# Supplementary materials for "Sample size calculation in hierarchical $2 \times 2$ factorial trials with unequal cluster sizes" by Tian et al.

## WEB APPENDIX A: LARGE-SAMPLE COVARIANCE MATRIX

We provide full details for deriving the explicit form of the $4 \times 4$ covariance matrix $\Sigma = \sigma_y^2 \left(\lim_{n\to\infty} n^{-1} \sum_{i=1}^n D_i^T R_i^{-1} D_i\right)^{-1}$. Recall that we have defined the design vector $D_{ij} = \left(1, (X_i - \pi_X), Z_{ij}, (X_i - \pi_X)Z_{ij}\right)^T$ and $D_i = (D_{i1}, \dots, D_{im_i})^T$. For each cluster $i$, the inverse of the compound symmetric correlation matrix can be obtained as

$$R_i^{-1} = \frac{1}{1-\rho} I_{m_i} - \frac{\rho}{(1-\rho)[1+(m_i-1)\rho]} J_{m_i} = \frac{1}{1-\rho}(I_{m_i} + c_i J_{m_i}),$$

where $c_i = -\rho/[1+(m_i-1)\rho]$. Therefore, we can write

$$\frac{1}{n}\sum_{i=1}^n D_i^T R_i^{-1} D_i = \frac{1}{n(1-\rho)}\sum_{i=1}^n D_i^T D_i + \frac{1}{n(1-\rho)}\sum_{i=1}^n c_i D_i^T J_{m_i} D_i. \tag{1}$$

Using the fact that $Z_{ij}^2 = Z_{ij}$ , we can expand one of the main components in expression (1),

$$\frac{1}{n}\sum_{i=1}^n D_i^T D_i = \frac{1}{n}\sum_{i=1}^n \begin{bmatrix} m_i & m_i(X_i-\pi_X) & \sum_{j=1}^{m_i} Z_{ij} & (X_i-\pi_X)\sum_{j=1}^{m_i} Z_{ij} \\ m_i(X_i-\pi_X) & m_i(X_i-\pi_X)^2 & (X_i-\pi_X)\sum_{j=1}^{m_i} Z_{ij} & (X_i-\pi_X)^2\sum_{j=1}^{m_i} Z_{ij} \\ \sum_{j=1}^{m_i} Z_{ij} & (X_i-\pi_X)\sum_{j=1}^{m_i} Z_{ij} & \sum_{j=1}^{m_i} Z_{ij} & (X_i-\pi_X)\sum_{j=1}^{m_i} Z_{ij} \\ (X_i-\pi_X)\sum_{j=1}^{m_i} Z_{ij} & (X_i-\pi_X)^2\sum_{j=1}^{m_i} Z_{ij} & (X_i-\pi_X)\sum_{j=1}^{m_i} Z_{ij} & (X_i-\pi_X)^2\sum_{j=1}^{m_i} Z_{ij} \end{bmatrix}.$$

Assuming the cluster sizes are non-informative, then the cluster size distribution $f(m_i)$ is independent of the assignment of randomized interventions. Define $\bar{m} = E(m_i)$ as the mean cluster size, $\sigma_X^2 = \pi_X(1-\pi_X)$ is the variance of cluster-level intervention indicator. Treating $\sum_{i=1}^{m_i} Z_{ij}$ as a random variable, then it has mean $m_i\pi_Z$ and variance $m_i\sigma_Z^2 = m_i\pi_Z(1-\pi_Z)$ due to binomial sampling. Invoking the Weak Law of Large Numbers (WLLN) for the independent but non-identically distributed random variable, we can write

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\left(\sum_{j=1}^{m_i} Z_{ij}\right) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n E\left(\sum_{j=1}^{m_i} Z_{ij}\right) = \pi_Z \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n m_i = \bar{m}\pi_Z. \tag{2}$$

For the first entry of the matrix above, it follows that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n m_i = E(m_i) = \bar{m}.$$

By the independence among the two interventions and $m_i$, the WLLN, and what is proved in equation (2), we can obtain the other entries in the matrix above

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n m_i(X_i-\pi_X) = E(m_i)E(X_i-\pi_X) = 0, \quad \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n m_i(X_i-\pi_X)^2 = E(m_i)\mathrm{Var}(X_i) = \bar{m}\sigma_X^2,$$

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\left\{(X_i-\pi_X)\sum_{j=1}^{m_i} Z_{ij}\right\} = E(X_i-\pi_X)\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\left(\sum_{j=1}^{m_i} Z_{ij}\right) = 0,$$

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\left\{(X_i-\pi_X)^2\sum_{j=1}^{m_i} Z_{ij}\right\} = \mathrm{Var}(X_i)\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\left(\sum_{j=1}^{m_i} Z_{ij}\right) = \sigma_X^2\bar{m}\pi_Z.$$

This will allow us to write

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n D_i^T D_i = \bar{m}\begin{bmatrix} 1 & 0 & \pi_Z & 0 \\ 0 & \sigma_X^2 & 0 & \pi_Z\sigma_X^2 \\ \pi_Z & 0 & \pi_Z & 0 \\ 0 & \pi_Z\sigma_X^2 & 0 & \pi_Z\sigma_X^2 \end{bmatrix}.$$

Similarly, we can expand the other component in equation (1),

$$\frac{1}{n}\sum_{i=1}^{n} c_i D_i^T J_{m_i} D_i = \frac{1}{n}\sum_{i=1}^{n} c_i \left(\sum_{j=1}^{m_i} D_{ij}\right)\left(\sum_{j=1}^{m_i} D_{ij}^T\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{-\rho}{1+(m_i-1)\rho}\right\}\begin{bmatrix} m_i^2 & m_i^2(X_i-\pi_X) & m_i\sum_{j=1}^{m_i} Z_{ij} & m_i(X_i-\pi_X)\sum_{j=1}^{m_i} Z_{ij} \\ m_i^2(X_i-\pi_X) & m_i^2(X_i-\pi_X)^2 & m_i(X_i-\pi_X)\sum_{j=1}^{m_i} Z_{ij} & m_i(X_i-\pi_X)^2\sum_{j=1}^{m_i} Z_{ij} \\ m_i\sum_{j=1}^{m_i} Z_{ij} & m_i(X_i-\pi_X)\sum_{j=1}^{m_i} Z_{ij} & \left(\sum_{j=1}^{m_i} Z_{ij}\right)^2 & (X_i-\pi_X)\left(\sum_{j=1}^{m_i} Z_{ij}\right)^2 \\ m_i(X_i-\pi_X)\sum_{j=1}^{m_i} Z_{ij} & m_i(X_i-\pi_X)^2\sum_{j=1}^{m_i} Z_{ij} & (X_i-\pi_X)\left(\sum_{j=1}^{m_i} Z_{ij}\right)^2 & (X_i-\pi_X)^2\left(\sum_{j=1}^{m_i} Z_{ij}\right)^2 \end{bmatrix}.$$

Define the following two expectations of the functions of cluster sizes,

$$\bar{\eta}_1 = E\left\{\frac{-m_i\rho}{1+(m_i-1)\rho}\right\}, \qquad \bar{\eta}_2 = E\left\{\frac{-m_i^2\rho}{1+(m_i-1)\rho}\right\}.$$

We show the derivations of some representative entries in the matrix above, by independence and WLLN,

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{-\rho}{1+(m_i-1)\rho}m_i\sum_{j=1}^{m_i} Z_{ij}\right\} &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{-m_i\rho}{1+(m_i-1)\rho}E\left(\sum_{j=1}^{m_i} Z_{ij}\right)\right\} \\ &= \pi_Z\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{-m_i^2\rho}{1+(m_i-1)\rho}\right\} \\ &= \bar{\eta}_2\pi_Z, \\ \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{-\rho}{1+(m_i-1)\rho}\left(\sum_{j=1}^{m_i} Z_{ij}\right)^2\right\} &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{-\rho}{1+(m_i-1)\rho}\mathrm{E}\left[\left(\sum_{j=1}^{m_i} Z_{ij}\right)^2\right]\right\} \\ &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{-\rho}{1+(m_i-1)\rho}\left[\mathrm{Var}\left(\sum_{j=1}^{m_i} Z_{ij}\right)+\left\{\mathrm{E}\left(\sum_{j=1}^{m_i} Z_{ij}\right)\right\}^2\right]\right\} \\ &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{-\rho}{1+(m_i-1)\rho}\left[m_i\sigma_Z^2+m_i^2\pi_Z^2\right]\right\} \\ &= \bar{\eta}_2\pi_Z^2+\bar{\eta}_1\sigma_Z^2.\end{aligned}$$

These allow us to write

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} c_i D_i^T J_{m_i} D_i = \begin{bmatrix} \bar{\eta}_2 & 0 & \bar{\eta}_2\pi_Z & 0 \\ 0 & \bar{\eta}_2\sigma_X^2 & 0 & \bar{\eta}_2\pi_Z\sigma_X^2 \\ \bar{\eta}_2\pi_Z & 0 & \bar{\eta}_2\pi_Z^2+\bar{\eta}_1\sigma_Z^2 & 0 \\ 0 & \bar{\eta}_2\pi_Z\sigma_X^2 & 0 & \bar{\eta}_2\pi_Z^2\sigma_X^2+\bar{\eta}_1\sigma_Z^2\sigma_X^2 \end{bmatrix}.$$

Combining the two components based on equation (1), we can write

$$\lim_{n\to\infty}\frac{1-\rho}{n}\sum_{i=1}^{n} D_i^T R_i^{-1} D_i = \begin{bmatrix}\Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22}\end{bmatrix} = \begin{bmatrix} \bar{m}+\bar{\eta}_2 & 0 & (\bar{m}+\bar{\eta}_2)\pi_Z & 0 \\ 0 & (\bar{m}+\bar{\eta}_2)\sigma_X^2 & 0 & (\bar{m}+\bar{\eta}_2)\pi_Z\sigma_X^2 \\ (\bar{m}+\bar{\eta}_2)\pi_Z & 0 & \bar{m}\pi_Z+\bar{\eta}_2\pi_Z^2+\bar{\eta}_1\sigma_Z^2 & 0 \\ 0 & (\bar{m}+\bar{\eta}_2)\pi_Z\sigma_X^2 & 0 & \bar{m}\pi_Z\sigma_X^2+\bar{\eta}_2\pi_Z^2\sigma_X^2+\bar{\eta}_1\sigma_Z^2\sigma_X^2 \end{bmatrix}.$$

We can observe that $\Omega_{12} = \Omega_{21} = \pi_Z\Omega_{11}$. Using block matrix inversion, we can obtain

$$\begin{bmatrix}\Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22}\end{bmatrix}^{-1} = \begin{bmatrix}\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22}\end{bmatrix} = \begin{bmatrix} (\Omega_{11}-\Omega_{12}\Omega_{22}^{-1}\Omega_{21})^{-1} & -\Omega_{11}^{-1}\Omega_{12}(\Omega_{22}-\Omega_{21}\Omega_{11}^{-1}\Omega_{12})^{-1} \\ -\Omega_{22}^{-1}\Omega_{21}(\Omega_{11}-\Omega_{12}\Omega_{22}^{-1}\Omega_{21})^{-1} & (\Omega_{22}-\Omega_{21}\Omega_{11}^{-1}\Omega_{12})^{-1}\end{bmatrix}.$$

We can compute that

$$\Sigma_{11} = (\Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21})^{-1} = \begin{bmatrix} \frac{(\bar{m}+\bar{\eta}_1)+(\bar{\eta}_2-\bar{\eta}_1)\pi_Z}{(\bar{m}+\bar{\eta}_2)(\bar{m}+\bar{\eta}_1)(1-\pi_Z)} & 0 \\ 0 & \frac{(\bar{m}+\bar{\eta}_1)+(\bar{\eta}_2-\bar{\eta}_1)\pi_Z}{(\bar{m}+\bar{\eta}_2)(\bar{m}+\bar{\eta}_1)(1-\pi_Z)\sigma_X^2} \end{bmatrix},$$

$$\Sigma_{22} = (\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12})^{-1} = (\Omega_{22} - \pi_Z\Omega_{12})^{-1} = \begin{bmatrix} \frac{1}{(\bar{m}+\bar{\eta}_1)\pi_Z(1-\pi_Z)} & 0 \\ 0 & \frac{1}{(\bar{m}+\bar{\eta}_1)\pi_Z(1-\pi_Z)\sigma_X^2} \end{bmatrix},$$

$$\Sigma_{12} = \Sigma_{21}^T = -\Omega_{11}^{-1}\Omega_{12}(\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12})^{-1} = \begin{bmatrix} -\frac{1}{(\bar{m}+\bar{\eta}_1)(1-\pi_Z)} & 0 \\ 0 & -\frac{1}{(\bar{m}+\bar{\eta}_1)(1-\pi_Z)\sigma_X^2} \end{bmatrix}.$$

Therefore, we can obtain the expression of $\Sigma$ or $n\text{Var}(\hat{b})$ as

$$\Sigma = \sigma_y^2 \left( \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} D_i^T R_i^{-1} D_i \right)^{-1} = \sigma_y^2(1-\rho) \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

## WEB APPENDIX B: ACCOUNTING FOR UNEQUAL CLUSTER SIZES WITH TAYLOR SERIES APPROXIMATION

We show the full details for the approximations accounting for the varying cluster sizes. Define CV to be the coefficient of variation of cluster sizes, then we have $\text{CV} = \sigma_m/\bar{m}$, where $\sigma_m$ is the standard deviation of the cluster sizes. Following the power series strategy applied in van Breukelen et al.,we define $d = m_i - \bar{m}$, and it follows that

$$\begin{aligned} E\left\{\frac{m_i\rho}{1+(m_i-1)\rho}\right\} &= E\left\{\frac{\bar{m}+d}{\bar{m}+d+(1-\rho)/\rho}\right\} \\ &= E\left\{\left(\frac{\bar{m}+d}{\bar{m}+(1-\rho)/\rho}\right)\left(\frac{1}{1+\frac{d}{\bar{m}+(1-\rho)/\rho}}\right)\right\} \\ &= E\left\{\left(\frac{\bar{m}+d}{\bar{m}+(1-\rho)/\rho}\right)\sum_{q=0}^{\infty}\left(\frac{-d}{\bar{m}+(1-\rho)/\rho}\right)^q\right\}. \end{aligned}$$

Expanding the series and discarding all terms $d^q$ with $q > 2$ in the equation above, we have the approximation that

$$E\left\{\frac{m_i\rho}{1+(m_i-1)\rho}\right\} \approx \frac{\bar{m}}{\bar{m}+(1-\rho)/\rho} + \frac{E(d)}{\bar{m}+(1-\rho)/\rho} - \frac{\bar{m}E(d)}{(\bar{m}+(1-\rho)/\rho)^2} - \frac{E(d^2)}{(\bar{m}+(1-\rho)/\rho)^2} + \frac{\bar{m}E(d^2)}{(\bar{m}+(1-\rho)/\rho)^3}.$$

Since it also follows that $E(d) = 0$ and $E(d^2) = \sigma_m^2 = \text{CV}^2\bar{m}^2$, we can write

$$E\left\{\frac{m_i\rho}{1+(m_i-1)\rho}\right\} \approx \frac{\bar{m}}{\bar{m}+(1-\rho)/\rho} - \frac{\text{CV}^2\bar{m}^2}{(\bar{m}+(1-\rho)/\rho)^2} + \frac{\text{CV}^2\bar{m}^3}{(\bar{m}+(1-\rho)/\rho)^3} = \frac{\bar{m}}{\bar{m}+(1-\rho)/\rho}\left\{1 - \text{CV}^2\frac{\bar{m}(1-\rho)/\rho}{\{\bar{m}+(1-\rho)/\rho\}^2}\right\}.$$

Therefore, we can use this key approximation to further derive $\bar{\eta}_1$, $\bar{\eta}_2$, and other required expressions in $\Sigma$,

$$\bar{\eta}_1 \approx -\frac{\bar{m}}{\bar{m}+(1-\rho)/\rho}\left\{1 - \text{CV}^2\frac{\bar{m}(1-\rho)/\rho}{\{\bar{m}+(1-\rho)/\rho\}^2}\right\} = -\frac{\bar{m}\rho}{1+(\bar{m}-1)\rho}\left\{1 - \text{CV}^2\frac{\bar{m}\rho(1-\rho)}{\{1+(\bar{m}-1)\rho\}^2}\right\},$$

then, it follows that

$$\bar{m}+\bar{\eta}_1 \approx \bar{m} - \frac{\bar{m}\rho}{1+(\bar{m}-1)\rho} + \text{CV}^2\frac{\bar{m}^2\rho^2(1-\rho)}{\{1+(\bar{m}-1)\rho\}^3} = \frac{\bar{m}\{1+(\bar{m}-2)\rho\}\{1+(\bar{m}-1)\rho\}^2 + \text{CV}\bar{m}^2\rho^2(1-\rho)}{\{1+(\bar{m}-1)\rho\}^3}.$$

We continue to derive the expression of $\bar{\eta}_2$ by simple mathematical manipulations,

$$\bar{\eta}_2 = E\left\{\frac{-m_i^2\rho}{1+(m_i-1)\rho}\right\} = -\bar{m} + E\left\{\frac{-m_i^2\rho}{1+(m_i-1)\rho} + m_i\right\} = -\bar{m} + E\left\{\frac{m_i(1-\rho)}{1+(m_i-1)\rho}\right\} = -\bar{m} - \frac{1-\rho}{\rho}\bar{\eta}_1,$$

and we can then use the approximation above to write

$$\bar{m}+\bar{\eta}_2 = -\frac{1-\rho}{\rho}\bar{\eta}_1 \approx \frac{\bar{m}(1-\rho)}{1+(\bar{m}-1)\rho}\left\{1 - \text{CV}^2\frac{\bar{m}\rho(1-\rho)}{\{1+(\bar{m}-1)\rho\}^2}\right\}.$$

## WEB APPENDIX C: CONTROLLED EFFECTS IN THE PRESENCE OF THE OTHER TREATMENT AND THE CORRESPONDING SAMPLE SIZE REQUIREMENTS

For comparison, we define a counterpart effect estimand of the controlled effect defined in the main text, that is, the controlled effect of a single treatment in the presence of the other treatment (CE$^*$):

$$\text{CE}_X^* = E[Y_{ij}(1,1) - Y_{ij}(0,1)],$$
$$\text{CE}_Z^* = E[Y_{ij}(1,1) - Y_{ij}(1,0)].$$

Assuming the linear mixed model in the main text and under factorial randomization, we further write

$$\text{CE}_X^* = E[Y_{ij}|X_i = 1, Z_{ij} = 1] - E[Y_{ij}|X_i = 0, Z_{ij} = 1] = (\beta_1 + \beta_2 + \beta_3 + \beta_4) - (\beta_1 + \beta_3) = \beta_2 + \beta_4,$$
$$\text{CE}_Z^* = E[Y_{ij}|X_i = 1, Z_{ij} = 1] - E[Y_{ij}|X_i = 1, Z_{ij} = 0] = (\beta_1 + \beta_2 + \beta_3 + \beta_4) - (\beta_1 + \beta_2) = \beta_3 + \beta_4.$$

This type of counterpart controlled effect is of interest when the investigators want to study the effect of one treatment in a population exposed to the other treatment in a $2 \times 2$ factorial trial.

