Web Appendix for "Accounting for unequal cluster sizes in designing cluster randomized trials to detect treatment effect heterogeneity" by Tong et al.

A. DERIVATION OF VARIANCE EXPRESSIONS FOR HTE AND AVERAGE TREATMENT EFFECT ANALYSES WITH UNEQUAL CLUSTER SIZES

A.1 Variance expression for the HTE estimator

We resume the notation in Section 2.2 of the main text. Recall the exchangeable o-ICC matrix for cluster i is written as $\boldsymbol{R}_i = (1-\rho_{y|x})\boldsymbol{I}_{m_i} + \rho_{y|x}\boldsymbol{J}_{m_i}$, where $\boldsymbol{I}_{m_i}$ and $\boldsymbol{J}_{m_i}$ are $m_i \times m_i$ identity and matrix of ones, respectively. The inverse of this correlation matrix is,

$$\boldsymbol{R}_i^{-1} = \frac{1}{1-\rho_{y|x}}\boldsymbol{I}_{m_i} - \frac{\rho_{y|x}}{(1-\rho_{y|x})\{1+(m_i-1)\rho_{y|x}\}}\boldsymbol{J}_{m_i} = \frac{1}{1-\rho_{y|x}}(\boldsymbol{I}_{m_i} + c_i\boldsymbol{J}_{m_i}),$$

where $c_i = -\rho_{y|x}/\{1+(m_i-1)\rho_{y|x}\}$. Recall that

$$\frac{1}{n}\boldsymbol{U}_n = \left(\frac{1}{1-\rho_{y|x}}\right)\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{Z}_i^T\boldsymbol{Z}_i - \left(\frac{1}{1-\rho_{y|x}}\right)\frac{1}{n}\sum_{i=1}^{n}c_i\boldsymbol{Z}_i^T\boldsymbol{J}_{m_i}\boldsymbol{Z}_i = \left(\frac{1}{1-\rho_{y|x}}\right)\frac{1}{n}(\boldsymbol{S}_n + \boldsymbol{T}_n),$$

where

$$\frac{1}{n}\boldsymbol{S}_n = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\boldsymbol{Z}_{ij}\boldsymbol{Z}_{ij}^T = \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} m_i & m_i(W_i-\overline{W}) & \sum_{j=1}^{m_i}X_{ij} & (W_i-\overline{W})\sum_{j=1}^{m_i}X_{ij} \\ m_i(W_i-\overline{W}) & m_i(W_i-\overline{W})^2 & (W_i-\overline{W})\sum_{j=1}^{m_i}X_{ij} & (W_i-\overline{W})^2\sum_{j=1}^{m_i}X_{ij} \\ \sum_{j=1}^{m_i}X_{ij} & (W_i-\overline{W})\sum_{j=1}^{m_i}X_{ij} & \sum_{j=1}^{m_i}X_{ij}^2 & (W_i-\overline{W})\sum_{j=1}^{m_i}X_{ij}^2 \\ (W_i-\overline{W})\sum_{j=1}^{m_i}X_{ij} & (W_i-\overline{W})^2\sum_{j=1}^{m_i}X_{ij} & (W_i-\overline{W})\sum_{j=1}^{m_i}X_{ij}^2 & (W_i-\overline{W})^2\sum_{j=1}^{m_i}X_{ij}^2 \end{bmatrix},$$

and

$$\frac{1}{n}\boldsymbol{T}_n = \frac{1}{n}\sum_{i=1}^{n}c_i\left(\sum_{j=1}^{m_i}\boldsymbol{Z}_{ij}\right)\left(\sum_{j=1}^{m_i}\boldsymbol{Z}_{ij}^T\right) =$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{-\rho_{y|x}}{1+(m_i-1)\rho_{y|x}}\begin{bmatrix} m_i^2 & m_i^2(W_i-\bar{W}) & m_i\sum_{j=1}^{m_i}X_{ij} & m_i(W_i-\overline{W})\sum_{j=1}^{m_i}X_{ij} \\ m_i^2(W_i-\overline{W}) & m_i^2(W_i-\overline{W})^2 & m_i(W_i-\overline{W})\sum_{j=1}^{m_i}X_{ij} & m_i(W_i-\overline{W})^2\sum_{j=1}^{m_i}X_{ij} \\ m_i\sum_{j=1}^{m_i}X_{ij} & m_i(W_i-\overline{W})\sum_{j=1}^{m_i}X_{ij} & \left(\sum_{j=1}^{m_i}X_{ij}\right)^2 & (W_i-\overline{W})\left(\sum_{j=1}^{m_i}X_{ij}\right)^2 \\ m_i(W_i-\overline{W})\sum_{j=1}^{m_i}X_{ij} & m_i(W_i-\overline{W})^2\sum_{j=1}^{m_i}X_{ij} & (W_i-\overline{W})\left(\sum_{j=1}^{m_i}X_{ij}\right)^2 & (W_i-\overline{W})^2\left(\sum_{j=1}^{m_i}X_{ij}\right)^2 \end{bmatrix}.$$

Assuming the cluster sizes are non-informative such that m_i is a random draw from some population density $f(m_i)$ with bounded first and second moments. We further define and derive the following population-level parameters:

1. $\overline{m} = E(m_i)$ is the mean of cluster sizes;
2. The conditional expectation of X_{ij} in cluster i is μ_{1i}, and the population or marginal expectation is $\mu_1 = E(\mu_{1i})$. We then treat $\sum_{j=1}^{m_i}X_{ij}$ as a random variable, which has conditional expectation of $m_i\mu_i$ in cluster i. Based on the Law of Iterated Expectation, we obtain $\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\left(\sum_{j=1}^{m_i}X_{ij}\right) = \overline{m}\mu_1$;
3. Assuming cluster-specific conditional expectation of X_{ij}^2 is μ_{2i} and the population or marginal expectation is $\mu_2 = E(\mu_{2i})$. Again we treat $\sum_{j=1}^{m_i}X_{ij}^2$ as a random variable with cluster-specific mean $m_i\mu_2$, and $\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\left(\sum_{j=1}^{m_i}X_{ij}^2\right) = \overline{m}\mu_2$;
4. Define the following expectation of functions of cluster sizes:

$$\overline{q} = E\left\{\frac{-m_i^2\rho_{y|x}}{1+(m_i-1)\rho_{y|x}}\right\}, \qquad \overline{p} = E\left\{\frac{-m_i\rho_{y|x}}{1+(m_i-1)\rho_{y|x}}\right\}.$$

Since we assume the cluster sizes are non-informative and therefore uncorrelated with the effect modifier, we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{-\rho_{y|x}}{1+(m_i-1)\rho_{y|x}}\right\}m_i\left(\sum_{j=1}^{m_i}X_{ij}\right) = \overline{q}\mu_1.$$

5. Define the conditional covariance between X_{ij} and X_{ik} in cluster i is κ_i, and therefore the marginal expectation $E(X_{ij}X_{ik}) = E(\kappa_i) = \kappa$. By definition, $\kappa = \rho_x(\mu_2 - \mu_1^2) + \mu_1^2 = \rho_x\mu_2 + (1-\rho_x)\mu_1^2$, and $\sigma_x^2 = \mu_2 - \mu_1^2$. Therefore,

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{-\rho_{y|x}}{1+(m_i-1)\rho_{y|x}}\right\}\left(\sum_{j=1}^{m_i}X_{ij}\right)^2 &= \overline{p}\mu_2 + (\overline{q}-\overline{p})\kappa \\
&= \{\overline{p} + (\overline{q}-\overline{p})\rho_x\}\mu_2 + (\overline{q}-\overline{p})(1-\rho_x)\mu_1^2 \\
&= \overline{p}\mu_2 + (\overline{q}-\overline{p})\rho_x\mu_2 + (\overline{q}-\overline{p})\mu_1^2 - (\overline{q}-\overline{p})\rho_x\mu_1^2 \\
&= \overline{q}\mu_1^2 + \overline{p}\sigma_x^2 + (\overline{q}-\overline{p})\rho_x\sigma_x^2 \\
&= \overline{q}\mu_2 - (\overline{q}-\overline{p})\sigma_x^2(1-\rho_x) \\
&= \overline{q}\eta_2,
\end{aligned}$$

where we define $\eta_2 = \mu_2 - (1-\overline{p}/\overline{q})\sigma_x^2(1-\rho_x)$ to simplify the subsequent notation;
6. Define $\sigma_w^2 = \overline{W}(1-\overline{W})$ as the variance of the treatment assignment.

Then we have the limit form of $\boldsymbol{S}_n$ and $\boldsymbol{T}_n$ given in Section 2.2,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{S}_n = \overline{m}\begin{bmatrix}1 & 0 & \mu_1 & 0\\ 0 & \sigma_w^2 & 0 & \mu_1\sigma_w^2\\ \mu_1 & 0 & \mu_2 & 0\\ 0 & \mu_1\sigma_w^2 & 0 & \mu_2\sigma_w^2\end{bmatrix},$$

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{T}_n = \overline{q}\begin{bmatrix}1 & 0 & \mu_1 & 0\\ 0 & \sigma_w^2 & 0 & \mu_1\sigma_w^2\\ \mu_1 & 0 & \eta_2 & 0\\ 0 & \mu_1\sigma_w^2 & 0 & \eta_2\sigma_w^2\end{bmatrix},$$

and furthermore,

$$\lim_{n\to\infty}\frac{(1-\rho_{y|x})}{n}\boldsymbol{U}_n = \begin{bmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{C} & \boldsymbol{D}\end{bmatrix} = \begin{bmatrix}\boldsymbol{A} & \mu_1\boldsymbol{A}\\ \mu_1\boldsymbol{A} & \boldsymbol{D}\end{bmatrix} = \begin{bmatrix}\overline{m}+\overline{q} & 0 & (\overline{m}+\overline{q})\mu_1 & 0\\ 0 & (\overline{m}+\overline{q})\sigma_w^2 & 0 & (\overline{m}+\overline{q})\mu_1\sigma_w^2\\ (\overline{m}+\overline{q})\mu_1 & 0 & \overline{m}\mu_2+\overline{q}\eta_2 & 0\\ 0 & (\overline{m}+\overline{q})\mu_1\sigma_w^2 & 0 & (\overline{m}\mu_2+\overline{q}\eta_2)\sigma_w^2\end{bmatrix}$$

Denote the inverse of the above matrix as,

$$\begin{bmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{C} & \boldsymbol{D}\end{bmatrix}^{-1} = \begin{bmatrix}\boldsymbol{E} & \boldsymbol{F}\\ \boldsymbol{G} & \boldsymbol{H}\end{bmatrix}$$

By block matrix inversion and some algebra, we have,

$$\begin{aligned}
\boldsymbol{E} = (\boldsymbol{A}-\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1} &= \frac{\overline{m}\mu_2+\overline{q}\eta_2}{(\overline{m}+\overline{q})\sigma_x^2\{\overline{m}+(1-\rho_x)\overline{p}+\rho_x\overline{q}\}}\begin{bmatrix}1 & 0\\ 0 & \sigma_w^{-2}\end{bmatrix}\\
\boldsymbol{H} = (\boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} = (\boldsymbol{D}-\mu_1\boldsymbol{B})^{-1} &= \frac{1}{\sigma_x^2\{\overline{m}+(1-\rho_x)\overline{p}+\rho_x\overline{q}\}}\begin{bmatrix}1 & 0\\ 0 & \sigma_w^{-2}\end{bmatrix}\\
\boldsymbol{F} = \boldsymbol{G} = -\boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} = -\mu_1\boldsymbol{H} &= \frac{-\mu_1}{\sigma_x^2\{\overline{m}+(1-\rho_x)\overline{p}+\rho_x\overline{q}\}}\begin{bmatrix}1 & 0\\ 0 & \sigma_w^{-2}\end{bmatrix}
\end{aligned}$$

Therefore, large-sample variance of $\sqrt{n}\hat{\boldsymbol{b}}$ is

$$\lim_{n\to\infty}\boldsymbol{\Sigma}_n = \sigma_{y|x}^2(\lim_{n\to\infty}n^{-1}\boldsymbol{U}_n)^{-1} = \sigma_{y|x}^2(1-\rho_{y|x})\begin{bmatrix}\boldsymbol{E} & \boldsymbol{F}\\ \boldsymbol{G} & \boldsymbol{H}\end{bmatrix}$$

Denote CV as the coefficient of variance of cluster size. Based on the Taylor series expansion result in van Breukelen et al.[1], we have

$$\overline{p} = E\left\{\frac{-m_i\rho_{y|x}}{1+(m_i-1)\rho_{y|x}}\right\} = -E\left\{\frac{m_i}{m_i+(1-\rho_{y|x})/\rho_{y|x}}\right\} \approx -\rho_{y|x}\left[\frac{\overline{m}}{1+(\overline{m}-1)\rho_{y|x}}\right]\left[1-\mathrm{CV}^2\frac{\overline{m}\rho_{y|x}(1-\rho_{y|x})}{\{1+(\overline{m}-1)\rho_{y|x}\}^2}\right]$$

$$\begin{aligned}\overline{q} = E\left\{\frac{-m_i^2\rho_{y|x}}{1+(m_i-1)\rho_{y|x}}\right\} &= -\overline{m} + E\left\{\frac{-m_i^2\rho_{y|x}}{1+(m_i-1)\rho_{y|x}}+m_i\right\} = -\overline{m} - \left(\frac{1-\rho_{y|x}}{\rho_{y|x}}\right)E\left\{\frac{-m_i\rho_{y|x}}{1+(m_i-1)\rho_{y|x}}\right\}\\ &= -\overline{m} - \left(\frac{1-\rho_{y|x}}{\rho_{y|x}}\right)\overline{p}\end{aligned}$$

We also notice that $\overline{m}+\overline{q} = -\left(\frac{1-\rho_{y|x}}{\rho_{y|x}}\right)\overline{p}$. One common component for $\boldsymbol{E}$, $\boldsymbol{F}$, $\boldsymbol{G}$, and $\boldsymbol{H}$ is $\overline{m}+(1-\rho_x)\overline{p}+\rho_x\overline{q}$, which can be further simplified by using the above Taylor series approximations,

$$\begin{aligned}\overline{m}+(1-\rho_x)\overline{p}+\rho_x\overline{q} =& \overline{m}+(1-\rho_x)\overline{p}-\rho_x\overline{m}-\rho_x\left(\frac{1-\rho_{y|x}}{\rho_{y|x}}\right)\overline{p} = (1-\rho_x)\overline{m}+\left(\frac{\rho_{y|x}-\rho_x}{\rho_{y|x}}\right)\overline{p}\\ \approx&(1-\rho_x)\overline{m}-\left(\frac{\rho_{y|x}-\rho_x}{\rho_{y|x}}\right)\rho_{y|x}\left\{\frac{\overline{m}}{1+(\overline{m}-1)\rho_{y|x}}\right\}\left[1-\mathrm{CV}^2\frac{\overline{m}\rho_{y|x}(1-\rho_{y|x})}{\{1+(\overline{m}-1)\rho_{y|x}\}^2}\right]\\ =&(1-\rho_x)\overline{m}-\left\{\frac{\overline{m}(\rho_{y|x}-\rho_x)}{1+(\overline{m}-1)\rho_{y|x}}\right\}\left[1-\mathrm{CV}^2\frac{\overline{m}\rho_{y|x}(1-\rho_{y|x})}{\{1+(\overline{m}-1)\rho_{y|x}\}^2}\right]\\ =&\frac{\overline{m}\{1+(\overline{m}-2)\rho_{y|x}-(\overline{m}-1)\rho_x\rho_{y|x}\}\{1+(\overline{m}-1)\rho_{y|x}\}^2+\overline{m}^2\mathrm{CV}^2\rho_{y|x}(1-\rho_{y|x})(\rho_{y|x}-\rho_x)}{\{1+(\overline{m}-1)\rho_{y|x}\}^3},\end{aligned}$$

which suggests that the variance of HTE estimator has the form,

$$\begin{aligned}n\mathrm{Var}(\hat{\beta}_4) =& \frac{\sigma_y^2(1-\rho_{y|x})}{\sigma_w^2\sigma_x^2\{\overline{m}+(1-\rho_x)\overline{p}+\rho_x\overline{q}\}}\\ =&\frac{\sigma_y^2(1-\rho_{y|x})\{1+(\overline{m}-1)\rho_{y|x}\}^3}{\sigma_w^2\sigma_x^2\overline{m}\left[\{1+(\overline{m}-2)\rho_{y|x}-(\overline{m}-1)\rho_x\rho_{y|x}\}\{1+(\overline{m}-1)\rho_{y|x}\}^2+\overline{m}\mathrm{CV}^2\rho_{y|x}(1-\rho_{y|x})(\rho_{y|x}-\rho_x)\right]}.\end{aligned}$$

This gives rise to the sample size formula (7) in Section 2.2.

