To Adaptively Randomize or to Rerandomize: A Comparison of Covariate-Adaptive Randomization and Rerandomization

Ziji Qin, Feifang Hu, and Yang Liu

The George Washington University and Renmin University of China

Supplementary Materials

The Supplementary Materials are organized as follows. Section S1 restates the assumptions and main theoretical results for completeness. Section S2 provides detailed proofs of all theoretical results. Sections S3 and S4 present additional simulation studies, tables, and figures that complement Sections 5 and 6 of the main text, respectively.

S1 Assumptions and Main Results

We begin by restating the key assumptions and the main theoretical results.

Assumption 3.1. 1. *$\{\boldsymbol{X}_i, Y_i(1), Y_i(0)\}_{i=1}^n$ are i.i.d. copies of $\{\boldsymbol{X}, Y(1), Y(0)\}$, with $\sigma^2_{Y(t)} = \mathbb{V}[Y(t)] < \infty$ and $\mathbb{E}[\mathbb{V}[Y(t)|\boldsymbol{X}]] > 0$ for $t \in \{0, 1\}$.*

2. *The feature map $\psi(\boldsymbol{X})$ satisfies $\mathbb{E}[\|\psi(\boldsymbol{X})\|^r] < \infty$ for some $r > 5$, $\|\Sigma_{Y(t)\psi}\| < \infty$, and $\Sigma_{\psi\psi}$ is invertible.*

Assumption 3.2. *Suppose that the distribution of $\psi(\boldsymbol{X})$ is Γ_ψ and that there exist an integer n_c and a constant $1 \geq c_\nu > 0$ such that $\Gamma_\psi^{n_c}(A) \geq c_\nu \int_A \nu(y)\, dy$ for any Borel set A, where $\nu(y)$ is a density function satisfying $\inf_{y \in O} \nu(y) > 0$ for an open set O, and Γ_ψ^k is the k-fold convolution of Γ_ψ.*

Proposition 3.1. (i) *Under Assumption 3.1 and CAR, $\mathbb{E}[\|\Lambda_n(\psi)\|^2] = O(n^{1/(c-1)})$ and thus $\Lambda_n(\psi) = O_p(n^{1/(2c-2)}) = o_p(n^{1/2})$ for $c > 2$. If Assumption 3.2 also holds, $\{\Lambda_n(\psi)\}_{n \geq 1}$ is a positive Harris recurrent Markov chain, and $\Lambda_n(\psi) = O_p(1)$.*

(ii) *Suppose Assumption 3.1 holds and RR is adopted with $r_n = nr$. Let $v(r, q) = F_{\chi^2_{q+2}}(r)/F_{\chi^2_q}(r)$ denote the efficiency factor. Then*

$$n^{-1} Imb_n(\tilde{\psi}) \xrightarrow{\mathbf{D}} \Lambda_{\mathsf{RR}}^\top \Lambda_{\mathsf{RR}} \quad \text{and} \quad n^{-1/2}\Lambda_n(\psi) \xrightarrow{\mathbf{D}} \Sigma_{\psi\psi}^{1/2} \Lambda_{\mathsf{RR}}, \tag{S1.1}$$

where $\boldsymbol{Z}_q = (Z_1, \ldots, Z_q)^\top$ follows $\mathcal{N}(\mathbf{0}, I_q)$, $\Lambda_{\mathsf{RR}} \overset{\mathbf{D}}{=} \boldsymbol{Z}_q | \boldsymbol{Z}_q^\top \boldsymbol{Z}_q < r$, and Λ_{RR} satisfies $\mathbb{E}[\Lambda_{\mathsf{RR}}] = \mathbf{0}_q$ and $\mathbb{V}[\Lambda_{\mathsf{RR}}] = v(r, q) I_q$.

Proposition 3.2. 1. *Suppose Assumptions 3.1 and 3.2 hold and CAR is employed. There exists $\sigma_m^2 \geq 0$ such that*

$$n^{-1/2}\Lambda_n(m_1+m_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^n (2T_i-1)[m_1(\boldsymbol{X}_i)+m_0(\boldsymbol{X}_i)] \xrightarrow{\mathbf{D}} \mathcal{N}(0,\sigma_m^2),$$

$$\text{and} \quad \sqrt{n}(\hat{\tau}-\tau) \xrightarrow{\mathbf{D}} \mathcal{N}(0,\sigma_Y^2(1-\gamma_{Y\psi,\mathsf{CAR}})), \tag{S1.2}$$

where $\gamma_{Y\psi,\mathsf{CAR}} = \sigma_Y^{-2}(\mathbb{V}[m_1(\boldsymbol{X})+m_0(\boldsymbol{X})]-\sigma_m^2)$.

2. *Suppose Assumption 3.1 holds and RR is implemented with $r_n = nr$. Let Z_0 be a standard normal random variable independent of $\boldsymbol{Z}_q$ defined in Proposition 3.1(ii). Then*

$$\sqrt{n}(\hat{\tau}-\tau) \xrightarrow{\mathbf{D}} \sigma_Y\left(\sqrt{\gamma_{Y\psi,\mathsf{RR}}}\, L_{q,r} + \sqrt{1-\gamma_{Y\psi,\mathsf{RR}}}\, Z_0\right), \tag{S1.3}$$

where $\gamma_{Y\psi,\mathsf{RR}} = \sigma_Y^{-2}\Sigma_{Y\psi}\Sigma_{\psi\psi}^{-1}\Sigma_{\psi Y}$, $L_{q,r} \overset{\mathbf{D}}{=} Z_1 \mid \boldsymbol{Z}_q^\top \boldsymbol{Z}_q < r$, and $L_{q,r}$ is symmetric about zero with $\mathbb{V}[L_{q,r}] = v(r,q)$.

Theorem 3.3. *Let $h(x) = x-1-\log x$, and suppose $q\to\infty$. Then the following statements hold.*

1. *Suppose $r = cq+o(1)$ with $c\neq 1$, $c>0$. (i) If $0<c<1$, then $F_{\chi_q^2}(r) \sim [(1-c)\sqrt{\pi q}]^{-1}\exp\{-2^{-1}qh(c)\}$, $\log\mathbb{E}[N] = 2^{-1}qh(c)+2^{-1}\log(\pi q)+\log(1-c)+o(1)$, and $v(r,q)\to c$. (ii) If $c>1$, then $F_{\chi_q^2}(r)\sim 1-[(c-1)\sqrt{\pi q}]^{-1}\exp\{-2^{-1}qh(c)\}$, $\mathbb{E}[N]\to 1$, and $v(r,q)\to 1$.*

2. *Suppose* $r = q + o(\sqrt{q})$*, then* $F_{\chi^2_q}(r) \to \Phi(0)$, $\mathbb{E}[N] \to [\Phi(0)]^{-1} = 2$*, and* $v(r,q) \to 1$*, where* $\Phi(x)$ *is the cumulative distribution function of a standard normal random variable.*

3. *Suppose* $r = (q+2)n^{-\alpha} \to 0$ *for some* $\alpha > 0$*. Then* $F_{\chi^2_q}(r) \sim \pi^{-1/2} \exp\{2^{-1}q(1-\alpha\log n) - 2^{-1}\log q + 1\}$, $\log \mathbb{E}[N] = 2^{-1}\alpha q \log n - 2^{-1}q + 2^{-1}\log(\pi q) - 1 + o(1)$*, and* $v(r,q) \sim r/(q+2) = n^{-\alpha} \to 0$*.*