The null hypotheses about these two effects can be given as (A1$^*$) $H_0^{A1*}$ : $\text{CE}_X^* = \beta_2 + \beta_4 = 0$ and (A2$^*$) $H_0^{A2*}$ : $\text{CE}_Z^* = \beta_3 + \beta_4 = 0$. Similar to the notations in the main text, define $\delta_2^* = \delta_2 + \delta_4$ and $\delta_3^* = \delta_3 + \delta_4$ as the effect sizes for the counterpart controlled effects of cluster-level and individual-level treatments, respectively. Then, for testing $H_0^{A1*}$, the total required number of clusters based on a two-sided Wald $z$-test is given by

$$n_{A1^*} = \frac{(z_{1-\alpha/2} + z_{1-\lambda})^2 \omega_2^*}{\delta_2^{*2}},$$

where $\alpha$ and $\lambda$ define the prescribed type I and type II error rates, and $\omega_2^* = n\text{Var}(\hat{\beta}_2 + \hat{\beta}_4)$. Likewise, for testing $H_0^{A2*}$, the total required number of clusters with a nominal test size $\alpha$ and power $1 - \lambda$ is given by

$$n_{A2^*} = \frac{(z_{1-\alpha/2} + z_{1-\lambda})^2 \omega_3^*}{\delta_3^{*2}},$$

where $\omega_3 = n\text{Var}(\hat{\beta}_3 + \hat{\beta}_4)$. Based on the same derivation procedures as in the main text, we can obtain

$$\begin{aligned}\omega_2^* &= n\text{Var}(\hat{\beta}_2 + \hat{\beta}_4) = n\text{Var}(\hat{b}_2) + nVar(\hat{b}_4) + 2n\text{Cov}(\hat{b}_2, \hat{b}_4)\\ &= \frac{\sigma_y^2(1-\rho)}{(\bar{m} + \bar{\eta}_2)\sigma_X^2} + \frac{\pi_Z \sigma_y^2(1-\rho)}{(\bar{m} + \bar{\eta}_1)\sigma_X^2(1-\pi_Z)} + \frac{\sigma_y^2(1-\rho)}{(\bar{m} + \bar{\eta}_1)\sigma_X^2\sigma_Z^2} - 2\frac{\sigma_y^2(1-\rho)}{(\bar{m} + \bar{\eta}_1)\sigma_X^2(1-\pi_Z)} = \frac{\sigma_y^2(1-\rho)}{(\bar{m} + \bar{\eta}_2)\sigma_X^2} + \frac{\sigma_y^2(1-\rho)}{(\bar{m} + \bar{\eta}_1)\sigma_X^2}\frac{1-\pi_Z}{\pi_Z},\end{aligned}$$

$$\begin{aligned}\omega_3^* &= n\text{Var}(\hat{\beta}_3 + \hat{\beta}_4) = n\text{Var}(\hat{\beta}_3) + n\text{Var}(\hat{\beta}_4) + 2n\text{Cov}(\hat{\beta}_3, \hat{\beta}_4) = n\text{Var}(\hat{\beta}_3) + n\text{Var}(\hat{b}_4) + 2nCov(\hat{b}_3, \hat{b}_4) - 2\pi_X n\text{Var}(\hat{b}_4)\\ &= \frac{\sigma_y^2(1-\rho)}{(\bar{m} + \bar{\eta}_1)(1-\pi_X)\pi_Z(1-\pi_Z)} + (1 - 2\pi_X)\frac{\sigma_y^2(1-\rho)}{(\bar{m} + \bar{\eta}_1)\sigma_X^2\sigma_Z^2} = \frac{\sigma_y^2(1-\rho)}{(\bar{m} + \bar{\eta}_1)\pi_Z(1-\pi_Z)\pi_X}.\end{aligned}$$

The general structures of $\omega_2^*$ and $\omega_3^*$ are similar to those of $\omega_2$ and $\omega_3$, respectively, with modifications based the two treatment allocation probability $\pi_X$ and $\pi_Z$. Based on the large-sample variances, we can similarly express the sample size formulas for the separate tests regarding these two controlled effects.

Specifically, the sample size formula for testing $H_0^{A1*}$ is given by

$$n_{A1}^* = n_{A1}^{*I} + n_{A1}^{*II},$$

where

$$n_{A1}^{*I} \approx \frac{(z_{1-\alpha/2} + z_{1-\lambda})^2 \sigma_y^2 \{1 + (\bar{m} - 1)\rho\}}{(\delta_2 + \delta_4)^2 \pi_X (1-\pi_X)\bar{m}} \left[1 - \text{CV}^2 \frac{\bar{m}\rho(1-\rho)}{\{1 + (\bar{m} - 1)\rho\}^2}\right]^{-1},$$
$$n_{A1}^{*II} \approx \frac{(1-\pi_Z)(z_{1-\alpha/2} + z_{1-\lambda})^2 \sigma_y^2 (1-\rho)\{1 + (\bar{m} - 1)\rho\}^3}{(\delta_2 + \delta_4)^2 \pi_Z \pi_X (1-\pi_X)\bar{m}[\{1 + (\bar{m} - 2)\rho\}\{1 + (\bar{m} - 1)\rho\}^2 + \text{CV}^2 \bar{m}\rho^2(1-\rho)]}.$$

Recall the comparative sample size $n_{A1}$ in the main text, we can find $n_{A1}^*$ has a similar two-component structure, with the first component almost the same as $n_{A1}$ (with the change in effect size) and the second component $n_{A1}^{*II} = n_{A1}^{II}\{(1-\pi_Z)^2/\pi_Z^2\} \times \delta_2^2/(\delta_2 + \delta_4)^2$. Due to their similar forms, the required number of clusters for testing the controlled effect with the other treatment present is equally sensitive to the cluster size variability compared with the test for the cluster-level treatment effect with the other treatment absent. And the second component $n_{A1}^{*II}$ can similarly be seen as the extra cost due to the additional individual-level treatment, compared with the traditional two-arm cluster randomized trials.

On the other hand, the required number of clusters for testing $H_0^{A2*}$ is given by

$$n_{A2}^* \approx \frac{(z_{1-\alpha/2} + z_{1-\lambda})^2 \sigma_y^2 (1-\rho)\{1+(\bar{m}-1)\rho\}^3}{(\delta_3+\delta_4)^2 \pi_Z (1-\pi_Z)\pi_X \bar{m}[\{1+(\bar{m}-2)\rho\}\{1+(\bar{m}-1)\rho\}^2 + \mathrm{CV}^2 \bar{m}\rho^2(1-\rho)]}.$$

Comparing the result with $n_{A2}$ in the main text, we find that $n_{A2}^* = n_{A2}\{(1-\pi_X)/\pi_X\} \times \delta_3^2/(\delta_3+\delta_4)^2$. Hence, similar to the test of the individual-level main effect, the required number of clusters $n_{A2}^*$ is not too sensitive to the cluster size variability.

## WEB APPENDIX D: SAMPLE SIZE REQUIREMENTS FOR THE JOINT TEST AND INTERSECTION-UNION TEST BASED ON THE CONTROLLED EFFECTS

In this appendix, we develop the sample size procedures of the joint test and intersection-union (I-U) test for the controlled effects of the two treatments.

### D.1 Joint test

For the joint test, we are interested in the null hypothesis of no effect for both treatments, that is, (D1) $H_0^{D1}$: $\mathrm{CE}_X = \mathrm{CE}_Z = 0$, or $\beta_2 = \beta_3 = 0$. The scientific interpretation of the null is, both the net effect of the CC program (in the absence of the CBT-SP program) and the net effect of the CBT-SP program (in the absence of the CC program) are zero. To obtain the sample size requirements for such joint tests, we need to obtain the covariance between the two effect estimators. Based on the results in Section 2.2 of the main manuscript, we can show

$$\begin{aligned} n\,\mathrm{Cov}(\hat{\beta}_2, \hat{\beta}_3) &= n\,\mathrm{Cov}(\hat{b}_2, \hat{b}_3 - \pi_X \hat{b}_4) = -\pi_X n\,\mathrm{Cov}(\hat{b}_2, \hat{b}_4) \\ &= \frac{\sigma_y^2(1-\rho)}{(\bar{m}+\bar{\eta}_1)(1-\pi_Z)(1-\pi_X)} \approx \frac{\sigma_y^2(1-\rho)\{1+(\bar{m}-1)\rho\}^3}{(1-\pi_Z)(1-\pi_X)\bar{m}[\{1+(\bar{m}-2)\rho\}\{1+(\bar{m}-1)\rho\}^2 + \mathrm{CV}^2\bar{m}\rho^2(1-\rho)]}, \end{aligned}$$

Define $\omega_{23} = n\mathrm{Cov}(\hat{\beta}_2, \hat{\beta}_3)$. Then, from the property of the FGLS estimator, the scaled vectors of the controlled effect estimators converge to a bivariate normal distribution,

$$\sqrt{n}\begin{bmatrix}\hat{\beta}_2\\ \hat{\beta}_3\end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix}\delta_2\\ \delta_3\end{bmatrix}, \begin{bmatrix}\omega_2 & \omega_{23}\\ \omega_{23} & \omega_3\end{bmatrix}\right),$$

where

$$\omega_2 \approx \frac{\sigma_y^2[1+(\bar{m}-1)\rho]}{\pi_X(1-\pi_X)\bar{m}}\left[\frac{[1+(\bar{m}-1)\rho]^2}{[1+(\bar{m}-1)\rho]^2 - \mathrm{CV}^2\bar{m}\rho(1-\rho)} + \frac{\pi_Z(1-\rho)[1+(\bar{m}-1)\rho]^2}{(1-\pi_Z)\{[1+(\bar{m}-2)\rho][1+(\bar{m}-1)\rho]^2 + \mathrm{CV}^2\bar{m}\rho^2(1-\rho)\}}\right],$$

$$\omega_3 \approx \frac{\sigma_y^2(1-\rho)\{1+(\bar{m}-1)\rho\}^3}{\pi_Z(1-\pi_Z)(1-\pi_X)\bar{m}[\{1+(\bar{m}-2)\rho\}\{1+(\bar{m}-1)\rho\}^2 + \mathrm{CV}^2\bar{m}\rho^2(1-\rho)]}, \quad \text{and} \quad \omega_{23} = \pi_Z\omega_3.$$

This motivates a Wald test statistic $J = n\hat{\beta}^T\hat{\Omega}^{-1}\hat{\beta}$, where $\hat{\beta} = (\hat{\beta}_2, \hat{\beta}_3)^T$ and $\hat{\Omega} = \begin{bmatrix}\widehat{\mathrm{Var}}(\hat{\beta}_2) & \widehat{\mathrm{Cov}}(\hat{\beta}_2, \hat{\beta}_3)\\ \widehat{\mathrm{Cov}}(\hat{\beta}_2, \hat{\beta}_3) & \widehat{\mathrm{Var}}(\hat{\beta}_3)\end{bmatrix}$. Furthermore, $J$ asymptotically follows a Chi-square distribution with 2 degrees of freedom and a non-centrality parameter $\theta = n\delta^T\hat{\Omega}^{-1}\delta$, where the target effect size vector $\delta = (\delta_2, \delta_3)^T$. Therefore, given the effect sizes, the power equation of the joint test is

$$1-\lambda = \int_{\chi^2_{1-\alpha}(2)}^{\infty} l\,(x; 2, \theta)\,dx, \tag{3}$$

where $\chi^2_{1-\alpha}(2)$ is the upper-$\alpha$ quantile of the Chi-square distribution with 2 degrees of freedom and $l(x; 2, \theta)$ is the probability density function (PDF) of the non-central Chi-square distribution with non-centrality parameter $\theta$. To estimate the sample size for the joint test, one could first fix the values of ICC, mean cluster sizes, CV, and the two effect sizes, and then specify a series of increasing integers $n$. The required sample size $n_{D1}$ can then be obtained by searching the minimum among the integers that provides $(1-\lambda)$ power according to (3)

## D.2 Intersection-union test

While the joint test rejects the null when at least one treatment has an effect on the outcome, investigators may conclude the "success" of a trial only when both treatments are effective. In this particular case, for the controlled treatment effects, the alternative hypothesis is formulated as $H_1^{E1}$: $\text{CE}_X \neq 0$ and $\text{CE}_Z \neq 0$, while the composite null hypothesis holds when at most one treatment has an controlled effect on the outcome. The I-U test is often used to test this composite null hypothesis, and has been previously applied in trials with multiple co-primary endpoints. While these previous applications focused on one-sided alternatives, we expand this approach to test a two-sided alternative $H_1^{E1}$. Compared to the joint test which answers whether there exists at least one effective treatment, the I-U test examines whether both treatments are effective in the absence of the other treatment.

For the controlled effects specifically, the I-U test considers a bivariate test statistic, $W = \left(W_2, W_3\right)^T$, where $W_2 = \hat{\beta}_2/\sqrt{\widehat{\text{Var}}(\hat{\beta}_2)}$ and $W_3 = \hat{\beta}_3/\sqrt{\widehat{\text{Var}}(\hat{\beta}_3)}$. $W$ follows a bivariate normal distribution

$$W \xrightarrow{d} N\left(\begin{bmatrix}\sqrt{n}\delta_2/\sqrt{\omega_2}\\ \sqrt{n}\delta_3/\sqrt{\omega_3}\end{bmatrix}, \Omega = \begin{bmatrix}1 & \omega_{23}/\sqrt{\omega_2\omega_3}\\ \omega_{23}/\sqrt{\omega_2\omega_3} & 1\end{bmatrix}\right),$$

Given the effect sizes $\delta_2$ and $\delta_3$, the power formula for the two-sided I-U test of the controlled treatment effects can be written as

$$\begin{aligned}1-\lambda &= P\left[\left\{|W_2| > z_{1-\alpha/2}\right\} \cap \left\{|W_3| > z_{1-\alpha/2}\right\}\right]\\ &= \int_{z_{1-\alpha/2}}^{\infty}\int_{z_{1-\alpha/2}}^{\infty} u(W_2, W_3) dW_2 dW_3 + \int_{z_{1-\alpha/2}}^{\infty}\int_{-\infty}^{z_{\alpha/2}} u(W_2, W_3) dW_2 dW_3\\ &\quad + \int_{-\infty}^{z_{\alpha/2}}\int_{z_{1-\alpha/2}}^{\infty} u(W_2, W_3) dW_2 dW_3 + \int_{-\infty}^{z_{\alpha/2}}\int_{-\infty}^{z_{\alpha/2}} u(W_2, W_3) dW_2 dW_3,\end{aligned}$$

where $z_{1-\alpha/2}$ is the upper-$\alpha/2$ quantile of the standard normal distribution, $u(W_2, W_3)$ is the PDF of the bivariate normal distribution of $W$ as discussed above. This I-U power formula can be used to numerically estimate the required sample size. Specifically, the investigators need to specify the values of ICC, mean and CV of cluster sizes, and the effect sizes. Then, a series of increasing integers $n$ can be plugged into the equation to compute the power. The smallest integer $n_{E1}$ that corresponds to no smaller than $(1-\lambda)$ power is then given as the estimated number of clusters to power the I-U test with the composite null $H_0^{E1}$. Finally, because the I-U test rejects $H_0^{E1}$ only when $W_2$ and $W_3$ both fall beyond the critical value, it is straightforward to see that the I-U test requires a sample size at least as large as that required by either test for the controlled effect of a single treatment in Section 3.1. In other words, $n_{E1} \geq \max\{n_{A1}, n_{A2}\}$.

## D.3 Finite-sample considerations

Considering the joint test for the two controlled effects, we focus on the same omnibus test statistic $J$. A simple modification for the Chi-square test to improve finite-sample performance is to assume $J$ follows an $F$-distribution with degrees of freedom $(2, n-2)$. The power equation then becomes

$$1-\lambda = \int_{F_{1-\alpha}(2,n-2)}^{\infty} l^*\left(x; 2, n-2; \theta\right) dx,$$

where $F_{1-\alpha}(2, n-2)$ is the upper-$\alpha$ quantile of the $F$-distribution with degrees of freedom $(2, n-2)$ and $l^*(x; 2, n-2; \theta)$ is the PDF of the non-central $F$-distribution with non-centrality parameter $\theta$, as defined in Web Appendix D.1.

For the finite-sample-corrected I-U test for the controlled effects of the two treatments, we now consider the test statistic $W$ follows a bivariate $t$-distribution with $n-2$ degrees of freedom, with mean $(\sqrt{n}\delta_2/\sqrt{\omega_2}, \sqrt{n}\delta_3/\sqrt{\omega_3})^T$ and covariance matrix $\Omega$ as indicated in the previous section. Hence, we alternatively consider rejecting $H_0^{E1}$ when $|W_2| > t_{n-2;1-\alpha/2}$ and

$|W_3| > t_{n-2;1-\alpha/2}$, where $t_{n-2;1-\alpha/2}$ is the upper-$\alpha/2$ quantile of the $t$-distribution with $n-2$ degrees of freedom. With this finite-sample consideration, the power formula will be given by

$$\begin{aligned}1-\lambda &= P\left[\left\{|W_2| > t_{n-2,1-\alpha/2}\right\} \cap \left\{|W_3| > t_{n-2,1-\alpha/2}\right\}\right] \\ &= \int_{t_{n-2,1-\alpha/2}}^{\infty} \int_{t_{n-2,1-\alpha/2}}^{\infty} u^*(W_2, W_3) dW_2 dW_3 + \int_{t_{n-2,1-\alpha/2}}^{\infty} \int_{-\infty}^{t_{n-2,\alpha/2}} u^*(W_2, W_3) dW_2 dW_3 \\ &+ \int_{-\infty}^{t_{n-2,\alpha/2}} \int_{t_{n-2,1-\alpha/2}}^{\infty} u^*(W_2, W_3) dW_2 dW_3 + \int_{-\infty}^{t_{n-2,\alpha/2}} \int_{-\infty}^{t_{n-2,\alpha/2}} u^*(W_2, W_3) dW_2 dW_3,\end{aligned}$$

where $u^*(W_2, W_3)$ is the PDF of the bivariate $t$-distribution. Having obtained the power formulas for both joint test and I-U test of the controlled effects, the same iterative algorithm described at the end of Web Appendix D.1 and D.2 can be used to search for the smallest value that satisfies the power equation as the result of sample size predictions.

## WEB APPENDIX E: SAMPLE SIZE REQUIREMENTS FOR THE JOINT TEST AND INTERSECTION-UNION TEST BASED ON THE MARGINAL EFFECTS

In this appendix, we develop the sample size procedures of the joint test and I-U test for the marginal effects of the two treatments.