A.2 Variance expression for the adjusted average treatment effect estimator

To test for the average treatment effect, we focus on deriving the variance of $\hat{\beta}_2$ when the effect modifier is mean centered (e.g., $\mu_1 = 0$). Observe that

$$\begin{aligned}n\mathrm{Var}(\hat{\beta}_2) =& \sigma_{y|x}^2(1-\rho_{y|x})\left[\frac{\overline{m}\mu_2+\overline{q}\eta_2}{(\overline{m}+\overline{q})\sigma_x^2\sigma_w^2\{\overline{m}+(1-\rho_x)\overline{p}+\rho_x\overline{q}\}}\right]\\ =&\sigma_{y|x}^2(1-\rho_{y|x})\left[\frac{\overline{m}(\mu_2-\mu_1^2)+\overline{q}(\eta_2-\mu_1^2)}{(\overline{m}+\overline{q})\sigma_x^2\sigma_w^2\{\overline{m}+(1-\rho_x)\overline{p}+\rho_x\overline{q}\}}\right]\\ =&\frac{\sigma_{y|x}^2[1+(\overline{m}-1)\rho_{y|x}]}{\sigma_w^2\overline{m}}\left[1-\mathrm{CV}^2\frac{\overline{m}\rho_{y|x}(1-\rho_{y|x})}{\{1+(\overline{m}-1)\rho_{y|x}\}^2}\right]^{-1},\end{aligned}$$

where the last equation holds because $\overline{m}(\mu_2-\mu_1^2)+\overline{q}(\eta_2-\mu_1^2) = \sigma_x^2\{\overline{m}+(1-\rho_x)\overline{p}+\rho_x\overline{q}\}$ and

$$\overline{m}+\overline{q} = \left(\frac{1-\rho_{y|x}}{\rho_{y|x}}\right)\rho_{y|x}\left\{\frac{\overline{m}}{1+(\overline{m}-1)\rho_{y|x}}\right\}\left[1-\mathrm{CV}^2\frac{\overline{m}\rho_{y|x}(1-\rho_{y|x})}{\{1+(\overline{m}-1)\rho_{y|x}\}^2}\right].$$

B. DERIVATION OF VARIANCE EXPRESSIONS FOR HTE AND AVERAGE TREATMENT EFFECT ANALYSES WITH UNEQUAL CLUSTER SIZES AND MULTIVARIATE EFFECT MODIFIERS

B.1 Variance-covariance matrix for the HTE estimator

We resume the notation in Section 3.1 of the main text, and write the reparameterized linear mixed model as

$$Y_{ij} = b_1 + b_2(W_i - \overline{W}) + \boldsymbol{b}_3^T \boldsymbol{X}_{ij} + \boldsymbol{b}_4^T (W_i - \overline{W})\boldsymbol{X}_{ij} + \lambda_i + \epsilon_{ij}, \quad \lambda_i \sim \mathcal{N}(0, \sigma_\lambda^2), \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2).$$

where $b_1 = \beta_1 + \beta_2 \overline{W}$, $b_2 = \beta_2$, $\boldsymbol{b}_3 = \boldsymbol{\beta}_3 + \boldsymbol{\beta}_4 \overline{W}$, and $\boldsymbol{b}_4 = \boldsymbol{\beta}_4$. Define the collection of design points $\boldsymbol{Z}_{ij} = (1, (W_i - \overline{W}), \boldsymbol{X}_{ij}^T, (W_i - \overline{W})\boldsymbol{X}_{ij}^T)^T$ and $\boldsymbol{Z}_i = (\boldsymbol{Z}_{i1}, \ldots, \boldsymbol{Z}_{im_i})^T$. Given the value of σ_λ^2 and σ_ϵ^2, the maximum likelihood estimator, $\sqrt{n}(\hat{\boldsymbol{b}} - \boldsymbol{b})$ is asymptotically normal with mean zero and variance equal to $\lim_{n\to\infty} \boldsymbol{\Sigma}_n = \sigma_{y|x}^2 (\lim_{n\to\infty} n^{-1}\boldsymbol{U}_n)^{-1}$, where $\boldsymbol{U}_n = \Sigma_{i=1}^n \boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i$. Extending the derivations in Appendix A, we have

$$\frac{1}{n}\boldsymbol{U}_n = \left(\frac{1}{1-\rho_{y|x}}\right)\frac{1}{n}\sum_{i=1}^n \boldsymbol{Z}_i^T \boldsymbol{Z}_i - \left(\frac{1}{1-\rho_{y|x}}\right)\frac{1}{n}\sum_{i=1}^n c_i \boldsymbol{Z}_i^T \boldsymbol{J}_{m_i} \boldsymbol{Z}_i = \left(\frac{1}{1-\rho_{y|x}}\right)\frac{1}{n}(\boldsymbol{S}_n + \boldsymbol{T}_n),$$

where

$$\begin{aligned}\frac{1}{n}\boldsymbol{S}_n =& \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^{m_i} \boldsymbol{Z}_{ij}\boldsymbol{Z}_{ij}^T \\ =& \frac{1}{n}\sum_{i=1}^n \begin{bmatrix} m_i & m_i(W_i - \overline{W}) & \sum_{j=1}^{m_i} \boldsymbol{X}_{ij}^T & (W_i - \overline{W})\sum_{j=1}^{m_i} \boldsymbol{X}_{ij}^T \\ m_i(W_i - \overline{W}) & m_i(W_i - \overline{W})^2 & (W_i - \overline{W})\sum_{j=1}^{m_i} \boldsymbol{X}_{ij}^T & (W_i - \overline{W})^2\sum_{j=1}^{m_i} \boldsymbol{X}_{ij}^T \\ \sum_{j=1}^{m_i} \boldsymbol{X}_{ij} & (W_i - \overline{W})\sum_{j=1}^{m_i} \boldsymbol{X}_{ij} & \sum_{j=1}^{m_i} \boldsymbol{X}_{ij}\boldsymbol{X}_{ij}^T & (W_i - \overline{W})\sum_{j=1}^{m_i} \boldsymbol{X}_{ij}\boldsymbol{X}_{ij}^T \\ (W_i - \overline{W})\sum_{j=1}^{m_i} \boldsymbol{X}_{ij} & (W_i - \overline{W})^2\sum_{j=1}^{m_i} \boldsymbol{X}_{ij} & (W_i - \overline{W})\sum_{j=1}^{m_i} \boldsymbol{X}_{ij}\boldsymbol{X}_{ij}^T & (W_i - \overline{W})^2\sum_{j=1}^{m_i} \boldsymbol{X}_{ij}\boldsymbol{X}_{ij}^T \end{bmatrix},\end{aligned}$$

and

$$\frac{1}{n}\boldsymbol{T}_n = \frac{1}{n}\sum_{i=1}^n c_i \left(\sum_{j=1}^{m_i} \boldsymbol{Z}_{ij}\right)\left(\sum_{j=1}^{m_i} \boldsymbol{Z}_{ij}^T\right) =$$

$$\frac{1}{n}\sum_{i=1}^n c_i \begin{bmatrix} m_i^2 & m_i^2(W_i - \bar{W}) & m_i \sum_{j=1}^{m_i} \boldsymbol{X}_{ij}^T & m_i(W_i - \overline{W})\sum_{j=1}^{m_i} \boldsymbol{X}_{ij}^T \\ m_i^2(W_i - \overline{W}) & m_i^2(W_i - \overline{W})^2 & m_i(W_i - \overline{W})\sum_{j=1}^{m_i} \boldsymbol{X}_{ij}^T & m_i(W_i - \overline{W})^2\sum_{j=1}^{m_i} \boldsymbol{X}_{ij}^T \\ m_i \sum_{j=1}^{m_i} \boldsymbol{X}_{ij}^T & m_i(W_i - \overline{W})\sum_{j=1}^{m_i} \boldsymbol{X}_{ij}^T & \left(\sum_{j=1}^{m_i} \boldsymbol{X}_{ij}\right)\left(\sum_{j=1}^{m_i} \boldsymbol{X}_{ij}\right)^T & (W_i - \overline{W})\left(\sum_{j=1}^{m_i} \boldsymbol{X}_{ij}\right)\left(\sum_{j=1}^{m_i} \boldsymbol{X}_{ij}\right)^T \\ m_i(W_i - \overline{W})\sum_{j=1}^{m_i} \boldsymbol{X}_{ij}^T & m_i(W_i - \overline{W})^2\sum_{j=1}^{m_i} \boldsymbol{X}_{ij}^T & (W_i - \overline{W})\left(\sum_{j=1}^{m_i} \boldsymbol{X}_{ij}\right)\left(\sum_{j=1}^{m_i} \boldsymbol{X}_{ij}\right)^T & (W_i - \overline{W})^2\left(\sum_{j=1}^{m_i} \boldsymbol{X}_{ij}\right)\left(\sum_{j=1}^{m_i} \boldsymbol{X}_{ij}\right)^T \end{bmatrix}$$

Both $\boldsymbol{S}_n$ and $\boldsymbol{T}_n$ are matrices with the size of $(2p+2)\times(2p+2)$. We can obtain the limit forms of $\boldsymbol{S}_n$ and $\boldsymbol{T}_n$ by noting the following results:

1. Define $\overline{m} = E(m_i)$ as the mean cluster size;
2. The conditional expectation of $\boldsymbol{X}_{ij}$ in cluster i is $\boldsymbol{\mu}_{1i}$, and the population or marginal expectation is $\boldsymbol{\mu}_1 = E(\boldsymbol{\mu}_{1i})$. We then treat $\sum_{j=1}^{m_i} \boldsymbol{X}_{ij}$ as a random vector, which has conditional expectation of $m_i\boldsymbol{\mu}_i$ in cluster i. Based on the Law of Iterated Expectation, we obtain $\lim_{n\to\infty} n^{-1}\sum_{i=1}^n \left(\sum_{j=1}^{m_i} \boldsymbol{X}_{ij}\right) = \overline{m}\boldsymbol{\mu}_1$;
3. Assuming cluster-specific conditional expectation of $\boldsymbol{X}_{ij}\boldsymbol{X}_{ij}^T$ is $\boldsymbol{\mu}_{2i}$ and the population or marginal expectation is $\boldsymbol{\mu}_2 = E(\boldsymbol{\mu}_{2i})$. Again we treat $\sum_{j=1}^{m_i} \boldsymbol{X}_{ij}\boldsymbol{X}_{ij}^T$ as a random matrix with cluster-specific mean $m_i\boldsymbol{\mu}_2$, and its limit $\lim_{n\to\infty} n^{-1}\sum_{i=1}^n \left(\sum_{j=1}^{m_i} \boldsymbol{X}_{ij}\boldsymbol{X}_{ij}^T\right) = \overline{m}\boldsymbol{\mu}_2$;
4. Define the following expectation of functions of cluster sizes:

$$\overline{q} = E\left\{\frac{-m_i^2\rho_{y|x}}{1+(m_i-1)\rho_{y|x}}\right\}, \qquad \overline{p} = E\left\{\frac{-m_i\rho_{y|x}}{1+(m_i-1)\rho_{y|x}}\right\}.$$

Since we assume the cluster sizes are non-informative and therefore uncorrelated with the effect modifier, we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \left\{\frac{-\rho_{y|x}}{1+(m_i-1)\rho_{y|x}}\right\} m_i \left(\sum_{j=1}^{m_i} \boldsymbol{X}_{ij}\right) = \overline{q}\boldsymbol{\mu}_1.$$

5. Define the conditional covariance between $\boldsymbol{X}_{ij}$ and $\boldsymbol{X}_{ik}$ in cluster i is κ_i, and therefore the marginal expectation $E(\boldsymbol{X}_{ij}\boldsymbol{X}_{ik}) = E(\boldsymbol{\kappa}_i) = \boldsymbol{\kappa}$. By definition, $\boldsymbol{\kappa}_i = \boldsymbol{\Lambda}_x^{1/2}\boldsymbol{\Gamma}_x^0\boldsymbol{\Lambda}_x^{1/2} + \boldsymbol{\mu}_1\boldsymbol{\mu}_1^T$, with $\boldsymbol{\Lambda}_x = \text{diag}(\sigma^2_{x_1}, \dots, \sigma^2_{x_p})$. Therefore,

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{-\rho_{y|x}}{1+(m_i-1)\rho_{y|x}}\right\}\left(\sum_{j=1}^{m_i}\boldsymbol{X}_{ij}\right)\left(\sum_{j=1}^{m_i}\boldsymbol{X}_{ij}\right)^T \\
=&\overline{p}\boldsymbol{\mu}_2 + (\overline{q}-\overline{p})\boldsymbol{\kappa} \\
=&\overline{p}\boldsymbol{\mu}_2 + (\overline{q}-\overline{p})\left(\boldsymbol{\Lambda}_x^{1/2}\boldsymbol{\Gamma}_x^0\boldsymbol{\Lambda}_x^{1/2} + \boldsymbol{\mu}_1\boldsymbol{\mu}_1^T\right) \\
=&\overline{p}\left(\boldsymbol{\Lambda}_x^{1/2}\boldsymbol{\Gamma}_x^1\boldsymbol{\Lambda}_x^{1/2} + \boldsymbol{\mu}_1\boldsymbol{\mu}_1^T\right) + (\overline{q}-\overline{p})\left(\boldsymbol{\Lambda}_x^{1/2}\boldsymbol{\Gamma}_x^0\boldsymbol{\Lambda}_x^{1/2} + \boldsymbol{\mu}_1\boldsymbol{\mu}_1^T\right) \\
=&\boldsymbol{\Lambda}_x^{1/2}\left(\overline{p}\,\boldsymbol{\Gamma}_x^1 + (\overline{q}-\overline{p})\boldsymbol{\Gamma}_x^0\right)\boldsymbol{\Lambda}_x^{1/2} + \overline{q}\boldsymbol{\mu}_1\boldsymbol{\mu}_1^T \\
=&\overline{q}\boldsymbol{\eta}_2,
\end{aligned}$$

where we define $\boldsymbol{\eta}_2 = \overline{q}^{\,-1}\boldsymbol{\Lambda}_x^{1/2}\left(\overline{p}\,\boldsymbol{\Gamma}_x^1 + (\overline{q}-\overline{p})\boldsymbol{\Gamma}_x^0\right)\boldsymbol{\Lambda}_x^{1/2} + \boldsymbol{\mu}_1\boldsymbol{\mu}_1^T$ to simplify the subsequent notation.