Corollary 3.1. *Suppose RR uses* $r_n \sim (q+2)n^{1-\alpha}$ *for some* $\alpha > 0$ *and* q *is sufficiently large, then* $v(r,q) \sim n^{-\alpha}$*,*

$$\Lambda_n(\psi) = o_p(\sqrt{n}), \quad \sqrt{n}(\hat{\tau} - \tau) \xrightarrow{\mathbf{D}} \mathcal{N}\left(0, \sigma_Y^2(1 - \gamma_{Y\psi,\mathsf{RR}})\right),$$

and the number of computations needed satisfies $\log \mathbb{E}[N] \sim O(\alpha \log n \cdot q)$*.*

Corollary 3.2. *Suppose Assumptions 3.1 and 3.2 hold and* q *is sufficiently large. Under either CAR or RR with* $\Lambda_n(\tilde{\psi})$ *and* $r \sim (q+2)n^{-\alpha}$*, we have* $n^{-1/2}\Lambda_n(\psi) = o_p(1)$*. Furthermore, if*

$$m_1(\boldsymbol{X}) + m_0(\boldsymbol{X}) = \{\psi(\boldsymbol{X}) - \mathbb{E}[\psi(\boldsymbol{X})]\}^\top \boldsymbol{\eta} + m_{\psi_\perp}(\boldsymbol{X}) \tag{S1.4}$$

for some $\boldsymbol{\eta} \in \mathbb{R}^q$*, such that* $\mathbb{E}[\boldsymbol{\psi}(\boldsymbol{X}) m_{\psi_\perp}(\boldsymbol{X})] = 0$ *and* $\mathbb{E}[m_{\psi_\perp}(\boldsymbol{X})] = 0$*, then* $\gamma_{Y\psi} = \gamma_{Y\psi,\mathsf{CAR}} = \gamma_{Y\psi,\mathsf{RR}} = \sigma_Y^{-2}\boldsymbol{\eta}^\top \Sigma_{\psi\psi}\boldsymbol{\eta}$ *and*

$$\sqrt{n}(\hat{\tau} - \tau) \xrightarrow{\mathbf{D}} \mathcal{N}(0, \sigma_Y^2(1 - \gamma_{Y\psi})) \tag{S1.5}$$

holds for both CAR and RR.

S2 Proofs of the Main Results

We begin with the following result for complete randomization, which is a direct consequence of the central limit theorem.

Proposition 1. *Under Assumption 3.1 and complete randomization, we have*

$$\begin{pmatrix} n^{-1/2}\Lambda_n(\psi) \\ \sqrt{n}\,(\hat{\tau}-\tau) \end{pmatrix} \xrightarrow{\mathbf{D}} \mathcal{N}\left(0, \begin{pmatrix} \Sigma_{\psi\psi} & \Sigma_{\psi Y} \\ \Sigma_{Y\psi} & \sigma_Y^2 \end{pmatrix}\right),$$

$$\text{and} \;\; n^{-1/2}\Lambda_n(\tilde{\psi}) \xrightarrow{\mathbf{D}} \mathcal{N}(0, I_q),$$

where $\tilde{\psi}(\boldsymbol{X}) = \Sigma_{\psi\psi}^{-1/2}(\psi(\boldsymbol{X}) - \mu_\psi)$, $\Sigma_{\psi Y} = \Sigma_{Y\psi}^\top = \mathrm{Cov}[Y(1)+Y(0), \psi(\boldsymbol{X})]$, $\sigma_Y^2 = 2(\sigma_{Y(1)}^2 + \sigma_{Y(0)}^2)$, *and* I_q *is the* $q \times q$ *identity matrix.*

S2.1 Proof of Proposition 3.1

Proposition 3.1 (i) follows from Theorems 3.1 and 3.4 of Ma et al. (2024). Proposition 3.1 (ii) follows from a derivation similar to that in Morgan et al. (2012).

S2.2 Proof of Proposition 3.2

Proposition 3.2 (i) follows from Theorem 4.4 of Ma et al. (2024). However, we use a different representation of the asymptotic variance of $\hat{\tau}$. Notice that

$$\begin{aligned}
\sigma_Y^2 &= 2(\sigma_{Y(1)}^2 + \sigma_{Y(0)}^2) \\
&= \mathbb{V}[Y(1) + Y(0)] + \mathbb{V}[Y(1) - Y(0)] \\
&= 2(\mathbb{V}[\epsilon(1)] + \mathbb{V}[\epsilon(0)]) + \mathbb{V}[m_1(\boldsymbol{X}) - m_0(\boldsymbol{X})] + \mathbb{V}[m_1(\boldsymbol{X}) + m_0(\boldsymbol{X})], \\
\text{and} \quad \Sigma_{Y(t)\psi} &= \mathrm{Cov}[Y(t), \psi(\boldsymbol{X})] \\
&= \mathrm{Cov}[m_t(\boldsymbol{X}) + \epsilon(t), \psi(\boldsymbol{X})] \\
&= \mathrm{Cov}[m_t(\boldsymbol{X}), \psi(\boldsymbol{X})] = \Sigma_{m_t\psi},
\end{aligned}$$

then $\gamma_{Y\psi,\mathsf{RR}} = \sigma_Y^{-2}\Sigma_{Y\psi}\Sigma_{\psi\psi}^{-1}\Sigma_{\psi Y} = \sigma_Y^{-2}\Sigma_{m\psi}\Sigma_{\psi\psi}^{-1}\Sigma_{\psi m}$ and

$$\begin{aligned}
\sigma_Y^2(1 - \gamma_{Y\psi,\mathsf{CAR}}) &= \sigma_Y^2 - \mathbb{V}[m_1(\boldsymbol{X}) + m_0(\boldsymbol{X})] + \sigma_m^2 \\
&= 2(\mathbb{V}[\epsilon(1)] + \mathbb{V}[\epsilon(0)]) + \mathbb{V}[m_1(\boldsymbol{X}) - m_0(\boldsymbol{X})] + \sigma_m^2.
\end{aligned}$$

In addition, given Proposition 1, Proposition 3.2 (ii) follows from a similar derivation as Li and Ding (2020).

S2.3 Proof of Theorem 3.3

We start the proof with some preparations. First, let $a = \frac{q}{2}$, the CDF for χ^2_q is

$$F_{\chi^2_q}(r) = \frac{1}{\Gamma(a)} \int_0^{r/2} t^{a-1} e^{-t}\, dt,$$

then the change of variables $t = as$ gives

$$F_{\chi^2_q}(r) = \frac{a^a}{\Gamma(a)} \int_0^{r/q} s^{-1} e^{ag(s)}\, ds, \qquad g(s) = \log s - s, \tag{S2.1}$$

and the Stirling's formula yields

$$\frac{a^a}{\Gamma(a)} = \frac{e^a \sqrt{a}}{\sqrt{2\pi}} \{1 + o(1)\}. \tag{S2.2}$$

Using the exact incomplete-gamma recursion, i.e., $\gamma(a+1, x) = a\gamma(a, x) - x^a e^{-x}$, we have

$$F_{\chi^2_{q+2}}(r) = F_{\chi^2_q}(r) - \frac{(r/2)^a e^{-r/2}}{\Gamma(a+1)}.$$

Consequently,

$$v(r, q) = 1 - \frac{(r/2)^{q/2} e^{-r/2}}{\Gamma(q/2+1) F_{\chi^2_q}(r)}. \tag{S2.3}$$

1. (a) For part **1. (i)**, suppose $q^{-1}r \to c \in (0, 1)$. Since $g'(s) = \frac{1}{s} - 1 > 0$, for $0 < s < 1$, the function g is increasing on $(0, c]$. The endpoint

Laplace approximation gives

$$\begin{aligned}\int_0^{r/q} s^{-1} e^{ag(s)}\, ds &\sim \frac{c^{-1}}{ag'(c)} e^{ag(c)} \\ &= \frac{1}{a(1-c)} e^{a(\log c - c)}.\end{aligned} \tag{S2.4}$$

