Efficient Covariate-Adaptive Randomization with Smallest Selection Bias in Clinical Trials

Li-Xin Zhang

Zhejiang Gongshang University

Supplementary Material

The supplementary material provides the theoretical proof of all the theorems, extended discussion on the selection bias, and the simulation study of the new CAR with a specific allocation function g. The results on the power are shown in Section S1. The properties of the CAR procedures are derived in Section S2. The selection bias of the Pocock and Simon (1975) procedure and its generalization is derived in Section S3, while further discussion on the selection bias and randomness is given in Section S4. The simulation study is presented in Section S5.

S1 Deriving the asymptotic power

In this section, we prove the results on the power of hypothesis testing to compare treatment effects which are given in Section 2.

Proof of Theorem 1. Write $W_n = (D_n/\sqrt{n})^2 + Q_n^2$. Then

$$n - \ell_n^2 = W_n + O(1)\frac{W_n^4}{n} = W_n\big(1 + o(1)\big).$$

Let $H(x) = \ln F(t_{1-\alpha}(\nu); \nu, \sqrt{x})$. Then

$$H'(x) = \frac{1}{2}\left.\frac{\frac{d}{d\delta}F(t_{1-\alpha}(\nu); \nu, \delta)}{\delta F(t_{1-\alpha}(\nu); \nu, \delta)}\right|_{\delta=\sqrt{x}}.$$

We first show that

$$\frac{\frac{d}{d\delta}F(t; \nu, \delta)}{\delta F(t; \nu, \delta)} \to -1 \text{ as } \delta \to \infty \text{ and } \nu \to \infty \tag{S1.1}$$

uniformly in t on a bounded interval. Notice that

$$F(t; \nu, \delta) = \mathsf{E}\left[\Phi\big(t\sqrt{\chi^2(\nu)/\nu} - \delta\big)\right],$$

$$\frac{d}{d\delta}F(t; \nu, \delta) = -\mathsf{E}\left[\varphi\big(t\sqrt{\chi^2(\nu)/\nu} - \delta\big)\right],$$

where $\chi^2(\nu)$ is a random variable which has a χ^2 distribution with degree of freedom ν. Choose a sequence ϵ_ν such that $0 < \epsilon_\nu \to 0$, $\epsilon_\nu\delta^2 \to \infty$ and $\epsilon_\nu\nu \to \infty$. Let $E = \{\chi^2(\nu)/\nu \le \epsilon_\nu\delta^2\}$. Note

$$\varphi(\delta)\frac{\delta}{1+\delta^2} \le 1 - \Phi(\delta) = \Phi(-\delta) \le \varphi(\delta)\frac{1}{\delta}, \quad \varphi(-\delta) = \varphi(\delta).$$

On the event E,

$$\frac{\varphi\big(t\sqrt{\chi^2(\nu)/\nu} - \delta\big)}{\delta\Phi\big(t\sqrt{\chi^2(\nu)/\nu} - \delta\big)} \to 1$$

uniformly in t on a bounded interval and ω. Hence

$$\mathsf{E}\left[\varphi\big(t\sqrt{\chi^2(\nu)/\nu} - \delta\big)I_E\right] = \big(1 + o(1)\big)\delta\mathsf{E}\left[\Phi\big(t\sqrt{\chi^2(\nu)/\nu} - \delta\big)I_E\right]$$

uniformly in t on a bounded interval. On the other hand,

$$\begin{aligned}\mathsf{P}(E^c) \le & e^{-\nu\epsilon_\nu\delta^2/4}\mathsf{E}\left[e^{\chi^2(\nu)/4}\right] \le e^{-\nu(\epsilon_\nu\delta^2/4-1)} \le e^{-2\delta^2} \\ \le & e^{-\delta^2}\delta\Phi(-\delta) \le e^{-\delta^2}\delta F(t;\nu,\delta)\end{aligned}$$

when δ and ν are large enough. Hence

$$\mathsf{E}\left[\varphi\big(t\sqrt{\chi^2(\nu)/\nu}-\delta\big)\right] = \big(1+o(1)\big)\delta\mathsf{E}\left[\Phi\big(t\sqrt{\chi^2(\nu)/\nu}-\delta\big)\right]$$

uniformly in t on a bounded interval. (S1.1) is proved, which implies

$$H'(x) = \frac{1}{2}\left.\frac{\frac{d}{d\delta}F(t_{1-\alpha}(\nu);\nu,\delta)}{\delta F(t_{1-\alpha}(\nu);\nu,\delta)}\right|_{\delta=\sqrt{x}} = -\frac{1}{2}\big(1+o(1)\big)$$

as ν and x tend to infinity. So, for $0 \le y_n \le x_n \to \infty$ with $y_n/x_n \to 0$, we have

$$H(x_n^2) - H(x_n^2 - y_n^2) = y_n^2 H'(z_n) = -\frac{y_n^2}{2}\left(1+o(1)\right),$$

where $z_n \in (x_n^2 - y_n^2, x_n^2)$. Hence

$$\frac{F\left(t_{1-\alpha}(\nu);\nu,x_n\right)}{F\left(t_{1-\alpha}(\nu);\nu,\sqrt{x_n^2-y_n^2}\right)} = \exp\left\{-\frac{y_n^2}{2}\left(1+o(1)\right)\right\}. \tag{S1.2}$$

Now, let $x_n^2 = \frac{\mu^2}{4\sigma_\epsilon^2}n$, $y_n^2 = \frac{\mu^2}{4\sigma_\epsilon^2}(n-\ell_n^2)$. Then $y_n^2 = \frac{\mu^2}{2\sigma_\epsilon^2}W_n^2(1+o(1))$. It follows that

$$LossP_{T,n}(\mu|\boldsymbol{X}) = \exp\left\{-\frac{u^2}{8\sigma_\epsilon^2}W_n\big(1+o(1)\big)\right\} \tag{S1.3}$$

Finally, note $\overline{\boldsymbol{X}}_n \to 0 \ \ a.s.$ and

$$\begin{aligned}\frac{S_{n,xx}}{n} =& \frac{1}{n}\left\{\sum_{i=1}^{n}(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_n)^T(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_n)\right. \\ & \left. - N_{n,1}\big(\overline{\boldsymbol{X}}_{n,1}^T\overline{\boldsymbol{X}}_{n,1} - \overline{\boldsymbol{X}}_n^T\overline{\boldsymbol{X}}_n\big) - N_{n,2}\big(\overline{\boldsymbol{X}}_{n,2}^T\overline{\boldsymbol{X}}_{n,2} - \overline{\boldsymbol{X}}_n^T\overline{\boldsymbol{X}}_n\big)\right\} \\ \to & \mathrm{Var}(\boldsymbol{X}) = diag(\sigma_{x,1}^2, \cdots, \sigma_{x,I}^2) \ \ a.s.\end{aligned}$$

We have

$$\begin{aligned}W_n =& (D_n/\sqrt{n})^2 + \big(\boldsymbol{D}_n^x/\sqrt{n}\big)^T\big[\mathrm{Var}(\boldsymbol{X})\big]^{-1}\big(\boldsymbol{D}_n^x/\sqrt{n}\big) \\ & + o(1)\left((D_n/\sqrt{n})^2 + \big\|\boldsymbol{D}_n^x/\sqrt{n}\big\|^2\right) \\ =& V_n\big(1 + o(1)\big).\end{aligned}$$

This completes the proof. $\square$

Proof of Theorem 2. Notice that $\mathsf{E}[X_k|d_k(X_k)]$ is a function of $\tilde{X}_k = d_k(X_k)$. We write it as $f_k(\tilde{X}_k)$. Then $X_{i,k} = \delta_{i,k} + f_k(\tilde{X}_{i,k})$. Under the condition that $\frac{D_n(k;t_k)}{\sqrt{n}} \to 0$, $t_k = 1, \ldots, m_k$, $k = 1, \ldots, p$, we have

$$\frac{\sum_{i=1}^n (2T_i - 1) f_k(\tilde{X}_{i,k})}{\sqrt{n}} = \frac{\sum_{t_k=1}^{m_k} D_n(k;t_k) f_k(\tilde{x}_k^{t_k})}{\sqrt{n}} \to 0.$$

Because the probability of $T_i = 1$ is a function of $\tilde{\boldsymbol{X}}_1, \ldots, \tilde{\boldsymbol{X}}_i$, the sequence $(2T_i - 1)\big(\delta_{i,1}, \ldots, \delta_{i,I}\big)$, $i = 1, 2, \ldots$, is a sequence of martingale vector differences with

$$\mathrm{Var}\left((2T_i - 1)\big(\delta_{i,1}, \ldots, \delta_{i,I}\big)\big|\mathscr{F}_{i-1}\right) = diag\left(\sigma_{\delta,1}^2, \ldots, \sigma_{\delta,I}^2\right),$$

where $\mathscr{F}_i$ is the history σ-field generated by $T_1, \ldots, T_{i-1}$, $\tilde{\boldsymbol{X}}_1, \ldots, \tilde{\boldsymbol{X}}_{i-1}$. By the central limit theorem of martingales, we have

$$\frac{\sum_{i=1}^{n}(2T_i - 1)\big(\delta_{i,1}, \ldots, \delta_{i,I}\big)^T}{\sqrt{n}} \xrightarrow{d} N_I\left(\mathbf{0}, diag\left(\sigma_{\delta,1}^2, \ldots, \sigma_{\delta,I}^2\right)\right).$$

Hence

$$V_n \xrightarrow{d} \sum_{k=1}^{I} \frac{\sigma_{\delta,k}^2}{\sigma_{x,k}^2} \chi_k^2(1).$$