6. Define $\sigma_w^2 = \overline{W}(1-\overline{W})$ as the variance of the treatment assignment.

Based on the above intermediate results, we can write

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{S}_n = \overline{m}\begin{bmatrix} 1 & 0 & \boldsymbol{\mu}_1^T & \boldsymbol{0} \\ 0 & \sigma_w^2 & \boldsymbol{0} & \boldsymbol{\mu}_1^T\sigma_w^2 \\ \boldsymbol{\mu}_1 & \boldsymbol{0} & \boldsymbol{\mu}_2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\mu}_1\sigma_w^2 & \boldsymbol{0} & \boldsymbol{\mu}_2\sigma_w^2 \end{bmatrix},$$

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{T}_n = \overline{q}\begin{bmatrix} 1 & 0 & \boldsymbol{\mu}_1^T & \boldsymbol{0} \\ 0 & \sigma_w^2 & \boldsymbol{0} & \boldsymbol{\mu}_1^T\sigma_w^2 \\ \boldsymbol{\mu}_1 & \boldsymbol{0} & \boldsymbol{\eta}_2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\mu}_1\sigma_w^2 & \boldsymbol{0} & \boldsymbol{\eta}_2\sigma_w^2 \end{bmatrix},$$

and furthermore,

$$\lim_{n\to\infty}\frac{(1-\rho_{y|x})}{n}\boldsymbol{U}_n = \begin{bmatrix}\boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D}\end{bmatrix} = \begin{bmatrix}\boldsymbol{A} & \boldsymbol{\mu}_1^T\boldsymbol{A} \\ \boldsymbol{\mu}_1\boldsymbol{A} & \boldsymbol{D}\end{bmatrix} = \begin{bmatrix} \overline{m}+\overline{q} & 0 & (\overline{m}+\overline{q})\boldsymbol{\mu}_1^T & \boldsymbol{0} \\ 0 & (\overline{m}+\overline{q})\sigma_w^2 & \boldsymbol{0} & (\overline{m}+\overline{q})\boldsymbol{\mu}_1^T\sigma_w^2 \\ (\overline{m}+\overline{q})\boldsymbol{\mu}_1 & \boldsymbol{0} & \overline{m}\boldsymbol{\mu}_2+\overline{q}\boldsymbol{\eta}_2 & \boldsymbol{0} \\ \boldsymbol{0} & (\overline{m}+\overline{q})\boldsymbol{\mu}_1\sigma_w^2 & \boldsymbol{0} & (\overline{m}\boldsymbol{\mu}_2+\overline{q}\boldsymbol{\eta}_2)\sigma_w^2 \end{bmatrix}$$

Denote the inverse of the above matrix as,

$$\begin{bmatrix}\boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D}\end{bmatrix}^{-1} = \begin{bmatrix}\boldsymbol{E} & \boldsymbol{F} \\ \boldsymbol{G} & \boldsymbol{H}\end{bmatrix}$$

By block matrix inversion and some algebra, we have,

$$\begin{aligned}
\boldsymbol{E} &= (\boldsymbol{A}-\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1} = \frac{1}{(\overline{m}+\overline{q})\left\{1-(\overline{m}+\overline{q})\boldsymbol{\mu}_1^T(\overline{m}\boldsymbol{\mu}_2+\overline{q}\boldsymbol{\eta}_2)^{-1}\boldsymbol{\mu}_1\right\}}\begin{bmatrix}1 & 0 \\ 0 & \sigma_w^{-2}\end{bmatrix} \\
\boldsymbol{H} &= (\boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} = (\boldsymbol{D}-(\boldsymbol{I}_2\otimes\boldsymbol{\mu}_1)\boldsymbol{B})^{-1} = \begin{bmatrix}1 & 0 \\ 0 & \sigma_w^{-2}\end{bmatrix}\otimes\left\{\overline{m}(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1\boldsymbol{\mu}_1^T)+\overline{q}(\boldsymbol{\eta}_2-\boldsymbol{\mu}_1\boldsymbol{\mu}_1^T)\right\}^{-1} \\
\boldsymbol{F} = \boldsymbol{G} &= -\boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} = -(\boldsymbol{I}_2\otimes\boldsymbol{\mu}_1^T)\boldsymbol{H} = -\begin{bmatrix}1 & 0 \\ 0 & \sigma_w^{-2}\end{bmatrix}\otimes\boldsymbol{\mu}_1^T\left\{\overline{m}(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1\boldsymbol{\mu}_1^T)+\overline{q}(\boldsymbol{\eta}_2-\boldsymbol{\mu}_1\boldsymbol{\mu}_1^T)\right\}^{-1}
\end{aligned}$$

Therefore, large-sample variance of $\sqrt{n}\hat{\boldsymbol{b}}$ is

$$\lim_{n\to\infty}\boldsymbol{\Sigma}_n = \sigma_{y|x}^2(\lim_{n\to\infty}n^{-1}\boldsymbol{U}_n)^{-1} = \sigma_{y|x}^2(1-\rho_{y|x})\begin{bmatrix}\boldsymbol{E} & \boldsymbol{F} \\ \boldsymbol{G} & \boldsymbol{H}\end{bmatrix}.$$

The variance-covariance for $\hat{\boldsymbol{b}}_4$ is

$$\begin{aligned} n\mathrm{Var}(\hat{\boldsymbol{b}}_4) =& \frac{\sigma^2_{y|x}(1-\rho_{y|x})}{\sigma^2_w} \left\{\overline{m}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1\boldsymbol{\mu}_1^T) + \overline{q}(\boldsymbol{\eta}_2 - \boldsymbol{\mu}_1\boldsymbol{\mu}_1^T)\right\}^{-1} \\ =& \frac{\sigma^2_{y|x}(1-\rho_{y|x})}{\sigma^2_w} \boldsymbol{\Lambda}_x^{-1/2}\left\{(\overline{p}+\overline{m})\boldsymbol{\Gamma}_x^1 + (\overline{q}-\overline{p})\boldsymbol{\Gamma}_x^0\right\}^{-1}\boldsymbol{\Lambda}_x^{-1/2} \\ =& \frac{\sigma^2_{y|x}(1-\rho_{y|x})}{\sigma^2_w} \boldsymbol{\Lambda}_x^{-1/2}\left\{\overline{m}\,\boldsymbol{\Gamma}_x^1 - \overline{m}\,\boldsymbol{\Gamma}_x^0 + \overline{p}\,\boldsymbol{\Gamma}_x^1 - \frac{1}{\rho_{y|x}}\overline{p}\,\boldsymbol{\Gamma}_x^0\right\}^{-1}\boldsymbol{\Lambda}_x^{-1/2}, \end{aligned}$$

where the last equation holds because

$$\overline{q} - \overline{p} = -\overline{m} - \left(\frac{1-\rho_{y|x}}{\rho_{y|x}}\right)\overline{p} - \overline{p} = -\overline{m} - \frac{1}{\rho_{y|x}}\overline{p}.$$

Further notice that

$$\begin{aligned} &\overline{m}\,\boldsymbol{\Gamma}_x^1 - \overline{m}\,\boldsymbol{\Gamma}_x^0 + \overline{p}\,\boldsymbol{\Gamma}_x^1 - \frac{1}{\rho_{y|x}}\overline{p}\,\boldsymbol{\Gamma}_x^0 \\ =& \overline{m}\left[\boldsymbol{\Gamma}_x^1 - \boldsymbol{\Gamma}_x^0 - \rho_{y|x}\left\{\frac{\theta_2^{-1}(\mathrm{CV})}{1+(\overline{m}-1)\rho_{y|x}}\right\}\boldsymbol{\Gamma}_x^1 + \left\{\frac{\theta_2^{-1}(\mathrm{CV})}{1+(\overline{m}-1)\rho_{y|x}}\right\}\boldsymbol{\Gamma}_x^0\right] \\ =& \frac{\overline{m}}{1+(\overline{m}-1)\rho_{y|x}}\left[\{1+(\overline{m}-1)\rho_{y|x}\}\boldsymbol{\Gamma}_x^1 - \{1+(\overline{m}-1)\rho_{y|x}\}\boldsymbol{\Gamma}_x^0 - \rho_{y|x}\theta_2^{-1}(\mathrm{CV})\boldsymbol{\Gamma}_x^1 + \theta_2^{-1}(\mathrm{CV})\boldsymbol{\Gamma}_x^0\right] \\ =& \frac{\overline{m}}{1+(\overline{m}-1)\rho_{y|x}}\left[\boldsymbol{\Gamma}_x^1 + (\overline{m}-2)\rho_{y|x}\boldsymbol{\Gamma}_x^1 - (\overline{m}-1)\rho_{y|x}\boldsymbol{\Gamma}_x^0 + (1-\theta_2^{-1}(\mathrm{CV}))\left(\rho_{y|x}\boldsymbol{\Gamma}_x^1 - \boldsymbol{\Gamma}_x^0\right)\right] \\ =& \frac{\overline{m}}{1+(\overline{m}-1)\rho_{y|x}}\left[\boldsymbol{\Gamma}_x^1 + (\overline{m}-2)\rho_{y|x}\boldsymbol{\Gamma}_x^1 - (\overline{m}-1)\rho_{y|x}\boldsymbol{\Gamma}_x^0 + \mathrm{CV}^2\frac{\overline{m}\rho_{y|x}(1-\rho_{y|x})(\rho_{y|x}\boldsymbol{\Gamma}_x^1 - \boldsymbol{\Gamma}_x^0)}{\{1+(\overline{m}-1)\rho_{y|x}\}^2}\right], \end{aligned}$$

where $\theta_2(\mathrm{CV})$ is defined in Section 2.2 of the main text. Because

$$\left[\boldsymbol{\Gamma}_x^1 + (\overline{m}-2)\rho_{y|x}\boldsymbol{\Gamma}_x^1 - (\overline{m}-1)\rho_{y|x}\boldsymbol{\Gamma}_x^0 + \mathrm{CV}^2\frac{\overline{m}\rho_{y|x}(1-\rho_{y|x})(\rho_{y|x}\boldsymbol{\Gamma}_x^1 - \boldsymbol{\Gamma}_x^0)}{\{1+(\overline{m}-1)\rho_{y|x}\}^2}\right] = \left\{\boldsymbol{\Gamma}_x^1 + (\overline{m}-2)\rho_{y|x}\boldsymbol{\Gamma}_x^1 - (\overline{m}-1)\rho_{y|x}\boldsymbol{\Gamma}_x^0\right\}\boldsymbol{\Theta}^{-1}(\mathrm{CV}),$$

we obtain expression (13) in Section 3.1 of the main text.

B.2 Variance expression for the adjusted average treatment effect estimator

To test for the average treatment effect, we focus on deriving the variance of $\hat{\beta}_2$ when the effect modifiers are mean centered (e.g., $\boldsymbol{\mu}_1 = 0$). Observe that

$$n\mathrm{Var}(\hat{\beta}_2) = \frac{\sigma^2_{y|x}(1-\rho_{y|x})}{\sigma^2_w} \times \frac{1}{(\overline{m}+\overline{q})\left\{1-(\overline{m}+\overline{q})\boldsymbol{\mu}_1^T(\overline{m}\boldsymbol{\mu}_2 + \overline{q}\boldsymbol{\eta}_2)^{-1}\boldsymbol{\mu}_1\right\}} = \frac{\sigma^2_{y|x}(1-\rho_{y|x})}{\sigma^2_w} \times \frac{1}{\overline{m}+\overline{q}}.$$

Because

$$\overline{m}+\overline{q} = \left(\frac{1-\rho_{y|x}}{\rho_{y|x}}\right)\rho_{y|x}\left\{\frac{\overline{m}}{1+(\overline{m}-1)\rho_{y|x}}\right\}\left[1-\mathrm{CV}^2\frac{\overline{m}\rho_{y|x}(1-\rho_{y|x})}{\{1+(\overline{m}-1)\rho_{y|x}\}^2}\right],$$

we have

$$n\mathrm{Var}(\hat{\beta}_2) = \frac{\sigma^2_{y|x}[1+(\overline{m}-1)\rho_{y|x}]}{\sigma^2_w\overline{m}} \times \theta_2(\mathrm{CV}),$$

C. VARIANCE EXPRESSION FOR HTE ANALYSIS WITH A UNIVARIATE EFFECT MODIFIER ALLOWING FOR BETWEEN- AND WITHIN-CLUSTER EFFECTS

The derivations in Web Appendix B can be applied to address the HTE analysis with a univariate effect modifier but allowing for between- and within-cluster effects (or sometimes called the contextual effects in CRTs). We provide the full derivation of such results to support the theoretical presentation in Section 3.2 of the main text. Recall that the model is considered as

$$Y_{ij} = \beta_1 + \beta_2 W_i + \beta_{31}^* \overline{X}_i + \beta_{32} X_{ij} + \beta_{41}^* \overline{X}_i W_i + \beta_{42} X_{ij} W_i + \lambda_i + \epsilon_{ij}, \quad \lambda_i \sim \mathcal{N}(0, \sigma_\lambda^2), \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2),$$

In what follows, we focus on expressing the concentration matrix, $\boldsymbol{\Omega}_4^{-1} = \left\{n\text{Var}(\hat{\boldsymbol{\beta}}_4)\right\}^{-1}$ with $\boldsymbol{\beta}_4 = (\beta_{41}^*, \beta_{42}^*)^T$. To apply the results in Web Appendix B, we need to compute the variance matrix $\boldsymbol{\Gamma}_x^1$ and $\boldsymbol{\Gamma}_x^0$ with $\boldsymbol{X}_{ij} = (\overline{X}_i, X_{ij})^T$. Clearly, we have $\text{Var}(X_{ij}) = \sigma_x^2$, and

$$\text{Var}(\overline{X}_i) = E\left\{\text{Var}(\overline{X}_i|m_i)\right\} + \text{Var}\left\{E(\overline{X}_i|m_i)\right\} = \sigma_x^2 E\left\{\frac{1 + (m_i - 1)\rho_x}{m_i}\right\},$$

where $E(\overline{X}_i|m_i) = 0$ as we assume the effect modifier is global mean centered. Using the approximation of van Breukelen et al.[1] in Appendix B, we can obtain $E(1/m_i) \approx (1/\overline{m})(1 + \text{CV}^2)$, and therefore

$$\text{Var}(\overline{X}_i) = \sigma_x^2\rho_x + \sigma_x^2(1-\rho_x)E\left(\frac{1}{m_i}\right) = \sigma_x^2\rho_x + (1-\rho_x)\frac{\sigma_x^2}{\overline{m}}(1+\text{CV}^2) \approx \sigma_x^2\left\{\frac{1 + (\overline{m}-1)\rho_x + \text{CV}^2(1-\rho_x)}{\overline{m}}\right\} = \sigma_x^2 r(\text{CV}).$$

This allows us to write

$$\boldsymbol{\Lambda}_x = \sigma_x^2 \begin{bmatrix} r(\text{CV}) & 0 \\ 0 & 1 \end{bmatrix}.$$

Similarly, we obtain

$$\begin{aligned} E(\overline{X}_i^2) =& E\left\{\frac{m_i\sigma_x^2 + m_i(m_i-1)\rho_x\sigma_x^2}{m_i^2}\right\} = \sigma_x^2 E\left\{\frac{1 + (m_i - 1)\rho_x}{m_i}\right\} \\ E(\overline{X}_i X_{ij}) =& E\left\{\frac{\sigma_x^2 + (m_i-1)\rho_x\sigma_x^2}{m_i}\right\} = \sigma_x^2 E\left\{\frac{1 + (m_i - 1)\rho_x}{m_i}\right\} \end{aligned}$$

and

$$E(\boldsymbol{X}_{ij}\boldsymbol{X}_{ij}^T) = E\begin{bmatrix} \overline{X}_i^2 & \overline{X}_i X_{ij} \\ X_{ij}\overline{X}_i & X_{ij}^2 \end{bmatrix} = \sigma_x^2 E\left\{\frac{1 + (m_i - 1)\rho_x}{m_i}\right\}\begin{bmatrix} 1 & 1 \\ 1 & \left[E\left\{\frac{1 + (m_i - 1)\rho_x}{m_i}\right\}\right]^{-1} \end{bmatrix}$$

Because the effect modifier is assumed to have zero marginal mean ($\mu_1 = 0$), by definition we have

$$\boldsymbol{\Gamma}_x^1 = \boldsymbol{\Lambda}_x^{-1/2}\left\{E(\boldsymbol{X}_{ij}\boldsymbol{X}_{ij}^T) - E(\boldsymbol{X}_{ij})E(\boldsymbol{X}_{ij})^T\right\}\boldsymbol{\Lambda}_x^{-1/2} = \begin{bmatrix} 1 & \left[E\left\{\frac{1 + (m_i - 1)\rho_x}{m_i}\right\}\right]^{1/2} \\ \left[E\left\{\frac{1 + (m_i - 1)\rho_x}{m_i}\right\}\right]^{1/2} & 1 \end{bmatrix}$$

Furthermore, notice $E(X_{ij}X_{ik}) = \sigma_x^2\rho_x$, and

$$E(\boldsymbol{X}_{ij}\boldsymbol{X}_{ik}^T) = E\begin{bmatrix} \overline{X}_i^2 & \overline{X}_i X_{ik} \\ X_{ij}\overline{X}_i & X_{ij}X_{ik} \end{bmatrix} = \sigma_x^2 E\left\{\frac{1 + (m_i - 1)\rho_x}{m_i}\right\}\begin{bmatrix} 1 & 1 \\ 1 & \rho_x\left[E\left\{\frac{1 + (m_i - 1)\rho_x}{m_i}\right\}\right]^{-1} \end{bmatrix},$$

which gives

$$\boldsymbol{\Gamma}_x^0 = \boldsymbol{\Lambda}_x^{-1/2}\left\{E(\boldsymbol{X}_{ij}\boldsymbol{X}_{ik}^T) - E(\boldsymbol{X}_{ij})E(\boldsymbol{X}_{ij})^T\right\}\boldsymbol{\Lambda}_x^{-1/2} = \begin{bmatrix} 1 & \left[E\left\{\frac{1 + (m_i - 1)\rho_x}{m_i}\right\}\right]^{1/2} \\ \left[E\left\{\frac{1 + (m_i - 1)\rho_x}{m_i}\right\}\right]^{1/2} & \rho_x \end{bmatrix}$$