Note that $h(c) = -g(c) - 1 = c - \log c - 1$, then combining (S2.1), (S2.2), and (S2.4), we have

$$\begin{aligned}F_{\chi_q^2}(r) &\sim \frac{e^a \sqrt{a}}{\sqrt{2\pi}} \frac{1}{a(1-c)} e^{a(\log c - c)} \\ &= \frac{1}{(1-c)\sqrt{2\pi a}} \exp\{a(1 + \log c - c)\} \\ &= \frac{1}{(1-c)\sqrt{\pi q}} \exp\left\{-\frac{q}{2} h(c)\right\}.\end{aligned} \tag{S2.5}$$

Taking negative logarithms gives

$$\log \mathbb{E}[N] = \frac{q}{2} h(c) + \frac{1}{2} \log(\pi q) + \log(1-c) + o(1)$$

Moreover, since $h'(c) = 1 - \frac{1}{c}$, $h''(c) = \frac{1}{c^2} > 0$, and $h(1) = 0$, we have $h(c) > 0$ for $c \neq 1$. Next, by Stirling's formula,

$$\begin{aligned}\frac{(ac)^a e^{-ac}}{\Gamma(a+1)} &\sim \frac{(ac)^a e^{-ac}}{\sqrt{2\pi a}(a/e)^a} \\ &= \frac{1}{\sqrt{\pi q}} \exp\left\{-\frac{q}{2} h(c)\right\}.\end{aligned} \tag{S2.6}$$

Combining (S2.6) and (S2.5), we have

$$\frac{(ac)^a e^{-ac}}{\Gamma(a+1)F_{\chi^2_q}(r)} \to 1-c.$$

Therefore, it follows from (S2.3) that $v(r,q) \to 1-(1-c) = c$.

(b) Next, for **1. (ii)**, suppose $q^{-1}r \to c > 1$. It is convenient to analyze the upper tail:

$$1 - F_{\chi^2_q}(r) = \frac{a^a}{\Gamma(a)} \int_{r/q}^{\infty} s^{-1} e^{ag(s)}\, ds.$$

For $s > 1$, $g'(s) = s^{-1} - 1 < 0$, so g is decreasing. Then the endpoint Laplace's method yields

$$\begin{aligned}\int_{r/q}^{\infty} s^{-1} e^{ag(s)}\, ds &\sim \frac{c^{-1}}{a|g'(c)|} e^{ag(c)} \\ &= \frac{1}{a(c-1)} e^{a(\log c - c)}.\end{aligned} \tag{S2.7}$$

Therefore, we have

$$1 - F_{\chi^2_q}(r) \sim \frac{1}{(c-1)\sqrt{\pi q}} \exp\left\{-\frac{q}{2}h(c)\right\}.$$

Since $h(c) > 0$, the right-hand side tends to zero. Consequently, we obtain $F_{\chi_q^2}(r) \to 1$ and $\mathbb{E}[N] \to 1$. By (S2.6), it follows that

$$\frac{(r/2)^{q/2}e^{-r/2}}{\Gamma(q/2+1)F_{\chi_q^2}(r)} \to 0.$$

Equation (S2.3) therefore gives $v(r,q) \to 1$.

2. To prove **2.**, we first assume $(2q)^{-1/2}(r-q) \to z$ for some $z \in \mathbb{R}$. Let $X_q \sim \chi_q^2$, then by CLT, $(2q)^{-1/2}(X_q - q)$ converges to a standard normal random variable in distribution, that is

$$F_{\chi_q^2}(r) = \Pr\left\{\frac{X_q - q}{\sqrt{2q}} \le \frac{r-q}{\sqrt{2q}}\right\} \to \Phi(z), \quad \text{as} \quad \frac{r-q}{\sqrt{2q}} \to z,$$

where $\Phi(z)$ represents the cumulative distribution function of a standard normal random variable.

Now, since we assume $r = q + o(\sqrt{q})$, then $q^{-1/2}(r-q) \to 0$ and $\mathbb{E}[N] \to \Phi(0)^{-1}$. It remains to prove $v(r,q) \to 1$. It follows from the CLT and (S2.3) that

$$\frac{(r/2)^{q/2}e^{-r/2}}{\Gamma(q/2+1)} = \sqrt{\frac{2}{q}}\,\phi(z) + o(q^{-1/2}),$$
$$\text{and thus} \quad v(r,q) = 1 - \sqrt{\frac{2}{q}}\frac{\phi(z)}{\Phi(z)} + o(q^{-1/2}) \to 1 \quad \text{as } z \to 0, q \to \infty,$$

where $\phi(z) = (2\pi)^{-1/2}e^{-z^2/2}$ is the probablity density function of a standard normal random variable.

3. For **3.**, suppose $r = (q+2)n^{-\alpha} \to 0$. Let $a = q/2$. Choose $0 \le t \le r/2$, then it is easy to see that

$$\int_0^{r/2} t^{a-1} e^{-r/2}\, dt \le \int_0^{r/2} t^{a-1} e^{-t}\, dt \le \int_0^{r/2} t^{a-1}\, dt,$$

and thus

$$e^{-r/2} \frac{(r/2)^a}{\Gamma(a+1)} \le F_{\chi_q^2}(r) \le \frac{(r/2)^a}{\Gamma(a+1)}.$$

Since $e^{-r/2} \to 1$ as $r \to 0$, it follows that $F_{\chi_q^2}(r) \sim [\Gamma(q/2+1)]^{-1}(r/2)^{q/2}$. Using $r = (q+2)n^{-\alpha}$ and the Stirling approximation, i.e., $\Gamma\left(\frac{q}{2}+1\right) = \sqrt{\pi q}\,[q/(2e)]^{q/2}\,\{1+o(1)\}$, we obtain

$$\begin{aligned} F_{\chi_q^2}(r) &\sim \frac{1}{\sqrt{\pi q}} \left\{q^{-1} e(q+2)\right\}^{q/2} n^{-\alpha q/2} \\ &= \pi^{-1/2} \exp\left\{\frac{1}{2}\left[q(1-\alpha\log n) - \log(q)\right] + 1 + o(1)\right\}, \end{aligned} \tag{S2.8}$$

since $2^{-1} q \log\left(1+2q^{-1}\right) = 1+o(1)$ and $\log(q+2) - \log q = o(1)$. Taking negative logarithms, we have

$$\log \mathbb{E}[N] = \frac{1}{2}\left[\alpha q \log n - q + \log(\pi q))\right] - 1 + o(1).$$

Since $r = (q+2)n^{-\alpha} \to 0$, $\alpha \log n - \log(q) \to \infty$ and therefore $\log \mathbb{E}[N] \sim$

$2^{-1}\alpha q \log n$.

Furthermore, it follows from (S2.8) again that

$$\begin{aligned} v(r,q) &= \frac{F_{\chi^2_{q+2}}(r)}{F_{\chi^2_q}(r)} \\ &\sim \frac{(r/2)^{q/2+1}}{\Gamma(q/2+2)} \frac{\Gamma(q/2+1)}{(r/2)^{q/2}} \\ &= \frac{r/2}{q/2+1} = \frac{r}{q+2} = n^{-\alpha}. \end{aligned}$$

The proof for this theorem is now complete.

S2.4 Proof of Corollary 3.1

Since $v(r,q) \sim n^{-\alpha} = o(1)$, $L_{q,r}$ converges to zero in probability as $r \downarrow 0$. Then the rest of the results follow from Proposition 3.1 (ii), Proposition 3.2 (ii), and Theorem 3.3.

S2.5 Proof of Corollary 3.2

(S1.5) is a direct consequence of (S1.4) and Proposition 3.2.