The result follows. $\square$

Proof of Remark 1. We first consider the two-sided t-test. Notice that the distribution function of T^2 is a F-distribution function $F(t; m, \nu, \delta)$ with degree of freedom $m = 1$ and $\nu = n - p - 2$, and the non-central parameter $\delta = \frac{\mu^2}{4\sigma_\epsilon^2}\ell_n^2$, it is sufficient to show that

$$\frac{\frac{d}{d\delta}F(t; m, \nu, \delta)}{F(t; m, \nu, \delta)} = -\frac{1}{2}\big(1 + o(1)\big) \tag{S1.4}$$

uniformly in $t \in [0, t_0]$ as $\nu \to \infty$ and $\delta \to \infty$. The non-central F-distribution has the following expansion

$$\begin{aligned} F(t; m, \nu, \delta) &= \sum_{j=0}^{\infty} e^{-\delta/2} \frac{(\delta/2)^j}{j!} F\left(\frac{m}{m + 2j}t; m + 2j, \nu, 0\right) \\ &= \sum_{j=0}^{\infty} e^{-\delta/2} \frac{(\delta/2)^j}{j!} \int_0^t f(x; m, \nu, j)dx, \end{aligned}$$

where

$$f(x; m, \nu, j) = \frac{\Gamma\big((m + \nu)/2 + j\big)}{\Gamma(m/2 + j)\Gamma(\nu)} \frac{\nu^{\nu/2} m^{m/2+j} x^{m/2-1+j}}{(\nu + mx)^{(\nu+m)/2+j}}.$$

So,

$$\begin{aligned}
&\frac{dF(t;m,\nu,\delta)}{d\delta}+\frac{1}{2}F(t;m,\nu,\delta)\\
&=\frac{1}{2}\sum_{j=0}^{\infty}e^{-\delta/2}\frac{(\delta/2)^j}{j!}\int_0^t f(x;m,\nu,j+1)dx\\
&=\frac{1}{2}\sum_{j=0}^{\infty}e^{-\delta/2}\frac{(\delta/2)^j}{j!}F\left(\frac{m}{m+2(j+1)}t;m+2(j+1),\nu,0\right).
\end{aligned}$$

It is obvious that

$$\begin{aligned}
&e^{-\delta/2}\frac{(\delta/2)^j}{j!}F\left(\frac{m}{m+2(j+1)}t;m+2(j+1),\nu,0\right)\\
=&\frac{2(j+1)}{\delta}e^{-\delta/2}\frac{(\delta/2)^{j+1}}{(j+1)!}F\left(\frac{m}{m+2(j+1)}t;m+2(j+1),\nu,0\right)\\
\leq&\frac{2(j+1)}{\delta}F(t;m,\nu,\delta)
\end{aligned}$$

and

$$\int_0^t f(x;m,\nu,j+1)dx\leq\frac{\nu+m+2j}{m+2j}\frac{mt}{\nu+mt}\int_0^t f(x;m,\nu,j)dx.$$

It follows that

$$\begin{aligned}
0\leq&\frac{dF(t;m,\nu,\delta)}{d\delta}+\frac{1}{2}F(t;m,\nu,\delta)\\
\leq&\frac{1}{2}\sum_{j=0}^{Km}\frac{2(j+1)}{\delta}F(t;m,\nu,\delta)\\
&+\frac{1}{2}\sum_{j=Km}^{\infty}e^{-\delta/2}\frac{(\delta/2)^j}{j!}\left(\frac{mt}{\nu}+\frac{mt}{m+2j}\right)\int_0^t f(x;m,\nu,j)dx\\
\leq&\left(\frac{2(Km+1)^2}{\delta}+\frac{mt}{\nu}+\frac{t}{2K+1}\right)F(t;m,\nu,\delta).
\end{aligned}$$

This proves (S1.4).

For the one-sided z-test for

$$H_0 : \mu = 0 \text{ versus } H_A : \mu > 0.$$

where $\mu = \mu_1 - \mu_2$ is the treatment effect difference, the test statistic is

$$U = \frac{\boldsymbol{L}\hat{\boldsymbol{\beta}}}{(\sigma_\epsilon^2 \boldsymbol{L}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{L}^T)^{1/2}}. \tag{S1.5}$$

The power function is

$$\beta_{U,n}(\mu|\boldsymbol{X}) = 1 - \Phi(z_{1-\alpha} - \sqrt{\delta}) \text{ with } \delta = \frac{u^2}{4\sigma_\epsilon^2}\ell_n^2.$$

It is obvious that

$$\frac{\frac{d}{d\delta}\Phi(z_{1-\alpha} - \sqrt{\delta})}{\Phi(z_{1-\alpha} - \sqrt{\delta})} \to -\frac{1}{2} \text{ as } \delta \to \infty.$$

We obtain a result similar to (S1.4).

For the two-sided z-test for

$$H_0 : \mu = 0 \text{ versus } H_A : \mu \neq 0,$$

the test statistic is U and the power function is

$$\beta_{|U|,n}(\mu|\boldsymbol{X}) = 1 - \Phi(z_{1-\alpha/2} - \sqrt{\delta}) + \Phi(-z_{1-\alpha/2} - \sqrt{\delta}) \text{ with } \delta = \frac{u^2}{4\sigma_\epsilon^2}\ell_n^2.$$

It is easily seen that

$$\begin{aligned}
&\frac{\frac{d}{d\delta}\left[\Phi(z_{1-\alpha/2} - \sqrt{\delta}) - \Phi(-z_{1-\alpha/2} - \sqrt{\delta})\right]}{\Phi(z_{1-\alpha/2} - \sqrt{\delta}) - \Phi(-z_{1-\alpha/2} - \sqrt{\delta})} \\
&= -\frac{1}{2\sqrt{\delta}}\frac{\varphi(z_{1-\alpha/2} - \sqrt{\delta}) - \varphi(-z_{1-\alpha/2} - \sqrt{\delta})}{\Phi(z_{1-\alpha/2} - \sqrt{\delta}) - \Phi(-z_{1-\alpha/2} - \sqrt{\delta})} \\
&\sim -\frac{1}{2\sqrt{\delta}}\frac{\varphi(z_{1-\alpha/2} - \sqrt{\delta})}{\Phi(z_{1-\alpha/2} - \sqrt{\delta})} \to -\frac{1}{2} \quad \text{as } \delta \to \infty.
\end{aligned}$$

We obtain a results similar to (S1.4). This completes the proof. □

S2 Deriving the properties of the CAR procedures

In this section, we prove the results on covariate adaptive randomization procedures which are given in Section 4.

Recall that $\boldsymbol{t} = (t_1, \ldots, t_I)$, $t_i = 1, \ldots, m_i, i = 1, \ldots, I$, has $M = \prod_{k=1}^{I} m_k$ values. Our purpose is to study the properties of $\boldsymbol{D}_n = [D_n(\boldsymbol{t})]$. Besides $\boldsymbol{D}_n$, we will also consider the weighted average of the imbalances $\Lambda_{n-1}(\boldsymbol{t})$ as in (2.1). Let

$$\begin{aligned}\Lambda_n(\boldsymbol{t}) =& w_o D_n + \sum_{i=1}^{I} w_{m,i} D_n(i; t_i) + w_s D_n(\boldsymbol{t}),\\ \boldsymbol{\Lambda}_n =& [\Lambda_n(\boldsymbol{t})]_{1 \le t_1 \le m_1, \ldots, 1 \le t_I \le m_I}.\end{aligned}$$

Also let $\mathscr{F}_{n-1}$ be the history σ-field generated by the covariates $\tilde{\boldsymbol{X}}_1, \ldots, \tilde{\boldsymbol{X}}_{n-1}$ and results of allocation $T_1, \ldots, T_{n-1}$. Then the allocation probability of the nth patient is

$$\mathsf{P}\left(T_n = 1 \middle| \mathscr{F}_{n-1}, \tilde{\boldsymbol{X}}_n = (\tilde{x}_1^{t_1}, \ldots, \tilde{x}_I^{t_I})\right) = g_n\left(4\Lambda_{n-1}(\boldsymbol{t})\right).$$

is a function of $\boldsymbol{\Lambda}_{n-1}$. It is obvious that $\boldsymbol{\Lambda}_n = \boldsymbol{L}(\boldsymbol{D}_n) : \boldsymbol{D}_n \to \boldsymbol{\Lambda}_n$ is a linear transform of $\boldsymbol{D}_n$. The following proposition gives the relation between $\boldsymbol{D}_n$ and $\boldsymbol{\Lambda}_n$ and tells us that both $(\boldsymbol{D}_n)_{n\ge1}$ and $(\boldsymbol{\Lambda}_n)_{n\ge1}$ are Markov chains.

Proposition S2.1. *(i) If $w_s > 0$, then $\mathbf{\Lambda}_n = \boldsymbol{L}(\boldsymbol{D}_n)$ is a one to one linear map; If $w_s + w_{m,i} > 0$, then each $D_n(i;t_i) = D_{i;t_i}(\mathbf{\Lambda}_n)$ is a linear transform of $\mathbf{\Lambda}_n$; For any case, $D_n = D(\mathbf{\Lambda}_n)$ is a linear transform of $\mathbf{\Lambda}_n$.*

(ii) $(\boldsymbol{D}_n)_{n\geq 1}$ is a non-homogeneous Markov chain on the space $\mathbb{Z}^m$ with period 2;

(iii) $(\mathbf{\Lambda}_n)_{n\geq 1}$ is a non-homogeneous Markov chain on the space $\boldsymbol{L}(\mathbb{Z}^m)$ with period 2.

Proof. For (i), taking the summation of $\Lambda_n(\boldsymbol{t})$ over all $\boldsymbol{t}$ yields

$$\sum_{\boldsymbol{t}} \Lambda_n(\boldsymbol{t}) = \big(w_s + \sum_{i=1}^{I} w_{m,i} \prod_{j\neq i} m_j + w_o M\big) D_n.$$

So D_n is a linear transform of $\mathbf{\Lambda}_n$. Taking the summation of $\Lambda_n(\boldsymbol{t})$ over all $t_1, \ldots, t_I$ except t_i yields

$$\begin{aligned}
&\sum_{t_1,\ldots,t_{i-1},t_{i+1},\ldots,t_I} \Lambda_n(t_1, \ldots, t_I) \\
=&\big(w_s + w_{m,i} \prod_{j\neq i} m_j\big) D_n(i;t_i) + \big(\sum_{l\neq i} w_{m,l} \prod_{j\neq i,l} m_j + w_o \prod_{j\neq i} m_j\big) D_n.
\end{aligned}$$

Hence, when $w_s + w_{m,i} > 0$, each $D_n(i;t_i)$ is a linear transform of $\mathbf{\Lambda}_n$ and D_n, and so it is a linear transform of $\mathbf{\Lambda}_n$. Finally, when $w_s > 0$, it is obvious that each $D_n(\boldsymbol{t})$ is a linear transform of $\Lambda_n(\boldsymbol{t})$, $D_n(1;t_1), \ldots, D_n(I;t_I)$ and D_n, and so it is a linear transform of $\mathbf{\Lambda}_n$. Hence, when $w_s > 0$, $\mathbf{\Lambda}_n = \boldsymbol{L}(\boldsymbol{D}_n)$ is a one to one linear map.