Based on the approximation $E(1/m_i) \approx (1/\overline{m})(1+\text{CV}^2)$, we obtain

$$\boldsymbol{\Gamma}_x^1 + (\overline{m}-2)\rho_{y|x}\boldsymbol{\Gamma}_x^1 - (\overline{m}-1)\rho_{y|x}\boldsymbol{\Gamma}_x^0 = \begin{bmatrix} (1-\rho_{y|x}) & \sqrt{r(\text{CV})}(1-\rho_{y|x}) \\ \sqrt{r(\text{CV})}(1-\rho_{y|x}) & 1+(\overline{m}-2)\rho_{y|x}-(\overline{m}-1)\rho_x\rho_{y|x} \end{bmatrix},$$

and

$$\boldsymbol{\Gamma}_x^0 - \rho_{y|x}\boldsymbol{\Gamma}_x^1 = \begin{bmatrix} (1-\rho_{y|x}) & \sqrt{r(\text{CV})}(1-\rho_{y|x}) \\ \sqrt{r(\text{CV})}(1-\rho_{y|x}) & \rho_x-\rho_{y|x} \end{bmatrix},$$

which can be plugged in the expression in Web Appendix B and obtain an explicit expression of the concentration matrix of the maximum likelihood HTE estimator as

$$\boldsymbol{\Omega}_4^{-1} = \frac{\overline{m}\sigma_w^2\sigma_x^2}{\sigma_{y|x}^2(1-\rho_{y|x})\left\{1+(\overline{m}-1)\rho_{y|x}\right\}}\begin{bmatrix} r(\text{CV})(1-\rho_{y|x})\theta_2^{-1}(\text{CV}) & r(\text{CV})(1-\rho_{y|x})\theta_2^{-1}(\text{CV}) \\ r(\text{CV})(1-\rho_{y|x})\theta_2^{-1}(\text{CV}) & \left\{1+(\overline{m}-2)\rho_{y|x}-(\overline{m}-1)\rho_x\rho_{y|x}\right\}\theta_1^{-1}(\text{CV}) \end{bmatrix}.$$

To more intuitively understand the impact of unequal cluster sizes in the HTE analysis based on model with contextual effects, we define the relative change in the determinant of the covariance matrix (or equivalently, concentration matrix), $\det(\boldsymbol{\Omega}_4^{-1}) = 1/\det(\boldsymbol{\Omega}_4)$, under unequal versus equal cluster sizes when average cluster sizes. Mathematically, the relative change in determinant due to unequal cluster sizes is computed as

$$\frac{\det(\boldsymbol{\Omega}_4)}{\det(\boldsymbol{\Omega}_4)|_{\text{CV}=0}} = \frac{\left\{1+(\overline{m}-2)\rho_{y|x}-(\overline{m}-1)\rho_x\rho_{y|x}\right\} - r(0)(1-\rho_{y|x})}{\left\{1+(\overline{m}-2)\rho_{y|x}-(\overline{m}-1)\rho_x\rho_{y|x}\right\}\theta_1^{-1}(\text{CV})\theta_2^{-1}(\text{CV}) - r(\text{CV})(1-\rho_{y|x})\theta_2^{-2}(\text{CV})}.$$

To simplify this expression, notice that $r(0) = \{1+(\overline{m}-1)\rho_x\}/\overline{m}$, and set

$$t = \frac{1+(\overline{m}-2)\rho_{y|x}-(\overline{m}-1)\rho_x\rho_{y|x}}{1-\rho_{y|x}},$$

and the relative change in determinant is simplified as

$$\frac{\det(\boldsymbol{\Omega}_4)}{\det(\boldsymbol{\Omega}_4)|_{\text{CV}=0}} = \theta_2(\text{CV}) \times \frac{t-r(0)}{t\theta_1^{-1}(\text{CV}) - r(\text{CV})},$$

where

$$\theta_2^{-1}(\text{CV}) = 1 - \text{CV}^2\frac{\overline{m}\rho_{y|x}(1-\rho_{y|x})}{\{1+(\overline{m}-1)\rho_{y|x}\}^2}, \qquad \theta_1^{-1}(\text{CV}) = 1 - \text{CV}^2\frac{\overline{m}\rho_{y|x}(1-\rho_{y|x})(\rho_x-\rho_{y|x})}{\{1+(\overline{m}-2)\rho_{y|x}-(\overline{m}-1)\rho_x\rho_{y|x}\}\{1+(\overline{m}-1)\rho_{y|x}\}^2}.$$

Observe that

$$t - r(0) = (1-\rho_x) + \frac{(\overline{m}-1)\rho_{y|x}(1-\rho_x)}{1-\rho_{y|x}} - \frac{1-\rho_x}{\overline{m}} = \frac{(\overline{m}-1)(1-\rho_x)}{\overline{m}(1-\rho_{y|x})}\left\{1+(\overline{m}-1)\rho_{y|x}\right\}$$

After some algebra, the denominator can be written as

$$t\theta_1^{-1}(\text{CV}) - r = \left[t - \text{CV}^2\frac{\overline{m}\rho_{y|x}(1-\rho_{y|x})}{\{1+(\overline{m}-1)\rho_{y|x}\}^2}\right] - \left\{\frac{1+(\overline{m}-1)\rho_x+\text{CV}^2(1-\rho_x)}{\overline{m}}\right\}\left[1-\text{CV}^2\frac{\overline{m}\rho_{y|x}(1-\rho_{y|x})}{\{1+(\overline{m}-1)\rho_{y|x}\}^2}\right].$$

Multiplying the above expression by $\{1+(\overline{m}-1)\rho_{y|x}\}^2$,

$$\begin{aligned}
&\{1+(\overline{m}-1)\rho_{y|x}\}^2\{t\theta_1^{-1}(\text{CV}) - r\} \\
=&\{1+(\overline{m}-1)\rho_{y|x}\}^2 t - \text{CV}^2\overline{m}\rho_{y|x}(1-\rho_{y|x}) - \left\{\frac{1+(\overline{m}-1)\rho_x+\text{CV}^2(1-\rho_x)}{\overline{m}}\right\}\{1+(\overline{m}-1)\rho_{y|x}\}^2 \\
&+\left\{\frac{1+(\overline{m}-1)\rho_x+\text{CV}^2(1-\rho_x)}{\overline{m}}\right\}\text{CV}^2\overline{m}\rho_{y|x}(1-\rho_{y|x}) \\
=&\{1+(\overline{m}-1)\rho_{y|x}\}^2\{t-r(0)\} + \text{CV}^2\rho_{y|x}(1-\rho_x)\{1+(\overline{m}-1)\rho_{y|x}\} - \frac{\text{CV}^2(1-\rho_x)}{\overline{m}}\left[\{1+(\overline{m}-1)\rho_{y|x}\}^2 - \text{CV}^2\overline{m}\rho_{y|x}(1-\rho_{y|x})\right] \\
=&\{1+(\overline{m}-1)\rho_{y|x}\}^2\{t-r(0)\} - \frac{\text{CV}^2(1-\rho_x)(1-\rho_{y|x})}{\overline{m}}\left\{1+(\overline{m}-1)\rho_{y|x} - \text{CV}^2\overline{m}\rho_{y|x}\right\}.
\end{aligned}$$

Explicitly writing out the ratio between the numerator and denominator, we can obtain

$$
\begin{aligned}
&\frac{\det(\mathbf{\Omega}_4)}{\det(\mathbf{\Omega}_4)|_{\mathrm{CV}=0}} = \theta_2(\mathrm{CV}) \times \frac{t - r(0)}{t\theta_1^{-1}(\mathrm{CV}) - r(\mathrm{CV})} \\
=&\{1+(\overline{m}-1)\rho_{y|x}\}^3\theta_2(\mathrm{CV})\left[\{1+(\overline{m}-1)\rho_{y|x}\}^3 - \mathrm{CV}^2\frac{(1-\rho_{y|x})^2\{1+(\overline{m}-1)\rho_{y|x}\}}{\overline{m}-1}\left\{1-\mathrm{CV}^2\frac{\overline{m}\rho_{y|x}}{1+(\overline{m}-1)\rho_{y|x}}\right\}\right]^{-1} \\
=&\{1+(\overline{m}-1)\rho_{y|x}\}^2\theta_2(\mathrm{CV})\left[\{1+(\overline{m}-1)\rho_{y|x}\}^2 - \mathrm{CV}^2\frac{(1-\rho_{y|x})^2}{\overline{m}-1}\left\{1-\left(1-\theta_2^{-1}(\mathrm{CV})\right)\frac{\{1+(\overline{m}-1)\rho_{y|x}\}}{1-\rho_{y|x}}\right\}\right]^{-1} \\
=&\{1+(\overline{m}-1)\rho_{y|x}\}^2\theta_2(\mathrm{CV})\left[\{1+(\overline{m}-1)\rho_{y|x}\}^2 - \mathrm{CV}^2\frac{(1-\rho_{y|x})}{\overline{m}-1}\left\{(1-\rho_{y|x})-\left(1-\theta_2^{-1}(\mathrm{CV})\right)\{1+(\overline{m}-1)\rho_{y|x}\}\right\}\right]^{-1} \\
=&\{1+(\overline{m}-1)\rho_{y|x}\}\theta_2(\mathrm{CV})\left[\{1+(\overline{m}-1)\rho_{y|x}\} - \mathrm{CV}^2\frac{(1-\rho_{y|x})}{\overline{m}-1}\left\{\frac{(1-\rho_{y|x})}{\{1+(\overline{m}-1)\rho_{y|x}\}}-\left(1-\theta_2^{-1}(\mathrm{CV})\right)\right\}\right]^{-1} \\
=&\theta_2(\mathrm{CV})\left[1 - \mathrm{CV}^2\frac{(1-\rho_{y|x})}{(\overline{m}-1)\{1+(\overline{m}-1)\rho_{y|x}\}}\left\{\frac{(1-\rho_{y|x})}{\{1+(\overline{m}-1)\rho_{y|x}\}}-\left(1-\theta_2^{-1}(\mathrm{CV})\right)\right\}\right]^{-1} \\
=&\theta_2(\mathrm{CV})\left[1 - \mathrm{CV}^2\frac{(1-\rho_{y|x})}{(\overline{m}-1)\{1+(\overline{m}-1)\rho_{y|x}\}}\left\{\frac{(1-\rho_{y|x})}{\{1+(\overline{m}-1)\rho_{y|x}\}}-\mathrm{CV}^2\frac{\overline{m}\rho_{y|x}(1-\rho_{y|x})}{\{1+(\overline{m}-1)\rho_{y|x}\}^2}\right\}\right]^{-1} \\
=&\theta_2(\mathrm{CV})\left[1 - \mathrm{CV}^2\frac{(1-\rho_{y|x})^2}{(\overline{m}-1)\{1+(\overline{m}-1)\rho_{y|x}\}^2}\left\{1-\mathrm{CV}^2\frac{\overline{m}\rho_{y|x}}{\{1+(\overline{m}-1)\rho_{y|x}\}}\right\}\right]^{-1}.
\end{aligned}
$$

D. QUADRATIC HTE MODEL AND THE ASSOCIATED VARIANCE EXPRESSION OF THE HTE ESTIMATOR

The quadratic HTE model (with age as an individual-level effect modifier) used in the sensitivity analysis in Section 5 of the main manuscript can be written as

$$Y_{ij} = \beta_1 + \beta_2 W_i + \beta_{31}X_{ij} + \beta_{32}X_{ij}^2 + \beta_{41}X_{ij}W_i + \beta_{42}X_{ij}^2W_i + \lambda_i + \epsilon_{ij}, \quad \lambda_i \sim \mathcal{N}(0,\sigma_\lambda^2), \quad \epsilon_{ij} \sim \mathcal{N}(0,\sigma_\epsilon^2).$$

To apply the approach derived in Section 3.1 of the main manuscript, we need to obtain three matrices, $\mathbf{\Lambda}_x$, $\mathbf{\Gamma}_x^1$ and $\mathbf{\Gamma}_x^0$. Without further external information, we assume that the collection of X_{ij} in each cluster is multivariate normal with mean zero, marginal variance σ_x^2 and common ICC ρ_x. To derive $\mathbf{\Lambda}_x$, we obtain $\mathrm{Var}(X_{ij}^2) = E(X_{ij}^4) - E(X_{ij}^2)^2 = E(X^4) - \{\mathrm{Var}(X_{ij}) + E(X_{ij})^2\}^2 = E(X_{ij}^4) - \mathrm{Var}(X_{ij})^2 = 3\sigma_x^4 - \sigma_x^4 = 2\sigma_x^4$. We therefore have

$$\mathbf{\Lambda}_x = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & 2\sigma_x^4 \end{bmatrix}.$$

To obtain $\mathbf{\Gamma_x^1}$ which is the marginal correlation matrix between the 2 covariates X_{ij} and X_{ij}^2, we first compute

$$\mathrm{corr}(X_{ij}, X_{ij}^2) = \frac{\mathrm{cov}(X_{ij}, X_{ij}^2)}{\sqrt{\mathrm{Var}(X_{ij})\mathrm{Var}(X_{ij}^2)}} = \frac{E(X_{ij}^3) - E(X_{ij})E(X_{ij}^2)}{\sqrt{\mathrm{Var}(X_{ij})\mathrm{Var}(X_{ij}^2)}} = 0,$$

under the mean-zero normality assumption. It is then immediate that

$$\mathbf{\Gamma_x^1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \boldsymbol{I}_{2\times 2}.$$

Finally, to obtain $\mathbf{\Gamma}_x^0$ which is the multivariate counterpart of the covariate-ICC in the univariate case, we need compute the ICC for X^2 and the cross-covariate ICC between X and X^2. These quantities require us to obtain the cross-moments of X_{ij} and X_{ik} under a bivariate normal model. To proceed, we recall that the joint moment generating function (MGF) for the assumed bivariate normal distribution of (X_{ij}, X_{ik}) is

$$M(t_1, t_2) = \exp\left\{\frac{1}{2}\sigma_x^2(t_1^2 + t_2^2 + 2\rho_x t_1 t_2)\right\}.$$

After some algebra, we obtain

$$E(X_{ij}X_{ik}^2) = \frac{\partial M(t_1, t_2)}{\partial t_1 \partial t_2^2}\Big|_{t_1=0, t_2=0} = 0$$

$$E(X_{ij}^2X_{ik}^2) = \frac{\partial M(t_1, t_2)}{\partial t_1^2 \partial t_2^2}\Big|_{t_1=0, t_2=0} = 2\rho_x^2\sigma_x^4 + \sigma_x^4.$$

Therefore, the common ICC between X_{ij}^2 and X_{ik}^2 is

$$\text{ICC}(X^2) = \frac{E(X_{ij}^2X_{ik}^2) - E(X_{ij}^2)E(X_{ik}^2)}{\sqrt{\text{Var}(X_{ij}^2)\text{Var}(X_{ik}^2)}} = \frac{E(X_{ij}^2X_{ik}^2) - \sigma_x^4}{2\sigma_x^4} = \frac{2\rho_x^2\sigma_x^4 + \sigma_x^4 - \sigma_x^4}{2\sigma_x^4} = \rho_x^2,$$

and the common cross-covariate ICC between X_{ij} and X_{ik}^2 is

$$\text{ICC}(X, X^2) = \text{ICC}(X^2, X) = \frac{E(X_{ij}X_{ik}^2) - E(X_{ij})E(X_{ik}^2)}{\sqrt{\text{Var}(X_{ij})\text{Var}(X_{ik}^2)}} = \frac{E(X_{ij}X_{ik}^2)}{\sqrt{2}\sigma_x^3} = 0.$$

This gives

$$\mathbf{\Gamma}_x^0 = \begin{bmatrix} \rho_x & 0 \\ 0 & \rho_x^2 \end{bmatrix}.$$