### E.1 Joint test

For the joint test of marginal effects, we are interested in the null hypothesis (D2) $H_0^{D2}$: $\mathrm{ME}_X = \mathrm{ME}_Z = 0$. We similarly find the covariance between the two marginal effect estimators. Based on the results in Section 2.2, we can show

$$n\,\mathrm{Cov}(\widehat{\mathrm{ME}}_X, \widehat{\mathrm{ME}}_Z) = n\,\mathrm{Cov}(\hat{\beta}_2 + \pi_X \hat{\beta}_4, \hat{\beta}_3 + \pi_Z \hat{\beta}_4) = n\,\mathrm{Cov}(\hat{b}_2 + \pi_X \hat{b}_4, \hat{b}_3) = 0,$$

which indicates that the two marginal treatment effect estimators are asymptotically orthogonal. From the property of the FGLS estimator, the scaled vector of the marginal treatment effect estimators converges to a bivariate normal distribution,

$$\sqrt{n}\begin{bmatrix}\widehat{\mathrm{ME}}_X \\ \widehat{\mathrm{ME}}_Z\end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix}\delta_X \\ \delta_Z\end{bmatrix}, \Omega' = \begin{bmatrix}\omega_X & 0 \\ 0 & \omega_Z\end{bmatrix}\right),$$

where

$$\begin{aligned}\omega_X &\approx \frac{\sigma_y^2\{1+(\bar{m}-1)\rho\}}{\pi_X(1-\pi_X)\bar{m}}\left[1-\mathrm{CV}^2\frac{\bar{m}\rho(1-\rho)}{\{1+(\bar{m}-1)\rho\}^2}\right]^{-1}, \\ \omega_Z &\approx \frac{\sigma_y^2(1-\rho)\{1+(\bar{m}-1)\rho\}^3}{\pi_Z(1-\pi_Z)\bar{m}\left[\{1+(\bar{m}-2)\rho\}\{1+(\bar{m}-1)\rho\}^2+\mathrm{CV}^2\bar{m}\rho^2(1-\rho)\right]}.\end{aligned}$$

This motivates a much simpler Wald test statistic than in the controlled effect case, $J^* = \widehat{\mathrm{ME}}_X^2/\widehat{\mathrm{Var}}(\widehat{\mathrm{ME}}_X) + \widehat{\mathrm{ME}}_Z^2/\widehat{\mathrm{Var}}(\widehat{\mathrm{ME}}_Z)$. Similarly, $J^*$ asymptotically follows a Chi-square distribution with 2 degrees of freedom and a non-centrality parameter $\theta' = n\left(\omega_X^{-1}\delta_X^2 + \omega_Z^{-1}\delta_Z^2\right)$. Therefore, given the target effect sizes, the power equation of the joint test is

$$1-\lambda = \int_{\chi^2_{1-\alpha}(2)}^{\infty} v\left(x; 2, \theta'\right) dx,$$

where $\chi^2_{1-\alpha}(2)$ is the upper-$\alpha$ quantile of the Chi-square distribution with 2 degrees of freedom and $v(x; 2, \theta')$ is the PDF of the non-central Chi-square distribution with non-centrality parameter $\theta'$. Based on this power formula, similar iterative searching procedure as described in Web Appendix D.1 could be used to calculate the sample size.

## E.2 Intersection-union test

The I-U test for the two marginal treatment effects considers the bivariate test statistic $\left(W_X, W_Z\right)^T$, where $W_X = \widehat{\mathrm{ME}}_X/\sqrt{\widehat{\mathrm{Var}}(\widehat{\mathrm{ME}}_X)}$ and $W_Z = \widehat{\mathrm{ME}}_Z/\sqrt{\widehat{\mathrm{Var}}(\widehat{\mathrm{ME}}_X)}$. Based on the asymptotic independence between the two marginal effect estimators, it follows that $\left(W_X, W_Z\right)^T$ approximately follows a bivariate normal distribution with mean $\left(\sqrt{n}\delta_X/\sqrt{\omega_X}, \sqrt{n}\delta_Z/\sqrt{\omega_Z}\right)^T$ and a covariance matrix equal to a $2 \times 2$ identity matrix, and therefore the I-U test rejects $H_0^{E2}$ when both $|W_X| > z_{1-\alpha/2}$ and $|W_Z| > z_{1-\alpha/2}$. Given the effect sizes $\delta_X$ and $\delta_Z$, the power formula for the two-sided I-U test can be written as

$$
\begin{aligned}
1-\lambda =& P\left[\left\{|W_X| > z_{1-\alpha/2}\right\} \cap \left\{|W_Z| > z_{1-\alpha/2}\right\}\right] \\
=& \Phi\left(z_{\alpha/2}; \delta_X/\sqrt{\omega_X/n}, 1\right) \Phi\left(z_{\alpha/2}; \delta_Z/\sqrt{\omega_Z}/n, 1\right) \\
&+ \Phi\left(z_{\alpha/2}; \delta_X/\sqrt{\omega_X/n}, 1\right)\left\{1 - \Phi\left(z_{1-\alpha/2}; \delta_Z/\sqrt{\omega_Z/n}, 1\right)\right\} \\
&+ \left\{1 - \Phi\left(z_{1-\alpha/2}; \delta_X/\sqrt{\omega_X/n}, 1\right)\right\} \Phi\left(z_{\alpha/2}; \delta_Z/\sqrt{\omega_Z/n}, 1\right) \\
&+ \left\{1 - \Phi\left(z_{1-\alpha/2}; \delta_X/\sqrt{\omega_X/n}, 1\right)\right\}\left\{1 - \Phi\left(z_{1-\alpha/2}; \delta_Z/\sqrt{\omega_Z/n}, 1\right)\right\} \\
=& \left\{1 + \Phi\left(z_{\alpha/2}; \delta_X/\sqrt{\omega_X/n}, 1\right) - \Phi\left(z_{1-\alpha/2}; \delta_X/\sqrt{\omega_X/n}, 1\right)\right\} \\
&\times \left\{1 + \Phi\left(z_{\alpha/2}; \delta_Z/\sqrt{\omega_Z/n}, 1\right) - \Phi\left(z_{1-\alpha/2}; \delta_Z/\sqrt{\omega_Z/n}, 1\right)\right\},
\end{aligned}
$$

where $\Phi(\bullet; \mu, \sigma^2)$ is the cumulative distribution function (CDF) corresponding to a normal distribution with mean $\mu$ and variance $\sigma^2$. Iterative searching procedures could be conducted to calculate the sample size based on this power formula.

## E.3 Finite-sample considerations

For testing $H_0^{D2}$ based on the omnibus statistics $J^*$ defined in Web Appendix E.1, it is also possible to simply replace the Chi-square distribution with 2 degrees of freedom with distribution $F(2, n-2)$ as in Web Appendix D.3. However, based on our preliminary simulation results, this may not be an efficient approach and might require more clusters than needed (in other words, it will be conservative). Due to asymptotic independence between $\widehat{\mathrm{ME}}_X$ and $\widehat{\mathrm{ME}}_Z$, a more precise correction can be made as follows: recall that $J^* = \widehat{\mathrm{ME}}_X^2/\widehat{\mathrm{Var}}(\widehat{\mathrm{ME}}_X) + \widehat{\mathrm{ME}}_Z^2/\widehat{\mathrm{Var}}(\widehat{\mathrm{ME}}_Z) = W_X^2 + W_Z^2$ is the sum of two quadratic statistics corresponding to the marginal effect of the cluster-level treatment and that of the individual-level treatment. In other words, the distribution of the sum of $W_X^2$ and $W_Z^2$ under the null and alternative will determine the distribution of $J^*$. With finite-sample corrections and under the null, $W_X^2$ follows an $F$-distribution with 1 and $n-2$ degrees of freedom, while $W_Z^2$ follows a Chi-square distribution with 1 degree of freedom. By asymptotic independence, the null distribution of $J^*$ can now be approximated by $F(1, n-2) + \chi^2(1)$, which is the mixed central $F$-$\chi^2$ distribution (i.e., distribution for the sum of an independent central $F$-random variable and an independent central Chi-square random variable). Because the critical value of this null distribution is not directly available, we draw $10,000$ simulations from $F(1, n-2)$ and $\chi^2(1)$, and numerically identify the upper-$\alpha$ quantile to form the associated rejection region. Under the alternative, the omnibus test statistic $J^*$ approximately follows $F(1, n-2, n\omega_X^{-1}\delta_X^2) + \chi^2(1, n\omega_Z^{-1}\delta_Z^2)$, which is the mixed noncentral $F$-$\chi^2$ distribution (i.e., distribution for the sum of an independent noncentral $F$-random variable and an independent noncentral Chi-square random variable). For each candidate $n$, we then draw $10,000$ simulations from $F(1, n-2, n\omega_X^{-1}\delta_X^2) + \chi^2(1, n\omega_Z^{-1}\delta_Z^2)$ and compute the power as the proportion of draws that fall beyond the critical value of the mixed central $F$-$\chi^2$ distribution. The required sample size $n_{D2}$ is identified as the smallest number of clusters such that the power is at least $1-\lambda$.

Finally, for testing $H_0^{E2}$, we can replace the $z$-based I-U test of the marginal effects in Web Appendix E.2 with a mixed $t$- and $z$-based I-U test to improve the validity for testing the cluster-level treatment effect. The power formula for the mixed $t$- and $z$-based I-U test could be written as

$$
\begin{aligned}
1-\lambda =& \left\{1 + \Psi_{n-2}\left(t_{\alpha/2,n-2}; \delta_X/\sqrt{\omega_X/n}\right) - \Psi_{n-2}\left(t_{1-\alpha/2,n-2}; \delta_X/\sqrt{\omega_X/n}\right)\right\} \\
&\times \left\{1 + \Phi\left(z_{\alpha/2}; \delta_Z/\sqrt{\omega_Z/n}, 1\right) - \Phi\left(z_{1-\alpha/2}; \delta_Z/\sqrt{\omega_Z/n}, 1\right)\right\},
\end{aligned}
$$

where $\Phi(\bullet;\mu,\sigma^2)$ is defined earlier, and $\Psi_{n-2}(\bullet;\theta)$ is the CDF corresponding to a $t$-distribution with $n-2$ degrees of freedom and noncentrality parameter $\theta'$. The target required number of clusters $n_{E2}$ for the corrected I-U test of the marginal effects can be obtained by appealing to the same iterative searching algorithm as discussed earlier.

## WEB APPENDIX F: ADDITIONAL SIMULATION RESULTS REGARDING SEPARATE CONTROLLED OR MARGINAL EFFECT TESTS AND INTERACTION TEST

**WEB TABLE 1** Required number of clusters $n_{A1}$, empirical type I error $\psi$, empirical power $\phi$, and predicted power $\hat{\phi}$ corresponding to the test for the controlled effect of the cluster-level treatment with and without finite-sample correction. The controlled effect size of the cluster-level treatment is $\delta_2 = 0.2$. Notation: $\bar{m}$ is the mean cluster size, $\rho$ is the ICC, CV is the coefficient of variation of cluster sizes. The results were based on 5,000 simulations.

| | | | $z$-test | | | | $t$-test | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | CV | $n_{A1}$ | $\psi$ | $\phi$ | $\hat{\phi}$ | $n_{A1}$ | $\psi$ | $\phi$ | $\hat{\phi}$ |
| $\bar{m} = 20$ | $\rho = 0.02$ | 0 | 94 | 0.05 | 0.81 | 0.80 | 96 | 0.05 | 0.80 | 0.80 |
| | | 0.3 | 96 | 0.05 | 0.81 | 0.81 | 98 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 98 | 0.05 | 0.81 | 0.80 | 100 | 0.05 | 0.81 | 0.80 |
| | | 0.9 | 104 | 0.05 | 0.81 | 0.80 | 106 | 0.05 | 0.80 | 0.80 |
| | $\rho = 0.05$ | 0 | 116 | 0.05 | 0.81 | 0.80 | 118 | 0.05 | 0.80 | 0.80 |
| | | 0.3 | 118 | 0.05 | 0.81 | 0.80 | 120 | 0.05 | 0.82 | 0.81 |
| | | 0.6 | 124 | 0.05 | 0.80 | 0.81 | 126 | 0.05 | 0.81 | 0.81 |
| | | 0.9 | 136 | 0.05 | 0.83 | 0.81 | 136 | 0.05 | 0.82 | 0.80 |
| | $\rho = 0.10$ | 0 | 152 | 0.05 | 0.81 | 0.80 | 154 | 0.05 | 0.81 | 0.80 |
| | | 0.3 | 154 | 0.05 | 0.81 | 0.80 | 156 | 0.05 | 0.82 | 0.80 |
| | | 0.6 | 160 | 0.05 | 0.81 | 0.80 | 162 | 0.05 | 0.80 | 0.80 |
| | | 0.9 | 176 | 0.05 | 0.80 | 0.80 | 176 | 0.05 | 0.80 | 0.80 |
| $\bar{m} = 50$ | $\rho = 0.02$ | 0 | 48 | 0.06 | 0.81 | 0.81 | 50 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 48 | 0.05 | 0.81 | 0.81 | 50 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 50 | 0.05 | 0.80 | 0.80 | 52 | 0.05 | 0.81 | 0.80 |
| | | 0.9 | 56 | 0.06 | 0.82 | 0.81 | 58 | 0.05 | 0.83 | 0.81 |
| | $\rho = 0.05$ | 0 | 70 | 0.05 | 0.81 | 0.80 | 72 | 0.05 | 0.81 | 0.80 |
| | | 0.3 | 72 | 0.06 | 0.81 | 0.81 | 74 | 0.05 | 0.80 | 0.81 |
| | | 0.6 | 74 | 0.05 | 0.80 | 0.80 | 76 | 0.05 | 0.80 | 0.80 |
| | | 0.9 | 80 | 0.05 | 0.81 | 0.80 | 82 | 0.05 | 0.80 | 0.80 |
| | $\rho = 0.10$ | 0 | 108 | 0.05 | 0.80 | 0.80 | 110 | 0.04 | 0.82 | 0.80 |
| | | 0.3 | 110 | 0.05 | 0.81 | 0.81 | 112 | 0.05 | 0.82 | 0.81 |
| | | 0.6 | 112 | 0.05 | 0.80 | 0.80 | 114 | 0.05 | 0.80 | 0.80 |
| | | 0.9 | 118 | 0.05 | 0.79 | 0.80 | 120 | 0.05 | 0.79 | 0.80 |
| $\bar{m} = 100$ | $\rho = 0.02$ | 0 | 32 | 0.05 | 0.80 | 0.81 | 34 | 0.04 | 0.82 | 0.81 |
| | | 0.3 | 32 | 0.05 | 0.81 | 0.80 | 34 | 0.05 | 0.80 | 0.80 |
| | | 0.6 | 34 | 0.06 | 0.80 | 0.81 | 36 | 0.04 | 0.81 | 0.81 |
| | | 0.9 | 38 | 0.06 | 0.82 | 0.82 | 40 | 0.05 | 0.82 | 0.82 |
| | $\rho = 0.05$ | 0 | 56 | 0.05 | 0.81 | 0.81 | 58 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 56 | 0.05 | 0.81 | 0.81 | 58 | 0.05 | 0.80 | 0.81 |
| | | 0.6 | 58 | 0.06 | 0.81 | 0.81 | 60 | 0.06 | 0.81 | 0.81 |
| | | 0.9 | 60 | 0.06 | 0.78 | 0.80 | 62 | 0.06 | 0.79 | 0.80 |
| | $\rho = 0.10$ | 0 | 94 | 0.05 | 0.81 | 0.81 | 96 | 0.06 | 0.80 | 0.81 |
| | | 0.3 | 94 | 0.05 | 0.80 | 0.80 | 96 | 0.05 | 0.80 | 0.80 |
| | | 0.6 | 96 | 0.05 | 0.80 | 0.80 | 98 | 0.05 | 0.81 | 0.80 |
| | | 0.9 | 100 | 0.05 | 0.80 | 0.81 | 102 | 0.05 | 0.80 | 0.81 |

**WEB TABLE 2** Required number of clusters $n_{A2}$, empirical type I error $\psi$, empirical power $\phi$, and predicted power $\hat{\phi}$ corresponding to the test for the controlled effect of the individual-level treatment. Notation: $\delta_3$ is the controlled effect size of the individual-level treatment, $\bar{m}$ is the mean cluster size, $\rho$ is the ICC, CV is the coefficient of variation of cluster sizes. The results were based on 5,000 simulations.