For (ii) and (iii), notice

$$D_n(\boldsymbol{t}) = D_{n-1}(\boldsymbol{t}) + 2\big(T_n - \frac{1}{2}\big)\mathbb{I}\{\tilde{\boldsymbol{X}}_n = (\tilde{x}_1^{t_1}, \ldots, \tilde{x}_I^{t_I})\}.$$

Then

$$\begin{aligned}
\mathsf{P}(\Delta D_n(\boldsymbol{t}) = 1|\mathscr{F}_{n-1}) =& g_n(4\Lambda_{n-1}(\boldsymbol{t}))p(\boldsymbol{t}),\\
\mathsf{P}(\Delta D_n(\boldsymbol{t}) = -1|\mathscr{F}_{n-1}) =& \big[1 - g_n\big(4\Lambda_{n-1}(\boldsymbol{t})\big)\big]p(\boldsymbol{t}),\\
\mathsf{P}(\Delta D_n(\boldsymbol{t}) = 0|\mathscr{F}_{n-1}) =& 1 - p(\boldsymbol{t}),
\end{aligned}$$

where $p(\boldsymbol{t}) = p(t_1, \ldots, t_I) = \mathsf{P}(\tilde{\boldsymbol{X}}_n = (\tilde{x}_1^{t_1}, \ldots, \tilde{x}_I^{t_I}))$, $\boldsymbol{p} = \big(p(\boldsymbol{t}), \boldsymbol{t} \in \mathbb{Z}^{m_1 \times \ldots \times m_I}\big)$. Let $\Delta\mathscr{D}$ be the state space of $\Delta\boldsymbol{D}_n = \boldsymbol{D}_n - \boldsymbol{D}_{n-1}$, i.e., each $\boldsymbol{d} \in \Delta\mathscr{D}$ has only one non-zero element which is 1 or -1. For two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ on $\mathbb{Z}^{m_1 \times \ldots \times m_I}$, we write $\boldsymbol{x} \cdot \boldsymbol{y} = \sum_{\boldsymbol{t}} x(\boldsymbol{t})y(\boldsymbol{t})$, $|\boldsymbol{x}| = (|x(\boldsymbol{t})|)$. The conditional probability above can be written in the following form,

$$\begin{aligned}
\mathsf{P}(\Delta\boldsymbol{D}_n = \boldsymbol{d}|\mathscr{F}_{n-1}) &= \left(g_n(4|\boldsymbol{d}| \cdot \boldsymbol{\Lambda}_{n-1}) - \frac{1}{2}\right)\boldsymbol{d} \cdot \boldsymbol{p} + \frac{1}{2}|\boldsymbol{d} \cdot \boldsymbol{p}| \qquad \text{(S2.1)}\\
&= \left(g_n\left(4|\boldsymbol{d}| \cdot \boldsymbol{L}(\boldsymbol{D}_{n-1})\right) - \frac{1}{2}\right)\boldsymbol{d} \cdot \boldsymbol{p} + \frac{1}{2}|\boldsymbol{d} \cdot \boldsymbol{p}|,\\
&\quad \boldsymbol{d} \in \Delta\mathscr{D},
\end{aligned}$$

which depends only on $\boldsymbol{\Lambda}_{n-1} = \boldsymbol{L}(\boldsymbol{D}_{n-1})$. So, conditional on $\boldsymbol{D}_{n-1}$, $\boldsymbol{D}_n$ is conditionally independent of $(\boldsymbol{D}_1, \ldots, \boldsymbol{D}_{n-2})$. It follows that $(\boldsymbol{D}_n)_{n\geq 1}$ is a Markov chain on $\mathbb{Z}^m$.

For (iii), notice that for any point $\boldsymbol{e}$ in the state space $\{\boldsymbol{L}(\boldsymbol{d}) : \boldsymbol{d} \in \Delta\mathscr{D}\}$

of $\Delta\mathbf{\Lambda}_n$,

$$\mathsf{P}(\Delta\mathbf{\Lambda}_n = \boldsymbol{e}|\mathscr{F}_{n-1}) =$$
$$\sum_{\boldsymbol{d}:\boldsymbol{L}(\boldsymbol{d})=\boldsymbol{e},\boldsymbol{d}\in\Delta\mathscr{D}} \left\{ \left(g_n(4|\boldsymbol{d}|\cdot\mathbf{\Lambda}_{n-1}) - \frac{1}{2} \right) \boldsymbol{d}\cdot\boldsymbol{p} + \frac{1}{2}|\boldsymbol{d}\cdot\boldsymbol{p}| \right\}, \tag{S2.2}$$

which depends only on $\mathbf{\Lambda}_{n-1}$. So, given $\mathbf{\Lambda}_{n-1}$, $\mathbf{\Lambda}_n$ is conditionally independent of $(\mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_{n-2})$. It follows that $\mathbf{\Lambda}_n$ is a Markov chain. This completes the proof of Proposition S2.1. $\square$

Proof of Theorem 4. Recall

$$M_n = \sum_{\boldsymbol{t}} w_s D_n^2(\boldsymbol{t}) + \sum_{i=1}^{I}\sum_{t_i=1}^{m_i} w_{m,i} D_n^2(i;t_i) + w_o D_n^2.$$

By Proposition S2.1 (i), $M_n = M^*(\mathbf{\Lambda}_n)$ is a function of $\mathbf{\Lambda}_n$, and $M_n \leq C\|\mathbf{\Lambda}_n\|^2$. On the other hand,

$$\begin{aligned}
|\Lambda_n(\boldsymbol{t})|^2 &\leq \big(w_o|D_n| + \sum_{i=1}^{I} w_{m,i}|D_n(i;t_i)| + w_s|D_n(\boldsymbol{t})|\big)^2 \\
&\leq \big(w_o|D_n|^2 + \sum_{i=1}^{I} w_{m,i}|D_n(i;t_i)|^2 + w_s|D_n(\boldsymbol{t})|^2\big)(w_o + \sum_{i=1}^{I} w_{m,i} + w_s), \\
&= w_o|D_n|^2 + \sum_{i=1}^{I} w_{m,i}|D_n(i;t_i)|^2 + w_s|D_n(\boldsymbol{t})|^2,
\end{aligned}$$

which implies that $\|\mathbf{\Lambda}_n\|^2 \leq M \cdot M_n$. We will prove the theorem via two steps. First, we will show that under condition

$$g_n(x) \leq 1/2 \leq g_n(-x) \text{ when } x \geq 0, \tag{S2.3}$$

we have

$$M_n = O(n) \ \text{ in } L_r \ \ \forall r > 0. \tag{S2.4}$$

In the second step, we will show that under conditions (S2.3) and

$$\frac{|1/2 - g_n(x_n)|}{|x_n|/n} \to +\infty \ \text{ whenever } \ 0 < \frac{|x_n|}{n} \to 0, \tag{S2.5}$$

we have

$$M_n = o(n) \quad \text{in probability}, \tag{S2.6}$$

which, together with (S2.4), implies that $M_n = o(n)$ in L_r for any $r > 0$. (i)-(iii) follow from immediately from this. Finally,

$$\mathsf{E}\left[p_n \vee (1-p_n)\right] = \frac{1}{2} + \sum_{\boldsymbol{t}} p(\boldsymbol{t})\mathsf{E}\left|\frac{1}{2} - g_n\big(4\Lambda_{n-1}(\boldsymbol{t})\big)\right| \to \frac{1}{2}$$

due to the condition that

$$g_n(x_n) \to \frac{1}{2} \text{ whenever } \ \frac{x_n}{n} \to 0, \tag{S2.7}$$

and the fact that $\frac{\|\boldsymbol{\Lambda}_n\|}{n} \leq C\frac{M_n^{1/2}}{n} \to 0$ in probability. Hence

$$SB_n = \frac{1}{n}\sum_{m=1}^{n} \mathsf{E}\left[p_m \vee (1-p_m)\right] \to \frac{1}{2}.$$

Now, we begin the proofs of (S2.4) and (S2.6). Given $\tilde{\boldsymbol{X}}_n = (\tilde{x}_1^{t_1}, \ldots, \tilde{x}_I^{t_I})$, if $T_n = 1$, then

$$\begin{aligned} M_n - M_{n-1} =& w_s \left\{ \left(D_{n-1}(\boldsymbol{t}) + 1\right)^2 - D_{n-1}^2(\boldsymbol{t}) \right\} \\ & + \sum_{i=1}^{I} w_{m,i} \left\{ \left(D_{n-1}(i;t_i) + 1\right)^2 - D_{n-1}^2(i;t_i) \right\} \\ & + w_o \left\{ (D_{n-1} + 1)^2 - D_{n-1}^2 \right\} \\ =& 2\Lambda_{n-1}(\boldsymbol{t}) + 1, \end{aligned}$$

while, if $T_n = 0$, then $M_n - M_{n-1} = -2\Lambda_{n-1}(\boldsymbol{t}) + 1$. So,

$$M_n - M_{n-1} = 2\Lambda_{n-1}(\boldsymbol{t})\,(2T_n - 1)\,\mathbb{I}\{\tilde{\boldsymbol{X}}_n = (\tilde{x}_1^{t_1}, \ldots, \tilde{x}_I^{t_I})\} + 1. \tag{S2.8}$$

Hence

$$\begin{aligned} & \mathsf{E}\left[M_n - M_{n-1} \middle| \tilde{\boldsymbol{X}}_n = (x_1^{t_1}, \ldots, x_I^{t_I}), \mathscr{F}_{n-1} \right] \\ =& 2\Lambda_{n-1}(\boldsymbol{t})\,[2g_n(4\Lambda_{n-1}(\boldsymbol{t})) - 1] + 1 \\ =& -4\left|\Lambda_{n-1}(\boldsymbol{t})\right| \cdot \left| \frac{1}{2} - g_n\left(4\Lambda_{n-1}(\boldsymbol{t})\right) \right| + 1, \end{aligned}$$

due to the condition that $g_n(-x) \leq 1/2 \leq g_n(x)$ when $x \geq 0$. It follows that

$$\mathsf{E}\left[M_n \middle| \mathscr{F}_{n-1} \right] - M_{n-1} = -4S_n(\boldsymbol{\Lambda}_{n-1}) + 1, \tag{S2.9}$$

where

$$S_n(\boldsymbol{\Lambda}_{n-1}) = \sum_{\boldsymbol{t}} \left|\Lambda_{n-1}(\boldsymbol{t})\right| \cdot \left| \frac{1}{2} - g_n\left(4\Lambda_{n-1}(\boldsymbol{t})\right) \right| p(\boldsymbol{t}) \tag{S2.10}$$

is a nonnegative function of $\boldsymbol{\Lambda}_{n-1}$.