S3 Additional Numerical Results

In this section, we present detailed specifications of the parameters used in CAR and RR and supplementary results on the distributions of covariate imbalance and $\hat{\tau}$ for the two studies considered in Section 5.

For the two studies in Section 5, we set the biased-coin probability to $\rho = 0.9$ for all CAR designs and consider both a *fixed* threshold, where r is chosen as the 5th percentile of a χ^2_{q+2} distribution, and an *adaptive* threshold of the form $r = (q+2)n^{-\alpha}$, with $\alpha \in \{0.1, 0.3, 0.5, 0.7\}$, for each RR design.

S3.1 Additional Results for Section 5.1

Figure S1 demonstrates the superior performance of HH in balancing discrete covariates. While both HH and RR reduce the magnitude of imbalance for X_1 relative to CR, HH achieves the most effective balance by efficiently leveraging both within-stratum and marginal imbalances through ψ_i, regardless of whether $k_0 = 2$ or $k_0 = 5$. Moreover, across both values of k_0, the distributions of $\sum_{i=1}^n (2T_i - 1)X_{i,1}$ under HH are the most tightly concentrated around zero.

For RR, our findings can be summarized in two main aspects. First, adopting a tightening threshold of the form $r = (q+2)n^{-\alpha}$ with $\alpha = 0.5$ yields substantially finer covariate balance than a fixed threshold. This improvement is evidenced by the markedly more concentrated distributions of $\sum_{i=1}^n (2T_i - 1)X_{i,1}$ under $\mathrm{RR}_{\mathsf{m,adp},0.5}$ and $\mathrm{RR}_{\mathsf{s,adp},0.5}$, relative to their fixed-

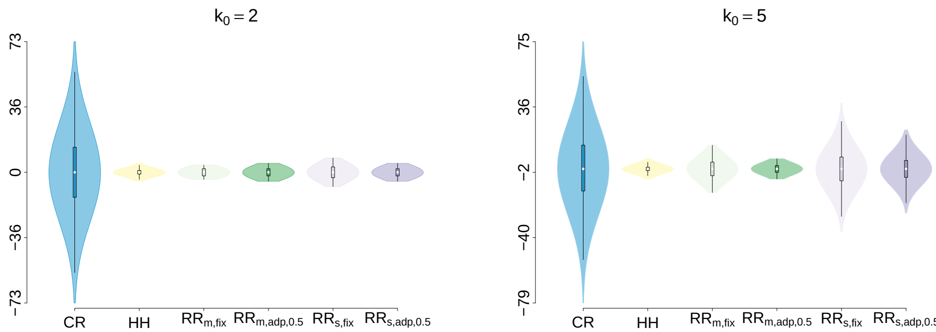

Figure S1: Distributions of $\sum_{i=1}^{n}(2T_i - 1)X_{i,1}$ under CR, HH, $\mathrm{RR}_{\mathsf{m,fix}}$, $\mathrm{RR}_{\mathsf{m,adp,0.5}}$, $\mathrm{RR}_{\mathsf{s,fix}}$, $\mathrm{RR}_{\mathsf{s,adp,0.5}}$ and the two settings $k_0 = 2$ and $k_0 = 5$ when $n = 800$ for Study 1.

threshold counterparts $\mathrm{RR}_{\mathsf{m,fix}}$ and $\mathrm{RR}_{\mathsf{s,fix}}$, respectively.

Second, stratified RR, i.e., $\mathrm{RR}_{\mathsf{s,fix}}$ and $\mathrm{RR}_{\mathsf{s,adp,0.5}}$, exhibits clear performance degradation when the number of strata is relatively large, whereas marginal RR performs well in both settings. As illustrated in Figure S1, the distributions of $\sum_{i=1}^{n}(2T_i - 1)X_{i,1}$ under $\mathrm{RR}_{\mathsf{s,fix}}$ and $\mathrm{RR}_{\mathsf{s,adp,0.5}}$ are noticeably more dispersed than those under $\mathrm{RR}_{\mathsf{m,fix}}$ and $\mathrm{RR}_{\mathsf{m,adp,0.5}}$, indicating poorer covariate balance under the stratified RR designs. Comparing $k_0 = 2$ with $k_0 = 5$ further shows that the performance of stratified RR deteriorates as k_0 increases. In contrast, the distributions of $\sum_{i=1}^{n}(2T_i - 1)X_{i,1}$ under $\mathrm{RR}_{\mathsf{m,fix}}$ and $\mathrm{RR}_{\mathsf{m,adp,0.5}}$ remain on comparable scales across these settings, underscoring the robustness of marginal RR when the number of covariates is relatively large.

Figure S2 further illustrates how improvements in covariate balance

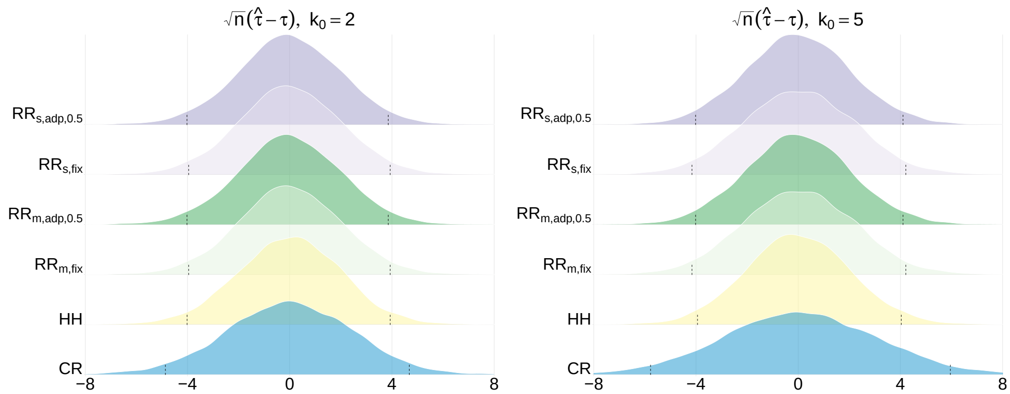

Figure S2: *Distributions of* $\sqrt{n}(\hat{\tau} - \tau)$ *under CR, HH,* $RR_{\mathsf{m,fix}}$*,* $RR_{\mathsf{m,adp,0.5}}$*,* $RR_{\mathsf{s,fix}}$*,* $RR_{\mathsf{s,adp,0.5}}$ *and the two settings* $k_0 = 2$ *and* $k_0 = 5$ *when* $n = 800$ *for Study 1.*

translate into more precise treatment-effect estimation. For both HH and the various RR designs, the distributions of $\sqrt{n}(\hat{\tau} - \tau)$ are more tightly concentrated around zero than under CR.

S3.2 Evaluation of $\mathbb{E}[N]$ with Respect to Feature Dimension q, Threshold α, and Sample Size n in Study 1

We further examine part (3) of Theorem 3.3 under the same configuration adopted in Study 1 for the RR designs. It is worth noting that Theorem 3.3 is established under an asymptotic regime in which the feature dimension q tends to infinity, whereas the values of q considered for the RR designs in our simulations are finite. Consequently, the numerical results should be interpreted as finite-sample approximations to the asymptotic behavior characterized by the theorem.

Table S1 reports the empirical estimates of $\mathbb{E}[N]$ as a function of α for $\mathrm{RR}_{\mathsf{m,adp},\alpha}$ and $\mathrm{RR}_{\mathsf{s,adp},\alpha}$ with $n = 400$ and $k_0 = 4$. The results indicate that, for $\mathrm{RR}_{\mathsf{s,adp},\alpha}$, $\mathbb{E}[N]$ increases sharply with α and rapidly approaches the imposed upper bound of 200,000. This behavior reflects the substantial computational burden induced by stratified RR when the balance threshold becomes more stringent. In light of this limitation, we henceforth concentrate on evaluating the operating characteristics of $\mathrm{RR}_{\mathsf{m,adp},\alpha}$, for which the expected number of randomization draws remains computationally feasible across $\alpha \in [0.2, 0.7]$.