| | | | $\delta_3 = 0.15$ | | | | $\delta_3 = 0.3$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | CV | $n_{A2}$ | $\psi$ | $\phi$ | $\hat{\phi}$ | $n_{A2}$ | $\psi$ | $\phi$ | $\hat{\phi}$ |
| $\bar{m} = 20$ | $\rho = 0.02$ | 0 | 140 | 0.05 | 0.81 | 0.80 | 36 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 140 | 0.04 | 0.81 | 0.80 | 36 | 0.05 | 0.83 | 0.81 |
| | | 0.6 | 140 | 0.05 | 0.81 | 0.80 | 36 | 0.05 | 0.81 | 0.81 |
| | | 0.9 | 140 | 0.05 | 0.81 | 0.80 | 36 | 0.05 | 0.80 | 0.82 |
| | $\rho = 0.05$ | 0 | 138 | 0.06 | 0.81 | 0.81 | 36 | 0.05 | 0.84 | 0.82 |
| | | 0.3 | 136 | 0.05 | 0.81 | 0.80 | 34 | 0.06 | 0.79 | 0.80 |
| | | 0.6 | 136 | 0.05 | 0.80 | 0.80 | 34 | 0.05 | 0.80 | 0.80 |
| | | 0.9 | 136 | 0.05 | 0.80 | 0.80 | 34 | 0.05 | 0.79 | 0.80 |
| | $\rho = 0.10$ | 0 | 132 | 0.05 | 0.81 | 0.81 | 34 | 0.05 | 0.83 | 0.82 |
| | | 0.3 | 130 | 0.04 | 0.81 | 0.80 | 34 | 0.06 | 0.81 | 0.82 |
| | | 0.6 | 130 | 0.05 | 0.81 | 0.80 | 34 | 0.05 | 0.82 | 0.82 |
| | | 0.9 | 130 | 0.05 | 0.81 | 0.80 | 34 | 0.05 | 0.81 | 0.82 |
| $\bar{m} = 50$ | $\rho = 0.02$ | 0 | 56 | 0.05 | 0.81 | 0.81 | 14 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 56 | 0.05 | 0.81 | 0.81 | 14 | 0.05 | 0.80 | 0.81 |
| | | 0.6 | 56 | 0.05 | 0.79 | 0.81 | 14 | 0.05 | 0.78 | 0.81 |
| | | 0.9 | 56 | 0.05 | 0.81 | 0.81 | 14 | 0.05 | 0.77 | 0.81 |
| | $\rho = 0.05$ | 0 | 54 | 0.05 | 0.82 | 0.80 | 14 | 0.05 | 0.83 | 0.82 |
| | | 0.3 | 54 | 0.05 | 0.80 | 0.80 | 14 | 0.05 | 0.81 | 0.82 |
| | | 0.6 | 54 | 0.05 | 0.81 | 0.80 | 14 | 0.05 | 0.80 | 0.82 |
| | | 0.9 | 54 | 0.05 | 0.80 | 0.80 | 14 | 0.05 | 0.78 | 0.82 |
| | $\rho = 0.10$ | 0 | 52 | 0.05 | 0.82 | 0.81 | 14 | 0.05 | 0.85 | 0.83 |
| | | 0.3 | 52 | 0.05 | 0.82 | 0.81 | 14 | 0.05 | 0.83 | 0.83 |
| | | 0.6 | 52 | 0.05 | 0.81 | 0.81 | 14 | 0.05 | 0.82 | 0.83 |
| | | 0.9 | 52 | 0.05 | 0.80 | 0.81 | 14 | 0.05 | 0.80 | 0.84 |
| $\bar{m} = 100$ | $\rho = 0.02$ | 0 | 28 | 0.06 | 0.81 | 0.81 | 8 | 0.05 | 0.85 | 0.86 |
| | | 0.3 | 28 | 0.05 | 0.81 | 0.81 | 8 | 0.05 | 0.85 | 0.86 |
| | | 0.6 | 28 | 0.05 | 0.80 | 0.81 | 8 | 0.05 | 0.83 | 0.86 |
| | | 0.9 | 28 | 0.05 | 0.79 | 0.81 | 8 | 0.05 | 0.80 | 0.86 |
| | $\rho = 0.05$ | 0 | 28 | 0.06 | 0.82 | 0.82 | 8 | 0.05 | 0.86 | 0.87 |
| | | 0.3 | 28 | 0.05 | 0.82 | 0.82 | 8 | 0.06 | 0.86 | 0.87 |
| | | 0.6 | 28 | 0.05 | 0.81 | 0.82 | 8 | 0.05 | 0.84 | 0.87 |
| | | 0.9 | 28 | 0.05 | 0.80 | 0.82 | 8 | 0.05 | 0.81 | 0.87 |
| | $\rho = 0.10$ | 0 | 26 | 0.05 | 0.81 | 0.81 | 8 | 0.05 | 0.88 | 0.88 |
| | | 0.3 | 26 | 0.05 | 0.80 | 0.81 | 8 | 0.06 | 0.87 | 0.88 |
| | | 0.6 | 26 | 0.05 | 0.80 | 0.81 | 8 | 0.05 | 0.86 | 0.88 |
| | | 0.9 | 26 | 0.05 | 0.78 | 0.81 | 8 | 0.05 | 0.82 | 0.88 |

**WEB TABLE 3** Required number of clusters $n_{B1}$, empirical type I error $\psi$, empirical power $\phi$, and predicted power $\hat{\phi}$ corresponding to the test for the marginal effect of the cluster-level treatment with and without finite-sample correction. The marginal effect size of the cluster-level treatment is $\delta_X = 0.2$. Notation: $\bar{m}$ is the mean cluster size, $\rho$ is the ICC, CV is the coefficient of variation of cluster sizes. The results were based on $5,000$ simulations.

| | | | $z$-test | | | | $t$-test | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | CV | $n_{B1}$ | $\psi$ | $\phi$ | $\hat{\phi}$ | $n_{B1}$ | $\psi$ | $\phi$ | $\hat{\phi}$ |
| $\bar{m} = 20$ | $\rho = 0.02$ | 0 | 56 | 0.05 | 0.81 | 0.81 | 58 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 56 | 0.05 | 0.81 | 0.81 | 58 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 60 | 0.06 | 0.81 | 0.81 | 62 | 0.05 | 0.81 | 0.81 |
| | | 0.9 | 66 | 0.05 | 0.82 | 0.81 | 68 | 0.05 | 0.81 | 0.81 |
| | $\rho = 0.05$ | 0 | 78 | 0.05 | 0.80 | 0.81 | 80 | 0.05 | 0.80 | 0.81 |
| | | 0.3 | 80 | 0.06 | 0.81 | 0.81 | 82 | 0.05 | 0.82 | 0.81 |
| | | 0.6 | 86 | 0.06 | 0.80 | 0.81 | 88 | 0.05 | 0.80 | 0.81 |
| | | 0.9 | 96 | 0.05 | 0.82 | 0.80 | 98 | 0.05 | 0.82 | 0.80 |
| | $\rho = 0.10$ | 0 | 114 | 0.06 | 0.79 | 0.80 | 116 | 0.05 | 0.81 | 0.80 |
| | | 0.3 | 118 | 0.05 | 0.80 | 0.81 | 118 | 0.05 | 0.80 | 0.80 |
| | | 0.6 | 124 | 0.05 | 0.80 | 0.80 | 126 | 0.05 | 0.81 | 0.80 |
| | | 0.9 | 138 | 0.06 | 0.81 | 0.80 | 140 | 0.04 | 0.81 | 0.80 |
| $\bar{m} = 50$ | $\rho = 0.02$ | 0 | 32 | 0.06 | 0.80 | 0.81 | 34 | 0.04 | 0.82 | 0.81 |
| | | 0.3 | 32 | 0.06 | 0.79 | 0.80 | 34 | 0.05 | 0.80 | 0.80 |
| | | 0.6 | 36 | 0.06 | 0.82 | 0.82 | 38 | 0.05 | 0.82 | 0.82 |
| | | 0.9 | 40 | 0.06 | 0.83 | 0.81 | 42 | 0.06 | 0.82 | 0.81 |
| | $\rho = 0.05$ | 0 | 56 | 0.05 | 0.83 | 0.81 | 58 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 56 | 0.06 | 0.80 | 0.81 | 58 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 60 | 0.06 | 0.81 | 0.81 | 62 | 0.05 | 0.81 | 0.81 |
| | | 0.9 | 66 | 0.06 | 0.81 | 0.81 | 68 | 0.05 | 0.82 | 0.81 |
| | $\rho = 0.10$ | 0 | 94 | 0.05 | 0.79 | 0.81 | 96 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 94 | 0.05 | 0.79 | 0.80 | 96 | 0.05 | 0.80 | 0.80 |
| | | 0.6 | 98 | 0.05 | 0.80 | 0.80 | 100 | 0.05 | 0.80 | 0.80 |
| | | 0.9 | 104 | 0.05 | 0.80 | 0.80 | 106 | 0.05 | 0.80 | 0.80 |
| $\bar{m} = 100$ | $\rho = 0.02$ | 0 | 24 | 0.06 | 0.81 | 0.81 | 26 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 24 | 0.06 | 0.80 | 0.80 | 26 | 0.05 | 0.80 | 0.80 |
| | | 0.6 | 26 | 0.06 | 0.81 | 0.81 | 28 | 0.05 | 0.81 | 0.81 |
| | | 0.9 | 30 | 0.06 | 0.82 | 0.82 | 32 | 0.05 | 0.82 | 0.82 |
| | $\rho = 0.05$ | 0 | 48 | 0.06 | 0.81 | 0.81 | 50 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 48 | 0.06 | 0.80 | 0.81 | 50 | 0.05 | 0.82 | 0.81 |
| | | 0.6 | 50 | 0.05 | 0.81 | 0.81 | 52 | 0.05 | 0.81 | 0.81 |
| | | 0.9 | 54 | 0.06 | 0.79 | 0.81 | 56 | 0.06 | 0.80 | 0.81 |
| | $\rho = 0.10$ | 0 | 86 | 0.06 | 0.79 | 0.80 | 88 | 0.05 | 0.80 | 0.80 |
| | | 0.3 | 88 | 0.05 | 0.81 | 0.81 | 90 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 88 | 0.05 | 0.79 | 0.80 | 90 | 0.05 | 0.81 | 0.80 |
| | | 0.9 | 92 | 0.06 | 0.78 | 0.80 | 94 | 0.06 | 0.79 | 0.80 |

**WEB TABLE 4** Required number of clusters $n_{B2}$, empirical type I error $\psi$, empirical power $\phi$, and predicted power $\hat{\phi}$ corresponding to the test for the marginal effect of the individual-level treatment. Notation: $\delta_Z$ is the marginal effect size of the individual-level treatment, $\bar{m}$ is the mean cluster size, $\rho$ is the ICC, CV is the coefficient of variation of cluster sizes. The results were based on 5,000 simulations.

| | | | $\delta_Z = 0.1$ | | | | $\delta_Z = 0.15$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | CV | $n_{A2}$ | $\psi$ | $\phi$ | $\hat{\phi}$ | $n_{A2}$ | $\psi$ | $\phi$ | $\hat{\phi}$ |
| $\bar{m} = 20$ | $\rho = 0.02$ | 0 | 158 | 0.05 | 0.81 | 0.80 | 70 | 0.05 | 0.80 | 0.80 |
| | | 0.3 | 158 | 0.05 | 0.81 | 0.80 | 70 | 0.05 | 0.80 | 0.80 |
| | | 0.6 | 156 | 0.04 | 0.80 | 0.80 | 70 | 0.04 | 0.80 | 0.80 |
| | | 0.9 | 156 | 0.05 | 0.80 | 0.80 | 70 | 0.05 | 0.81 | 0.80 |
| | $\rho = 0.05$ | 0 | 154 | 0.05 | 0.82 | 0.80 | 70 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 154 | 0.05 | 0.81 | 0.80 | 68 | 0.05 | 0.81 | 0.80 |
| | | 0.6 | 154 | 0.05 | 0.80 | 0.80 | 68 | 0.05 | 0.80 | 0.80 |
| | | 0.9 | 154 | 0.04 | 0.81 | 0.80 | 68 | 0.05 | 0.80 | 0.80 |
| | $\rho = 0.10$ | 0 | 148 | 0.05 | 0.81 | 0.80 | 66 | 0.05 | 0.83 | 0.81 |
| | | 0.3 | 148 | 0.05 | 0.82 | 0.80 | 66 | 0.05 | 0.83 | 0.81 |
| | | 0.6 | 146 | 0.04 | 0.82 | 0.80 | 66 | 0.05 | 0.82 | 0.81 |
| | | 0.9 | 146 | 0.05 | 0.80 | 0.80 | 66 | 0.05 | 0.81 | 0.81 |
| $\bar{m} = 50$ | $\rho = 0.02$ | 0 | 64 | 0.04 | 0.81 | 0.81 | 28 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 64 | 0.05 | 0.82 | 0.81 | 28 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 64 | 0.04 | 0.82 | 0.81 | 28 | 0.05 | 0.80 | 0.81 |
| | | 0.9 | 64 | 0.05 | 0.82 | 0.81 | 28 | 0.05 | 0.78 | 0.81 |
| | $\rho = 0.05$ | 0 | 62 | 0.05 | 0.82 | 0.81 | 28 | 0.05 | 0.83 | 0.82 |
| | | 0.3 | 62 | 0.05 | 0.81 | 0.81 | 28 | 0.05 | 0.82 | 0.82 |
| | | 0.6 | 62 | 0.05 | 0.81 | 0.81 | 28 | 0.05 | 0.81 | 0.82 |
| | | 0.9 | 62 | 0.05 | 0.80 | 0.81 | 28 | 0.05 | 0.79 | 0.82 |
| | $\rho = 0.10$ | 0 | 58 | 0.05 | 0.81 | 0.80 | 26 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 58 | 0.05 | 0.81 | 0.80 | 26 | 0.05 | 0.82 | 0.81 |
| | | 0.6 | 58 | 0.05 | 0.80 | 0.80 | 26 | 0.05 | 0.81 | 0.81 |
| | | 0.9 | 58 | 0.05 | 0.81 | 0.80 | 26 | 0.05 | 0.80 | 0.81 |
| $\bar{m} = 100$ | $\rho = 0.02$ | 0 | 32 | 0.05 | 0.81 | 0.81 | 14 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 32 | 0.05 | 0.82 | 0.81 | 14 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 32 | 0.05 | 0.81 | 0.81 | 14 | 0.05 | 0.80 | 0.81 |
| | | 0.9 | 32 | 0.05 | 0.80 | 0.81 | 14 | 0.05 | 0.76 | 0.81 |
| | $\rho = 0.05$ | 0 | 32 | 0.05 | 0.83 | 0.82 | 14 | 0.05 | 0.83 | 0.82 |
| | | 0.3 | 32 | 0.05 | 0.83 | 0.82 | 14 | 0.05 | 0.82 | 0.82 |
| | | 0.6 | 32 | 0.05 | 0.82 | 0.82 | 14 | 0.05 | 0.81 | 0.82 |
| | | 0.9 | 32 | 0.05 | 0.81 | 0.82 | 14 | 0.05 | 0.77 | 0.82 |
| | $\rho = 0.10$ | 0 | 30 | 0.05 | 0.82 | 0.82 | 14 | 0.05 | 0.85 | 0.84 |
| | | 0.3 | 30 | 0.05 | 0.83 | 0.82 | 14 | 0.05 | 0.84 | 0.84 |
| | | 0.6 | 30 | 0.05 | 0.81 | 0.82 | 14 | 0.05 | 0.83 | 0.84 |
| | | 0.9 | 30 | 0.05 | 0.81 | 0.82 | 14 | 0.05 | 0.80 | 0.84 |

# WEB APPENDIX G: A SIMULATION STUDY FOR JOINT AND INTERSECTION-UNION TESTING OF THE CONTROLLED OR MARGINAL EFFECTS

## G.1 Simulation designs

As mentioned in Section 4 of the main paper, we also carried out simulation studies to evaluate the sample size formulas regarding the joint tests and I-U tests for both controlled and marginal effects in a hierarchical $2\times2$ factorial trial with equal randomization ($\pi_X = \pi_Z = 1/2$). Similar to other sample size formulas discussed in the main text, the number of clusters is also determined by nominal type I error rate ($\alpha$), power ($1-\lambda$), total variance ($\sigma_y^2$), ICC ($\rho$), mean cluster size ($\bar{m}$), CV of cluster sizes, and the effect sizes specified in different hypotheses ($\delta_2$, $\delta_3$, $\delta_X$, or $\delta_Z$). Throughout, we fixed the total variance $\sigma_y^2$ at 1, nominal type I error $\alpha$ at 0.05 and the desired power level $1-\lambda$ at 0.8, and varied the remaining parameters. We considered three levels of mean cluster sizes $\bar{m} \in \{20, 50, 100\}$, and three levels of ICC $\rho \in \{0.02, 0.05, 0.1\}$, based on the range commonly reported in the cluster randomized design literature. The CV of cluster sizes were chosen from CV $\in \{0, 0.3, 0.6, 0.9\}$ with CV $= 0$ representing equal cluster sizes. To ensure a realistic range of predicted sample sizes, we separately specified effect sizes for each type of hypothesis. We chose $(\delta_2, \delta_3) \in \{(0.2, 0.15), (0.4, 0.3)\}$ for testing $H_0^{D1}$, $(\delta_X, \delta_Z) \in \{(0.2, 0.1), (0.25, 0.15)\}$ for testing $H_0^{D2}$, $(\delta_2, \delta_3) \in \{(0.2, 0.15), (0.4, 0.3)\}$ for testing $H_0^{E1}$, and $(\delta_X, \delta_Z) \in \{(0.2, 0.1), (0.4, 0.2)\}$ for testing $H_0^{E2}$. For each parameter combination, we estimated the number of clusters $n$ that gives at least 80% power and rounded to the nearest even integer above to ensure an exactly equal randomization. We used the predicted cluster number $n$ to simulate correlated outcomes and obtain the empirical power to validate the accuracy of the closed-form power prediction.

**WEB TABLE 5** Specification of regression parameters for generating correlated outcome data in different simulation scenarios. (D1) represents the joint test of the controlled effects of the two treatments, (D2) represents the joint test of the marginal effects of the two treatments, (E1) represents the intersection-union (I-U) test of the two controlled effects, and (E2) represents the I-U test of the two marginal effects. $\beta_2$, $\beta_3$, and $\beta_4$ stands for the true regression parameters corresponding to the main effect of the cluster-level treatment, the main effect of the individual-level treatment, and the interaction effect, respectively.

| Test label | Hypothesis | $\beta_2$ | $\beta_3$ | $\beta_4$ | Effect size |
|---|---|---|---|---|---|
| (D1) | Null ($H_0^{D1}$) | 0 | 0 | 0.05 | $\text{CE}_X = \text{CE}_Z = 0$ |
| | Alternative ($H_1^{D1}$) | $\delta_2$ | $\delta_3$ | 0.05 | $\text{CE}_X = \delta_2, \text{CE}_Z = \delta_3$ |
| (D2) | Null ($H_0^{D2}$) | 0.15 | 0.15 | $-0.3$ | $\text{ME}_X = \text{ME}_Z = 0$ |
| | Alternative ($H_1^{D2}$) | 0.15 | $\delta_Z - \delta_X + 0.15$ | $2(\delta_X - 0.15)$ | $\text{ME}_X = \delta_X, \text{ME}_Z = \delta_Z$ |
| (E1) | Null ($H_0^{E1}$) | 0 | $\delta_3$ | 0.05 | $\text{CE}_X = 0, \text{CE}_Z = \delta_3$ |
| | Alternative ($H_1^{E1}$) | $\delta_2$ | $\delta_3$ | 0.05 | $\text{CE}_X = \delta_2, \text{CE}_Z = \delta_3$ |
| (E2) | Null ($H_0^{E2}$) | 0.15 | $\delta_Z + 0.15$ | $-0.3$ | $\text{ME}_X = 0, \text{ME}_Z = \delta_Z$ |
| | Alternative ($H_1^{E2}$) | 0.15 | $\delta_Z - \delta_X + 0.15$ | $2(\delta_X - 0.15)$ | $\text{ME}_X = \delta_X, \text{ME}_Z = \delta_Z$ |

We generated correlated outcome data based on model (1) in the main paper. We fixed $\beta_1 = 1$ for simplicity. Web Table 5 summarized the specification of regression parameters in each simulation scenario to match the assumed controlled or marginal effect sizes. For instance, when designing the simulations for testing $H_0^{D2}$ and $H_0^{E2}$, we fixed $\beta_2 = 0.15$ and solved $\beta_3$ and $\beta_4$ based on the corresponding linear constraints put by the effect sizes. Given values of $\bar{m}$ and CV, we simulate varying cluster sizes using $m_i \sim \text{Gamma}(g, h)$, where the shape parameter $g = \text{CV}^{-2}$ and the rate parameter $h = \bar{m}^{-1}\text{CV}^{-2}$. The simulated cluster size $m_i$ was rounded to the nearest integer and ensured to be at least 2 for computational stability. Finally, the cluster-specific random intercept $a_i$ was randomly generated from $\mathcal{N}(0, \rho)$, and the random error $\epsilon_{ij}$ was independently generated from $\mathcal{N}(0, 1-\rho)$. For each parameter combination, we generated $5,000$ hypothetical factorial trials to evaluate empirical type I error under the null and power under the alternative.