From (S2.9) and by noting $M_0 = 0$, it follows that

$$\mathsf{E}\left[M_n\right] \leq n. \tag{S2.11}$$

Further, by (S2.8) we have

$$\begin{aligned} M_n =& M_{n-1} + 1 + 2\Lambda_{n-1}(\boldsymbol{t})(2T_n - 1)\mathbb{I}\{\tilde{\boldsymbol{X}}_n = (\tilde{x}_1^{t_1}, \ldots, \tilde{x}_I^{t_I})\} \\ \widehat{=}& M_{n-1} + 1 + \xi. \end{aligned}$$

It is obvious that

$$|\xi| = 2|\Lambda_{n-1}(\boldsymbol{t})| \leq 2\sqrt{M_{n-1}}, \quad \mathsf{E}[\xi|\mathscr{F}_{n-1}] = -4S_n(\boldsymbol{\Lambda}_{n-1}).$$

It follows that for positive integer r,

$$\begin{aligned} M_n^{r+1} - M_{n-1}^{r+1} =& (r+1)(M_{n-1}+1)^r\xi \\ &+ \left\{(M_{n-1}+1)^{r+1} - M_{n-1}^{r+1} + \sum_{k=2}^{r+1}\binom{r+1}{k}\xi^k(M_{n-1}+1)^{r+1-k}\right\} \\ &\leq (r+1)(M_{n-1}+1)^r\xi + C_r(M_{n-1}+1)^r, \end{aligned}$$

where C_r is a constant which depends on r. It follows that

$$\begin{aligned} &\mathsf{E}[M_n^{r+1}|\mathscr{F}_{n-1}] - M_{n-1}^{r+1} \\ \leq& -4(r+1)(M_{n-1}+1)^r S_n(\boldsymbol{\Lambda}_{n-1}) + C_r(M_{n-1}+1)^r \qquad \text{(S2.12)} \\ \leq& C_r(M_{n-1}+1)^r. \end{aligned}$$

By (S2.11) and the induction we conclude (S2.4). Obviously, (S2.4) implies

$$\frac{\|\mathbf{\Lambda}_n\|}{n} \to 0 \text{ in probability.}$$

Next, we consider (S2.6). By (S2.9),

$$\mathsf{E}\left[M_n\right] - \mathsf{E}\left[M_{n-1}\right] = -4\mathsf{E}\left[S_n(\mathbf{\Lambda}_{n-1})\right] + 1. \tag{S2.13}$$

Let $m := m_n \in \{0, 1, \ldots, n\}$ be the last one for which $1 - 4\mathsf{E}\left[S_{m+1}(\mathbf{\Lambda}_m)\right] \geq 0$. Then

$$\mathsf{E}\left[M_n\right] \leq \mathsf{E}\left[M_{m+1}\right] \leq \mathsf{E}\left[M_m\right] + 1$$

and

$$4\mathsf{E}\left[S_{m+1}(\mathbf{\Lambda}_m)\right] \leq 1.$$

If $m = m_n$ is bounded, the proof is completed. Assume $m \to \infty$. Recall (S2.10). Notice that the condition (S2.5) implies

$$\frac{S_{m+1}(\mathbf{\Lambda}_m)}{\frac{1}{m}\sum_{\boldsymbol{t}} |\Lambda_m(\boldsymbol{t})|^2 p(\boldsymbol{t})} \to +\infty \;\; as \;\; \frac{\|\mathbf{\Lambda}_m\|}{m} \to 0.$$

On the other hand,

$$M_m \geq \sum_{\boldsymbol{t}} |\Lambda_m(\boldsymbol{t})|^2 p(\boldsymbol{t}) \geq c_0 M_m.$$

So for any $0 < \epsilon < 1$, there exists a $\delta > 0$ such that

$$S_{m+1}(\mathbf{\Lambda}_m) I\Big\{\frac{\|\mathbf{\Lambda}_m\|}{m} \leq \delta\Big\} \geq \frac{1}{\epsilon}\frac{1}{m} M_m I\Big\{\frac{\|\mathbf{\Lambda}_m\|}{m} \leq \delta\Big\}.$$

So

$$\mathsf{E}\Big[M_m I\Big\{\frac{\|\boldsymbol{\Lambda}_m\|}{m} \le \delta\Big\}\Big] \le \epsilon m \mathsf{E}\Big[S_{m+1}(\boldsymbol{\Lambda}_m)\Big] \le \frac{\epsilon m}{4}.$$

Also,

$$\begin{aligned}\mathsf{E}\Big[M_m I\Big\{\frac{\|\boldsymbol{\Lambda}_m\|}{m} > \delta\Big\}\Big] \le & \Big\{\mathsf{E}\Big[M_m^2\Big]\Big\}^{1/2}\Big\{\mathsf{P}\Big(\frac{\|\boldsymbol{\Lambda}_m\|}{m} > \delta\Big)\Big\}^{1/2} \\ \le & cm\mathsf{P}^{1/2}\Big(\frac{\|\boldsymbol{\Lambda}_m\|}{m} > \delta\Big).\end{aligned}$$

It follows that

$$\mathsf{E}\left[M_n\right] \le n\Big\{\frac{\epsilon}{4} + c\mathsf{P}^{1/2}\Big(\frac{\|\boldsymbol{\Lambda}_m\|}{m} > \delta\Big)\Big\} + 1.$$

By the fact $\mathsf{P}\Big(\frac{\|\boldsymbol{\Lambda}_m\|}{m} > \delta\Big) \to 0$, we conclude that $\frac{\mathsf{E}[M_n]}{n} \to 0$. This proves (S2.6). The proof of Theorem 4 is completed. □

Proof of Theorem 5. Notice that, by (S2.12),

$$\mathsf{E}[M_n^{r+1}] - \mathsf{E}[M_{n-1}^{r+1}] \le -4(r+1)\mathsf{E}\left[(M_{n-1}+1)^r S_n(\boldsymbol{\Lambda}_{n-1})\right] + C_r\mathsf{E}\left[(M_{n-1}+1)^r\right].$$

Let $m := m_n \in \{0, 1, \ldots, n\}$ be the last one for which

$$-4(r+1)\mathsf{E}\left[(M_m+1)^r S_{m+1}(\boldsymbol{\Lambda}_m)\right] + C_r\mathsf{E}\left[(M_m+1)^r\right] \ge 0. \qquad \text{(S2.14)}$$

Then

$$\mathsf{E}[M_n^{r+1}] \le \mathsf{E}[M_{m+1}^{r+1}] \le \mathsf{E}[M_m^{r+1}] + C_r\mathsf{E}\left[(M_m+1)^r\right]. \qquad \text{(S2.15)}$$

Recall (S2.10). Note for $\delta > 0$ small enough, on the event $E = E(\boldsymbol{t}) =$

$\left\{\frac{|\Lambda_m(\boldsymbol{t})|}{m^\gamma} \le \delta\right\}$,

$$S_{m+1}(\boldsymbol{\Lambda}_m) \ge \frac{-g'(0)}{2}\frac{1}{m^\gamma}|\Lambda_m(\boldsymbol{t})|^2 p(\boldsymbol{t}) \ge c_0 \frac{1}{m^\gamma}|\Lambda_m(\boldsymbol{t})|^2,$$

and on the event E^c,

$$S_{m+1}(\boldsymbol{\Lambda}_m) \ge \big(\frac{1}{2} - g(c\delta)\big) \vee \big(\frac{1}{2} - g(-c\delta)\big)|\Lambda_m(\boldsymbol{t})|p(\boldsymbol{t}) \ge c_0|\Lambda_m(\boldsymbol{t})|.$$

$C^{-1}M_n \le \|\boldsymbol{\Lambda}_n\|^2 \le C \cdot M_n$. Then by (S2.14),

$$\mathsf{E}[|\Lambda_m(\boldsymbol{t})|^{2r+2}E] \le Cm^\gamma \mathsf{E}\left[(M_m+1)^r S_{m+1}(\boldsymbol{\Lambda}_m)\right] \le C_r m^\gamma \mathsf{E}\left[(M_m+1)^r\right],$$

and

$$\mathsf{E}[|\Lambda_m(\boldsymbol{t})|^{2r+1}E^c] \le C\mathsf{E}\left[(M_m+1)^r S_{m+1}(\boldsymbol{\Lambda}_m)\right] \le C_r \mathsf{E}\left[(M_m+1)^r\right].$$

By the Hölder inequality,

$$\begin{aligned}
\mathsf{E}[|\Lambda_m(\boldsymbol{t})|^{2r+2}E^c] &= \mathsf{E}\left[|\Lambda_m(\boldsymbol{t})|^{\frac{2r+1}{p}}|\Lambda_m(\boldsymbol{t})|^{\frac{2r+1}{q}+1}E^C\right] \\
&\le \left(\mathsf{E}\left[|\Lambda_m(\boldsymbol{t})|^{2r+1}E^c\right]\right)^{1/p}\left(\mathsf{E}\left[|\Lambda_m(\boldsymbol{t})|^{2r+1+q}E^c\right]\right)^{1/q} \\
&\le C_r\left(\mathsf{E}\left[(M_m+1)^r\right]\right)^{1/p}\left(\mathsf{E}\left[M_m^{r+\frac{1}{2}+\frac{1}{2}q}\right]\right)^{1/q} \\
&\le C_r\left(\mathsf{E}\left[(M_m+1)^r\right]\right)^{1/p}\left(\mathsf{E}\left[M_m^{2r+1+q}\right]\right)^{1/2q},
\end{aligned}$$

where $p, q > 1$, $1/p + 1/q = 1$. It follows that

$$\begin{aligned}
\mathsf{E}[M_n^{r+1}] \le& \mathsf{E}[M_m^{r+1}] + C_r\mathsf{E}\left[(M_m+1)^r\right] \le c_r\sum_{\boldsymbol{t}}\mathsf{E}[|\Lambda_m(\boldsymbol{t})|^{2r+2}] + C_r\mathsf{E}\left[(M_m+1)^r\right] \\
\le& C_r m^\gamma\mathsf{E}\left[(M_m+1)^r\right] + C_r\left(\mathsf{E}\left[(M_m+1)^r\right]\right)^{1/p}\left(\mathsf{E}\left[M_m^{2r+1+q}\right]\right)^{1/2q}.
\end{aligned} \tag{S2.16}$$

We will use the induction method to obtain the moment estimates of $\mathsf{E}[M_n^r]$ by inducting both in n and r. We assume

$$\mathsf{E}[M_n^s] \leq C_s n^{\beta s} \text{ for all integers } n, s \geq 0, \tag{S2.17}$$

and

$$\mathsf{E}[M_n^r] \leq C_r n^{\alpha r} \text{ for all } n, \tag{S2.18}$$

where $0 < \beta/2 < \alpha \leq \beta \leq 2$.