Table S1: Evaluation of $\mathbb{E}[N]$ with respect to α when $n = 400$ and $k_0 = 4$.

α	0.2	0.3	0.4	0.5	0.6	0.7
$\mathrm{RR}_{\mathsf{m,adp},\alpha}$	4	11	32	99	329	1,149
$\mathrm{RR}_{\mathsf{s,adp},\alpha}$	120	4107	129,457	198,723	199,999	200,000

Table S2: Evaluation of $\mathbb{E}[N]$ with respect to feature dimension q, sample size n for Study 1.

	n	q				
		2	3	4	5	6
$\mathrm{RR}_{\mathsf{m,adp},0.5}$	200	8	20	51	131	337
	400	10	34	99	305	889
	600	14	42	148	482	1,554
	800	14	54	190	686	2,392
	1000	16	65	238	892	3,276
$\mathrm{RR}_{\mathsf{m,adp},0.7}$	200	22	82	390	1,778	7,493
	400	31	200	1,149	5,724	29,237
	600	47	327	1,879	11,392	65,734
	800	62	420	2,753	18,710	98,457
	1000	74	528	3,619	27,569	123,257

Table S2 and Figures S3 and S4 examine the behavior of $\mathbb{E}[N]$ as func-

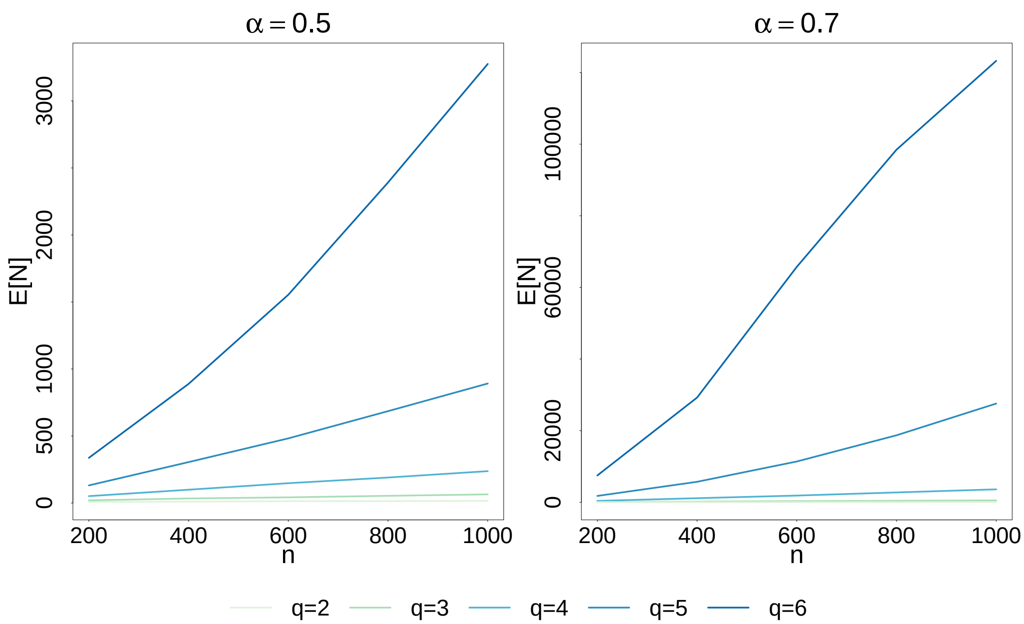

Figure S3: Evaluation of $\mathbb{E}[N]$ with respect to sample size n for $\mathrm{RR}_{\mathsf{m,adp},\alpha}$ when $q = k_0$.

tions of the feature dimension q and the sample size n, respectively, while Figure S5 further investigates the dependence of $\mathbb{E}[N]$ on the threshold parameter α for $\mathrm{RR}_{\mathsf{m,adp},\alpha}$. Collectively, these results provide empirical support for the relationship $\log \mathbb{E}[N] \sim O(q\alpha \log n)$ when $r = (q+2)n^{-\alpha}$ is used for $\mathrm{RR}_{\mathsf{m,adp},\alpha}$.

This conclusion is substantiated by the following observations. First, Figure S3 suggests that $\mathbb{E}[N]$ increases approximately linearly with n on the original scale when $\alpha q \approx 1$, consistent with a logarithmic dependence on n after transformation. Second, Figure S4 demonstrates that $\mathbb{E}[N]$ grows exponentially as q increases, indicating a linear dependence of $\log \mathbb{E}[N]$ on

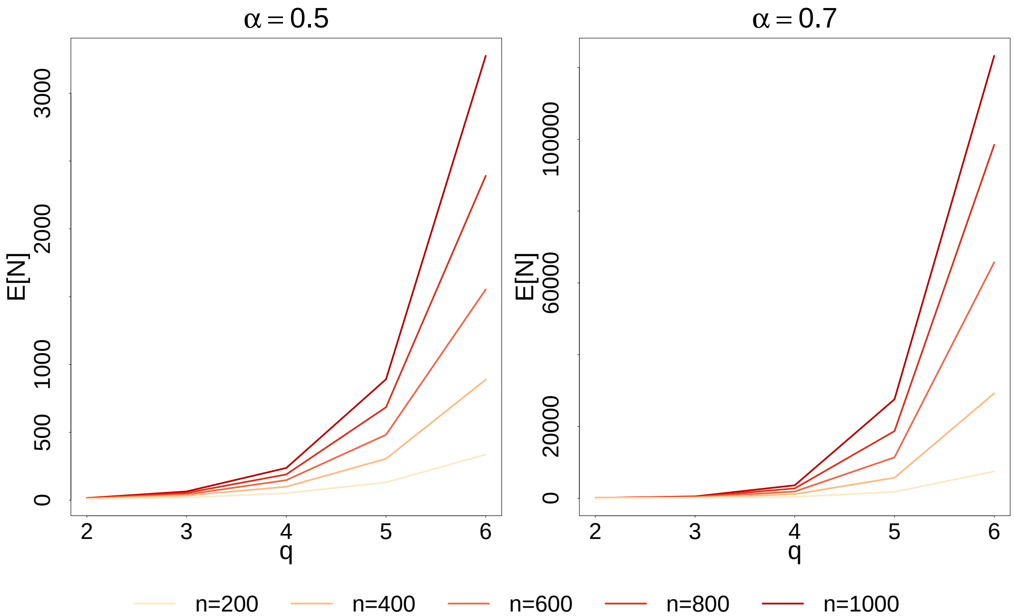

Figure S4: Evaluation of $\mathbb{E}[N]$ with respect to feature dimension $q = k_0$ for $\mathrm{RR}_{\mathsf{m,adp},\alpha}$ in Study 1.

q. Third, Figure S5 shows an exponential increase of $\mathbb{E}[N]$ with respect to α, implying a linear dependence of $\log \mathbb{E}[N]$ on α. These empirical patterns are further corroborated by the numerical summaries reported in Table S2 and Table 2 in the main text.