For each simulated hypothetical factorial trial, we fitted the linear mixed model (1) using the restricted maximum likelihood estimation (REML) and carried out the corresponding test for inference. Under each null, the empirical type I error rate was computed as the proportion of false rejections among the $5,000$ trials. Under the alternative, the empirical power was calculated as the proportion of correct rejections among the $5,000$ trials, and was compared with the power prediction based on our proposed formulas. Finally, we replicated the simulations based on the modified sample size methods discussed in Web Appendix D and E, to assess the potential improvement of type I error rate. Our simulations were carried out in R (version 4.0.3) and the linear mixed model was fitted using the `nlme` package.

## G.2 Simulation results

Web Tables 6 and 7 present the estimated required number of clusters ($n_{D1}$), empirical type I error ($\psi$), empirical power ($\phi$) and predicted power ($\hat{\phi}$) corresponding to the joint test for the cluster-level and the individual-level controlled effects with two levels of effect sizes. For the simulation parameters we considered, the estimated sample sizes are generally similar between the large-sample Chi-square test and the $F$-test, with the latter one always requiring a few more clusters. The $F$-test controls for the type I error rate inflation when the estimated sample size is smaller than 30. Except for a handful of cases where the predicted power is slightly lower than the empirical power (suggesting the proposed approach is slightly conservative), the empirical power of the $F$-test generally agrees with the analytical results.

Web Table 8 and 9 present the estimated required number of clusters ($n_{E1}$), empirical type I error, empirical power and predicted power corresponding to the I-U test for the two controlled effects with two levels of effect sizes. Similar to the sample size inflation observed for the joint test for the controlled effects, the finite-sample corrected test also requires a few more clusters. Overall, our sample size procedure accurately predicted the power for both the $z$-based and the $t$-based I-U test, with the latter maintaining an accurate type I error rate with fewer than 20 clusters.

Web Tables 10 and 11 present the estimated required number of clusters ($n_{D2}$), empirical type I error, empirical power and predicted power corresponding to the joint test for the cluster-level and the individual-level marginal effects with two levels of effect sizes. Furthermore, Web Table 12 and 13 present the estimated required number of clusters ($n_{E2}$), empirical and predicted results corresponding to the I-U test for the two marginal effects with two levels of effect sizes. The general patterns are similar to those for studying the controlled effects. In particular, the mixed $F$-$\chi^2$ test for the joint test and the mixed $z$- and $t$-based test for the I-U test both provided empirical power in full agreement with the prediction; they also both prevent the type I error rate inflation when the estimated sample size is smaller than about 30.

**WEB TABLE 6** Required number of clusters $n_{D1}$, empirical type I error $\psi$, empirical power $\phi$, and predicted power $\hat{\phi}$ corresponding to the joint test of the controlled effects with and without finite-sample correction. The controlled effect size of the cluster-level treatment is $\delta_2 = 0.2$ and controlled effect size of the individual-level treatment is $\delta_3 = 0.1$. Notation: $\bar{m}$ is the mean cluster size, $\rho$ is the ICC, CV is the coefficient of variation of cluster sizes. The results were based on 5,000 simulations.

| | | | $\chi^2$-test | | | | $F$-test | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | CV | $n_{D1}$ | $\psi$ | $\phi$ | $\hat{\phi}$ | $n_{D1}$ | $\psi$ | $\phi$ | $\hat{\phi}$ |
| $\bar{m} = 20$ | $\rho = 0.02$ | 0 | 114 | 0.05 | 0.81 | 0.80 | 118 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 116 | 0.05 | 0.82 | 0.81 | 118 | 0.05 | 0.81 | 0.80 |
| | | 0.6 | 118 | 0.05 | 0.79 | 0.80 | 122 | 0.05 | 0.81 | 0.80 |
| | | 0.9 | 126 | 0.05 | 0.82 | 0.81 | 128 | 0.05 | 0.81 | 0.80 |
| | $\rho = 0.05$ | 0 | 136 | 0.05 | 0.81 | 0.81 | 138 | 0.05 | 0.80 | 0.80 |
| | | 0.3 | 136 | 0.05 | 0.80 | 0.80 | 140 | 0.05 | 0.82 | 0.80 |
| | | 0.6 | 142 | 0.06 | 0.81 | 0.80 | 144 | 0.05 | 0.81 | 0.80 |
| | | 0.9 | 152 | 0.06 | 0.82 | 0.81 | 154 | 0.05 | 0.83 | 0.80 |
| | $\rho = 0.10$ | 0 | 160 | 0.05 | 0.81 | 0.80 | 164 | 0.05 | 0.80 | 0.80 |
| | | 0.3 | 162 | 0.05 | 0.81 | 0.80 | 166 | 0.04 | 0.82 | 0.80 |
| | | 0.6 | 166 | 0.05 | 0.81 | 0.80 | 170 | 0.05 | 0.81 | 0.80 |
| | | 0.9 | 176 | 0.05 | 0.81 | 0.80 | 178 | 0.04 | 0.81 | 0.80 |
| $\bar{m} = 50$ | $\rho = 0.02$ | 0 | 56 | 0.05 | 0.82 | 0.81 | 58 | 0.04 | 0.80 | 0.80 |
| | | 0.3 | 56 | 0.05 | 0.81 | 0.81 | 60 | 0.04 | 0.81 | 0.81 |
| | | 0.6 | 58 | 0.05 | 0.80 | 0.81 | 62 | 0.04 | 0.82 | 0.81 |
| | | 0.9 | 62 | 0.05 | 0.81 | 0.81 | 66 | 0.05 | 0.82 | 0.81 |
| | $\rho = 0.05$ | 0 | 72 | 0.05 | 0.81 | 0.81 | 74 | 0.05 | 0.81 | 0.80 |
| | | 0.3 | 72 | 0.06 | 0.81 | 0.80 | 76 | 0.04 | 0.82 | 0.81 |
| | | 0.6 | 74 | 0.05 | 0.80 | 0.81 | 78 | 0.05 | 0.81 | 0.81 |
| | | 0.9 | 78 | 0.06 | 0.81 | 0.81 | 80 | 0.04 | 0.81 | 0.80 |
| | $\rho = 0.10$ | 0 | 86 | 0.05 | 0.82 | 0.80 | 90 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 86 | 0.05 | 0.80 | 0.80 | 90 | 0.05 | 0.81 | 0.80 |
| | | 0.6 | 88 | 0.05 | 0.82 | 0.80 | 92 | 0.05 | 0.82 | 0.81 |
| | | 0.9 | 90 | 0.05 | 0.80 | 0.80 | 92 | 0.05 | 0.80 | 0.80 |
| $\bar{m} = 100$ | $\rho = 0.02$ | 0 | 34 | 0.05 | 0.81 | 0.81 | 38 | 0.04 | 0.83 | 0.82 |
| | | 0.3 | 34 | 0.06 | 0.80 | 0.80 | 38 | 0.04 | 0.83 | 0.81 |
| | | 0.6 | 36 | 0.05 | 0.81 | 0.82 | 38 | 0.04 | 0.82 | 0.80 |
| | | 0.9 | 38 | 0.05 | 0.82 | 0.82 | 40 | 0.04 | 0.81 | 0.80 |
| | $\rho = 0.05$ | 0 | 46 | 0.06 | 0.82 | 0.82 | 48 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 46 | 0.05 | 0.82 | 0.81 | 48 | 0.05 | 0.80 | 0.81 |
| | | 0.6 | 46 | 0.05 | 0.81 | 0.81 | 50 | 0.04 | 0.82 | 0.82 |
| | | 0.9 | 48 | 0.05 | 0.80 | 0.82 | 50 | 0.04 | 0.81 | 0.81 |
| | $\rho = 0.10$ | 0 | 52 | 0.05 | 0.81 | 0.80 | 56 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 52 | 0.06 | 0.81 | 0.80 | 56 | 0.05 | 0.82 | 0.81 |
| | | 0.6 | 52 | 0.06 | 0.80 | 0.80 | 56 | 0.05 | 0.81 | 0.81 |
| | | 0.9 | 54 | 0.06 | 0.81 | 0.81 | 56 | 0.05 | 0.80 | 0.81 |

**WEB TABLE 7** Required number of clusters $n_{D1}$, empirical type I error $\psi$, empirical power $\phi$, and predicted power $\hat{\phi}$ corresponding to the joint test of the controlled effects with and without finite-sample correction. The controlled effect size of the cluster-level treatment is $\delta_2 = 0.5$ and controlled effect size of the individual-level treatment is $\delta_3 = 0.25$. Notation: $\bar{m}$ is the mean cluster size, $\rho$ is the ICC, CV is the coefficient of variation of cluster sizes. The results were based on 5,000 simulations.

| | | | $\chi^2$-test | | | | $F$-test | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | CV | $n_{D1}$ | $\psi$ | $\phi$ | $\hat{\phi}$ | $n_{D1}$ | $\psi$ | $\phi$ | $\hat{\phi}$ |
| $\bar{m} = 20$ | $\rho = 0.02$ | 0 | 20 | 0.05 | 0.84 | 0.84 | 22 | 0.03 | 0.83 | 0.82 |
| | | 0.3 | 20 | 0.05 | 0.83 | 0.84 | 22 | 0.04 | 0.83 | 0.81 |
| | | 0.6 | 20 | 0.06 | 0.82 | 0.82 | 22 | 0.04 | 0.80 | 0.80 |
| | | 0.9 | 20 | 0.05 | 0.79 | 0.80 | 24 | 0.04 | 0.83 | 0.82 |
| | $\rho = 0.05$ | 0 | 22 | 0.05 | 0.82 | 0.81 | 26 | 0.04 | 0.85 | 0.83 |
| | | 0.3 | 22 | 0.06 | 0.81 | 0.81 | 26 | 0.04 | 0.84 | 0.82 |
| | | 0.6 | 24 | 0.06 | 0.83 | 0.83 | 26 | 0.04 | 0.82 | 0.81 |
| | | 0.9 | 26 | 0.07 | 0.83 | 0.83 | 28 | 0.04 | 0.83 | 0.81 |
| | $\rho = 0.10$ | 0 | 26 | 0.06 | 0.82 | 0.81 | 30 | 0.04 | 0.84 | 0.82 |
| | | 0.3 | 26 | 0.05 | 0.81 | 0.80 | 30 | 0.04 | 0.83 | 0.82 |
| | | 0.6 | 28 | 0.06 | 0.82 | 0.82 | 30 | 0.04 | 0.82 | 0.81 |
| | | 0.9 | 28 | 0.06 | 0.80 | 0.80 | 32 | 0.05 | 0.83 | 0.82 |
| $\bar{m} = 50$ | $\rho = 0.02$ | 0 | 10 | 0.07 | 0.85 | 0.85 | 12 | 0.02 | 0.82 | 0.80 |
| | | 0.3 | 10 | 0.06 | 0.83 | 0.85 | 14 | 0.03 | 0.89 | 0.88 |
| | | 0.6 | 10 | 0.06 | 0.84 | 0.84 | 14 | 0.03 | 0.87 | 0.86 |
| | | 0.9 | 10 | 0.07 | 0.79 | 0.81 | 14 | 0.04 | 0.84 | 0.84 |
| | $\rho = 0.05$ | 0 | 12 | 0.06 | 0.82 | 0.82 | 16 | 0.04 | 0.88 | 0.85 |
| | | 0.3 | 12 | 0.07 | 0.83 | 0.82 | 16 | 0.03 | 0.86 | 0.85 |
| | | 0.6 | 12 | 0.07 | 0.81 | 0.81 | 16 | 0.03 | 0.85 | 0.84 |
| | | 0.9 | 14 | 0.07 | 0.84 | 0.85 | 16 | 0.04 | 0.83 | 0.82 |
| | $\rho = 0.10$ | 0 | 14 | 0.07 | 0.82 | 0.81 | 18 | 0.03 | 0.85 | 0.83 |
| | | 0.3 | 14 | 0.07 | 0.80 | 0.81 | 18 | 0.04 | 0.85 | 0.83 |
| | | 0.6 | 14 | 0.06 | 0.80 | 0.80 | 18 | 0.04 | 0.83 | 0.83 |
| | | 0.9 | 16 | 0.07 | 0.84 | 0.85 | 18 | 0.04 | 0.82 | 0.82 |
| $\bar{m} = 100$ | $\rho = 0.02$ | 0 | 6 | 0.09 | 0.85 | 0.85 | 10 | 0.02 | 0.91 | 0.88 |
| | | 0.3 | 6 | 0.08 | 0.84 | 0.84 | 10 | 0.02 | 0.91 | 0.88 |
| | | 0.6 | 6 | 0.08 | 0.83 | 0.83 | 10 | 0.03 | 0.88 | 0.87 |
| | | 0.9 | 6 | 0.08 | 0.78 | 0.81 | 10 | 0.03 | 0.86 | 0.85 |
| | $\rho = 0.05$ | 0 | 8 | 0.08 | 0.85 | 0.85 | 12 | 0.03 | 0.91 | 0.88 |
| | | 0.3 | 8 | 0.08 | 0.86 | 0.85 | 12 | 0.03 | 0.91 | 0.88 |
| | | 0.6 | 8 | 0.08 | 0.83 | 0.84 | 12 | 0.03 | 0.89 | 0.87 |
| | | 0.9 | 8 | 0.08 | 0.81 | 0.83 | 12 | 0.04 | 0.86 | 0.87 |
| | $\rho = 0.10$ | 0 | 10 | 0.08 | 0.88 | 0.87 | 12 | 0.04 | 0.86 | 0.82 |
| | | 0.3 | 10 | 0.08 | 0.88 | 0.87 | 12 | 0.03 | 0.85 | 0.82 |
| | | 0.6 | 10 | 0.08 | 0.86 | 0.87 | 12 | 0.03 | 0.84 | 0.82 |
| | | 0.9 | 10 | 0.08 | 0.84 | 0.87 | 12 | 0.03 | 0.81 | 0.82 |

**WEB TABLE 8** Required number of clusters $n_{E1}$, empirical type I error $\psi$, empirical power $\phi$, and predicted power $\hat{\phi}$ corresponding to the intersection-union test of the controlled effects with and without finite-sample correction. The controlled effect size of the cluster-level treatment is $\delta_2 = 0.25$ and the controlled effect size of the individual-level treatment is $\delta_3 = 0.15$. Notation: $\bar{m}$ is the mean cluster size, $\rho$ is the ICC, CV is the coefficient of variation of cluster sizes. The results were based on $5,000$ simulations.

| | | | $z$-based I-U test | | | | $t$-based I-U test | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | CV | $n_{E1}$ | $\psi$ | $\phi$ | $\hat{\phi}$ | $n_{E1}$ | $\psi$ | $\phi$ | $\hat{\phi}$ |
| $\bar{m} = 20$ | $\rho = 0.02$ | 0 | 140 | 0.04 | 0.80 | 0.80 | 142 | 0.03 | 0.82 | 0.80 |
| | | 0.3 | 142 | 0.03 | 0.81 | 0.81 | 142 | 0.03 | 0.81 | 0.80 |
| | | 0.6 | 142 | 0.04 | 0.80 | 0.81 | 144 | 0.04 | 0.82 | 0.81 |
| | | 0.9 | 142 | 0.04 | 0.80 | 0.80 | 144 | 0.04 | 0.80 | 0.80 |
| | $\rho = 0.05$ | 0 | 142 | 0.04 | 0.82 | 0.80 | 144 | 0.04 | 0.83 | 0.80 |
| | | 0.3 | 142 | 0.03 | 0.81 | 0.80 | 144 | 0.04 | 0.82 | 0.80 |
| | | 0.6 | 144 | 0.04 | 0.81 | 0.81 | 146 | 0.03 | 0.83 | 0.81 |
| | | 0.9 | 146 | 0.04 | 0.80 | 0.80 | 148 | 0.04 | 0.81 | 0.80 |
| | $\rho = 0.10$ | 0 | 148 | 0.04 | 0.82 | 0.81 | 148 | 0.03 | 0.81 | 0.80 |
| | | 0.3 | 148 | 0.04 | 0.82 | 0.80 | 150 | 0.03 | 0.83 | 0.80 |
| | | 0.6 | 150 | 0.04 | 0.82 | 0.80 | 152 | 0.04 | 0.81 | 0.80 |
| | | 0.9 | 156 | 0.04 | 0.82 | 0.81 | 156 | 0.04 | 0.81 | 0.80 |
| $\bar{m} = 50$ | $\rho = 0.02$ | 0 | 58 | 0.04 | 0.81 | 0.81 | 60 | 0.04 | 0.82 | 0.81 |
| | | 0.3 | 58 | 0.04 | 0.81 | 0.81 | 60 | 0.03 | 0.82 | 0.81 |
| | | 0.6 | 58 | 0.04 | 0.79 | 0.80 | 60 | 0.04 | 0.82 | 0.80 |
| | | 0.9 | 60 | 0.04 | 0.80 | 0.81 | 62 | 0.03 | 0.80 | 0.81 |
| | $\rho = 0.05$ | 0 | 64 | 0.04 | 0.81 | 0.81 | 66 | 0.04 | 0.82 | 0.81 |
| | | 0.3 | 64 | 0.04 | 0.81 | 0.81 | 66 | 0.04 | 0.82 | 0.81 |
| | | 0.6 | 66 | 0.05 | 0.81 | 0.81 | 68 | 0.04 | 0.82 | 0.81 |
| | | 0.9 | 68 | 0.04 | 0.80 | 0.81 | 70 | 0.04 | 0.82 | 0.81 |
| | $\rho = 0.10$ | 0 | 78 | 0.05 | 0.80 | 0.80 | 80 | 0.05 | 0.81 | 0.80 |
| | | 0.3 | 80 | 0.04 | 0.82 | 0.81 | 82 | 0.04 | 0.82 | 0.81 |
| | | 0.6 | 80 | 0.05 | 0.81 | 0.80 | 82 | 0.05 | 0.80 | 0.80 |
| | | 0.9 | 84 | 0.05 | 0.79 | 0.81 | 86 | 0.04 | 0.80 | 0.81 |
| $\bar{m} = 100$ | $\rho = 0.02$ | 0 | 32 | 0.04 | 0.82 | 0.82 | 34 | 0.04 | 0.84 | 0.82 |
| | | 0.3 | 32 | 0.04 | 0.83 | 0.82 | 34 | 0.04 | 0.83 | 0.82 |
| | | 0.6 | 32 | 0.04 | 0.81 | 0.81 | 34 | 0.04 | 0.82 | 0.81 |
| | | 0.9 | 34 | 0.05 | 0.80 | 0.82 | 36 | 0.04 | 0.82 | 0.82 |
| | $\rho = 0.05$ | 0 | 40 | 0.05 | 0.80 | 0.80 | 42 | 0.05 | 0.80 | 0.80 |
| | | 0.3 | 42 | 0.05 | 0.82 | 0.82 | 42 | 0.04 | 0.81 | 0.82 |
| | | 0.6 | 42 | 0.05 | 0.80 | 0.81 | 44 | 0.04 | 0.82 | 0.81 |
| | | 0.9 | 44 | 0.05 | 0.80 | 0.82 | 46 | 0.05 | 0.81 | 0.82 |
| | $\rho = 0.10$ | 0 | 62 | 0.06 | 0.82 | 0.81 | 64 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 62 | 0.05 | 0.81 | 0.81 | 64 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 62 | 0.05 | 0.80 | 0.80 | 64 | 0.05 | 0.80 | 0.80 |
| | | 0.9 | 64 | 0.06 | 0.78 | 0.80 | 66 | 0.04 | 0.79 | 0.80 |