Let q be an integer large enough such that

$$\frac{\alpha r}{p} + \beta \frac{2r+1}{2q} + \frac{1}{2}\beta = \alpha r + \frac{r(\beta - \alpha) + \beta/2}{q} + \frac{1}{2}\beta \leq (r+1)\alpha.$$

Then

$$\begin{aligned}\mathsf{E}[M_m^{r+1}] \leq & C_r m^\gamma m^{\alpha r} + C_r \left(m^{\alpha r}\right)^{1/p} \left(m^{\beta(2r+1+q)}\right)^{1/2q} \\ \leq & C_r \left(m^\gamma m^{\alpha r} + m^{\alpha(r+1)}\right).\end{aligned}$$

By (S2.15),

$$\mathsf{E}[M_n^{r+1}] \leq C_r \big(n^\gamma n^{\alpha r} + n^{\alpha(r+1)}\big) \leq C_{r+1} n^{\alpha(r+1)} \quad \text{for all } n, \tag{S2.19}$$

whenever $\alpha \geq \gamma$.

Obviously, (S2.17) holds for $\beta = 2$ since $M_n \leq n^2$, and (S2.18) is true for $r = 0$. So, by (S2.19) and the induction, (S2.18) holds for all integer r whenever $\alpha > 1$, which implies that (S2.17) holds for any β with $\beta \geq \gamma$ and $\beta > 1$. Repeating the above procedure i times concludes that (S2.17) holds for any β with $\beta \geq \gamma$ and $\beta > 1/2^{i-1}$. Stopping the procedure when $1/2^{i-1} < \gamma$, we conclude that

$$\mathsf{E} M_n^r \leq C_r n^{\gamma r} \quad \text{for all integers } n \text{ and } r.$$

Thus,

$$M_n = O(n^\gamma) \quad \text{in } L_r \text{ for all } r > 0, \tag{S2.20}$$

is proved. By (S2.20), for any $\epsilon, \delta > 0$,

$$\sum_n \mathsf{P}\left(M_n \geq \delta n^{\gamma+\epsilon}\right) \leq \sum_n \frac{\mathsf{E} M_n^Q}{\delta^Q n^{Q\gamma} n^{\epsilon Q}} \leq C \leq \sum_n n^{-\epsilon Q} < \infty$$

when $Q > 1/\epsilon$. By the Borel-Cantelli lemma,

$$M_n = o(n^{\gamma+\epsilon}) \ a.s. \text{ for any } \epsilon > 0 \tag{S2.21}$$

is proved. And so, (i)-(iii) follow from Proposition S2.1.

Finally, we consider the selection bias. Note

$$\left|\frac{1}{2} - g_{m+1}\left(4\Lambda_m(\boldsymbol{t})\right)\right| = -4g'(0)\frac{\left|\Lambda_m(\boldsymbol{t})\right|}{m^\gamma} + 4\theta_m(\boldsymbol{t})\frac{\left|\Lambda_m(\boldsymbol{t})\right|}{m^\gamma}, \tag{S2.22}$$

where

$$\theta_m(\boldsymbol{t}) = \frac{\left|\frac{1}{2} - g\left(4\frac{\Lambda_m(\boldsymbol{t})}{m^\gamma}\right)\right|}{\frac{4|\Lambda_m(\boldsymbol{t})|}{m^\gamma}} + 4g'(0)$$

is bounded and converges to 0 in probability due to the fact that $0 \leq g(x) \leq 1$, $g'(0)$ is finite and $\frac{|\Lambda_m(\boldsymbol{t})|}{m^\gamma} \to 0$ in probability. So, $\mathsf{E}\theta_m(\boldsymbol{t})^2 = o(1)$ as $m \to \infty$. By the Cauchy-Schwartz inequality,

$$\mathsf{E}\left|\theta_m(\boldsymbol{t})\frac{\left|\Lambda_m(\boldsymbol{t})\right|}{m^\gamma}\right| = o(1) \cdot \left(\mathsf{E}\left|\frac{\left|\Lambda_m(\boldsymbol{t})\right|}{m^\gamma}\right|^2\right)^{1/2}.$$

It follows that

$$\begin{aligned}
&\mathsf{E}\left[p_m \vee (1-p_m)\right] = \frac{1}{2} + \sum_{\boldsymbol{t}} p(\boldsymbol{t})\mathsf{E}\left|\frac{1}{2} - g_m\left(4\Lambda_{m-1}(\boldsymbol{t})\right)\right| \\
=&\frac{1}{2} + \sum_{\boldsymbol{t}} p(\boldsymbol{t})\mathsf{E}\left[-4g'(0)\frac{|\Lambda_{m-1}(\boldsymbol{t})|}{(m-1)^{\gamma}}\right] + \sum_{\boldsymbol{t}} p(\boldsymbol{t})o(1)\left(\mathsf{E}\left|\frac{|\Lambda_{m-1}(\boldsymbol{t})|}{(m-1)^{\gamma}}\right|^2\right)^{1/2} \\
=&\frac{1}{2} - 4g'(0)\sum_{\boldsymbol{t}} p(\boldsymbol{t})\frac{\mathsf{E}\left[|\Lambda_{m-1}(\boldsymbol{t})|\right]}{(m-1)^{\gamma}} + o\left(\frac{1}{(m-1)^{\gamma/2}}\right), \qquad \text{(S2.23)}
\end{aligned}$$

where, the last equality is due to the fact that $\|\boldsymbol{\Lambda}_m\| \le CM_m^{1/2} = O(m^{\gamma/2})$ in any L_r. Hence

$$\begin{aligned}
SB_n =&\frac{1}{n}\sum_{m=1}^{n}\mathsf{E}\left[p_m \vee (1-p_m)\right] \\
=&\frac{1}{2} - 4g'(0)\frac{1}{n}\sum_{m=1}^{n}\sum_{\boldsymbol{t}} p(\boldsymbol{t})\frac{\mathsf{E}\left[|\Lambda_m(\boldsymbol{t})|\right]}{m^{\gamma}} + o\left(\frac{1}{n^{\gamma/2}}\right) \\
=:&\frac{1}{2} + a_n n^{-\gamma/2} + o\left(n^{-\gamma/2}\right).
\end{aligned}$$

Next, we show a_n is bounded away from zero and infinity. From (S2.13), we conclude that

$$0 = \lim_{n\to\infty}\frac{\mathsf{E}[M_{n+1}]}{n} = -4\lim_{n\to\infty}\frac{1}{n}\sum_{m=1}^{n}\mathsf{E}[S_{m+1}(\boldsymbol{\Lambda}_m)] + 1.$$

By noting (S2.22), similar to (S2.23) we have

$$\begin{aligned}
\mathsf{E}[S_{m+1}(\boldsymbol{\Lambda}_m)] =& -g'(0)\sum_{\boldsymbol{t}} p(\boldsymbol{t})\frac{\mathsf{E}|\Lambda_m(\boldsymbol{t})|^2}{m^{\gamma}} + o(1)\sum_{\boldsymbol{t}} p(\boldsymbol{t})\mathsf{E}\left(\frac{\mathsf{E}|\Lambda_m(\boldsymbol{t})|^2}{m^{\gamma}}\right)^2 \\
=& -g'(0)\sum_{\boldsymbol{t}} p(\boldsymbol{t})\frac{\mathsf{E}|\Lambda_m(\boldsymbol{t})|^2}{m^{\gamma}} + o(1).
\end{aligned}$$

It follows that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{m=1}^{n}\sum_{\boldsymbol{t}} p(\boldsymbol{t})\frac{\mathsf{E}|\Lambda_m(\boldsymbol{t})|^2}{m^\gamma} = \frac{1}{-16g'(0)}.$$

Regarding the summation on the right-hand side above as $\mathsf{E}|Z_n|^2$ of a random variable $Z_n = \frac{\Lambda_\tau(\boldsymbol{\kappa})}{\tau^{\gamma/2}}$, where $\mathsf{P}(\tau = m) = \frac{1}{n}$, $\mathsf{P}(\boldsymbol{\kappa} = \boldsymbol{t}) = p(\boldsymbol{t})$, then

$$\frac{1}{n}\sum_{m=1}^{n}\sum_{\boldsymbol{t}} p(\boldsymbol{t})\frac{\mathsf{E}|\Lambda_m(\boldsymbol{t})|}{m^{\gamma/2}} = \mathsf{E}|Z_n| \le (\mathsf{E}|Z_n|^2)^{1/2} \sim \frac{1}{4\sqrt{-g'(0)}}.$$

Hence

$$\begin{aligned}
a_n n^{-\gamma/2} &= -4g'(0)\frac{1}{n}\sum_{m=1}^{n}\frac{m\mathsf{E}|Z_m| - (m-1)\mathsf{E}|Z_{m-1}|}{m^{\gamma/2}} \\
&\sim -4g'(0)\frac{1}{n}\left[\frac{n\mathsf{E}|Z_n|}{n^{\gamma/2}} + \sum_{m=1}^{n-1} m\mathsf{E}|Z_m|\left(\frac{1}{m^{\gamma/2}} - \frac{1}{(m+1)^{\gamma/2}}\right)\right] \qquad \text{(S2.24)} \\
&\le -4g'(0)\frac{1}{n}\left[\frac{n(\mathsf{E}|Z_n|^2)^{1/2}}{n^{\gamma/2}} + \sum_{m=1}^{n-1} m(\mathsf{E}|Z_m|^2)^{1/2}\left(\frac{1}{m^{\gamma/2}} - \frac{1}{(m+1)^{\gamma/2}}\right)\right] \\
&\sim \sqrt{|g'(0)|}\left[n^{-\gamma/2} + \frac{1}{n}\sum_{m=1}^{n-1} m\left(\frac{1}{m^{\gamma/2}} - \frac{1}{(m+1)^{\gamma/2}}\right)\right] \sim \frac{2\sqrt{|g'(0)|}}{2-\gamma} n^{-\gamma/2}.
\end{aligned}$$

Hence $\limsup_{n\to\infty} a_n \le 2\sqrt{|g'(0)|}/(2-\gamma)$.