S3.3 Additional Results for Section 5.2

Figure S6 examines several coordinates of $\Lambda_n(\psi)$, where

$$\psi_i = (X_{i,1}, X_{i,2}, X_{i,3}, X_{i,2}^2, X_{i,3}^2, X_{i,2}X_{i,3})^\top,$$

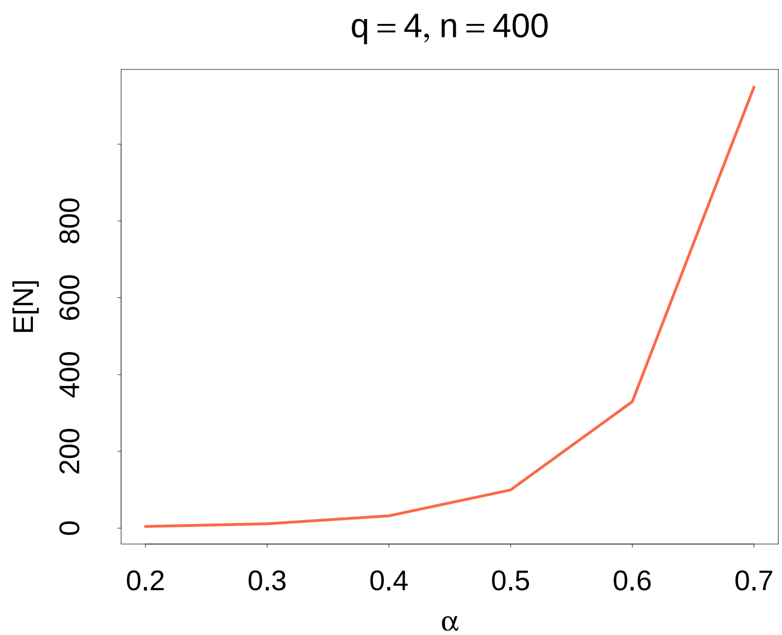

Figure S5: Evaluation of $\mathbb{E}[N]$ with respect to the threshold parameter α for $\mathrm{RR}_{\mathsf{m,adp},\alpha}$ when $q = k_0 = 4$ and $n = 400$ in Study 1.

$X_{i,1} \sim$ Bernoulli(1/2), and $X_{i,2}, X_{i,3} \sim \mathcal{N}(0,1)$ independently. First, although both HH and $\mathrm{RR}_{\mathsf{mar}}$ achieve satisfactory balance for the discrete covariate $X_{i,1}$, they provide less effective balance for the continuous covariates than procedures that explicitly incorporate continuous features. Second, among the designs that jointly account for discrete and continuous covariates, $\mathrm{CAR}_{\mathsf{mix}}$ exhibits superior performance: the imbalance measures associated with the continuous components are most tightly concentrated around zero under $\mathrm{CAR}_{\mathsf{mix}}$. Third, the more stringent rejection threshold used in $\mathrm{RR}_{\mathsf{adp},0.5}$ substantially improves balance across all coordinates of

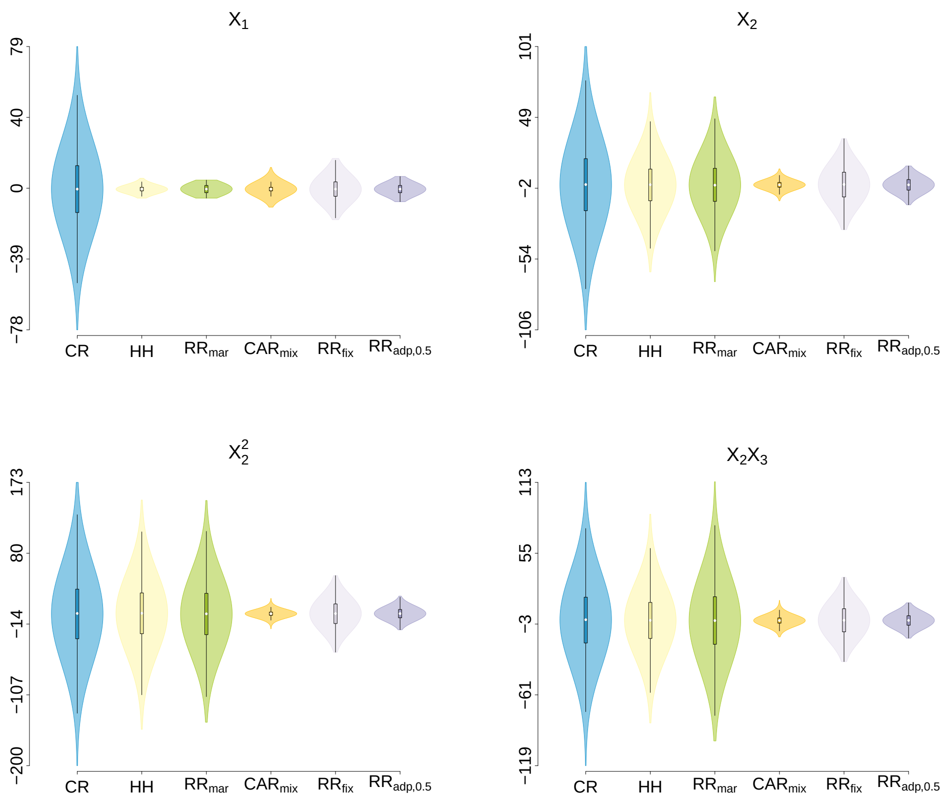

Figure S6: *Distributions of* $\sum_{i=1}^{n}(2T_i-1)X_{i,1}$ *(discrete),* $\sum_{i=1}^{n}(2T_i-1)X_{i,2}$ *(continuous),* $\sum_{i=1}^{n}(2T_i-1)X_{i,2}^2$*, and* $\sum_{i=1}^{n}(2T_i-1)X_{i,2}X_{i,3}$ *under CR, HH,* RR_{mar}*,* CAR_{mix}*,* RR_{fix}*, and* $RR_{\mathsf{adp},0.5}$ *for Models 1 and 2 in Study 2 with* $n = 800$*.*

$\Lambda_n(\psi)$ relative to $\mathrm{RR}_{\mathsf{fix}}$. The resulting imbalance distributions are on a scale comparable to those obtained under $\mathrm{CAR}_{\mathsf{mix}}$. Figure S7 assesses the performance of the considered designs in estimating the treatment effect. Analogous to the findings in Study 1, improved covariate balance reduces the variability of $\hat{\tau}$. Owing to its substantial improvement in covariate balance, $\mathrm{CAR}_{\mathsf{mix}}$ yields estimates of τ that are markedly more concentrated

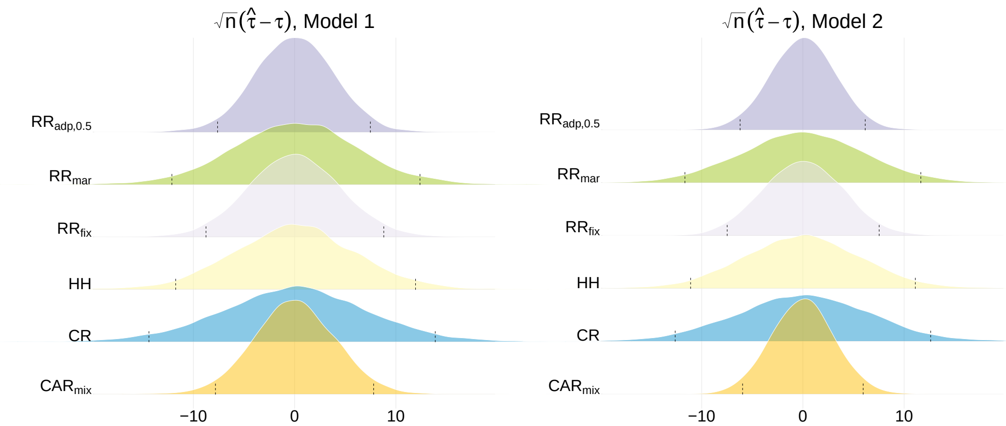

Figure S7: *Distributions of* $\sqrt{n}(\hat{\tau} - \tau)$ *under CR, HH,* RR_{mar}*,* CAR_{mix}*,* RR_{fix}*, and* $RR_{\mathsf{adp},0.5}$ *for Models 1 and 2 in Study 2 with* $n = 800$*.*

around the true effect than those obtained under the alternative designs. Moreover, because HH and $\mathrm{RR}_{\mathsf{mar}}$ fail to attain a level of covariate balance comparable to that achieved by $\mathrm{CAR}_{\mathsf{mix}}$ and $\mathrm{RR}_{\mathsf{adp},0.5}$, their distributions of $\hat{\tau}$ are noticeably more dispersed. By yielding finer covariate balance, $\mathrm{CAR}_{\mathsf{mix}}$ and $\mathrm{RR}_{\mathsf{adp},0.5}$ produce estimates of τ that are more tightly concentrated around the true effect. These qualitative patterns are consistent under Models 1 and 2, underscoring the desirable properties of both CAR and RR designs.