**WEB TABLE 9** Required number of clusters $n_{E1}$, empirical type I error $\psi$, empirical power $\phi$, and predicted power $\hat{\phi}$ corresponding to the intersection-union test of the controlled effects with and without finite-sample correction. The controlled effect size of the cluster-level treatment is $\delta_2 = 0.5$ and the controlled effect size of the individual-level treatment is $\delta_3 = 0.25$. Notation: $\bar{m}$ is the mean cluster size, $\rho$ is the ICC, CV is the coefficient of variation of cluster sizes. The results were based on 5,000 simulations.

| | | | $z$-based I-U test | | | | $t$-based I-U test | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | CV | $n_{E1}$ | $\psi$ | $\phi$ | $\hat{\phi}$ | $n_{E1}$ | $\psi$ | $\phi$ | $\hat{\phi}$ |
| $\bar{m} = 20$ | $\rho = 0.02$ | 0 | 50 | 0.04 | 0.81 | 0.80 | 52 | 0.03 | 0.82 | 0.80 |
| | | 0.3 | 50 | 0.03 | 0.79 | 0.80 | 52 | 0.03 | 0.82 | 0.80 |
| | | 0.6 | 50 | 0.04 | 0.80 | 0.80 | 52 | 0.03 | 0.81 | 0.80 |
| | | 0.9 | 50 | 0.04 | 0.79 | 0.80 | 52 | 0.03 | 0.81 | 0.80 |
| | $\rho = 0.05$ | 0 | 50 | 0.04 | 0.82 | 0.81 | 52 | 0.03 | 0.83 | 0.81 |
| | | 0.3 | 50 | 0.04 | 0.80 | 0.81 | 52 | 0.04 | 0.83 | 0.81 |
| | | 0.6 | 50 | 0.04 | 0.81 | 0.81 | 52 | 0.03 | 0.82 | 0.81 |
| | | 0.9 | 50 | 0.04 | 0.80 | 0.81 | 52 | 0.03 | 0.82 | 0.81 |
| | $\rho = 0.10$ | 0 | 50 | 0.04 | 0.83 | 0.82 | 52 | 0.04 | 0.85 | 0.82 |
| | | 0.3 | 50 | 0.04 | 0.81 | 0.81 | 52 | 0.04 | 0.84 | 0.81 |
| | | 0.6 | 50 | 0.04 | 0.82 | 0.81 | 52 | 0.04 | 0.83 | 0.81 |
| | | 0.9 | 50 | 0.04 | 0.80 | 0.81 | 52 | 0.03 | 0.83 | 0.81 |
| $\bar{m} = 50$ | $\rho = 0.02$ | 0 | 20 | 0.04 | 0.80 | 0.80 | 24 | 0.03 | 0.87 | 0.84 |
| | | 0.3 | 20 | 0.05 | 0.80 | 0.80 | 24 | 0.03 | 0.86 | 0.84 |
| | | 0.6 | 22 | 0.04 | 0.83 | 0.84 | 24 | 0.03 | 0.86 | 0.84 |
| | | 0.9 | 22 | 0.04 | 0.80 | 0.84 | 24 | 0.04 | 0.84 | 0.83 |
| | $\rho = 0.05$ | 0 | 22 | 0.05 | 0.84 | 0.83 | 24 | 0.04 | 0.87 | 0.83 |
| | | 0.3 | 22 | 0.05 | 0.83 | 0.83 | 24 | 0.04 | 0.86 | 0.83 |
| | | 0.6 | 22 | 0.05 | 0.82 | 0.83 | 24 | 0.04 | 0.85 | 0.83 |
| | | 0.9 | 22 | 0.04 | 0.79 | 0.82 | 24 | 0.04 | 0.83 | 0.82 |
| | $\rho = 0.10$ | 0 | 24 | 0.05 | 0.83 | 0.82 | 26 | 0.05 | 0.85 | 0.83 |
| | | 0.3 | 24 | 0.05 | 0.81 | 0.82 | 26 | 0.04 | 0.84 | 0.82 |
| | | 0.6 | 24 | 0.05 | 0.81 | 0.81 | 26 | 0.04 | 0.84 | 0.82 |
| | | 0.9 | 24 | 0.06 | 0.77 | 0.80 | 26 | 0.04 | 0.80 | 0.80 |
| $\bar{m} = 100$ | $\rho = 0.02$ | 0 | 12 | 0.05 | 0.87 | 0.86 | 14 | 0.03 | 0.92 | 0.86 |
| | | 0.3 | 12 | 0.05 | 0.86 | 0.86 | 14 | 0.03 | 0.90 | 0.86 |
| | | 0.6 | 12 | 0.05 | 0.84 | 0.86 | 14 | 0.03 | 0.89 | 0.86 |
| | | 0.9 | 12 | 0.05 | 0.81 | 0.86 | 14 | 0.03 | 0.86 | 0.85 |
| | $\rho = 0.05$ | 0 | 12 | 0.06 | 0.81 | 0.81 | 14 | 0.04 | 0.85 | 0.81 |
| | | 0.3 | 12 | 0.06 | 0.80 | 0.81 | 14 | 0.04 | 0.82 | 0.80 |
| | | 0.6 | 14 | 0.06 | 0.85 | 0.87 | 16 | 0.04 | 0.87 | 0.87 |
| | | 0.9 | 14 | 0.06 | 0.81 | 0.86 | 16 | 0.05 | 0.84 | 0.86 |
| | $\rho = 0.10$ | 0 | 16 | 0.07 | 0.80 | 0.80 | 20 | 0.04 | 0.85 | 0.85 |
| | | 0.3 | 18 | 0.06 | 0.84 | 0.85 | 20 | 0.04 | 0.86 | 0.85 |
| | | 0.6 | 18 | 0.06 | 0.83 | 0.84 | 20 | 0.05 | 0.84 | 0.84 |
| | | 0.9 | 18 | 0.06 | 0.80 | 0.83 | 20 | 0.05 | 0.82 | 0.83 |

**WEB TABLE 10** Required number of clusters $n_{D2}$, empirical type I error $\psi$, empirical power $\phi$, and predicted power $\hat{\phi}$ corresponding to the joint test of the marginal effects with and without finite-sample correction. The marginal effect size of the cluster-level treatment is $\delta_X = 0.2$ and marginal effect size of the individual-level treatment is $\delta_Z = 0.1$. Notation: $\bar{m}$ is the mean cluster size, $\rho$ is the ICC, CV is the coefficient of variation of cluster sizes. The results were based on 5,000 simulations.

| | | | $\chi^2$-test | | | | mixed $\chi^2$- and $F$-test | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | CV | $n_{D2}$ | $\psi$ | $\phi$ | $\hat{\phi}$ | $n_{D2}$ | $\psi$ | $\phi$ | $\hat{\phi}$ |
| $\bar{m} = 20$ | $\rho = 0.02$ | 0 | 50 | 0.06 | 0.80 | 0.81 | 52 | 0.05 | 0.81 | 0.80 |
| | | 0.3 | 52 | 0.05 | 0.82 | 0.82 | 54 | 0.06 | 0.82 | 0.83 |
| | | 0.6 | 54 | 0.05 | 0.81 | 0.81 | 54 | 0.05 | 0.80 | 0.81 |
| | | 0.9 | 58 | 0.06 | 0.82 | 0.81 | 58 | 0.05 | 0.81 | 0.80 |
| | $\rho = 0.05$ | 0 | 64 | 0.05 | 0.82 | 0.81 | 66 | 0.04 | 0.81 | 0.81 |
| | | 0.3 | 64 | 0.06 | 0.80 | 0.80 | 66 | 0.05 | 0.80 | 0.80 |
| | | 0.6 | 68 | 0.06 | 0.80 | 0.81 | 68 | 0.05 | 0.80 | 0.80 |
| | | 0.9 | 74 | 0.06 | 0.83 | 0.81 | 76 | 0.05 | 0.82 | 0.82 |
| | $\rho = 0.10$ | 0 | 80 | 0.05 | 0.81 | 0.81 | 82 | 0.06 | 0.81 | 0.81 |
| | | 0.3 | 80 | 0.05 | 0.81 | 0.80 | 82 | 0.05 | 0.82 | 0.81 |
| | | 0.6 | 84 | 0.05 | 0.81 | 0.81 | 84 | 0.05 | 0.81 | 0.80 |
| | | 0.9 | 88 | 0.05 | 0.81 | 0.81 | 90 | 0.05 | 0.82 | 0.81 |
| $\bar{m} = 50$ | $\rho = 0.02$ | 0 | 26 | 0.06 | 0.81 | 0.81 | 28 | 0.04 | 0.82 | 0.82 |
| | | 0.3 | 26 | 0.06 | 0.81 | 0.80 | 28 | 0.05 | 0.82 | 0.81 |
| | | 0.6 | 28 | 0.06 | 0.81 | 0.81 | 30 | 0.05 | 0.82 | 0.82 |
| | | 0.9 | 30 | 0.07 | 0.81 | 0.81 | 32 | 0.05 | 0.82 | 0.82 |
| | $\rho = 0.05$ | 0 | 36 | 0.06 | 0.81 | 0.81 | 36 | 0.05 | 0.80 | 0.80 |
| | | 0.3 | 36 | 0.06 | 0.80 | 0.81 | 38 | 0.05 | 0.81 | 0.82 |
| | | 0.6 | 38 | 0.05 | 0.82 | 0.82 | 38 | 0.05 | 0.80 | 0.81 |
| | | 0.9 | 40 | 0.06 | 0.82 | 0.82 | 40 | 0.05 | 0.80 | 0.81 |
| | $\rho = 0.10$ | 0 | 44 | 0.06 | 0.83 | 0.80 | 44 | 0.05 | 0.82 | 0.80 |
| | | 0.3 | 44 | 0.06 | 0.80 | 0.80 | 44 | 0.06 | 0.79 | 0.80 |
| | | 0.6 | 46 | 0.06 | 0.82 | 0.82 | 46 | 0.05 | 0.80 | 0.81 |
| | | 0.9 | 46 | 0.06 | 0.80 | 0.81 | 48 | 0.05 | 0.82 | 0.81 |
| $\bar{m} = 100$ | $\rho = 0.02$ | 0 | 18 | 0.07 | 0.84 | 0.84 | 18 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 18 | 0.07 | 0.83 | 0.83 | 18 | 0.05 | 0.80 | 0.81 |
| | | 0.6 | 18 | 0.06 | 0.82 | 0.82 | 20 | 0.05 | 0.82 | 0.83 |
| | | 0.9 | 20 | 0.07 | 0.83 | 0.84 | 20 | 0.05 | 0.80 | 0.81 |
| | $\rho = 0.05$ | 0 | 24 | 0.06 | 0.84 | 0.83 | 24 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 24 | 0.06 | 0.83 | 0.83 | 24 | 0.05 | 0.80 | 0.80 |
| | | 0.6 | 24 | 0.07 | 0.80 | 0.82 | 26 | 0.05 | 0.83 | 0.83 |
| | | 0.9 | 24 | 0.06 | 0.80 | 0.81 | 26 | 0.05 | 0.80 | 0.83 |
| | $\rho = 0.10$ | 0 | 28 | 0.06 | 0.82 | 0.83 | 28 | 0.05 | 0.80 | 0.81 |
| | | 0.3 | 28 | 0.06 | 0.83 | 0.83 | 28 | 0.06 | 0.81 | 0.81 |
| | | 0.6 | 28 | 0.05 | 0.83 | 0.82 | 28 | 0.05 | 0.81 | 0.81 |
| | | 0.9 | 28 | 0.05 | 0.81 | 0.82 | 28 | 0.04 | 0.79 | 0.81 |

**WEB TABLE 11** Required number of clusters $n_{D2}$, empirical type I error $\psi$, empirical power $\phi$, and predicted power $\hat{\phi}$ corresponding to the joint test of the marginal effects with and without finite-sample correction. The marginal effect size of the cluster-level treatment is $\delta_X = 0.25$ and marginal effect size of the individual-level treatment is $\delta_Z = 0.15$. Notation: $\bar{m}$ is the mean cluster size, $\rho$ is the ICC, CV is the coefficient of variation of cluster sizes. The results were based on 5,000 simulations.

| | | | $\chi^2$-test | | | | mixed $\chi^2$- and $F$-test | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | CV | $n_{D2}$ | $\psi$ | $\phi$ | $\hat{\phi}$ | $n_{D2}$ | $\psi$ | $\phi$ | $\hat{\phi}$ |
| $\bar{m} = 20$ | $\rho = 0.02$ | 0 | 30 | 0.05 | 0.83 | 0.82 | 30 | 0.04 | 0.81 | 0.81 |
| | | 0.3 | 30 | 0.06 | 0.81 | 0.82 | 30 | 0.05 | 0.79 | 0.80 |
| | | 0.6 | 30 | 0.06 | 0.80 | 0.80 | 32 | 0.05 | 0.81 | 0.81 |
| | | 0.9 | 32 | 0.06 | 0.80 | 0.80 | 34 | 0.05 | 0.81 | 0.81 |
| | $\rho = 0.05$ | 0 | 36 | 0.05 | 0.82 | 0.81 | 36 | 0.05 | 0.81 | 0.80 |
| | | 0.3 | 36 | 0.06 | 0.81 | 0.81 | 38 | 0.04 | 0.81 | 0.82 |
| | | 0.6 | 38 | 0.06 | 0.81 | 0.81 | 40 | 0.05 | 0.82 | 0.82 |
| | | 0.9 | 40 | 0.06 | 0.83 | 0.81 | 42 | 0.06 | 0.82 | 0.81 |
| | $\rho = 0.10$ | 0 | 44 | 0.06 | 0.83 | 0.82 | 44 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 44 | 0.06 | 0.82 | 0.81 | 44 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 44 | 0.05 | 0.82 | 0.80 | 46 | 0.05 | 0.81 | 0.81 |
| | | 0.9 | 46 | 0.05 | 0.82 | 0.80 | 48 | 0.05 | 0.82 | 0.81 |
| $\bar{m} = 50$ | $\rho = 0.02$ | 0 | 16 | 0.07 | 0.86 | 0.85 | 16 | 0.05 | 0.83 | 0.82 |
| | | 0.3 | 16 | 0.06 | 0.85 | 0.84 | 16 | 0.05 | 0.82 | 0.81 |
| | | 0.6 | 16 | 0.06 | 0.82 | 0.83 | 18 | 0.05 | 0.84 | 0.84 |
| | | 0.9 | 18 | 0.08 | 0.83 | 0.85 | 18 | 0.06 | 0.80 | 0.82 |
| | $\rho = 0.05$ | 0 | 20 | 0.06 | 0.84 | 0.83 | 22 | 0.05 | 0.85 | 0.84 |
| | | 0.3 | 20 | 0.06 | 0.83 | 0.83 | 22 | 0.05 | 0.84 | 0.84 |
| | | 0.6 | 20 | 0.06 | 0.81 | 0.82 | 22 | 0.05 | 0.83 | 0.84 |
| | | 0.9 | 20 | 0.07 | 0.80 | 0.80 | 22 | 0.05 | 0.80 | 0.81 |
| | $\rho = 0.10$ | 0 | 22 | 0.06 | 0.81 | 0.80 | 24 | 0.05 | 0.82 | 0.82 |
| | | 0.3 | 22 | 0.06 | 0.81 | 0.80 | 24 | 0.05 | 0.81 | 0.82 |
| | | 0.6 | 24 | 0.06 | 0.84 | 0.83 | 24 | 0.05 | 0.82 | 0.81 |
| | | 0.9 | 24 | 0.06 | 0.81 | 0.82 | 24 | 0.05 | 0.78 | 0.81 |
| $\bar{m} = 100$ | $\rho = 0.02$ | 0 | 10 | 0.08 | 0.85 | 0.85 | 10 | 0.05 | 0.79 | 0.80 |
| | | 0.3 | 10 | 0.08 | 0.84 | 0.85 | 10 | 0.06 | 0.79 | 0.80 |
| | | 0.6 | 10 | 0.08 | 0.83 | 0.84 | 12 | 0.05 | 0.85 | 0.86 |
| | | 0.9 | 10 | 0.09 | 0.79 | 0.82 | 12 | 0.06 | 0.81 | 0.84 |
| | $\rho = 0.05$ | 0 | 12 | 0.08 | 0.83 | 0.82 | 14 | 0.05 | 0.86 | 0.85 |
| | | 0.3 | 12 | 0.07 | 0.83 | 0.82 | 14 | 0.05 | 0.85 | 0.85 |
| | | 0.6 | 12 | 0.07 | 0.81 | 0.82 | 14 | 0.05 | 0.83 | 0.85 |
| | | 0.9 | 12 | 0.08 | 0.78 | 0.81 | 14 | 0.06 | 0.81 | 0.84 |
| | $\rho = 0.10$ | 0 | 14 | 0.07 | 0.86 | 0.84 | 14 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 14 | 0.07 | 0.85 | 0.84 | 14 | 0.05 | 0.81 | 0.80 |
| | | 0.6 | 14 | 0.07 | 0.83 | 0.84 | 14 | 0.05 | 0.79 | 0.81 |
| | | 0.9 | 14 | 0.07 | 0.81 | 0.84 | 14 | 0.05 | 0.77 | 0.80 |

**WEB TABLE 12** Required number of clusters $n_{E2}$, empirical type I error $\psi$, empirical power $\phi$, and predicted power $\hat{\phi}$ corresponding to the intersection-union test of the marginal effects with and without finite-sample correction. The marginal effect size of the cluster-level treatment is $\delta_X = 0.2$ and the marginal effect size of the individual-level treatment is $\delta_Z = 0.1$. Notation: $\bar{m}$ is the mean cluster size, $\rho$ is the ICC, CV is the coefficient of variation of cluster sizes. The results were based on 5,000 simulations.