On the other hand,

$$\mathsf{E}|Z_n|^3 = \frac{1}{n}\sum_{m=1}^{n}\sum_{\boldsymbol{t}} p(\boldsymbol{t})\frac{\mathsf{E}|\Lambda_m(\boldsymbol{t})|^3}{m^{3\gamma/2}} = O(1)$$

and $\mathsf{E}Z_n^2 \le (\mathsf{E}|Z_n|)^{1/2}(\mathsf{E}|Z_n|^3)^{1/2}$. It follows that

$$\begin{aligned}
\liminf_{n\to\infty} a_n &= \liminf_{n\to\infty}[-4g'(0)]\liminf_{n\to\infty}\mathsf{E}|Z_n| \\
&\ge \liminf_{n\to\infty}[-4g'(0)]\left(\frac{\mathsf{E}Z_n^2}{\mathsf{E}Z_n^3}\right)^2 \ge c_0 > 0
\end{aligned}$$

by (S2.24). Thus

$$SB_n = \frac{1}{2} + O(n^{-\gamma/2}). \tag{S2.25}$$

is proved. To show the constant c_0 does not depend on γ. It is sufficient to show that for any integer $r \geq 0$, there is a constant A_r which does not depend on γ such that

$$\limsup_{n\to\infty} \frac{\mathsf{E}M_n^r}{n^{\gamma r}} \leq A_r, \tag{S2.26}$$

which implies that $\limsup_{n\to\infty} \mathsf{E}[|Z_n|^3] \leq C$ for a constant C that does not depend on γ. We use induction, assume (S2.26) holds for r. We now consider $r+1$. Let $m = m_n$ such that (S2.14) and (S2.15) hold. Similar to (S2.23), we have

$$\begin{aligned}
\frac{c_r}{4(r+1)}\mathsf{E}(M_m+1)^r \geq & \, \mathsf{E}\left[(M_m+1)^r S_{m+1}(\boldsymbol{\Lambda}_m)\right] \\
= & - 4g'(0)\mathsf{E}\left[(M_m+1)^r \frac{\sum_{\boldsymbol{t}} \Lambda_m^2(\boldsymbol{t})p(\boldsymbol{t})}{m^\gamma}\right] \\
& + o(1)\left[\mathsf{E}\left((M_m+1)^r \frac{\sum_{\boldsymbol{t}} \Lambda_m^2(\boldsymbol{t})p(\boldsymbol{t})}{m^\gamma}\right)^2\right]^{1/2} \\
\geq & cm^{-\gamma}\mathsf{E}\left[(M_m+1)^r M_m\right] + o(1)m^{-\gamma}\left[\mathsf{E}M_m^{2(r+1)}\right]^{1/2} \\
= & cm^{-\gamma}\mathsf{E}\left[M_m^{r+1}\right] + o(1)m^{-\gamma}m^{\gamma(r+1)}.
\end{aligned}$$

Hence, by (S2.14),

$$\mathsf{E}[M_n^{r+1}] \leq m^\gamma \left(\frac{C_r}{4(r+1)c} + C_r\right)\mathsf{E}(M_m+1)^r + o(1)m^{\gamma(r+1)}.$$

It follows that

$$\limsup_{n\to\infty} \frac{\mathsf{E}[M_n^{r+1}]}{n^{\gamma(r+1)}} \le \left(\frac{C_r}{4(r+1)c} + C_r \right) A_r,$$

by the induction. This completes the proof. □

Proof of Corollary 3. Note $C^{-1}M_n \le \|\boldsymbol{\Lambda}_n\|^2 \le CM_n$. By the condition given in the corollary, when $4|\Lambda_n(\boldsymbol{t})| \ge x_0$,

$$\left| g_n(4\Lambda_n(\boldsymbol{t})) - \frac{1}{2} \right| \ge p_0 > 0.$$

Let $E = \{4|\Lambda_m(\boldsymbol{t})| < x_0\}$. Then on E^c,

$$S_{m+1}(\boldsymbol{\Lambda}_m) \ge c_0 |\Lambda_m(\boldsymbol{t})|.$$

It follows that

$$\mathsf{E}[|\Lambda_m(\boldsymbol{t})|^{2r+2} E] \le C,$$

and

$$\mathsf{E}[|\Lambda_m(\boldsymbol{t})|^{2r+1} E^c] \le C\mathsf{E}\left[|\Lambda_m(\boldsymbol{t})|^{2r} S_{m+1}(\boldsymbol{\Lambda}_m) \right] \le C\mathsf{E}\left[(M_m + 1)^r S_{m+1}(\boldsymbol{\Lambda}_m) \right].$$

Let $m \in \{0, 1, \ldots, n\}$ satisfy (S2.14) and (S2.15). Then (S2.16) remains true with $\gamma = 0$. The remaining proof is similar to that of (S2.20) and (S2.21) by repeating the induction procedure (S2.17)-(S2.19) i times until $1/2^{i-1} < \epsilon/2$. □

S3 Deriving the asymptotic selection bias

In this section we give the proofs of Theorem 3 in Section 3 and Remark 2 in Section 4.

Theorem 3 is a special case of Remark 2.

Proof of the Remark 2. The proof of that $\sup_n \mathsf{E}[M_n^r] < \infty$ can be found in Hu and Zhang (2020). For completeness of this paper, we give a direct proof here. The allocation function probability $g(4\Lambda_{n-1}(\boldsymbol{t}))$ is now only a function of $\boldsymbol{\Lambda}_{n-1}$, so the Markov chain $\boldsymbol{\Lambda}_n$ is homogeneous. Note $0 < g(x) < 1$ for all x. The transition probability in (S2.2) is always positive. So the Markov chain is irreducible. It can be verified that $S_n(\boldsymbol{\Lambda}_{n-1}) \to +\infty$ as $M_{n-1} \to \infty$. By (S2.12), it follows that there are positive constants b_r and c_r such that

$$\mathsf{E}\left[M_n^{r+1}|\mathscr{F}_{n-1}\right] - M_{n-1}^{r+1} \leq -\left(M_{n-1}+1\right)^r + b_r \mathbb{I}\left\{M_{n-1} \leq c_r\right\}.$$

Notice that $M_n = M^*(\boldsymbol{\Lambda}_n)$ is a function of $\boldsymbol{\Lambda}_n$. The above inequality is just

$$P_\lambda\left(M^*\right)^{r+1} - \left(M^*\right)^{r+1} \leq -\left(M^*+1\right)^r + b_r \mathbb{I}\left\{M^* \leq c_r\right\}, \tag{S3.1}$$

where $M^* = M^*(\boldsymbol{\Lambda})$, $(P_\lambda f)(\boldsymbol{\Lambda}) = \sum_{\boldsymbol{\Lambda}' \in \boldsymbol{L}(\mathbb{Z}^m)} P_\lambda(\boldsymbol{\Lambda}, \boldsymbol{\Lambda}') f(\boldsymbol{\Lambda}')$ is the transition expectation of the function f, $P_\lambda(\boldsymbol{\Lambda}, \boldsymbol{\Lambda}')$ is the transition probability from $\boldsymbol{\Lambda}$ to $\boldsymbol{\Lambda}'$. Then (S3.1) implies that the irreducible Markov chain $\boldsymbol{\Lambda}_n$ with period 2 is positive recurrent with an invariant distribution $\boldsymbol{\pi}$ and

$\sup_n \mathsf{E}\big[(M_n+1)^r\big] < \infty$. In fact, write

$$\overline{P}_\lambda^{(n)}(\mathbf{\Lambda}, \mathbf{\Lambda}') = \frac{1}{n}\sum_{m=0}^{n-1} P_\lambda^{(m)}(\mathbf{\Lambda}, \mathbf{\Lambda}'),$$

where $P_\lambda^{(m)}(\mathbf{\Lambda}, \mathbf{\Lambda}')$ is the m-step transition probability from $\mathbf{\Lambda}$ to $\mathbf{\Lambda}'$. (S3.1) implies that for any initial distribution $\boldsymbol{\pi}_0$,

$$\begin{aligned}\limsup_{n\to\infty} \boldsymbol{\pi}_0 \overline{P}_\lambda^{(n)}\big(M^*+1\big)^r \leq & b_r \limsup_{n\to\infty} \boldsymbol{\pi}_0 \overline{P}_\lambda^{(n)} I\big\{M^* \leq c_r\big\} + \limsup_{n\to\infty} \frac{\boldsymbol{\pi}_0\big(M^*\big)^{r+1}}{n} \\ \leq & b_r \limsup_{n\to\infty} \boldsymbol{\pi}_0 \overline{P}_\lambda^{(n)} I\big\{M^* \leq c_r\big\},\ \forall r > 0. \qquad \text{(S3.2)}\end{aligned}$$

If the Markov-Chain is not positive recurrent, then for any $\mathbf{\Lambda}$ and $\mathbf{\Lambda}'$, $\overline{P}_\lambda^{(n)}(\mathbf{\Lambda}, \mathbf{\Lambda}') \to 0$. For each $C > 0$, $\{M^* \leq C\}$ contains only a finite number of values of the Markov chain. So, $\boldsymbol{\pi}_0 \overline{P}_\lambda^{(n)} I\{M^* \leq C\} \to 0$. It follows that the limit in (S3.2) is zero, which is obviously a contradiction since $\boldsymbol{\pi}_0 \overline{P}_\lambda^{(n)}\big(M^*+1\big)^r \geq 1$. Hence $\mathbf{\Lambda}_n$ is a positive recurrent Markov Chain and has an invariant distribution $\boldsymbol{\pi}$. Now, (S3.2) implies

$$\mathsf{E}_\pi\left[\big(M^*+1\big)^r\right] = \boldsymbol{\pi}\overline{P}_\lambda^{(n)}\big(M^*+1\big)^r \leq b_r,\ \forall r > 0,$$

which implies $\sup_n \mathsf{E}[M_n^r] < \infty$ since $\frac{1}{2}\Big(\mathsf{E}[M_{2n}^r] + \mathsf{E}[M_{2n+1}^r]\Big) \to \mathsf{E}_{\boldsymbol{\pi}}[(M^*)^r]$.