S4 Additional Results for Section 6

We assume the outcomes are generated according to the following model:

$$
\begin{aligned}
\text{FinalHAMD}_i = T_i\mu_1 + (1 - T_i)\mu_0 + \beta_1\text{TreatedCD}_i + \beta_2\text{Race}_i \\
+ f_1(\text{BMI}_i) + f_2(\text{Age}_i) + f_3(\text{HAMD}_i) + \epsilon_i,
\end{aligned}
$$

where ϵ_i are i.i.d. $\mathcal{N}(0, 4)$, $\mu_1 = 12.48$, $\mu_0 = 17.33$, and $\tau = \mu_1 - \mu_0 = -4.85$.

For the discretization-based designs, we dichotomize each of BMI, Age, and HAMD into two levels: $d(\text{BMI}) = 1$ if $\text{BMI} > 23$ and $d(\text{BMI}) = 0$ otherwise; $d(\text{Age}) = 1$ if $\text{Age} > 45$ and $d(\text{Age}) = 0$ otherwise; and $d(\text{HAMD}) = 1$ if $\text{HAMD} > 23$ and $d(\text{HAMD}) = 0$ otherwise. Analogous to Study 2 in Section 5.2, we consider HH and RR_{mar} based on these discretized margins. For designs that incorporate continuous features, we consider $\text{CAR}_{\mathsf{mix}}$ and RR as described in Example 4.3 in the main text. The RR designs with fixed and adaptive thresholds are denoted by RR_{fix} and $\text{RR}_{\mathsf{adp},\alpha}$, respectively.

In this section, we compare coordinates of $\Lambda_n(\psi)$ that are not examined in the main text and further evaluate treatment-effect estimation using the difference-in-means estimator $\hat{\tau}$. Figures S9 and S10, together with Table S3, examine coordinates of $\Lambda_n(\psi)$ for $\psi_i = (\text{TreatedCD}_i, \text{Race}_i, \text{BMI}_i, \text{Age}_i, \text{HAMD}_i, \text{BMI}_i^2, \text{Age}_i^2, \text{HAMD}_i^2)^\top$. The results are evaluated under 10,000 repetitions.

The main conclusions can be summarized as follows. First, $\text{CAR}_{\mathsf{mix}}$ consistently outperforms the competing designs in achieving covariate balance across all components of ψ_i, as evidenced by the uniformly tighter concentration of the distributions of $\{\sum_{i=1}^{n}(2T_i - 1)\psi_{i,j} : 1 \leq j \leq 8\}$ around zero. Second, a comparison among $\mathrm{CAR}_{\mathsf{mix}}$, $\mathrm{RR}_{\mathsf{fix}}$, and $\mathrm{RR}_{\mathsf{adp},0.5}$ indicates that RR with an appropriately chosen threshold can attain performance close to that of CAR. In particular, Figures S9 and S10 show that the distributions of $\sum_{i=1}^{n}(2T_i - 1)\psi_{i,j}$ under $\mathrm{RR}_{\mathsf{adp},0.5}$ are close to those under $\mathrm{CAR}_{\mathsf{mix}}$. As shown in Table S3, the standard deviations under $\mathrm{RR}_{\mathsf{adp},\alpha}$ approach the corresponding values under $\mathrm{CAR}_{\mathsf{mix}}$ as α increases. Finally, the improved covariate balance delivered by CAR and RR enhances the performance of $\hat{\tau}$. The distributions of $\hat{\tau}$ under $\mathrm{RR}_{\mathsf{mar}}$ and HH are comparable but slightly more dispersed than those under $\mathrm{CAR}_{\mathsf{mix}}$ or $\mathrm{RR}_{\mathsf{adp},0.5}$, reflecting the superior balance achieved by the latter two procedures.

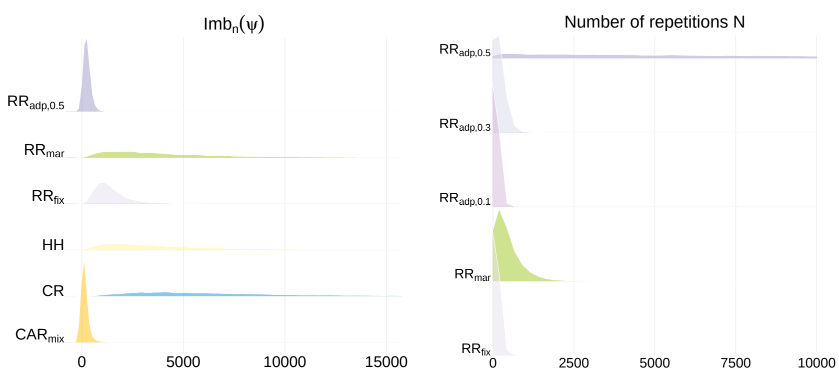

Figure S8: *The distributions of $Imb_n(\psi)$ (left) and N (right) for the study of Keller et al. (2000). The proportion of cap hits for $RR_{\mathsf{adp},0.7}$ is 88.03%; its distribution of N is therefore omitted.*