Sample size requirements based on the quadratic HTE model is then given by

$$1 - \theta \leq \int_{\chi^2_{1-\alpha}(2)}^{\infty} \varphi\left(u; 2, n\boldsymbol{\delta}^T\mathbf{\Omega}_4^{-1}\boldsymbol{\delta}\right) du,$$

where $\boldsymbol{\delta}$ is the effect size corresponding to $(\beta_{41}, \beta_{42})^T$, $\varphi(u; 2, \zeta)$ is the density of the non-central χ^2-distribution with 2 degrees of freedom and non-centrality parameter ζ, and the variance expression is given by

$$\mathbf{\Omega}_4 = \frac{\sigma_{y|x}^2(1 - \rho_{y|x})\left\{1 + (\overline{m} - 1)\rho_{y|x}\right\}}{\overline{m}\sigma_w^2}\mathbf{\Lambda}_x^{-1/2}\mathbf{\Theta}(\text{CV})\left\{\boldsymbol{I}_{2\times2} + (\overline{m} - 2)\rho_{y|x}\boldsymbol{I}_{2\times2} - (\overline{m} - 1)\rho_{y|x}\mathbf{\Gamma}_x^0\right\}^{-1}\mathbf{\Lambda}_x^{-1/2},$$

and

$$\mathbf{\Theta}(\text{CV}) = \left[\boldsymbol{I}_{2\times2} - \text{CV}^2\frac{\overline{m}\rho_{y|x}(1 - \rho_{y|x})}{\{1 + (\overline{m} - 1)\rho_{y|x}\}^2}\left\{\boldsymbol{I}_{2\times2} + (\overline{m} - 2)\rho_{y|x}\boldsymbol{I}_{2\times2} - (\overline{m} - 1)\rho_{y|x}\mathbf{\Gamma}_x^0\right\}^{-1}(\mathbf{\Gamma}_x^0 - \rho_{y|x}\mathbf{\Gamma}_x^1)\right]^{-1}.$$

Some algebra shows that $\mathbf{\Omega}_4 = \text{diag}\{s_1, s_2\}$, where

$$s_1 = \frac{\sigma_y^2(1 - \rho_{y|x})\{1 + (\overline{m} - 1)\rho_{y|x}\}^3}{\sigma_w^2\sigma_x^2\overline{m}\left[\{1 + (\overline{m} - 2)\rho_{y|x} - (\overline{m} - 1)\rho_x\rho_{y|x}\}\{1 + (\overline{m} - 1)\rho_{y|x}\}^2 + \overline{m}\text{CV}^2\rho_{y|x}(1 - \rho_{y|x})(\rho_{y|x} - \rho_x)\right]}$$

$$s_2 = \frac{\sigma_y^2(1 - \rho_{y|x})\{1 + (\overline{m} - 1)\rho_{y|x}\}^3}{2\sigma_w^2\sigma_x^4\overline{m}\left[\{1 + (\overline{m} - 2)\rho_{y|x} - (\overline{m} - 1)\rho_x^2\rho_{y|x}\}\{1 + (\overline{m} - 1)\rho_{y|x}\}^2 + \overline{m}\text{CV}^2\rho_{y|x}(1 - \rho_{y|x})(\rho_{y|x} - \rho_x^2)\right]}.$$

Interestingly, s_1 is simply the variance expression we have derived in Web Appendix A based on the HTE model with only a linear X_{ij} term. On the other hand, s_2 has the same structure as s_1 but includes the appropriate marginal variance $(2\sigma_x^4)$ and covariate-ICC (ρ_x^2) for the squared covariate X_{ij}^2.

E. WEB TABLES FOR SIMULATION RESULTS ON TESTING HTE

WEB TABLE 1 Estimated required number of clusters (n) for HTE test based on the proposed formula, empirical type I error rate of the Wald test for HTE (ψ percent), predicted power (ϕ^* percent) and empirical power (ϕ percent) of the HTE test with a continuous individual-level effect modifier. The effect size for power is set to be $\beta_4 = \delta = 0.15$.

		CV = 0				CV = 0.3				CV = 0.6				CV = 0.9			
$\overline{m}$	$(\rho_x, \rho_{y\|x})$	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ
	(0.10, 0.01)	72	5.3	80.1	79.9	72	4.8	80.1	80.4	72	4.5	80.5	79.4	72	5.4	80.3	79.0
	(0.10, 0.05)	72	5.1	80.2	80.4	72	4.9	80.2	79.4	72	4.9	80.1	79.4	74	5.1	80.5	81.2
	(0.10, 0.10)	70	5.3	80.1	79.6	70	5.2	80.1	80.0	70	5.0	80.1	80.1	70	5.2	80.1	79.4
	(0.25, 0.01)	74	5.2	80.2	81.4	74	5.1	80.1	79.9	74	5.2	80.3	79.7	76	5.1	80.3	79.8
20	(0.25, 0.05)	78	5.3	80.1	80.4	78	5.1	80.0	79.3	80	4.7	80.2	81.3	80	5.2	80.2	80.6
	(0.25, 0.10)	80	5.2	80.3	80.6	80	5.1	80.3	80.9	80	5.6	80.1	80.5	80	5.2	80.4	79.3
	(0.50, 0.01)	76	5.4	80.1	79.0	78	4.6	80.4	79.9	78	5.5	80.2	78.5	80	5.2	80.1	79.3
	(0.50, 0.05)	92	5.1	80.1	80.5	92	5.0	80.3	80.2	94	4.9	80.3	78.7	98	5.2	80.0	80.5
	(0.50, 0.10)	100	4.7	80.2	80.2	100	5.4	80.1	80.3	102	4.9	80.4	81.0	104	5.0	80.3	80.6
	(0.10, 0.01)	30	5.1	80.3	81.2	30	5.2	80.3	80.5	30	5.8	80.1	80.0	30	4.8	81.2	80.1
	(0.10, 0.05)	30	4.8	80.0	80.7	30	5.0	80.0	80.8	30	4.8	81.3	80.0	30	5.6	81.2	79.9
	(0.10, 0.10)	28	4.6	80.1	80.4	28	5.1	80.1	79.7	28	5.3	80.1	79.1	28	4.9	80.1	78.4
	(0.25, 0.01)	32	5.0	80.9	80.6	32	5.2	80.7	81.4	32	5.5	80.3	79.3	32	5.0	80.8	79.6
50	(0.25, 0.05)	34	5.4	80.2	81.0	34	5.0	80.2	81.6	34	5.6	80.0	79.7	34	5.1	80.9	79.0
	(0.25, 0.10)	34	4.7	80.7	82.1	34	4.9	80.7	82.0	34	4.9	80.7	82.4	34	5.3	80.6	80.1
	(0.50, 0.01)	34	5.3	80.8	79.3	34	5.4	80.5	79.3	36	5.7	80.6	80.1	36	5.1	80.1	78.2
	(0.50, 0.05)	44	5.6	80.9	81.0	44	5.4	80.7	82.1	44	5.5	80.3	80.6	44	5.0	80.4	78.7
	(0.50, 0.10)	46	5.0	80.7	80.7	46	4.5	80.6	80.6	46	5.4	80.5	79.8	46	5.4	80.2	79.9
	(0.10, 0.01)	16	5.8	81.0	83.6	16	4.8	81.0	83.6	16	5.3	80.8	81.3	16	4.9	80.6	80.5
	(0.10, 0.05)	16	5.2	81.1	83.7	16	5.1	81.1	81.8	16	5.0	81.1	81.4	16	5.3	81.0	80.0
	(0.10, 0.10)	14	5.1	80.1	80.6	14	5.1	80.1	78.9	14	4.9	80.1	78.8	14	4.7	80.1	75.9
	(0.25, 0.01)	16	5.6	80.3	80.4	16	5.4	80.2	79.5	18	5.5	82.2	82.6	18	4.6	81.6	80.8
100	(0.25, 0.05)	18	4.9	80.2	82.4	18	5.3	80.2	82.4	18	5.0	80.1	80.7	18	5.5	80.0	79.3
	(0.25, 0.10)	18	4.2	81.3	83.4	18	5.2	81.3	83.0	18	5.0	81.3	81.9	18	4.9	81.2	80.3
	(0.50, 0.01)	20	5.4	81.0	81.4	20	5.6	80.7	80.7	20	4.9	81.8	79.9	20	5.8	80.3	78.3
	(0.50, 0.05)	24	5.3	81.6	80.8	24	6.0	81.5	81.1	24	5.7	81.4	80.2	24	5.2	81.0	78.0
	(0.50, 0.10)	24	5.8	81.0	80.0	24	5.7	81.0	80.7	24	5.2	80.9	79.7	24	4.5	80.8	78.9

WEB TABLE 2 Estimated required number of clusters (n)for HTE test based on the proposed formula, empirical type I error rate of the Wald test for HTE (ψ percent), predicted power (ϕ^* percent) and empirical power (ϕ percent) of the HTE test with a continuous individual-level effect modifier. The effect size for power is set to be $\beta_4 = \delta = 0.10$.

		CV = 0				CV = 0.3				CV = 0.6				CV = 0.9			
$\overline{m}$	$(\rho_x, \rho_{y\|x})$	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ
20	(0.10, 0.01)	160	5.4	80.2	79.6	160	5.1	80.1	79.8	160	5.0	80.0	80.6	162	5.0	80.1	79.8
	(0.10, 0.05)	162	4.9	80.2	80.8	162	5.1	80.2	80.3	162	5.3	80.1	79.0	164	5.0	80.2	79.3
	(0.10, 0.10)	158	5.2	80.0	80.3	158	5.3	80.0	79.2	158	5.4	80.0	78.9	158	4.4	80.0	80.5
	(0.25, 0.01)	164	4.9	80.2	79.9	164	4.7	80.1	79.5	166	5.0	80.2	80.6	168	4.8	80.2	80.0
	(0.25, 0.05)	176	6.1	80.0	80.1	176	5.1	80.1	79.9	178	4.8	80.0	79.6	180	4.8	80.2	80.6
	(0.25, 0.10)	178	5.5	80.2	80.8	178	5.4	80.1	80.0	178	4.9	80.2	79.9	180	5.1	80.2	79.6
	(0.50, 0.01)	172	5.0	80.1	79.6	172	5.5	80.1	79.5	176	5.4	80.1	79.4	180	5.3	80.1	79.4
	(0.50, 0.05)	206	4.8	80.2	80.9	206	4.8	80.1	80.1	212	5.3	80.2	80.7	220	5.0	80.2	80.5
	(0.50, 0.10)	222	5.6	80.1	80.8	224	5.4	80.1	80.6	226	5.1	80.1	80.5	232	5.0	80.2	80.0
50	(0.10, 0.01)	66	5.0	80.2	80.3	66	4.9	80.1	80.4	66	5.1	80.6	80.4	66	4.9	80.3	78.6
	(0.10, 0.05)	66	4.9	80.5	80.8	66	5.4	80.4	79.5	66	5.2	80.4	78.8	66	4.9	80.4	79.2
	(0.10, 0.10)	64	4.7	80.1	80.8	64	5.3	80.1	81.0	64	5.0	80.1	79.6	64	5.4	80.1	79.1
	(0.25, 0.01)	70	5.1	80.4	80.1	70	5.3	80.3	79.6	70	5.3	80.4	80.3	72	4.9	80.3	79.8
	(0.25, 0.05)	74	4.9	80.1	80.1	74	5.2	80.0	79.9	76	4.9	80.4	80.6	76	5.1	80.2	80.3
	(0.25, 0.10)	74	5.3	80.1	79.7	74	5.2	80.1	80.3	74	4.9	80.0	80.0	74	4.8	80.5	79.9
	(0.50, 0.01)	76	5.6	80.0	79.4	76	5.7	80.2	79.1	78	5.4	80.3	79.0	82	5.9	80.1	79.5
	(0.50, 0.05)	96	4.9	80.2	80.1	96	5.0	80.0	80.9	98	5.4	80.4	81.1	98	5.1	80.0	79.8
	(0.50, 0.10)	100	5.4	80.2	80.0	100	5.2	80.1	79.2	102	4.7	80.4	80.7	102	5.2	80.1	80.3
100	(0.10, 0.01)	34	5.1	80.1	81.6	34	5.4	80.1	80.4	34	5.6	81.1	80.6	34	5.0	80.9	78.7
	(0.10, 0.05)	34	4.8	80.2	81.4	34	5.4	80.2	80.7	34	5.3	80.2	81.3	34	5.7	80.2	78.9
	(0.10, 0.10)	32	5.4	80.7	80.1	32	4.9	80.7	81.3	32	5.0	80.7	80.5	32	4.5	80.7	78.1
	(0.25, 0.01)	36	5.1	80.3	80.9	36	4.9	80.2	79.5	38	5.0	80.9	80.0	38	5.0	80.3	80.6
	(0.25, 0.05)	40	4.1	80.9	82.3	40	5.0	80.9	81.2	40	5.2	80.9	81.6	40	5.2	80.8	81.1
	(0.25, 0.10)	38	4.5	80.0	80.4	38	5.6	81.0	81.1	38	5.3	81.0	78.9	38	4.3	81.0	78.7
	(0.50, 0.01)	42	5.7	80.3	78.5	42	5.7	80.0	80.0	44	6.3	80.1	80.4	46	5.5	80.3	80.6
	(0.50, 0.05)	52	5.4	80.1	80.1	52	5.2	80.1	79.5	54	5.4	80.6	81.6	54	5.7	80.3	79.4
	(0.50, 0.10)	54	5.0	80.3	80.3	54	4.9	80.2	81.1	54	5.3	80.2	80.1	54	5.0	80.1	80.5

WEB TABLE 3 Estimated required number of clusters (n) for HTE test based on the proposed formula, empirical type I error rate of the Wald test for HTE (ψ percent), predicted power (ϕ^* percent) and empirical power (ϕ percent) of the HTE test with a continuous individual-level effect modifier. The effect size for power is set to be $\beta_4 = \delta = 0.25$.

		CV = 0				CV = 0.3				CV = 0.6				CV = 0.9			
$\overline{m}$	$(\rho_x, \rho_{y\|x})$	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ
	(0.10, 0.01)	26	5.7	80.8	80.2	26	5.3	80.7	79.7	26	4.8	80.6	78.9	26	5.2	80.4	78.0
	(0.10, 0.05)	26	5.2	80.3	80.2	26	4.8	80.3	79.5	26	5.0	80.2	78.8	26	4.7	80.1	77.6
	(0.10, 0.10)	26	4.9	81.3	81.7	26	4.8	81.3	81.3	26	5.3	81.3	80.0	26	5.0	81.3	78.7
	(0.25, 0.01)	28	5.1	81.3	80.7	28	5.4	81.2	80.8	28	5.8	80.9	79.3	28	4.8	80.3	79.0
20	(0.25, 0.05)	28	5.7	80.0	79.2	30	5.2	81.3	81.5	30	5.1	81.0	80.1	30	4.9	80.4	79.5
	(0.25, 0.10)	30	4.8	81.1	82.4	30	5.4	81.0	81.9	30	5.5	80.9	80.4	30	5.2	80.7	79.5
	(0.50, 0.01)	28	5.5	81.0	78.2	28	4.5	80.8	77.8	28	5.2	80.1	76.2	30	5.0	80.3	76.9
	(0.50, 0.05)	34	5.8	80.4	79.1	34	5.5	80.1	79.4	34	6.0	80.5	78.7	36	5.4	80.1	79.0
	(0.50, 0.10)	36	5.4	80.6	80.2	36	4.7	80.5	78.8	38	5.1	81.0	80.8	38	5.4	80.2	78.6
	(0.10, 0.01)	12	5.2	82.3	85.2	12	5.1	82.3	83.7	12	5.2	82.1	81.8	12	4.9	81.9	79.6
	(0.10, 0.05)	12	5.2	82.0	85.0	12	5.7	82.0	82.7	12	5.6	82.0	82.0	12	5.8	81.9	78.9
	(0.10, 0.10)	12	5.2	83.4	85.7	12	5.3	83.4	84.4	12	5.2	83.4	84.4	12	5.1	83.4	81.0
	(0.25, 0.01)	12	5.0	80.3	82.3	12	5.1	80.2	79.3	12	5.8	83.1	79.5	12	5.4	82.4	76.1
50	(0.25, 0.05)	12	5.8	80.6	80.7	12	5.0	80.6	79.4	12	5.2	80.4	78.0	12	5.7	80.2	75.2
	(0.25, 0.10)	12	5.9	81.1	80.4	12	5.2	81.1	80.9	12	5.2	81.1	77.3	12	4.8	81.0	76.2
	(0.50, 0.01)	12	5.5	80.0	75.0	14	5.3	82.8	81.0	14	6.0	81.8	79.4	14	5.7	80.2	77.2
	(0.50, 0.05)	16	5.8	82.1	80.9	16	5.6	82.0	80.3	16	5.4	81.6	79.0	16	5.6	80.8	77.0
	(0.50, 0.10)	16	5.3	80.2	78.9	16	5.1	80.1	79.1	18	5.5	82.3	81.7	18	5.8	82.0	79.7
	(0.10, 0.01)	6	4.9	84.9	83.2	6	5.6	84.9	81.8	6	5.1	84.7	79.7	6	5.2	84.6	73.4
	(0.10, 0.05)	6	4.7	84.9	84.9	6	5.4	84.9	84.1	6	5.5	84.9	79.7	6	5.0	84.9	75.0
	(0.10, 0.10)	6	5.3	86.5	86.2	6	5.3	86.5	84.6	6	4.9	86.5	82.1	6	5.1	86.5	75.8
	(0.25, 0.01)	6	5.9	81.9	79.6	6	5.4	81.8	78.5	6	5.4	81.4	75.7	6	5.9	80.8	68.7
100	(0.25, 0.05)	8	5.8	85.2	89.0	8	5.0	85.2	87.8	8	5.8	85.1	86.8	8	5.7	85.0	82.1
	(0.25, 0.10)	6	5.6	80.5	79.8	6	5.3	80.5	79.0	6	5.4	80.5	75.9	6	5.5	80.5	71.3
	(0.50, 0.01)	8	5.9	81.9	81.9	8	6.3	81.6	80.3	8	6.4	80.7	78.5	8	6.2	84.3	73.9
	(0.50, 0.05)	10	5.7	83.1	85.1	10	6.0	83.1	85.1	10	6.1	82.9	82.2	10	6.0	82.6	81.0
	(0.50, 0.10)	10	5.1	82.5	85.5	10	5.2	82.5	86.1	10	5.6	82.5	83.1	10	5.4	82.4	80.8

WEB TABLE 4 Estimated required number of clusters (n) for HTE test based on the proposed formula, empirical type I error rate of the Wald test for HTE (ψ percent), predicted power (ϕ^* percent) and empirical power (ϕ percent) of the HTE test with a binary individual-level effect modifier. The effect size for power is set to be $\beta_4 = \delta = 0.25$.