| | | | $z$-based I-U test | | | | $t$- and $z$-based I-U test | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | CV | $n_{E2}$ | $\psi$ | $\phi$ | $\hat{\phi}$ | $n_{E2}$ | $\psi$ | $\phi$ | $\hat{\phi}$ |
| $\bar{m} = 20$ | $\rho = 0.02$ | 0 | 158 | 0.05 | 0.80 | 0.80 | 158 | 0.04 | 0.80 | 0.80 |
| | | 0.3 | 158 | 0.04 | 0.81 | 0.80 | 158 | 0.04 | 0.81 | 0.80 |
| | | 0.6 | 158 | 0.04 | 0.81 | 0.80 | 158 | 0.04 | 0.81 | 0.80 |
| | | 0.9 | 160 | 0.04 | 0.81 | 0.80 | 160 | 0.04 | 0.81 | 0.80 |
| | $\rho = 0.05$ | 0 | 160 | 0.04 | 0.81 | 0.80 | 160 | 0.04 | 0.81 | 0.80 |
| | | 0.3 | 160 | 0.04 | 0.81 | 0.80 | 162 | 0.04 | 0.82 | 0.81 |
| | | 0.6 | 162 | 0.04 | 0.80 | 0.80 | 164 | 0.04 | 0.82 | 0.81 |
| | | 0.9 | 168 | 0.04 | 0.82 | 0.80 | 168 | 0.04 | 0.82 | 0.80 |
| | $\rho = 0.10$ | 0 | 172 | 0.05 | 0.81 | 0.80 | 174 | 0.05 | 0.82 | 0.80 |
| | | 0.3 | 174 | 0.04 | 0.81 | 0.80 | 174 | 0.04 | 0.81 | 0.80 |
| | | 0.6 | 178 | 0.05 | 0.81 | 0.80 | 180 | 0.04 | 0.81 | 0.81 |
| | | 0.9 | 186 | 0.04 | 0.82 | 0.80 | 188 | 0.05 | 0.81 | 0.80 |
| $\bar{m} = 50$ | $\rho = 0.02$ | 0 | 66 | 0.05 | 0.80 | 0.81 | 66 | 0.04 | 0.80 | 0.81 |
| | | 0.3 | 66 | 0.04 | 0.80 | 0.81 | 66 | 0.04 | 0.80 | 0.81 |
| | | 0.6 | 66 | 0.04 | 0.80 | 0.80 | 68 | 0.04 | 0.81 | 0.81 |
| | | 0.9 | 68 | 0.04 | 0.81 | 0.80 | 70 | 0.04 | 0.81 | 0.81 |
| | $\rho = 0.05$ | 0 | 76 | 0.05 | 0.81 | 0.80 | 78 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 78 | 0.05 | 0.82 | 0.81 | 78 | 0.04 | 0.82 | 0.81 |
| | | 0.6 | 80 | 0.04 | 0.81 | 0.81 | 80 | 0.04 | 0.81 | 0.81 |
| | | 0.9 | 84 | 0.05 | 0.81 | 0.81 | 84 | 0.04 | 0.80 | 0.81 |
| | $\rho = 0.10$ | 0 | 102 | 0.05 | 0.81 | 0.80 | 104 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 102 | 0.05 | 0.80 | 0.80 | 104 | 0.04 | 0.80 | 0.80 |
| | | 0.6 | 106 | 0.05 | 0.80 | 0.81 | 108 | 0.05 | 0.82 | 0.81 |
| | | 0.9 | 110 | 0.05 | 0.78 | 0.80 | 112 | 0.06 | 0.79 | 0.80 |
| $\bar{m} = 100$ | $\rho = 0.02$ | 0 | 36 | 0.06 | 0.80 | 0.80 | 38 | 0.04 | 0.82 | 0.82 |
| | | 0.3 | 38 | 0.05 | 0.82 | 0.82 | 38 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 38 | 0.05 | 0.81 | 0.81 | 40 | 0.04 | 0.83 | 0.83 |
| | | 0.9 | 40 | 0.05 | 0.81 | 0.81 | 42 | 0.05 | 0.81 | 0.83 |
| | $\rho = 0.05$ | 0 | 52 | 0.05 | 0.80 | 0.80 | 54 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 52 | 0.05 | 0.80 | 0.80 | 54 | 0.04 | 0.81 | 0.81 |
| | | 0.6 | 54 | 0.06 | 0.80 | 0.81 | 56 | 0.05 | 0.81 | 0.81 |
| | | 0.9 | 58 | 0.06 | 0.80 | 0.82 | 58 | 0.05 | 0.79 | 0.80 |
| | $\rho = 0.10$ | 0 | 86 | 0.06 | 0.79 | 0.80 | 88 | 0.05 | 0.80 | 0.80 |
| | | 0.3 | 88 | 0.05 | 0.81 | 0.81 | 90 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 90 | 0.05 | 0.81 | 0.81 | 92 | 0.04 | 0.80 | 0.81 |
| | | 0.9 | 92 | 0.06 | 0.78 | 0.80 | 94 | 0.06 | 0.79 | 0.80 |

**WEB TABLE 13** Required number of clusters $n_{E2}$, empirical type I error $\psi$, empirical power $\phi$, and predicted power $\hat{\phi}$ corresponding to the intersection-union test of the marginal effects with and without finite-sample correction. The marginal effect size of the cluster-level treatment is $\delta_X = 0.4$ and the marginal effect size of the individual-level treatment is $\delta_Z = 0.2$. Notation: $\bar{m}$ is the mean cluster size, $\rho$ is the ICC, CV is the coefficient of variation of cluster sizes. The results were based on 5,000 simulations.

| | | | $z$-based I-U test | | | | $t$- and $z$-based I-U test | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | CV | $n_{E2}$ | $\psi$ | $\phi$ | $\hat{\phi}$ | $n_{E2}$ | $\psi$ | $\phi$ | $\hat{\phi}$ |
| $\bar{m} = 20$ | $\rho = 0.02$ | 0 | 40 | 0.05 | 0.80 | 0.81 | 40 | 0.04 | 0.80 | 0.81 |
| | | 0.3 | 40 | 0.05 | 0.80 | 0.81 | 40 | 0.04 | 0.80 | 0.81 |
| | | 0.6 | 40 | 0.05 | 0.80 | 0.81 | 40 | 0.04 | 0.80 | 0.81 |
| | | 0.9 | 40 | 0.04 | 0.80 | 0.80 | 40 | 0.04 | 0.80 | 0.80 |
| | $\rho = 0.05$ | 0 | 40 | 0.05 | 0.80 | 0.80 | 42 | 0.05 | 0.83 | 0.82 |
| | | 0.3 | 40 | 0.05 | 0.80 | 0.80 | 42 | 0.04 | 0.82 | 0.82 |
| | | 0.6 | 42 | 0.05 | 0.82 | 0.82 | 42 | 0.04 | 0.82 | 0.81 |
| | | 0.9 | 42 | 0.05 | 0.80 | 0.80 | 44 | 0.05 | 0.82 | 0.82 |
| | $\rho = 0.10$ | 0 | 44 | 0.05 | 0.82 | 0.81 | 44 | 0.04 | 0.82 | 0.80 |
| | | 0.3 | 44 | 0.05 | 0.82 | 0.81 | 46 | 0.05 | 0.83 | 0.82 |
| | | 0.6 | 46 | 0.05 | 0.82 | 0.82 | 46 | 0.04 | 0.82 | 0.81 |
| | | 0.9 | 48 | 0.05 | 0.81 | 0.82 | 48 | 0.05 | 0.81 | 0.81 |
| $\bar{m} = 50$ | $\rho = 0.02$ | 0 | 18 | 0.05 | 0.84 | 0.85 | 18 | 0.04 | 0.84 | 0.84 |
| | | 0.3 | 18 | 0.06 | 0.84 | 0.84 | 18 | 0.04 | 0.83 | 0.83 |
| | | 0.6 | 18 | 0.06 | 0.82 | 0.84 | 18 | 0.04 | 0.81 | 0.83 |
| | | 0.9 | 18 | 0.07 | 0.79 | 0.83 | 18 | 0.05 | 0.77 | 0.81 |
| | $\rho = 0.05$ | 0 | 20 | 0.05 | 0.84 | 0.83 | 20 | 0.04 | 0.81 | 0.80 |
| | | 0.3 | 20 | 0.07 | 0.82 | 0.83 | 22 | 0.04 | 0.85 | 0.85 |
| | | 0.6 | 20 | 0.07 | 0.80 | 0.81 | 22 | 0.05 | 0.83 | 0.83 |
| | | 0.9 | 22 | 0.06 | 0.81 | 0.83 | 22 | 0.05 | 0.79 | 0.81 |
| | $\rho = 0.10$ | 0 | 26 | 0.06 | 0.81 | 0.81 | 28 | 0.04 | 0.82 | 0.82 |
| | | 0.3 | 26 | 0.06 | 0.81 | 0.81 | 28 | 0.05 | 0.82 | 0.82 |
| | | 0.6 | 28 | 0.06 | 0.83 | 0.83 | 28 | 0.05 | 0.81 | 0.80 |
| | | 0.9 | 28 | 0.06 | 0.79 | 0.81 | 30 | 0.05 | 0.81 | 0.81 |
| $\bar{m} = 100$ | $\rho = 0.02$ | 0 | 10 | 0.07 | 0.84 | 0.85 | 12 | 0.04 | 0.89 | 0.89 |
| | | 0.3 | 10 | 0.08 | 0.84 | 0.85 | 12 | 0.04 | 0.88 | 0.89 |
| | | 0.6 | 10 | 0.07 | 0.80 | 0.84 | 12 | 0.05 | 0.85 | 0.87 |
| | | 0.9 | 10 | 0.08 | 0.74 | 0.81 | 12 | 0.05 | 0.81 | 0.85 |
| | $\rho = 0.05$ | 0 | 14 | 0.07 | 0.84 | 0.84 | 16 | 0.05 | 0.84 | 0.85 |
| | | 0.3 | 14 | 0.07 | 0.82 | 0.84 | 16 | 0.05 | 0.84 | 0.84 |
| | | 0.6 | 14 | 0.07 | 0.80 | 0.82 | 16 | 0.05 | 0.82 | 0.83 |
| | | 0.9 | 16 | 0.07 | 0.82 | 0.86 | 16 | 0.06 | 0.77 | 0.81 |
| | $\rho = 0.10$ | 0 | 22 | 0.06 | 0.80 | 0.81 | 24 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 22 | 0.06 | 0.80 | 0.81 | 24 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 24 | 0.07 | 0.82 | 0.83 | 26 | 0.05 | 0.83 | 0.83 |
| | | 0.9 | 24 | 0.07 | 0.80 | 0.82 | 26 | 0.05 | 0.80 | 0.82 |

## WEB APPENDIX H: A SIMULATION STUDY ABOUT OTHER CLUSTER SIZE DISTRIBUTIONS

In this appendix, we assess the robustness of the simulation results for our proposed formulas in Section 3 under two additional cases of cluster size distributions. Given the values of $\bar{m}$ and CV, we simulated varying cluster sizes using: (H.1) $m_i \sim \mathcal{N}(\bar{m}, \bar{m}\text{CV})$, (H.2) $m_i \sim \text{Uniform}(\bar{m} - c, \bar{m} + c)$, where $c = \sqrt{3}\bar{m}\text{CV}$. In both cases, each simulated cluster size $m_i$ was rounded to the nearest integer and ensured to be at least 2 for computational stability. The results for the normal-distribution case were provided from Web Table 14 to 20, while the results for the uniform-distribution case were provided from Web Table 21 to 27. In general, as we pointed out in the main paper, the simulation results were insensitive to other cluster size distributions that we considered.

□

**WEB TABLE 14** Required number of clusters $n_{A1}$, empirical type I error $\psi$, empirical power $\phi$, and predicted power $\hat{\phi}$ corresponding to the test for the controlled effect of the cluster-level treatment with and without finite-sample correction. The controlled effect size of the cluster-level treatment is $\delta_2 = 0.2$. The unequal cluster sizes approximately follow a normal distribution. Notation: $\bar{m}$ is the mean cluster size, $\rho$ is the ICC, CV is the coefficient of variation of cluster sizes. The results were based on 5,000 simulations.

| | | | $z$-test | | | | $t$-test | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | CV | $n_{A1}$ | $\psi$ | $\phi$ | $\hat{\phi}$ | $n_{A1}$ | $\psi$ | $\phi$ | $\hat{\phi}$ |
| $\bar{m} = 20$ | $\rho = 0.02$ | 0 | 94 | 0.05 | 0.81 | 0.80 | 96 | 0.05 | 0.80 | 0.80 |
| | | 0.3 | 96 | 0.05 | 0.80 | 0.81 | 98 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 98 | 0.05 | 0.80 | 0.80 | 100 | 0.05 | 0.81 | 0.80 |
| | | 0.9 | 104 | 0.05 | 0.83 | 0.80 | 106 | 0.05 | 0.84 | 0.80 |
| | $\rho = 0.05$ | 0 | 116 | 0.05 | 0.81 | 0.80 | 118 | 0.05 | 0.80 | 0.80 |
| | | 0.3 | 118 | 0.05 | 0.81 | 0.80 | 120 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 124 | 0.05 | 0.82 | 0.81 | 126 | 0.05 | 0.82 | 0.81 |
| | | 0.9 | 136 | 0.05 | 0.84 | 0.81 | 136 | 0.05 | 0.84 | 0.80 |
| | $\rho = 0.10$ | 0 | 152 | 0.05 | 0.81 | 0.80 | 154 | 0.05 | 0.81 | 0.80 |
| | | 0.3 | 154 | 0.05 | 0.81 | 0.80 | 156 | 0.05 | 0.81 | 0.80 |
| | | 0.6 | 160 | 0.05 | 0.80 | 0.80 | 162 | 0.05 | 0.81 | 0.80 |
| | | 0.9 | 176 | 0.05 | 0.83 | 0.80 | 176 | 0.05 | 0.82 | 0.80 |
| $\bar{m} = 50$ | $\rho = 0.02$ | 0 | 48 | 0.06 | 0.81 | 0.81 | 50 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 48 | 0.06 | 0.80 | 0.81 | 50 | 0.05 | 0.80 | 0.81 |
| | | 0.6 | 50 | 0.05 | 0.81 | 0.80 | 52 | 0.05 | 0.80 | 0.80 |
| | | 0.9 | 56 | 0.05 | 0.83 | 0.81 | 58 | 0.05 | 0.84 | 0.81 |
| | $\rho = 0.05$ | 0 | 70 | 0.05 | 0.81 | 0.80 | 72 | 0.05 | 0.81 | 0.80 |
| | | 0.3 | 72 | 0.05 | 0.81 | 0.81 | 74 | 0.06 | 0.81 | 0.81 |
| | | 0.6 | 74 | 0.05 | 0.80 | 0.80 | 76 | 0.05 | 0.81 | 0.80 |
| | | 0.9 | 80 | 0.06 | 0.80 | 0.80 | 82 | 0.04 | 0.81 | 0.80 |
| | $\rho = 0.10$ | 0 | 108 | 0.05 | 0.80 | 0.80 | 110 | 0.04 | 0.82 | 0.80 |
| | | 0.3 | 110 | 0.05 | 0.80 | 0.81 | 112 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 112 | 0.05 | 0.78 | 0.80 | 114 | 0.05 | 0.79 | 0.80 |
| | | 0.9 | 118 | 0.05 | 0.79 | 0.80 | 120 | 0.05 | 0.79 | 0.80 |
| $\bar{m} = 100$ | $\rho = 0.02$ | 0 | 32 | 0.05 | 0.80 | 0.81 | 34 | 0.04 | 0.82 | 0.81 |
| | | 0.3 | 32 | 0.05 | 0.80 | 0.80 | 34 | 0.05 | 0.80 | 0.80 |
| | | 0.6 | 34 | 0.06 | 0.80 | 0.81 | 36 | 0.05 | 0.81 | 0.81 |
| | | 0.9 | 38 | 0.06 | 0.82 | 0.82 | 40 | 0.05 | 0.83 | 0.82 |
| | $\rho = 0.05$ | 0 | 56 | 0.05 | 0.81 | 0.81 | 58 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 56 | 0.05 | 0.81 | 0.81 | 58 | 0.05 | 0.82 | 0.81 |
| | | 0.6 | 58 | 0.06 | 0.79 | 0.81 | 60 | 0.05 | 0.80 | 0.81 |
| | | 0.9 | 60 | 0.06 | 0.78 | 0.80 | 62 | 0.05 | 0.79 | 0.80 |
| | $\rho = 0.10$ | 0 | 94 | 0.05 | 0.81 | 0.81 | 96 | 0.06 | 0.80 | 0.81 |
| | | 0.3 | 94 | 0.06 | 0.79 | 0.80 | 96 | 0.05 | 0.80 | 0.80 |
| | | 0.6 | 96 | 0.05 | 0.79 | 0.80 | 98 | 0.05 | 0.80 | 0.80 |
| | | 0.9 | 100 | 0.06 | 0.77 | 0.81 | 102 | 0.05 | 0.78 | 0.81 |

**WEB TABLE 15** Required number of clusters $n_{A1}$, empirical type I error $\psi$, empirical power $\phi$, and predicted power $\hat{\phi}$ corresponding to the test for the controlled effect of the cluster-level treatment with and without finite-sample correction. The controlled effect size of the cluster-level treatment is $\delta_2 = 0.4$. The unequal cluster sizes approximately follow a normal distribution. Notation: $\bar{m}$ is the mean cluster size, $\rho$ is the ICC, CV is the coefficient of variation of cluster sizes. The results were based on 5,000 simulations.