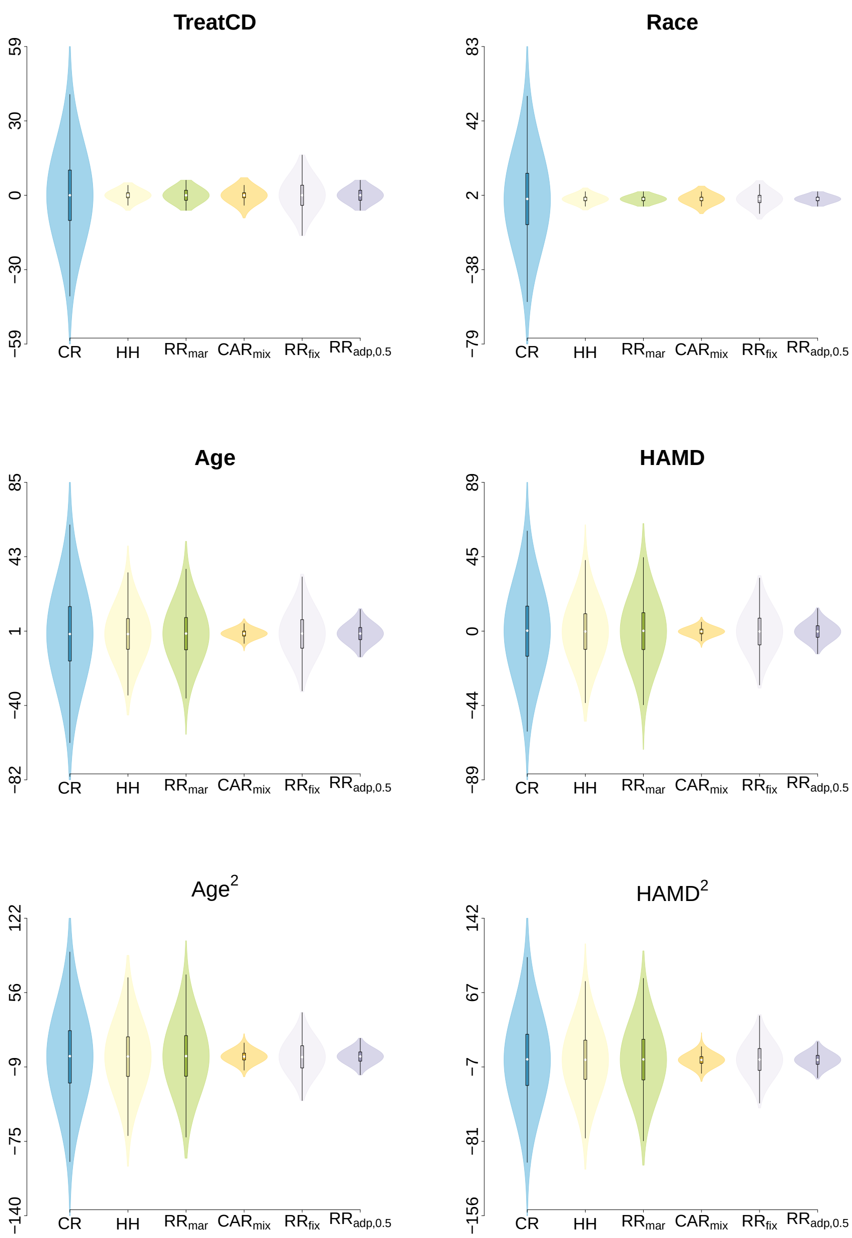

Figure S9: *Distributions of selected coordinates of $\Lambda_n(\psi)$ under CR, HH, RR_{mar}, CAR_{mix}, RR_{fix}, and $RR_{\mathsf{adp},0.5}$ for the study of Keller et al. (2000). The top row displays TreatedCD and Race, the second row displays Age and HAMD, and the bottom row displays Age^2 and $HAMD^2$.*

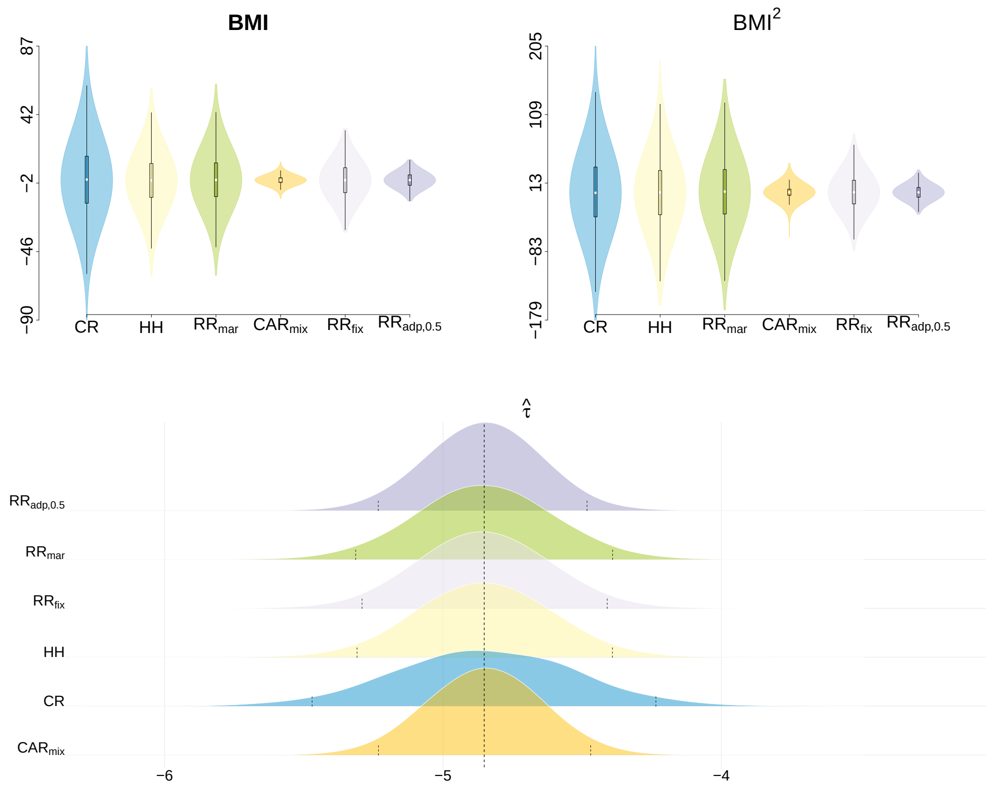

Figure S10: *Simulation results for the study of Keller et al. (2000). The top row displays the distributions of the covariate imbalances for BMI and* BMI^2*; the bottom panel displays the distributions of* $\hat{\tau}$*, with the vertical dotted line marking the true treatment effect* $\tau = -4.85$.

Table S3: Evaluation of covariate imbalance under different design approaches for the study of Keller et al. (2000).

Design	$\Lambda_n(\psi)$				
	ψ_j =Race	ψ_j =Age	ψ_j =HAMD	ψ_j = Age^2	ψ_j = HAMD^2
	mean/s.d.	mean/s.d.	mean/s.d.	mean/s.d.	mean/s.d.
CR	-0.09/21.21	-0.17/22.30	0.29/22.30	-0.20/34.40	0.25/37.82
HH	-0.02/ 1.41	-0.25/12.73	-0.09/15.73	-0.21/25.83	-0.10/28.96
$\mathrm{RR_{mar}}$	-0.01/ 1.43	-0.09/13.38	0.14/16.40	0.31/26.08	0.49/29.72
$\mathrm{CAR_{mix}}$	0.00/ 1.89	0.02/ 2.11	-0.01/ 2.16	-0.04/ 4.55	-0.07/ 5.10
$\mathrm{RR_{fix}}$	-0.00/ 3.46	-0.12/11.30	-0.04/11.46	-0.26/13.63	0.29/15.26
$\mathrm{RR_{adp,0.1}}$	-0.01/ 4.74	0.01/15.36	-0.13/15.52	-0.07/18.46	-0.10/20.87
$\mathrm{RR_{adp,0.3}}$	-0.03/ 2.63	0.04/ 8.64	-0.08/ 8.63	-0.01/10.51	-0.21/11.62
$\mathrm{RR_{adp,0.5}}$	-0.00/ 1.43	-0.04/ 4.73	-0.03/ 4.70	-0.03/ 5.63	0.01/ 6.25
$\mathrm{RR_{adp,0.7}}$	-0.01/ 1.07	0.08/ 3.54	-0.01/ 3.50	-0.05/ 4.19	-0.04/ 4.73

References

Keller, M. B., J. P. McCullough, D. N. Klein, et al. (2000). A comparison of nefazodone, the cognitive behavioral-analysis system of psychotherapy, and their combination for the treatment of chronic depression. *New England Journal of Medicine 342*(20), 1462–1470.

Li, X. and P. Ding (2020). Rerandomization and regression adjustment. *Journal of the Royal Statistical Society Series B: Statistical Methodology 82*(1), 241–268.

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.

Morgan, K. L., D. B. Rubin, et al. (2012). Rerandomization to improve covariate balance in experiments. *The Annals of Statistics 40*(2), 1263–1282.

Ziji Qin
Department of Statistics, The George Washington University
E-mail: zqin123@gwu.edu

Feifang Hu
Department of Statistics, The George Washington University
E-mail: feifang@gwu.edu

Yang Liu

Institute of Statistics and Big Data, Renmin University of China
E-mail: yangliu2022@ruc.edu.cn