		CV = 0				CV = 0.3				CV = 0.6				CV = 0.9			
$\overline{m}$	$(\rho_x, \rho_{y\|x})$	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ
	(0.10, 0.01)	122	5.4	80.2	79.7	122	5.1	80.2	79.8	122	4.9	80.1	79.8	124	5.2	80.2	80.0
	(0.10, 0.05)	124	5.5	80.1	79.4	124	5.2	80.1	80.0	124	5.0	80.3	80.0	124	5.3	80.2	78.7
	(0.10, 0.10)	120	5.4	80.1	79.9	120	4.8	80.1	80.1	120	5.1	80.1	78.6	120	4.9	80.1	80.1
	(0.25, 0.01)	126	5.1	80.2	79.3	126	4.9	80.1	80.2	126	4.5	80.1	80.0	128	5.5	80.2	79.5
20	(0.25, 0.05)	134	4.9	80.2	80.1	134	5.2	80.1	79.9	136	4.5	80.1	78.8	138	4.7	80.1	80.6
	(0.25, 0.10)	136	5.1	80.2	79.5	136	5.0	80.2	80.5	136	4.6	80.0	80.0	136	4.8	80.1	79.4
	(0.50, 0.01)	130	5.2	80.0	79.5	132	5.2	80.1	79.3	134	5.4	80.0	79.6	138	5.3	80.0	78.4
	(0.50, 0.05)	156	5.1	80.1	79.2	158	5.3	80.1	79.6	162	5.1	80.2	80.6	168	4.9	80.2	79.9
	(0.50, 0.10)	170	5.4	80.1	78.7	170	4.8	80.1	80.7	172	4.9	80.1	80.1	176	4.4	80.2	79.0
	(0.10, 0.01)	50	5.3	80.5	80.6	50	5.2	80.5	80.4	50	5.3	80.3	79.0	50	4.9	80.1	78.3
	(0.10, 0.05)	50	5.2	80.2	80.0	50	5.7	80.2	79.3	50	5.5	80.2	79.6	50	5.1	80.1	79.0
	(0.10, 0.10)	48	4.7	80.1	79.6	48	5.4	80.1	79.3	48	4.3	80.1	80.2	48	5.1	80.1	77.6
	(0.25, 0.01)	52	5.3	80.0	79.2	54	5.9	80.6	80.9	54	5.3	80.2	79.5	54	4.8	80.2	79.3
50	(0.25, 0.05)	58	4.6	80.5	80.0	58	5.0	80.5	80.6	58	4.9	80.3	79.5	58	4.9	80.1	78.7
	(0.25, 0.10)	56	5.2	80.3	80.0	56	5.1	80.3	79.9	56	5.4	80.3	79.3	56	4.8	80.2	79.4
	(0.50, 0.01)	58	5.3	80.6	79.5	58	4.5	80.3	79.0	60	5.3	80.6	79.6	62	5.1	80.3	78.8
	(0.50, 0.05)	74	4.5	80.5	80.5	74	5.0	80.3	80.8	74	5.0	80.4	79.5	76	5.0	80.2	80.3
	(0.50, 0.10)	76	5.0	80.1	80.5	76	5.6	80.0	78.7	78	4.9	80.4	79.6	78	5.1	80.1	79.2
	(0.10, 0.01)	26	4.8	81.4	81.1	26	5.6	81.4	81.7	26	5.8	81.3	79.6	26	5.0	81.1	77.6
	(0.10, 0.05)	26	4.8	81.5	81.3	26	4.6	81.5	81.0	26	5.1	81.5	79.9	26	5.2	81.5	78.7
	(0.10, 0.10)	24	5.5	80.1	79.9	24	5.3	80.1	79.1	24	5.3	80.1	78.1	24	5.3	80.1	77.5
	(0.25, 0.01)	28	5.8	81.1	80.1	28	5.5	81.0	79.9	28	5.1	80.6	78.6	28	4.9	80.0	77.9
100	(0.25, 0.05)	30	4.9	80.0	80.6	30	5.0	81.3	80.2	30	4.8	81.2	79.6	30	5.2	81.1	78.7
	(0.25, 0.10)	30	5.2	81.1	81.4	30	4.5	81.1	81.6	30	5.7	81.1	81.3	30	5.2	81.1	78.7
	(0.50, 0.01)	32	5.3	80.3	79.4	32	5.4	80.0	78.6	34	5.5	80.4	78.6	36	6.0	81.1	79.5
	(0.50, 0.05)	40	4.6	80.5	79.6	40	4.8	80.5	78.7	40	5.6	80.3	78.6	42	5.0	80.9	79.5
	(0.50, 0.10)	42	4.7	80.9	81.2	42	5.3	80.8	79.9	42	4.4	80.8	79.7	42	4.7	80.7	78.5

WEB TABLE 5 Estimated required number of clusters (n) for HTE test based on the proposed formula, empirical type I error rate of the Wald test for HTE (ψ percent), predicted power (ϕ^* percent) and empirical power (ϕ percent) of the HTE test with a binary individual-level effect modifier. The effect size for power is set to be $\beta_4 = \delta = 0.35$.

		CV = 0				CV = 0.3				CV = 0.6				CV = 0.9			
$\overline{m}$	$(\rho_x, \rho_{y\|x})$	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ
	(0.10, 0.01)	62	5.1	80.1	78.7	62	5.2	80.0	80.4	64	5.4	80.5	79.4	64	5.2	80.3	79.1
	(0.10, 0.05)	64	5.1	80.2	80.4	64	5.2	80.2	80.1	64	6.2	80.1	79.4	64	5.1	80.0	79.5
	(0.10, 0.10)	62	5.1	80.6	81.1	62	4.9	80.6	80.8	62	4.7	80.6	79.2	62	4.5	80.6	79.0
	(0.25, 0.01)	64	5.2	80.3	79.3	64	5.2	80.2	79.7	66	4.9	80.5	79.9	66	5.4	80.6	79.3
20	(0.25, 0.05)	68	4.3	80.0	79.5	70	5.2	80.5	79.5	70	5.2	80.1	80.5	70	4.8	80.2	78.1
	(0.25, 0.10)	70	5.1	80.3	80.2	70	5.1	80.2	80.3	70	4.9	80.1	80.2	70	4.9	80.4	78.8
	(0.50, 0.01)	68	5.1	80.4	79.5	68	4.7	80.2	79.0	68	5.8	80.1	77.9	70	5.4	80.1	78.7
	(0.50, 0.05)	80	5.3	80.3	79.1	80	5.5	80.1	78.9	82	5.0	80.2	79.3	86	5.0	80.1	80.5
	(0.50, 0.10)	88	5.0	80.4	80.3	88	4.8	80.2	79.8	88	4.8	80.2	79.6	90	5.4	80.2	79.9
	(0.10, 0.01)	26	5.5	81.3	81.4	26	5.2	81.2	81.1	26	5.4	81.1	78.6	26	4.9	80.8	78.1
	(0.10, 0.05)	26	4.8	81.0	80.6	26	4.8	81.0	80.1	26	4.8	80.9	79.2	26	5.0	80.9	78.7
	(0.10, 0.10)	26	4.1	80.9	83.3	26	5.1	80.9	82.9	26	4.7	80.9	80.9	26	5.5	80.9	80.4
	(0.25, 0.01)	28	5.0	80.7	81.1	28	5.2	80.6	80.2	28	5.4	80.1	80.2	28	5.0	80.8	78.2
50	(0.25, 0.05)	30	5.2	80.4	80.5	30	5.0	80.4	81.2	30	5.2	80.2	79.9	30	5.4	81.3	78.5
	(0.25, 0.10)	30	5.8	80.9	80.9	30	5.1	80.9	81.2	30	5.2	80.9	80.5	30	5.4	80.8	79.0
	(0.50, 0.01)	30	5.8	81.1	78.8	30	5.8	80.8	78.0	32	6.0	81.1	79.9	32	5.3	80.7	78.6
	(0.50, 0.05)	38	5.3	80.2	80.1	38	5.4	80.1	79.7	38	5.2	80.7	78.6	40	5.0	81.0	79.9
	(0.50, 0.10)	40	5.2	80.3	79.7	40	5.7	80.3	79.3	40	5.4	80.1	78.8	40	5.0	80.8	77.7
	(0.10, 0.01)	14	4.8	80.7	82.8	14	5.2	80.6	82.5	14	5.2	80.5	80.9	14	5.1	80.3	78.1
	(0.10, 0.05)	14	5.2	80.7	82.7	14	5.1	80.7	82.0	14	5.2	80.7	80.2	14	5.1	80.7	78.3
	(0.10, 0.10)	14	5.3	82.4	84.4	14	4.8	82.4	84.9	14	5.3	82.4	82.8	14	5.0	82.4	80.2
	(0.25, 0.01)	14	5.6	80.3	79.0	14	5.7	80.2	77.7	16	4.7	82.5	81.7	16	5.0	81.9	79.6
100	(0.25, 0.05)	16	4.8	80.5	81.7	16	5.2	80.5	81.9	16	5.2	80.5	79.9	16	5.4	80.4	77.9
	(0.25, 0.10)	16	5.0	81.6	82.5	16	5.3	81.6	82.5	16	5.3	81.6	80.0	16	5.0	81.6	78.6
	(0.50, 0.01)	18	6.0	81.9	80.9	18	5.9	81.6	80.0	18	6.2	80.7	77.6	18	5.8	81.4	74.8
	(0.50, 0.05)	22	5.4	81.6	81.2	22	6.1	81.6	79.9	22	4.6	81.4	79.5	22	5.0	81.1	76.5
	(0.50, 0.10)	22	5.2	81.0	80.8	22	5.0	81.0	81.3	22	5.8	80.9	78.0	22	4.9	80.8	76.9

WEB TABLE 6 Estimated required number of clusters (n) for HTE test based on the proposed formula, empirical type I error rate of the Wald test for HTE (ψ percent), predicted power (ϕ^* percent) and empirical power (ϕ percent) of the HTE test with a binary individual-level effect modifier. The effect size for power is set to be $\beta_4 = \delta = 0.45$.

		CV = 0				CV = 0.3				CV = 0.6				CV = 0.9			
$\overline{m}$	$(\rho_x, \rho_{y\|x})$	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ
20	(0.10, 0.01)	38	4.5	80.6	80.1	38	5.1	80.5	81.0	38	5.0	80.4	80.2	38	5.2	80.2	77.2
	(0.10, 0.05)	38	5.8	80.1	80.2	38	4.9	80.1	78.9	38	5.0	80.0	78.9	40	5.7	80.9	79.6
	(0.10, 0.10)	38	4.8	80.1	81.9	38	5.3	80.1	80.0	38	5.1	80.1	79.0	38	4.8	80.1	79.0
	(0.25, 0.01)	40	5.3	80.6	80.3	40	4.8	80.5	79.0	40	5.5	80.2	79.1	40	5.0	80.6	78.9
	(0.25, 0.05)	42	5.1	80.8	80.5	42	5.4	80.7	80.7	42	5.2	80.4	79.6	44	5.0	80.8	79.5
	(0.25, 0.10)	42	4.6	80.5	80.5	42	6.3	80.5	79.3	42	4.9	80.3	79.0	42	5.2	80.1	78.2
	(0.50, 0.01)	42	5.1	80.9	80.6	42	5.4	80.7	78.4	42	5.3	80.9	77.9	44	5.8	80.7	77.7
	(0.50, 0.05)	48	5.1	80.0	79.3	50	5.1	80.5	80.6	50	5.3	80.5	78.0	52	5.0	80.6	78.8
	(0.50, 0.10)	54	4.8	80.7	79.4	54	5.3	80.5	81.2	54	5.6	80.0	79.4	56	5.0	80.6	79.2
50	(0.10, 0.01)	16	5.2	81.9	80.4	16	4.8	81.9	81.3	16	4.7	81.7	79.1	16	5.3	81.5	77.0
	(0.10, 0.05)	16	5.9	81.6	80.5	16	4.8	81.6	79.2	16	4.9	81.6	80.0	16	5.3	81.5	77.0
	(0.10, 0.10)	16	5.6	80.6	82.1	16	4.6	80.6	82.4	16	5.5	80.6	81.4	16	5.1	80.6	78.7
	(0.25, 0.01)	18	5.0	82.2	82.7	18	5.0	82.1	82.4	18	4.7	81.7	80.7	18	5.7	81.0	77.4
	(0.25, 0.05)	18	5.9	81.4	80.1	18	5.3	81.3	79.5	18	5.3	81.2	78.9	18	5.6	81.0	75.8
	(0.25, 0.10)	18	5.2	81.9	80.1	18	5.2	81.9	79.5	18	5.8	81.8	79.5	18	5.4	81.8	76.4
	(0.50, 0.01)	18	5.5	80.8	76.8	18	5.1	80.5	77.7	20	6.2	81.6	78.7	20	6.1	80.0	76.0
	(0.50, 0.05)	24	4.9	81.3	80.1	24	5.1	81.1	79.5	24	5.6	80.7	77.2	24	5.5	81.6	77.2
	(0.50, 0.10)	24	5.0	81.0	79.0	24	5.4	80.9	77.6	24	5.1	80.8	77.7	24	5.3	80.5	74.0
100	(0.10, 0.01)	8	5.6	81.3	79.5	8	5.6	81.3	78.9	8	5.5	81.2	75.8	8	5.5	80.9	72.7
	(0.10, 0.05)	8	5.4	81.4	80.6	8	5.1	81.4	78.5	8	5.4	81.4	77.1	8	5.3	81.3	73.6
	(0.10, 0.10)	8	5.2	83.1	82.8	8	4.7	83.1	81.3	8	5.5	83.1	78.0	8	5.1	83.1	75.0
	(0.25, 0.01)	10	5.2	82.7	84.2	10	5.7	82.5	82.9	10	5.7	82.2	80.2	10	5.4	81.6	76.5
	(0.25, 0.05)	10	5.3	80.2	81.4	10	5.5	80.2	81.3	10	5.3	80.1	79.5	10	5.4	80.0	74.8
	(0.25, 0.10)	10	5.6	81.3	82.8	10	4.8	81.3	82.4	10	5.3	81.3	79.7	10	4.7	81.3	75.6
	(0.50, 0.01)	10	5.8	80.8	74.6	10	6.2	80.5	71.8	12	6.8	83.3	78.7	12	5.7	81.8	74.2
	(0.50, 0.05)	14	5.4	82.5	81.5	14	5.6	82.4	80.7	14	5.2	82.2	79.4	14	5.9	81.9	75.0
	(0.50, 0.10)	14	5.9	81.9	82.1	14	5.5	81.9	81.5	14	5.5	81.8	77.5	14	5.5	81.7	74.3

F. WEB TABLES FOR SIMULATION RESULTS ON TESTING AVERAGE TREATMENT EFFECT (ADJUSTING FOR THE COVARIATE)

WEB TABLE 7 Estimated required number of clusters (n) for the covariate-adjusted average treatment effect test based on the proposed formula, empirical type I error rate of the Wald test for average treatment effect (ψ percent), predicted power (ϕ^* percent) and empirical power (ϕ percent) of the adjusted average treatment effect test with a continuous individual-level effect modifier. The effect size for power of the average treatment effect is induced from the HTE model and obtained as 0.325.

