end{aligned}$$

For fixed $\alpha > 0$, the inequality is strict whenever $\mathrm{Var}(\mu_Z) > 0$. Thus, for designs with bounded within-stratum imbalances, the standard model-based Wald test is asymptotically anti-conservative when prognostic covariates are omitted.

Under Conditions 1–2,

$$A + \frac{(K+1)\sigma_h^2}{2} - W(m) = \alpha\,\mathrm{Var}(\mu_Z) + \frac{(K+1)\sigma_h^2}{2} \geq 0.$$

If α is estimated under a misspecified treatment-only model, its probability limit must be analyzed separately.

Bibliography

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

Hu, F. and L.-X. Zhang (2020). On the theory of covariate-adaptive designs. *arXiv preprint arXiv:2004.02994*.