Next, consider the selection bias SB_n. Notice that

$$\begin{aligned} SB_n =& \frac{1}{n}\sum_{m=1}^{n} \mathsf{E}\left[p_m \vee (1-p_m)\right] = \frac{1}{2} + \frac{1}{n}\sum_{m=1}^{n} \mathsf{E}\Big[\Big|p_m - \frac{1}{2}\Big|\Big] \\ =& \frac{1}{2} + \frac{1}{n}\sum_{m=1}^{n}\sum_{\boldsymbol{t}} \mathsf{E}\Big|g\big(4\Lambda_{m-1}(\boldsymbol{t})\big) - \frac{1}{2}\Big|p(\boldsymbol{t}) \\ \to& \frac{1}{2} + \sum_{\boldsymbol{t}} \mathsf{E}_{\boldsymbol{\pi}}\Big[\Big|g\big(4\Lambda(\boldsymbol{t})\big) - \frac{1}{2}\Big|\Big]p(\boldsymbol{t}) > 1/2, \end{aligned} \tag{S3.3}$$

by the ergodic theorem. Here the last inequality is due to $g(x) \not\equiv \frac{1}{2}$ and the fact that the Markov chain $\boldsymbol{\Lambda}_n$ is irreducible and so each possible state has a positive probability mass under the invariant distribution $\boldsymbol{\pi}$. This completes the proof. $\Box$.

S4 Further discussion on the randomness

Selection bias under partial information guessing strategy. When the clinical trials have many covariates, one may think that it is hard for the experimenter to know all the information so that he/or she can use an optimal guessing strategy, and so the selection bias can be ignored. We can show that this is not true at least for the Pocock and Simon's (1975) procedure. Write the allocation probability $p_m = g(\Lambda_{m-1}(\tilde{\boldsymbol{X}}_m))$ as a function of $\Lambda_{m-1}(\tilde{\boldsymbol{X}}_m)$. The guessing strategy is to guess treatment 1 when $G_m > 0$, treatment 2 when $G_m < 0$, and guess treatment 1 or 2 with probability 1/2 when $G_m = 0$, where G_m is the guessing factor. Then the expected proportion of correct guesses is

$$SB_n(g,G) = \frac{1}{2} + \frac{1}{n}\sum_{m=1}^{n} \mathsf{E}\left[\big(p_m - \frac{1}{2}\big)\mathrm{sgn}(G_m)\right],$$

where "g" stands for the allocation $g(x)$, "G" stands for the guessing strategy G_m. Because the experimenter uses part or all information of $\Lambda_{m-1}(\tilde{\boldsymbol{X}}_m)$, so $G_m = G(\Lambda_{m-1}(\tilde{\boldsymbol{X}}_m))$ is a function of $\Lambda_{m-1}(\tilde{\boldsymbol{X}}_m)$, where $G(\cdot)$ is a random function. We assume that under the guessing strategy, the allocation is guessed in a correct direction, i.e., for any point x, y, $\mathsf{P}\big(\{G(x) - G(y)\} \cdot \{g(x) - g(y)\} \geq 0\big) = P(x,y) \geq 1/2$, and there is at least a pair of points x_0 and y_0 such that $\mathsf{P}\big(G(x_0) > 0 > G(y_0), g(x_0) > g(y_0)\big) > 1/2$.

Theorem S4.1. *For the Pocock and Simon's procedure, Hu and Hu's procedure or more general procedures as in Remark 2, we have*

$$\lim_{n\to\infty} SB_n(g,G) > \frac{1}{2}.$$

Proof. Notice Remark 2. The sequence $\{\Lambda_n(\boldsymbol{t}); t_j = 1, \cdots, m_j, j = 1, \cdots, I\}_{n=1}^{\infty}$ is an irreducible positive recurrent multi-dimensional Markov Chain having an invariant distribution $\boldsymbol{\pi}$. Let $\Lambda^*(\boldsymbol{t}), \tilde{\boldsymbol{X}}^*$ be independent copies of $\Lambda(\boldsymbol{t}), \tilde{\boldsymbol{X}}$ and define $\eta = \Lambda(\tilde{\boldsymbol{X}})$, $\eta^* = \Lambda^*(\tilde{\boldsymbol{X}}^*)$. Then by the ergodic theorem,

$$
\begin{aligned}
&SB_n(g,G)-\frac{1}{2}=\frac{1}{n}\sum_{m=1}^{n}\mathsf{E}\left[\left(p_m-\frac{1}{2}\right)\mathrm{sgn}(G_m)\right]\\
\longrightarrow&\mathsf{E}_{\boldsymbol{\pi}}\left[\Big(g\big(\eta\big)\big)-\frac{1}{2}\Big)\mathrm{sgn}[G(\eta)]\right]\\
=&\mathsf{E}_{\boldsymbol{\pi}}\left[g\big(\eta\big)\big)-\frac{1}{2}\right]\mathsf{E}\left[\mathrm{sgn}[G(\eta)]\right]+\\
&\frac{1}{2}\mathsf{E}_{\boldsymbol{\pi}\times\boldsymbol{\pi}}\left[\big(g(\eta)-g(\eta^*)\big)\Big(\mathrm{sgn}(G(\eta))-\mathrm{sgn}(G(\eta^*))\Big)\right]\\
=&\mathsf{E}_{\boldsymbol{\pi}}\left[g\big(\eta\big)\big)-\frac{1}{2}\right]\mathsf{E}\left[\mathrm{sgn}[G(\eta)]\right]\\
&+\frac{1}{2}\mathsf{E}_{\boldsymbol{\pi}\times\boldsymbol{\pi}}\left[\Big|\big(g(\eta)-g(\eta^*)\big)\Big(\mathrm{sgn}(G(\eta))-\mathrm{sgn}(G(\eta^*))\Big)\Big|\left(2P(\eta,\eta^*)-1\right)\right].
\end{aligned}
$$

For the value of the limit, the second term is positive due to the assumption for guessing in a correct direction. The first term is zero because $\mathsf{E}_{\boldsymbol{\pi}}g(\eta)=1/2$ under the invariant distribution $\boldsymbol{\pi}$, which can be verified by

$$
0\leftarrow\frac{\mathsf{E}_{\boldsymbol{\pi}}[D_n]}{n}=\frac{1}{n}\sum_{j=1}^{n}\mathsf{E}_{\boldsymbol{\pi}}[2T_j-1]=\frac{1}{n}\sum_{j=1}^{n}2(\mathsf{E}_{\boldsymbol{\pi}}p_j-1)=2\mathsf{E}_{\boldsymbol{\pi}}g(\eta)-1.
$$

This completes the proof. $\square$

Entropy of a CAR procedure. Besides the selection bias, another measure of the randomness is the entropy. The entropy of a probability p is defined by

$$
E(p)=-p\log p-(1-p)\log(1-p).
$$

The entropy of the design with allocation probabilities p_ms is defined by

$$
Ent_n=\mathsf{E}\left[\frac{1}{n}\sum_{m=1}^{n}E(p_m)\right].
$$

It is obvious that $0 \leq Ent_n \leq \log 2$. Ent_n takes its largest value $\log 2$ when and only when the randomization is complete randomization in which $p_m = 1/2$ for all m. Ent_n takes its smallest value 0 when and only when each patient is assigned deterministically.

Theorem S4.2. *We have*

$$Ent_n \to \log 2 \ \ \textit{if and only if} \ \ SB_n \to \frac{1}{2}.$$

Proof. Since $SB_n = \frac{1}{2} + \frac{1}{n}\sum_{m=1}^{n} \mathsf{E}\big|p_m - \frac{1}{2}\big|$, it is sufficient to notice that

$$2\Big(p - \frac{1}{2}\Big)^2 \leq E\Big(\frac{1}{2}\Big) - E(p) \leq (2\log 2)\Big|p - \frac{1}{2}\Big|,$$

and

$$\frac{1}{n}\sum_{m=1}^{n} \mathsf{E}\big|p_m - \frac{1}{2}\big| \leq \left(\frac{1}{n}\sum_{m=1}^{n} \mathsf{E}\big|p_m - \frac{1}{2}\big|^2\right)^{1/2}. \qquad \square$$

Theorem S4.3. *For the Pocock and Simon's procedure, Hu and Hu's procedure or more general procedures as in Remark 2, we have*

$$\lim_{n\to\infty} Ent_n < \log 2.$$

Proof. Similar to (S3.3), we have

$$\begin{aligned} Ent_n =& \frac{1}{n}\sum_{m=1}^{n} \mathsf{E}\left[E(p_m)\right] = \frac{1}{n}\sum_{m=1}^{n}\sum_{\boldsymbol{t}} \mathsf{E}\left[E\Big(g\big(4\Lambda_{m-1}(\boldsymbol{t})\big)\Big)\right] p(\boldsymbol{t}) \\ &\to \sum_{\boldsymbol{t}} \mathsf{E}_{\boldsymbol{\pi}}\left[E\Big(g\big(4\Lambda(\boldsymbol{t})\big)\Big)\right] p(\boldsymbol{t}) < \log 2, \end{aligned}$$

by the ergodic theorem. Here the last inequality is due to $E(x) < \log 2$ when $x \neq 1/2$, $g(x) \not\equiv 1/2$ and the fact that each possible state of $\boldsymbol{\Lambda}_n$ has a positive probability mass under the invariant distribution $\boldsymbol{\pi}$. $\square$

S5 Simulation Study

In this section, we study the new CAR by simulation.