		CV = 0				CV = 0.3				CV = 0.6				CV = 0.9			
$\overline{m}$	$(\rho_x, \rho_{y\|x})$	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ
20	(0.10, 0.01)	20	4.0	80.4	78.2	22	4.3	84.0	82.9	22	4.0	82.6	79.6	24	4.5	83.6	81.2
	(0.10, 0.05)	32	4.9	81.3	81.1	32	5.1	80.4	81.0	34	4.8	80.1	80.7	40	5.7	81.7	83.3
	(0.10, 0.10)	46	5.1	80.8	79.4	46	5.3	80.0	79.1	50	5.7	81.1	80.8	56	5.0	81.4	81.0
	(0.25, 0.01)	20	3.5	80.4	78.5	22	3.5	84.0	81.4	22	4.4	82.6	80.0	24	4.6	83.6	80.0
	(0.25, 0.05)	32	4.9	81.3	80.9	32	4.8	80.4	80.4	34	6.0	80.1	79.5	40	6.0	81.7	82.5
	(0.25, 0.10)	46	5.4	80.8	80.6	46	5.1	80.0	79.9	50	5.0	81.1	81.4	56	5.2	81.4	80.9
	(0.50, 0.01)	20	4.1	80.4	77.4	22	4.1	84.0	81.1	22	4.5	82.6	78.1	24	5.0	83.6	79.6
	(0.50, 0.05)	32	5.1	81.3	79.8	32	5.6	80.4	79.7	34	5.6	80.1	78.9	40	5.3	81.7	81.8
	(0.50, 0.10)	46	5.5	80.8	79.6	46	4.9	80.0	79.3	50	5.4	81.1	81.4	56	4.9	81.4	81.0
50	(0.10, 0.01)	12	3.4	83.7	81.9	12	3.0	83.0	80.4	12	4.3	80.5	79.6	14	4.7	83.4	81.8
	(0.10, 0.05)	24	5.5	82.5	82.8	24	4.9	81.9	82.1	26	5.4	83.0	83.1	28	5.3	82.3	82.0
	(0.10, 0.10)	38	5.2	81.0	81.0	38	5.3	80.5	80.6	40	5.5	81.2	80.8	42	4.7	80.8	79.3
	(0.25, 0.01)	12	3.7	83.7	81.9	12	3.9	83.0	79.5	12	3.8	80.5	77.1	14	5.1	83.4	80.9
	(0.25, 0.05)	24	5.0	82.5	82.6	24	4.9	81.9	81.3	26	5.6	83.0	81.6	28	5.4	82.3	80.8
	(0.25, 0.10)	38	5.5	81.0	80.8	38	5.5	80.5	80.6	40	4.7	81.2	81.2	42	5.2	80.8	79.2
	(0.50, 0.01)	12	4.0	83.7	79.8	12	3.6	83.0	78.3	12	3.9	80.5	75.9	14	4.1	83.4	79.3
	(0.50, 0.05)	24	5.5	82.5	81.0	24	5.2	81.9	80.9	26	5.8	83.0	81.3	28	5.5	82.3	80.4
	(0.50, 0.10)	38	4.9	81.0	80.9	38	4.9	80.5	80.6	40	4.7	81.2	81.2	42	5.1	80.8	79.6
100	(0.10, 0.01)	10	3.7	89.1	88.7	10	4.1	88.4	86.5	10	4.8	86.2	84.8	10	5.2	81.4	79.8
	(0.10, 0.05)	20	5.1	80.4	80.3	22	5.1	84.0	83.9	22	5.6	82.6	82.7	22	5.4	80.1	79.2
	(0.10, 0.10)	36	4.9	81.8	82.5	36	5.0	81.6	80.6	36	5.3	80.8	80.6	38	5.1	81.6	80.9
	(0.25, 0.01)	10	4.2	89.1	88.4	10	4.7	88.4	87.1	10	4.8	86.2	82.9	10	5.2	81.4	79.0
	(0.25, 0.05)	20	5.5	80.4	79.9	22	4.8	84.0	84.0	22	5.3	82.6	81.7	22	4.9	80.1	77.9
	(0.25, 0.10)	36	5.0	81.8	81.8	36	5.1	81.6	81.5	36	5.1	80.8	80.3	38	5.2	81.6	80.6
	(0.50, 0.01)	10	3.9	89.1	86.8	10	4.0	88.4	85.9	10	4.4	86.2	82.3	10	5.1	81.4	77.1
	(0.50, 0.05)	20	5.1	80.4	79.0	22	5.1	84.0	82.9	22	5.3	82.6	81.1	22	4.9	80.1	78.4
	(0.50, 0.10)	36	4.5	81.8	81.3	36	5.3	81.6	81.0	36	5.0	80.8	79.3	38	5.2	81.6	80.2

WEB TABLE 8 Estimated required number of clusters (n) for covariate-adjusted average treatment effect test based on the proposed formula, empirical type I error rate of the Wald test for average treatment effect (ψ percent), predicted power (ϕ^* percent) and empirical power (ϕ percent) of the adjusted average treatment effect test with a continuous individual-level effect modifier. The effect size for power of the average treatment effect is induced from the HTE model and obtained as 0.30.

		CV = 0				CV = 0.3				CV = 0.6				CV = 0.9			
$\overline{m}$	$(\rho_x, \rho_{y\|x})$	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ
20	(0.10, 0.01)	24	4.2	82.1	81.2	24	3.9	81.6	79.0	24	4.0	80.1	77.9	26	4.2	80.9	78.6
	(0.10, 0.05)	38	5.3	82.2	82.4	38	5.1	81.3	80.8	40	5.1	80.6	80.5	46	5.5	81.2	83.7
	(0.10, 0.10)	54	5.1	81.1	80.6	54	5.0	80.3	80.4	58	5.0	80.8	80.3	64	5.1	80.5	80.2
	(0.25, 0.01)	24	3.7	82.1	80.5	24	3.5	81.6	79.9	24	4.6	80.1	78.5	26	4.2	80.9	79.5
	(0.25, 0.05)	38	5.2	82.2	81.7	38	5.3	81.3	81.0	40	5.2	80.6	81.2	46	5.5	81.2	81.9
	(0.25, 0.10)	54	5.5	81.1	79.1	54	5.2	80.3	81.0	58	5.3	80.8	80.6	64	5.2	80.5	81.0
	(0.50, 0.01)	24	4.1	82.1	79.7	24	3.9	81.6	79.0	24	3.9	80.1	77.4	26	4.6	80.9	78.4
	(0.50, 0.05)	38	5.1	82.2	80.2	38	5.0	81.3	80.7	40	6.0	80.6	79.5	46	4.8	81.2	81.8
	(0.50, 0.10)	54	5.1	81.1	80.0	54	4.6	80.3	79.5	58	4.6	80.8	80.7	64	5.5	80.5	80.4
50	(0.10, 0.01)	14	4.0	84.8	83.8	14	4.1	84.0	83.0	14	4.5	81.7	79.6	16	5.2	83.3	81.4
	(0.10, 0.05)	28	4.7	82.9	82.0	28	5.0	82.2	82.4	30	4.9	82.9	82.4	32	5.8	81.6	82.0
	(0.10, 0.10)	44	5.1	80.8	80.5	44	5.4	80.3	79.3	46	5.1	80.7	80.6	48	4.8	80.0	79.1
	(0.25, 0.01)	14	3.6	84.8	83.5	14	3.7	84.0	82.0	14	4.8	81.7	79.5	16	4.8	83.3	82.1
	(0.25, 0.05)	28	5.4	82.9	82.4	28	4.9	82.2	81.8	30	5.6	82.9	81.8	32	5.3	81.6	80.7
	(0.25, 0.10)	44	4.7	80.8	80.8	44	4.9	80.3	80.5	46	4.8	80.7	80.1	48	5.2	80.0	79.6
	(0.50, 0.01)	14	3.9	84.8	82.3	14	3.8	84.0	81.2	14	3.9	81.7	77.9	16	4.9	83.3	81.0
	(0.50, 0.05)	28	4.6	82.9	82.4	28	5.3	82.2	82.3	30	5.0	82.9	83.4	32	5.3	81.6	81.3
	(0.50, 0.10)	44	5.3	80.8	81.4	44	5.5	80.3	79.6	46	5.2	80.7	80.6	48	5.0	80.0	78.9
100	(0.10, 0.01)	10	4.0	83.9	82.9	10	3.8	83.1	82.5	10	4.2	80.3	79.3	12	6.1	84.3	82.8
	(0.10, 0.05)	24	5.0	82.1	82.7	24	5.0	81.6	81.0	24	4.9	80.2	80.5	26	4.8	81.1	79.9
	(0.10, 0.10)	42	5.0	81.9	82.0	42	5.2	81.7	82.0	42	4.9	80.9	80.0	44	5.1	81.4	80.6
	(0.25, 0.01)	10	4.2	83.9	82.1	10	4.5	83.1	81.2	10	4.8	80.3	78.2	12	5.4	84.3	83.2
	(0.25, 0.05)	24	4.9	82.1	81.9	24	5.5	81.6	81.4	24	5.5	80.2	79.9	26	5.6	81.1	80.1
	(0.25, 0.10)	42	5.4	81.9	82.3	42	4.8	81.7	81.4	42	4.6	80.9	81.3	44	4.8	81.4	79.7
	(0.50, 0.01)	10	4.2	83.9	81.8	10	4.1	83.1	80.3	10	4.9	80.3	76.6	12	5.1	84.3	81.2
	(0.50, 0.05)	24	5.3	82.1	81.1	24	5.1	81.6	82.0	24	4.9	80.2	79.6	26	5.8	81.1	78.8
	(0.50, 0.10)	42	5.3	81.9	81.6	42	4.9	81.7	81.5	42	4.6	80.9	81.0	44	4.5	81.4	80.1

WEB TABLE 9 Estimated required number of clusters (n) for covariate-adjusted average treatment effect test based on the proposed formula, empirical type I error rate of the Wald test for average treatment effect (ψ percent), predicted power (ϕ^* percent) and empirical power (ϕ percent) of the adjusted average treatment effect test with a continuous individual-level effect modifier. The effect size for power of the average treatment effect is induced from the HTE model and obtained as 0.375.

		CV = 0				CV = 0.3				CV = 0.6				CV = 0.9			
$\overline{m}$	$(\rho_x, \rho_{y\|x})$	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ
	(0.10,0.01)	16	3.4	81.6	79.7	16	3.5	81.1	78.3	18	3.4	84.7	82.7	18	4.3	82.2	77.9
	(0.10,0.05)	24	5.0	80.3	80.0	26	5.5	82.8	82.3	26	5.2	80.0	80.3	30	5.3	80.9	81.4
	(0.10,0.10)	36	5.0	81.9	81.5	36	4.9	81.1	81.2	38	4.5	81.0	79.9	42	5.5	80.8	81.1
	(0.25,0.01)	16	3.1	81.6	78.9	16	3.2	81.1	78.8	18	3.8	84.7	81.6	18	4.3	82.2	76.9
20	(0.25,0.05)	24	5.1	80.3	79.1	26	5.1	82.8	82.5	26	5.8	80.0	78.9	30	6.0	80.9	82.0
	(0.25,0.10)	36	5.1	81.9	81.6	36	4.6	81.1	80.5	38	5.3	81.0	80.2	42	4.7	80.8	81.3
	(0.50,0.01)	16	3.3	81.6	78.0	16	3.1	81.1	76.3	18	3.8	84.7	80.0	18	4.0	82.2	75.3
	(0.50,0.05)	24	4.5	80.3	79.2	26	4.7	82.8	81.8	26	5.1	80.0	78.4	30	5.5	80.9	80.6
	(0.50,0.10)	36	5.1	81.9	81.4	36	5.1	81.1	79.2	38	5.1	81.0	81.2	42	5.0	80.8	79.7
	(0.10,0.01)	10	3.1	85.4	84.4	10	2.9	84.7	83.6	10	3.4	82.4	79.3	12	4.4	86.7	83.8
	(0.10,0.05)	18	5.1	81.1	81.1	18	5.2	80.4	79.2	20	4.9	82.9	82.5	22	5.4	83.1	81.6
	(0.10,0.10)	30	5.2	82.3	81.6	30	5.4	81.8	81.6	30	5.1	80.4	80.3	32	4.9	80.7	79.2
	(0.25,0.01)	10	3.1	85.4	82.0	10	2.8	84.7	81.7	10	3.2	82.4	77.4	12	4.6	86.7	82.3
50	(0.25,0.05)	18	4.7	81.1	79.7	18	4.8	80.4	79.5	20	5.6	82.9	80.9	22	5.8	83.1	81.6
	(0.25,0.10)	30	5.0	82.3	82.1	30	5.2	81.8	81.3	30	4.8	80.4	79.9	32	5.3	80.7	79.4
	(0.50,0.01)	10	3.0	85.4	80.4	10	3.6	84.7	78.4	10	3.6	82.4	76.2	12	4.5	86.7	80.3
	(0.50,0.05)	18	5.2	81.1	79.6	18	4.9	80.4	79.4	20	5.6	82.9	81.0	22	5.1	83.1	81.6
	(0.50,0.10)	30	4.4	82.3	82.1	30	5.1	81.8	80.3	30	4.8	80.4	78.1	32	4.9	80.7	78.5
	(0.10,0.01)	8	3.4	88.1	86.4	8	3.5	87.4	86.0	8	4.0	85.1	82.5	8	3.9	80.1	77.2
	(0.10,0.05)	16	5.6	81.6	80.7	16	5.1	81.1	80.8	18	5.3	84.8	84.1	18	4.8	82.4	80.4
	(0.10,0.10)	28	5.2	82.4	82.4	28	5.2	82.2	82.2	28	4.7	81.4	81.7	28	5.0	80.0	79.4
	(0.25,0.01)	8	3.1	88.1	86.1	8	3.7	87.4	84.2	8	3.4	85.1	81.5	8	4.4	80.1	75.0
100	(0.25,0.05)	16	4.8	81.6	80.2	16	4.6	81.1	79.5	18	4.9	84.8	83.8	18	5.7	82.4	80.3
	(0.25,0.10)	28	5.2	82.4	81.3	28	5.3	82.2	81.6	28	5.5	81.4	81.0	28	4.9	80.0	78.5
	(0.50,0.01)	8	3.5	88.1	82.1	8	3.3	87.4	81.9	8	3.5	85.1	78.4	8	4.0	80.1	72.9
	(0.50,0.05)	16	4.6	81.6	80.5	16	5.1	81.1	80.2	18	4.7	84.8	82.8	18	5.4	82.4	78.6
	(0.50,0.10)	28	4.8	82.4	81.6	28	4.8	82.2	81.8	28	5.4	81.4	79.8	28	5.3	80.0	76.8

G. WEB TABLES FOR SIMULATION RESULTS ON TESTING AVERAGE TREATMENT EFFECT (WITHOUT ADJUSTING FOR THE COVARIATE)

WEB TABLE 10 Estimated required number of clusters (n) for the unadjusted average treatment effect test, empirical type I error rate of the Wald test for the average treatment effect (ψ percent), predicted power (ϕ^* percent) and empirical power (ϕ percent) of the unadjusted average treatment effect test, when the outcomes are simulated with a continuous individual-level effect modifier. The effect size for power of the average treatment effect is induced from the HTE model and obtained as 0.325.

		CV = 0				CV = 0.3				CV = 0.6				CV = 0.9			
$\overline{m}$	$(\rho_x, \rho_{y\|x})$	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ
	(0.10, 0.01)	22	4.0	81.3	80.9	22	3.5	80.7	79.8	24	3.7	82.6	82.0	26	4.8	82.8	81.5
	(0.10, 0.05)	34	4.9	81.8	81.2	34	5.2	80.9	81.4	36	5.3	80.5	80.6	42	5.4	81.7	81.8
	(0.10, 0.10)	48	5.0	81.1	80.4	48	5.2	80.4	80.9	52	4.9	81.3	81.0	58	5.5	81.4	81.5
	(0.25, 0.01)	24	4.1	82.0	81.1	24	4.2	81.4	80.0	26	4.7	82.6	82.3	28	5.0	81.8	81.2
20	(0.25, 0.05)	36	4.8	82.2	83.0	36	5.4	81.3	81.4	38	5.3	80.8	80.1	44	5.4	81.7	82.8
	(0.25, 0.10)	50	5.0	81.5	80.8	50	5.0	80.7	80.4	52	4.2	80.0	80.1	58	5.2	80.3	81.4
	(0.50, 0.01)	26	4.1	81.0	80.1	26	3.9	80.1	79.9	28	4.5	80.7	79.6	32	4.8	81.7	82.5
	(0.50, 0.05)	38	5.3	81.5	81.7	38	5.3	80.6	81.3	40	4.7	80.0	80.7	46	4.7	80.8	82.1
	(0.50, 0.10)	52	4.6	81.0	81.5	52	5.1	80.3	80.7	56	4.7	81.0	81.3	62	4.8	81.1	80.2
	(0.10, 0.01)	14	4.3	85.7	85.3	14	4.3	84.9	84.1	14	4.3	82.4	82.6	16	4.4	83.6	84.4
	(0.10, 0.05)	24	4.9	80.2	79.3	26	4.7	82.9	83.0	26	4.8	80.8	80.6	28	5.3	80.1	79.9
	(0.10, 0.10)	40	4.6	81.7	81.2	40	4.6	81.2	81.3	42	4.4	81.8	81.8	44	5.6	81.3	79.2
	(0.25, 0.01)	14	3.7	80.3	80.8	16	3.6	85.3	84.8	16	4.3	82.8	82.0	18	5.3	83.0	83.0
50	(0.25, 0.05)	26	5.3	80.9	80.2	26	4.9	80.3	80.6	28	4.9	81.3	81.2	30	5.5	80.6	81.6
	(0.25, 0.10)	40	5.1	80.0	79.3	42	4.6	81.6	83.3	42	5.0	80.3	81.1	46	5.0	81.6	80.4
	(0.50, 0.01)	18	4.8	83.8	83.7	18	4.5	83.0	82.9	18	4.6	80.3	80.9	20	4.7	80.0	80.8
	(0.50, 0.05)	30	5.0	82.8	82.1	30	5.1	82.2	81.5	30	5.0	80.3	79.9	34	5.0	82.1	82.5
	(0.50, 0.10)	44	4.6	81.3	80.8	44	4.7	80.9	81.5	46	4.9	81.5	81.5	48	5.1	81.0	80.1
	(0.10, 0.01)	10	4.2	83.2	83.0	10	4.3	82.3	80.4	12	4.4	88.1	88.2	12	4.9	83.8	83.8
	(0.10, 0.05)	22	5.6	82.1	81.5	22	5.0	81.6	81.9	22	5.1	80.2	78.8	24	5.3	81.5	80.2
	(0.10, 0.10)	36	5.4	80.4	79.7	36	4.9	80.2	79.4	38	5.4	81.6	81.6	38	4.9	80.2	78.9
	(0.25, 0.01)	12	4.2	84.4	83.6	12	4.5	83.6	83.4	12	5.4	81.2	80.3	14	5.2	83.9	83.7
100	(0.25, 0.05)	24	5.0	82.6	82.7	24	4.8	82.2	82.6	24	4.5	80.9	79.5	26	5.4	82.0	80.9
	(0.25, 0.10)	38	5.0	80.9	81.7	38	4.7	80.6	81.4	40	4.9	81.9	82.2	40	5.5	80.6	79.1
	(0.50, 0.01)	14	4.3	81.5	81.2	14	4.8	80.8	81.8	16	6.0	84.7	84.1	16	5.5	81.0	80.8
	(0.50, 0.05)	26	5.3	81.3	80.6	26	4.7	80.9	80.6	28	5.3	82.9	82.6	28	4.6	80.9	79.4
	(0.50, 0.10)	40	4.8	80.1	80.2	42	5.3	81.9	81.3	42	5.0	81.2	80.9	44	5.4	81.9	81.1

WEB TABLE 11 Estimated required number of clusters (n) for unadjusted average treatment effect test, empirical type I error rate of the Wald test for average treatment effect (ψ percent), predicted power (ϕ^* percent) and empirical power (ϕ percent) of the unadjusted average treatment effect test, when the outcomes are simulated with a continuous individual-level effect modifier. The effect size for power of the average treatment effect is induced from the HTE model and obtained as 0.300.

		CV = 0				CV = 0.3				CV = 0.6				CV = 0.9			
$\overline{m}$	$(\rho_x, \rho_{y\|x})$	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ
	(0.10, 0.01)	26	3.7	83.2	82.7	26	4.2	82.6	80.7	26	4.5	80.9	79.2	28	4.3	81.0	79.5
	(0.10, 0.05)	38	4.3	80.8	80.3	40	5.1	82.0	82.0	42	4.9	81.2	82.0	48	5.8	81.5	82.5
	(0.10, 0.10)	54	4.8	80.1	80.5	56	5.1	80.8	81.6	60	4.7	81.2	80.4	66	5.2	80.8	80.7
	(0.25, 0.01)	26	4.3	81.1	80.8	26	4.6	80.4	79.5	28	4.6	81.6	81.8	30	4.7	80.8	80.8
20	(0.25, 0.05)	40	5.3	81.5	80.9	40	5.0	80.6	79.4	44	5.0	81.8	81.7	48	5.4	80.2	82.0
	(0.25, 0.10)	56	5.3	80.7	80.2	58	4.9	81.3	80.4	60	5.0	80.4	80.4	68	5.1	81.2	81.0
	(0.50, 0.01)	28	4.2	80.9	80.0	28	3.9	80.2	79.5	30	5.3	80.7	80.3	34	4.6	81.6	81.6
	(0.50, 0.05)	42	5.0	81.4	81.6	42	4.5	80.5	80.7	46	5.2	81.5	82.0	52	5.3	81.4	83.6
	(0.50, 0.10)	58	4.7	80.6	81.1	60	4.7	81.2	79.9	62	4.9	80.3	80.0	70	4.7	81.1	81.9
	(0.10, 0.01)	14	3.9	81.3	80.5	14	4.3	80.4	79.5	16	4.8	83.9	82.9	18	5.4	84.4	83.6
	(0.10, 0.05)	28	5.0	81.3	81.2	28	4.8	80.6	81.4	30	5.3	81.3	80.2	32	5.0	80.1	80.3
	(0.10, 0.10)	46	4.8	81.6	81.0	46	5.1	81.2	81.7	48	4.8	81.5	82.0	50	4.6	80.7	80.0
	(0.25, 0.01)	16	4.3	83.3	82.6	16	4.8	82.4	82.3	18	4.2	84.8	84.9	20	4.7	84.5	84.5
50	(0.25, 0.05)	30	5.0	82.2	82.7	30	4.7	81.6	80.9	32	5.2	82.2	81.1	34	5.4	80.9	80.8
	(0.25, 0.10)	46	5.1	80.5	81.3	46	5.0	80.0	80.3	48	4.8	80.4	80.4	52	4.9	81.3	80.1
	(0.50, 0.01)	18	4.3	82.6	81.9	18	4.6	81.7	80.9	20	5.0	83.5	82.9	22	5.3	82.6	84.5
	(0.50, 0.05)	32	5.0	81.9	81.1	32	4.9	81.3	80.9	34	5.1	81.8	81.6	36	5.7	80.7	79.9
	(0.50, 0.10)	48	5.6	80.3	79.0	50	5.2	81.5	81.2	50	5.0	80.3	80.5	54	5.0	81.1	78.8
	(0.10, 0.01)	12	4.7	87.6	86.8	12	4.8	86.9	86.6	12	4.6	84.5	84.3	14	5.5	86.5	86.4
	(0.10, 0.05)	24	5.6	80.4	80.3	26	5.6	83.3	82.9	26	5.1	81.9	81.6	28	5.3	82.5	82.2
	(0.10, 0.10)	42	5.3	81.0	81.3	42	5.3	80.7	80.2	44	5.6	81.8	81.3	44	5.1	80.4	78.3
	(0.25, 0.01)	12	4.3	82.3	82.3	12	4.9	81.5	82.2	14	5.1	86.0	85.0	14	5.3	81.5	81.5
100	(0.25, 0.05)	26	5.1	81.5	82.5	26	5.4	81.1	81.2	28	5.2	82.8	83.1	28	4.5	80.5	78.6
	(0.25, 0.10)	44	5.2	81.6	82.0	44	5.2	81.4	80.6	44	5.3	80.6	80.7	46	5.3	81.1	81.0
	(0.50, 0.01)	14	4.5	81.5	81.4	14	4.7	80.8	80.6	16	4.8	84.5	84.7	16	5.5	80.3	80.4
	(0.50, 0.05)	28	5.0	81.2	80.6	28	5.0	80.8	80.7	30	4.9	82.4	81.9	30	5.2	80.3	79.2
	(0.50, 0.10)	46	5.4	81.4	81.6	46	4.8	81.2	81.2	46	5.0	80.4	80.6	48	5.5	80.9	79.9

WEB TABLE 12 Estimated required number of clusters (n) for unadjusted average treatment effect test, empirical type I error rate of the Wald test for average treatment effect (ψ percent), predicted power (ϕ^* percent) and empirical power (ϕ percent) of the unadjusted average treatment effect test, when the outcomes are simulated with a continuous individual-level effect modifier. The effect size for power of the average treatment effect is induced from the HTE model and obtained as 0.375.

		CV = 0				CV = 0.3				CV = 0.6				CV = 0.9			
$\overline{m}$	$(\rho_x, \rho_{y\|x})$	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ	n	ψ	ϕ^*	ϕ
20	(0.10, 0.01)	18	3.6	81.1	79.7	18	4.1	80.4	79.6	20	4.0	83.0	81.3	22	4.1	83.6	82.2
	(0.10, 0.05)	26	5.0	80.0	80.9	28	5.0	82.3	82.5	30	5.0	82.4	82.0	34	5.5	82.5	83.0
	(0.10, 0.10)	38	5.3	81.6	81.5	38	5.3	80.8	81.3	40	5.2	80.6	79.7	44	5.7	80.2	80.0
	(0.25, 0.01)	20	3.5	80.8	80.9	22	3.5	84.0	83.8	22	4.6	81.6	81.2	24	4.5	80.9	81.6
	(0.25, 0.05)	30	4.7	82.7	82.0	30	5.2	81.9	82.6	32	5.1	81.9	81.0	36	5.2	81.7	82.9
	(0.25, 0.10)	40	5.0	81.4	81.7	40	5.2	80.7	80.7	42	5.2	80.4	80.2	46	5.0	80.0	80.7
	(0.50, 0.01)	24	3.9	81.6	81.2	24	3.6	80.7	80.6	26	4.4	81.3	81.8	30	4.7	82.3	84.7
	(0.50, 0.05)	32	4.9	80.5	80.7	34	5.2	82.2	82.4	36	5.2	82.0	82.5	40	5.3	81.6	82.1
	(0.50, 0.10)	44	5.2	81.7	81.7	44	4.7	81.1	81.0	46	5.2	80.8	80.5	50	5.4	80.4	80.0
50	(0.10, 0.01)	12	3.8	85.6	85.3	12	4.3	84.8	83.9	12	4.0	82.2	81.0	14	4.6	84.4	84.0
	(0.10, 0.05)	20	5.1	81.6	81.3	20	4.9	80.9	80.8	22	4.8	82.9	82.9	24	5.2	82.8	82.5
	(0.10, 0.10)	32	5.4	82.5	82.2	32	4.5	82.0	82.0	32	4.9	80.6	80.7	34	4.6	80.7	80.8
	(0.25, 0.01)	14	3.6	84.2	84.4	14	3.9	83.4	83.2	14	3.8	80.7	81.0	16	5.3	81.8	82.5
	(0.25, 0.05)	22	4.6	81.1	80.7	22	4.7	80.5	80.9	24	5.5	82.3	82.3	26	5.1	82.2	81.6
	(0.25, 0.10)	34	5.0	82.1	81.6	34	4.5	81.7	81.8	34	5.2	80.4	80.2	36	5.0	80.5	80.3
	(0.50, 0.01)	18	4.0	84.2	84.5	18	4.4	83.5	82.6	18	4.4	81.1	81.2	20	4.6	81.4	82.0
	(0.50, 0.05)	26	4.6	81.6	81.8	26	4.9	81.1	81.2	28	5.3	82.5	82.1	30	5.4	82.4	82.2
	(0.50, 0.10)	36	5.2	80.1	80.0	38	4.6	81.9	82.0	38	4.9	80.8	81.0	40	4.8	80.9	79.5
100	(0.10, 0.01)	10	4.2	88.4	88.2	10	3.9	87.8	87.2	10	4.3	85.6	84.6	10	5.2	81.0	80.6
	(0.10, 0.05)	18	5.1	82.4	81.6	18	5.2	82.0	82.6	18	5.8	80.6	79.9	20	5.2	82.8	81.5
	(0.10, 0.10)	30	5.3	82.8	82.8	30	5.0	82.6	82.9	30	4.9	81.8	81.3	30	5.7	80.4	78.6
	(0.25, 0.01)	12	4.4	86.2	85.8	12	4.4	85.5	84.9	12	4.7	83.5	83.9	14	6.1	86.5	85.7
	(0.25, 0.05)	20	5.1	81.8	81.7	20	4.9	81.4	80.8	20	4.8	80.1	81.1	22	5.2	82.1	80.9
	(0.25, 0.10)	32	5.6	82.4	82.2	32	4.7	82.2	81.9	32	5.0	81.4	82.8	32	5.2	80.2	79.1
	(0.50, 0.01)	16	4.5	85.6	85.7	16	4.7	85.1	84.5	16	5.2	83.5	82.3	16	5.5	80.6	79.7
	(0.50, 0.05)	24	4.9	82.1	82.8	24	5.0	81.8	82.0	24	4.7	80.8	80.8	26	5.1	82.4	80.5
	(0.50, 0.10)	34	5.3	80.2	79.8	36	4.9	82.3	83.2	36	5.2	81.7	81.2	36	4.8	80.6	79.4

H. ADDITIONAL RESULTS FOR THE DATA EXAMPLE

WEB TABLE 13 Estimated required number of clusters (n) based on the quadratic HTE model with a continuous potential effect modifier (age) and the continuous outcome, concern about failing in the STRIDE study. The nominal type I error rate is 5% and the nominal power is 80%. The frame-highlighted cells indicate estimation based on the specified design assumptions, and the CV of cluster sizes, covariate-ICC and conditional outcome-ICC are varied as a sensitivity analysis for sample size. The true HTE for linear age term is zero.

		HTE for Age^2 = 0.1				HTE for Age^2 = 0.2				HTE for Age^2 = 0.3			
		CV of cluster sizes				CV of cluster sizes				CV of cluster sizes			
$\rho_{y\|x}$	ρ_x	0	0.25	0.5	0.75	0	0.25	0.5	0.75	0	0.25	0.5	0.75
0.01	0.01	76	76	76	76	22	22	22	22	12	12	12	12
	0.025	76	76	**76**	76	22	22	**22**	22	12	12	**12**	12
	0.05	76	76	76	76	22	22	22	22	12	12	12	12
	0.10	76	76	76	76	22	22	22	22	12	12	12	12
	0.20	76	76	76	78	22	22	22	22	12	12	12	12
0.05	0.01	74	74	74	74	20	20	20	20	10	10	10	10
	0.025	74	74	74	74	20	20	20	20	10	10	10	10
	0.05	74	74	74	74	20	20	20	20	10	10	10	10
	0.10	74	74	74	74	20	20	20	20	10	10	10	10
	0.20	76	76	76	76	22	22	22	22	12	12	12	12

References

1. Breukelen vGJ, Candel MJ, Berger MP. Relative efficiency of unequal versus equal cluster sizes in cluster randomized and multicentre trials. *Statistics in Medicine* 2007; 26(13): 2589–2603.

□