Model Setting: The underlying model is a linear model with independent covariates:

$$Y_i = T_i\mu_1 + (1 - T_i)\mu_2 + \sum_{j=1}^{3} \beta_j X_{i,j} + \epsilon_i,$$

where $\beta_j = 1$ for $j = 1, 2, 3$, $X_{i,1} \sim N(0, 1)$, $X_{i,2}, X_{i,3} \sim N(1, 1)$, and the covariates are mutually independent. The random error $\epsilon_i \sim N(0, 2^2)$ is independent of all $X_{i,j}$.

For randomization, we employ complete randomization (CR) and four covariate-adaptive randomization (CAR) procedures to assign patients to treatments: Pocock and Simon's procedure (PS) and the new CAR procedures (NEW) with $\gamma = 0.2$, 0.5, and 0.8, based on the discretized covariates $\widetilde{\boldsymbol{X}}_i = (d(X_{i,1}), d(X_{i,2}), d(X_{i,3}))$, where $d(X) = 0I\{X \leq 0\} + 1I\{0 < X < 2\} + 2I\{X \geq 2\}$. Pocock and Simon's procedure is the general CAR with imbalance measures

$$Imb_n^{(t)} = \frac{1}{I}\sum_{i=1}^{I}[D_n^{(t)}(i;t_i^*)]^2,$$

and the allocation probability

$$p_n = g_n(Imb_n^{(1)} - Imb_n^{(2)}) = \begin{cases} 1-\rho, & Imb_n^{(1)} - Imb_n^{(2)} > 0, \\ \frac{1}{2}, & Imb_n^{(1)} - Imb_n^{(2)} = 0, \\ \rho, & Imb_n^{(1)} - Imb_n^{(2)} < 0, \end{cases} \tag{S5.1}$$

where $1/2 < \rho < 1$. For the new CAR, we choose imbalance measures the same as that in Pocock and Simon's procedure, and the function g in (4.8) to be $g(x) = (1-\rho) \vee \big(\frac{1}{2} - \frac{1}{2}\,\mathrm{sgn}(x)(1 \wedge |x|)\big) \wedge \rho$, $0 < \gamma < 1$ and $1/2 < \rho < 1$, i.e., the allocation probability of the nth patients is

$$\begin{aligned} p_n &= g_n(Imb_n^{(1)} - Imb_n^{(2)}) \\ &= (1-\rho) \vee \left[\frac{1}{2} - \frac{1}{2}\,\mathrm{sgn}(Imb_n^{(1)} - Imb_n^{(2)})\left(1 \wedge \frac{|Imb_n^{(1)} - Imb_n^{(2)}|}{(n-1)^\gamma}\right)\right] \wedge \rho. \end{aligned} \tag{S5.2}$$

This CAR procedure has the property that, when $|Imb_n^{(1)} - Imb_n^{(2)}|/(n-1)^\gamma$ is large, the allocation probability is the same as that of Pocock and Simon's procedure, and when $|Imb_n^{(1)} - Imb_n^{(2)}|/(n-1)^\gamma$ is small, the allocation probability is a linear function of $(Imb_n^{(1)} - Imb_n^{(2)})/(n-1)^\gamma$:

$$p_n = g_n(Imb_n^{(1)} - Imb_n^{(2)}) = \frac{1}{2} - \frac{1}{2}\frac{Imb_n^{(1)} - Imb_n^{(2)}}{(n-1)^\gamma}.$$

Under each randomization procedure, we apply the two-sided t-test for

$$H_0 : \mu_1 = \mu_2 \quad \text{vs.} \quad H_A : \mu_1 \neq \mu_2,$$

using the test statistic defined in Equation (2.3) of the main text.

We evaluate the Type I error rates under H_0, power under $H_A : \mu_1 - \mu_2 = \delta/\sqrt{n}$ for $\delta = 5, 10, 15$, and selection bias for sample sizes $n \in \{50, 100, 200, 400\}$. Each simulation result is based on 5,000 replications.

The results are reported in Table 1. For complete randomization, the selection bias is fixed at 0.5, obtained analytically rather than through simulation.

First, across all settings, the Type I error rates are close to the nominal level of 5%, indicating that the hypothesis tests are well calibrated. Second, for relatively small sample sizes, both the new CAR procedures and the PS procedure yield slightly higher power than complete randomization; however, this power advantage gradually diminishes as the sample size increases. Third, for the same value of ρ, the selection bias under the PS procedure is consistently larger than that under the new CAR procedures. Furthermore, smaller values of ρ lead to lower selection bias for both the new CAR and PS procedures, and the larger values of γ lead to lower selection bias for the new CAR. Notably, the selection bias of the PS procedure does not decrease with increasing sample size, whereas the new CAR procedures

Table 1: Simulation results for selection bias, Type I error rates, and power under various randomization procedures and sample sizes.

n	Ran	Allocation Probability	SB	Type I	Power		
					$\delta = 5$	$\delta = 10$	$\delta = 15$
50	CR	-	0.5000	0.0530	0.2120	0.6440	0.9356
	PS	(S5.1) with $\rho = 0.9$	0.8334	0.0550	0.2260	0.6672	0.9500
	NEW	(S5.2) with $\gamma = 0.2, \rho = 0.9$	0.8186	0.0486	0.2328	0.6686	0.9502
	NEW	(S5.2) with $\gamma = 0.5, \rho = 0.9$	0.7391	0.0500	0.2280	0.6648	0.9496
	NEW	(S5.2) with $\gamma = 0.8, \rho = 0.9$	0.6544	0.0510	0.2284	0.6696	0.9466
	PS	(S5.1) with $\rho = 2/3$	0.6444	0.0496	0.2222	0.6646	0.9468
	NEW	(S5.2) with $\gamma = 0.2, \rho = 2/3$	0.6444	0.0506	0.2254	0.6664	0.9442
	NEW	(S5.2) with $\gamma = 0.5, \rho = 2/3$	0.6386	0.0518	0.2312	0.6596	0.9476
	NEW	(S5.2) with $\gamma = 0.8, \rho = 2/3$	0.6156	0.0510	0.2148	0.6696	0.9486
100	CR	-	0.5000	0.0620	0.2328	0.6776	0.9490
	PS	(S5.1) with $\rho = 0.9$	0.8254	0.0482	0.2330	0.6878	0.9578
	NEW	(S5.2) with $\gamma = 0.2, \rho = 0.9$	0.8051	0.0506	0.2348	0.6950	0.9580
	NEW	(S5.2) with $\gamma = 0.5, \rho = 0.9$	0.7098	0.0504	0.2304	0.7032	0.9562
	NEW	(S5.2) with $\gamma = 0.8, \rho = 0.9$	0.6209	0.0468	0.2356	0.6882	0.9576
	PS	(S5.1) with $\rho = 2/3$	0.6493	0.0504	0.2380	0.6966	0.9568
	NEW	(S5.2) with $\gamma = 0.2, \rho = 2/3$	0.6491	0.0512	0.2354	0.6922	0.9534
	NEW	(S5.2) with $\gamma = 0.5, \rho = 2/3$	0.6384	0.0486	0.2304	0.6916	0.9598
	NEW	(S5.2) with $\gamma = 0.8, \rho = 2/3$	0.6026	0.0508	0.2382	0.6894	0.9558
200	CR	-	0.5000	0.0492	0.2366	0.6910	0.9554
	PS	(S5.1) with $\rho = 0.9$	0.8336	0.0462	0.2424	0.7024	0.9616
	NEW	(S5.2) with $\gamma = 0.2, \rho = 0.9$	0.7994	0.0452	0.2298	0.7160	0.9626
	NEW	(S5.2) with $\gamma = 0.5, \rho = 0.9$	0.6856	0.0506	0.2382	0.6986	0.9608
	NEW	(S5.2) with $\gamma = 0.8, \rho = 0.9$	0.5960	0.0536	0.2426	0.7012	0.9608
	PS	(S5.1) with $\rho = 2/3$	0.6512	0.0520	0.2286	0.6880	0.9590
	NEW	(S5.2) with $\gamma = 0.2, \rho = 2/3$	0.6506	0.0472	0.2334	0.7004	0.9590
	NEW	(S5.2) with $\gamma = 0.5, \rho = 2/3$	0.6332	0.0508	0.2310	0.6938	0.9570
	NEW	(S5.2) with $\gamma = 0.8, \rho = 2/3$	0.5866	0.0520	0.2320	0.6908	0.9570
400	CR	-	0.5000	0.0498	0.2308	0.6914	0.9590
	PS	(S5.1) with $\rho = 0.9$	0.8295	0.0576	0.2376	0.7082	0.9612
	NEW	(S5.2) with $\gamma = 0.2, \rho = 0.9$	0.7903	0.0524	0.2408	0.7026	0.9614
	NEW	(S5.2) with $\gamma = 0.5, \rho = 0.9$	0.6596	0.0492	0.2384	0.7002	0.9628
	NEW	(S5.2) with $\gamma = 0.8, \rho = 0.9$	0.5743	0.0516	0.2356	0.6994	0.9588
	PS	(S5.1) with $\rho = 2/3$	0.6531	0.0504	0.2260	0.6918	0.9618
	NEW	(S5.2) with $\gamma = 0.2, \rho = 2/3$	0.6526	0.0462	0.2298	0.7056	0.9642
	NEW	(S5.2) with $\gamma = 0.5, \rho = 2/3$	0.6256	0.0502	0.2398	0.7116	0.9616
	NEW	(S5.2) with $\gamma = 0.8, \rho = 2/3$	0.5694	0.0528	0.2400	0.7056	0.9604

exhibit a decreasing selection bias as the sample size grows. In particular, the reduction is more pronounced for larger values of γ, indicating a faster convergence rate. This observation aligns with our theoretical results.

Acknowledgments

We are grateful to Mr. Yuhang Tao, a Ph.D. student under the author's supervision, for contributing the simulation data in Table 1.