| | | | $z$-test | | | | $t$-test | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | CV | $n_{A1}$ | $\psi$ | $\phi$ | $\hat{\phi}$ | $n_{A1}$ | $\psi$ | $\phi$ | $\hat{\phi}$ |
| $\bar{m} = 20$ | $\rho = 0.02$ | 0 | 24 | 0.05 | 0.81 | 0.81 | 26 | 0.04 | 0.82 | 0.81 |
| | | 0.3 | 24 | 0.06 | 0.80 | 0.81 | 26 | 0.04 | 0.81 | 0.81 |
| | | 0.6 | 26 | 0.05 | 0.82 | 0.82 | 28 | 0.04 | 0.83 | 0.82 |
| | | 0.9 | 26 | 0.05 | 0.82 | 0.80 | 30 | 0.04 | 0.86 | 0.83 |
| | $\rho = 0.05$ | 0 | 30 | 0.05 | 0.83 | 0.82 | 32 | 0.04 | 0.83 | 0.82 |
| | | 0.3 | 30 | 0.06 | 0.82 | 0.81 | 32 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 32 | 0.07 | 0.82 | 0.82 | 34 | 0.05 | 0.82 | 0.82 |
| | | 0.9 | 34 | 0.06 | 0.83 | 0.81 | 36 | 0.05 | 0.84 | 0.81 |
| | $\rho = 0.10$ | 0 | 38 | 0.05 | 0.81 | 0.80 | 40 | 0.05 | 0.80 | 0.80 |
| | | 0.3 | 40 | 0.05 | 0.82 | 0.82 | 42 | 0.05 | 0.81 | 0.82 |
| | | 0.6 | 40 | 0.05 | 0.81 | 0.80 | 42 | 0.05 | 0.81 | 0.80 |
| | | 0.9 | 44 | 0.06 | 0.82 | 0.80 | 46 | 0.05 | 0.82 | 0.80 |
| $\bar{m} = 50$ | $\rho = 0.02$ | 0 | 12 | 0.06 | 0.80 | 0.81 | 14 | 0.04 | 0.81 | 0.80 |
| | | 0.3 | 12 | 0.05 | 0.80 | 0.81 | 16 | 0.04 | 0.85 | 0.86 |
| | | 0.6 | 14 | 0.06 | 0.84 | 0.84 | 16 | 0.04 | 0.84 | 0.84 |
| | | 0.9 | 14 | 0.06 | 0.81 | 0.81 | 16 | 0.04 | 0.83 | 0.81 |
| | $\rho = 0.05$ | 0 | 18 | 0.06 | 0.82 | 0.81 | 20 | 0.04 | 0.81 | 0.81 |
| | | 0.3 | 18 | 0.07 | 0.80 | 0.81 | 20 | 0.04 | 0.80 | 0.81 |
| | | 0.6 | 20 | 0.06 | 0.83 | 0.83 | 22 | 0.05 | 0.83 | 0.83 |
| | | 0.9 | 20 | 0.07 | 0.80 | 0.80 | 24 | 0.05 | 0.84 | 0.84 |
| | $\rho = 0.10$ | 0 | 28 | 0.06 | 0.82 | 0.82 | 30 | 0.06 | 0.81 | 0.82 |
| | | 0.3 | 28 | 0.05 | 0.81 | 0.81 | 30 | 0.05 | 0.82 | 0.81 |
| | | 0.6 | 28 | 0.06 | 0.79 | 0.80 | 30 | 0.05 | 0.79 | 0.80 |
| | | 0.9 | 30 | 0.06 | 0.79 | 0.81 | 32 | 0.05 | 0.79 | 0.81 |
| $\bar{m} = 100$ | $\rho = 0.02$ | 0 | 8 | 0.07 | 0.80 | 0.81 | 12 | 0.04 | 0.90 | 0.88 |
| | | 0.3 | 8 | 0.07 | 0.79 | 0.80 | 12 | 0.04 | 0.87 | 0.87 |
| | | 0.6 | 10 | 0.07 | 0.84 | 0.87 | 12 | 0.04 | 0.85 | 0.86 |
| | | 0.9 | 10 | 0.08 | 0.82 | 0.84 | 12 | 0.05 | 0.83 | 0.83 |
| | $\rho = 0.05$ | 0 | 14 | 0.07 | 0.81 | 0.81 | 16 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 14 | 0.07 | 0.80 | 0.81 | 16 | 0.04 | 0.81 | 0.80 |
| | | 0.6 | 16 | 0.06 | 0.83 | 0.85 | 18 | 0.05 | 0.83 | 0.84 |
| | | 0.9 | 16 | 0.07 | 0.79 | 0.83 | 18 | 0.05 | 0.80 | 0.82 |
| | $\rho = 0.10$ | 0 | 24 | 0.07 | 0.82 | 0.81 | 26 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 24 | 0.06 | 0.81 | 0.81 | 26 | 0.05 | 0.80 | 0.81 |
| | | 0.6 | 24 | 0.06 | 0.78 | 0.80 | 26 | 0.05 | 0.79 | 0.80 |
| | | 0.9 | 26 | 0.06 | 0.79 | 0.82 | 28 | 0.05 | 0.78 | 0.82 |

**WEB TABLE 16** Required number of clusters $n_{A2}$, empirical type I error $\psi$, empirical power $\phi$, and predicted power $\hat{\phi}$ corresponding to the test for the controlled effect of the individual-level treatment. The unequal cluster sizes approximately follow a normal distribution. Notation: $\delta_3$ is the controlled effect size of the individual-level treatment, $\bar{m}$ is the mean cluster size, $\rho$ is the ICC, CV is the coefficient of variation of cluster sizes. The results were based on 5,000 simulations.

| | | | $\delta_3 = 0.15$ | | | | $\delta_3 = 0.3$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | CV | $n_{A2}$ | $\psi$ | $\phi$ | $\hat{\phi}$ | $n_{A2}$ | $\psi$ | $\phi$ | $\hat{\phi}$ |
| $\bar{m} = 20$ | $\rho = 0.02$ | 0 | 140 | 0.05 | 0.81 | 0.80 | 36 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 140 | 0.05 | 0.81 | 0.80 | 36 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 140 | 0.05 | 0.82 | 0.80 | 36 | 0.05 | 0.83 | 0.81 |
| | | 0.9 | 140 | 0.05 | 0.83 | 0.80 | 36 | 0.06 | 0.83 | 0.82 |
| | $\rho = 0.05$ | 0 | 138 | 0.06 | 0.81 | 0.81 | 36 | 0.05 | 0.84 | 0.82 |
| | | 0.3 | 136 | 0.06 | 0.81 | 0.80 | 34 | 0.05 | 0.81 | 0.80 |
| | | 0.6 | 136 | 0.05 | 0.82 | 0.80 | 34 | 0.05 | 0.81 | 0.80 |
| | | 0.9 | 136 | 0.05 | 0.83 | 0.80 | 34 | 0.05 | 0.82 | 0.80 |
| | $\rho = 0.10$ | 0 | 132 | 0.05 | 0.81 | 0.81 | 34 | 0.05 | 0.83 | 0.82 |
| | | 0.3 | 130 | 0.04 | 0.82 | 0.80 | 34 | 0.05 | 0.83 | 0.82 |
| | | 0.6 | 130 | 0.05 | 0.82 | 0.80 | 34 | 0.05 | 0.83 | 0.82 |
| | | 0.9 | 130 | 0.05 | 0.84 | 0.80 | 34 | 0.05 | 0.84 | 0.82 |
| $\bar{m} = 50$ | $\rho = 0.02$ | 0 | 56 | 0.05 | 0.81 | 0.81 | 14 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 56 | 0.05 | 0.81 | 0.81 | 14 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 56 | 0.05 | 0.81 | 0.81 | 14 | 0.05 | 0.80 | 0.81 |
| | | 0.9 | 56 | 0.05 | 0.82 | 0.81 | 14 | 0.06 | 0.80 | 0.81 |
| | $\rho = 0.05$ | 0 | 54 | 0.05 | 0.82 | 0.80 | 14 | 0.05 | 0.83 | 0.82 |
| | | 0.3 | 54 | 0.05 | 0.80 | 0.80 | 14 | 0.05 | 0.82 | 0.82 |
| | | 0.6 | 54 | 0.05 | 0.82 | 0.80 | 14 | 0.05 | 0.81 | 0.82 |
| | | 0.9 | 54 | 0.05 | 0.82 | 0.80 | 14 | 0.06 | 0.81 | 0.82 |
| | $\rho = 0.10$ | 0 | 52 | 0.05 | 0.82 | 0.81 | 14 | 0.05 | 0.85 | 0.83 |
| | | 0.3 | 52 | 0.05 | 0.82 | 0.81 | 14 | 0.05 | 0.84 | 0.83 |
| | | 0.6 | 52 | 0.05 | 0.82 | 0.81 | 14 | 0.05 | 0.83 | 0.83 |
| | | 0.9 | 52 | 0.05 | 0.82 | 0.81 | 14 | 0.06 | 0.83 | 0.84 |
| $\bar{m} = 100$ | $\rho = 0.02$ | 0 | 28 | 0.06 | 0.81 | 0.81 | 8 | 0.05 | 0.85 | 0.86 |
| | | 0.3 | 28 | 0.05 | 0.81 | 0.81 | 8 | 0.05 | 0.84 | 0.86 |
| | | 0.6 | 28 | 0.05 | 0.80 | 0.81 | 8 | 0.05 | 0.83 | 0.86 |
| | | 0.9 | 28 | 0.05 | 0.82 | 0.81 | 8 | 0.05 | 0.83 | 0.86 |
| | $\rho = 0.05$ | 0 | 28 | 0.06 | 0.82 | 0.82 | 8 | 0.05 | 0.86 | 0.87 |
| | | 0.3 | 28 | 0.05 | 0.82 | 0.82 | 8 | 0.05 | 0.85 | 0.87 |
| | | 0.6 | 28 | 0.05 | 0.82 | 0.82 | 8 | 0.05 | 0.84 | 0.87 |
| | | 0.9 | 28 | 0.05 | 0.84 | 0.82 | 8 | 0.05 | 0.84 | 0.87 |
| | $\rho = 0.10$ | 0 | 26 | 0.05 | 0.81 | 0.81 | 8 | 0.05 | 0.88 | 0.88 |
| | | 0.3 | 26 | 0.05 | 0.81 | 0.81 | 8 | 0.05 | 0.87 | 0.88 |
| | | 0.6 | 26 | 0.05 | 0.81 | 0.81 | 8 | 0.05 | 0.86 | 0.88 |
| | | 0.9 | 26 | 0.05 | 0.84 | 0.81 | 8 | 0.05 | 0.86 | 0.88 |

**WEB TABLE 17** Required number of clusters $n_{B1}$, empirical type I error $\psi$, empirical power $\phi$, and predicted power $\hat{\phi}$ corresponding to the test for the marginal effect of the cluster-level treatment with and without finite-sample correction. The marginal effect size of the cluster-level treatment is $\delta_X = 0.2$. The unequal cluster sizes approximately follow a normal distribution. Notation: $\bar{m}$ is the mean cluster size, $\rho$ is the ICC, CV is the coefficient of variation of cluster sizes. The results were based on 5,000 simulations.

| | | | $z$-test | | | | $t$-test | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | CV | $n_{B1}$ | $\psi$ | $\phi$ | $\hat{\phi}$ | $n_{B1}$ | $\psi$ | $\phi$ | $\hat{\phi}$ |
| $\bar{m} = 20$ | $\rho = 0.02$ | 0 | 56 | 0.05 | 0.81 | 0.81 | 58 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 56 | 0.06 | 0.80 | 0.81 | 58 | 0.05 | 0.80 | 0.81 |
| | | 0.6 | 60 | 0.06 | 0.80 | 0.81 | 62 | 0.05 | 0.82 | 0.81 |
| | | 0.9 | 66 | 0.05 | 0.85 | 0.81 | 68 | 0.05 | 0.84 | 0.81 |
| | $\rho = 0.05$ | 0 | 78 | 0.05 | 0.80 | 0.81 | 80 | 0.05 | 0.80 | 0.81 |
| | | 0.3 | 80 | 0.05 | 0.81 | 0.81 | 82 | 0.05 | 0.80 | 0.81 |
| | | 0.6 | 86 | 0.05 | 0.81 | 0.81 | 88 | 0.05 | 0.81 | 0.81 |
| | | 0.9 | 96 | 0.05 | 0.84 | 0.80 | 98 | 0.05 | 0.83 | 0.80 |
| | $\rho = 0.10$ | 0 | 114 | 0.06 | 0.79 | 0.80 | 116 | 0.05 | 0.81 | 0.80 |
| | | 0.3 | 118 | 0.06 | 0.81 | 0.81 | 118 | 0.05 | 0.80 | 0.80 |
| | | 0.6 | 124 | 0.05 | 0.80 | 0.80 | 126 | 0.05 | 0.80 | 0.80 |
| | | 0.9 | 138 | 0.05 | 0.82 | 0.80 | 140 | 0.05 | 0.82 | 0.80 |
| $\bar{m} = 50$ | $\rho = 0.02$ | 0 | 32 | 0.06 | 0.80 | 0.81 | 34 | 0.04 | 0.82 | 0.81 |
| | | 0.3 | 32 | 0.07 | 0.79 | 0.80 | 34 | 0.05 | 0.80 | 0.80 |
| | | 0.6 | 36 | 0.07 | 0.82 | 0.82 | 38 | 0.05 | 0.82 | 0.82 |
| | | 0.9 | 40 | 0.06 | 0.84 | 0.81 | 42 | 0.05 | 0.84 | 0.81 |
| | $\rho = 0.05$ | 0 | 56 | 0.05 | 0.83 | 0.81 | 58 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 56 | 0.05 | 0.80 | 0.81 | 58 | 0.05 | 0.80 | 0.81 |
| | | 0.6 | 60 | 0.05 | 0.79 | 0.81 | 62 | 0.05 | 0.79 | 0.81 |
| | | 0.9 | 66 | 0.06 | 0.82 | 0.81 | 68 | 0.06 | 0.81 | 0.81 |
| | $\rho = 0.10$ | 0 | 94 | 0.05 | 0.79 | 0.81 | 96 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 94 | 0.06 | 0.80 | 0.80 | 96 | 0.05 | 0.80 | 0.80 |
| | | 0.6 | 98 | 0.05 | 0.78 | 0.80 | 100 | 0.05 | 0.80 | 0.80 |
| | | 0.9 | 104 | 0.05 | 0.79 | 0.80 | 106 | 0.05 | 0.79 | 0.80 |
| $\bar{m} = 100$ | $\rho = 0.02$ | 0 | 24 | 0.06 | 0.81 | 0.81 | 26 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 24 | 0.07 | 0.80 | 0.80 | 26 | 0.06 | 0.79 | 0.80 |
| | | 0.6 | 26 | 0.07 | 0.80 | 0.81 | 28 | 0.05 | 0.81 | 0.81 |
| | | 0.9 | 30 | 0.06 | 0.83 | 0.82 | 32 | 0.05 | 0.82 | 0.82 |
| | $\rho = 0.05$ | 0 | 48 | 0.06 | 0.81 | 0.81 | 50 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 48 | 0.06 | 0.80 | 0.81 | 50 | 0.06 | 0.80 | 0.81 |
| | | 0.6 | 50 | 0.06 | 0.79 | 0.81 | 52 | 0.05 | 0.78 | 0.81 |
| | | 0.9 | 54 | 0.05 | 0.79 | 0.81 | 56 | 0.05 | 0.79 | 0.81 |
| | $\rho = 0.10$ | 0 | 86 | 0.06 | 0.79 | 0.80 | 88 | 0.05 | 0.80 | 0.80 |
| | | 0.3 | 88 | 0.05 | 0.80 | 0.81 | 90 | 0.05 | 0.80 | 0.81 |
| | | 0.6 | 88 | 0.05 | 0.79 | 0.80 | 90 | 0.05 | 0.80 | 0.80 |
| | | 0.9 | 92 | 0.05 | 0.76 | 0.80 | 94 | 0.05 | 0.77 | 0.80 |

**WEB TABLE 18** Required number of clusters $n_{B1}$, empirical type I error $\psi$, empirical power $\phi$, and predicted power $\hat{\phi}$ corresponding to the test for the marginal effect of the cluster-level treatment with and without finite-sample correction. The marginal effect size of the cluster-level treatment is $\delta_X = 0.4$. The unequal cluster sizes approximately follow a normal distribution. Notation: $\bar{m}$ is the mean cluster size, $\rho$ is the ICC, CV is the coefficient of variation of cluster sizes. The results were based on 5,000 simulations.

| | | | $z$-test | | | | $t$-test | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | CV | $n_{B1}$ | $\psi$ | $\phi$ | $\hat{\phi}$ | $n_{B1}$ | $\psi$ | $\phi$ | $\hat{\phi}$ |
| $\bar{m} = 20$ | $\rho = 0.02$ | 0 | 14 | 0.05 | 0.80 | 0.81 | 16 | 0.04 | 0.81 | 0.81 |
| | | 0.3 | 14 | 0.06 | 0.79 | 0.81 | 16 | 0.04 | 0.79 | 0.80 |
| | | 0.6 | 16 | 0.06 | 0.82 | 0.83 | 18 | 0.05 | 0.82 | 0.83 |
| | | 0.9 | 18 | 0.06 | 0.85 | 0.84 | 20 | 0.05 | 0.86 | 0.84 |
| | $\rho = 0.05$ | 0 | 20 | 0.07 | 0.82 | 0.82 | 22 | 0.04 | 0.82 | 0.82 |
| | | 0.3 | 20 | 0.06 | 0.80 | 0.81 | 22 | 0.05 | 0.80 | 0.81 |
| | | 0.6 | 22 | 0.07 | 0.82 | 0.82 | 24 | 0.05 | 0.82 | 0.82 |
| | | 0.9 | 24 | 0.07 | 0.82 | 0.80 | 28 | 0.05 | 0.85 | 0.83 |
| | $\rho = 0.10$ | 0 | 30 | 0.05 | 0.82 | 0.82 | 32 | 0.05 | 0.82 | 0.82 |
| | | 0.3 | 30 | 0.06 | 0.81 | 0.81 | 32 | 0.06 | 0.81 | 0.81 |
| | | 0.6 | 32 | 0.06 | 0.81 | 0.81 | 34 | 0.05 | 0.81 | 0.81 |
| | | 0.9 | 36 | 0.06 | 0.83 | 0.82 | 38 | 0.06 | 0.83 | 0.82 |
| $\bar{m} = 50$ | $\rho = 0.02$ | 0 | 8 | 0.08 | 0.80 | 0.81 | 12 | 0.04 | 0.88 | 0.88 |
| | | 0.3 | 8 | 0.08 | 0.78 | 0.80 | 12 | 0.04 | 0.88 | 0.87 |
| | | 0.6 | 10 | 0.09 | 0.84 | 0.86 | 12 | 0.05 | 0.83 | 0.85 |
| | | 0.9 | 10 | 0.08 | 0.81 | 0.81 | 14 | 0.05 | 0.88 | 0.87 |
| | $\rho = 0.05$ | 0 | 14 | 0.07 | 0.80 | 0.81 | 16 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 14 | 0.08 | 0.80 | 0.81 | 16 | 0.06 | 0.80 | 0.80 |
| | | 0.6 | 16 | 0.07 | 0.82 | 0.84 | 18 | 0.06 | 0.83 | 0.83 |
| | | 0.9 | 18 | 0.08 | 0.82 | 0.84 | 20 | 0.06 | 0.84 | 0.84 |
| | $\rho = 0.10$ | 0 | 24 | 0.06 | 0.82 | 0.81 | 26 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 24 | 0.06 | 0.81 | 0.81 | 26 | 0.05 | 0.80 | 0.81 |
| | | 0.6 | 26 | 0.06 | 0.80 | 0.83 | 28 | 0.05 | 0.81 | 0.83 |
| | | 0.9 | 26 | 0.06 | 0.78 | 0.80 | 28 | 0.05 | 0.77 | 0.80 |
| $\bar{m} = 100$ | $\rho = 0.02$ | 0 | 6 | 0.11 | 0.79 | 0.81 | 10 | 0.05 | 0.88 | 0.89 |
| | | 0.3 | 6 | 0.11 | 0.79 | 0.80 | 10 | 0.05 | 0.89 | 0.89 |
| | | 0.6 | 8 | 0.09 | 0.85 | 0.88 | 10 | 0.05 | 0.85 | 0.87 |
| | | 0.9 | 8 | 0.12 | 0.81 | 0.84 | 10 | 0.06 | 0.82 | 0.83 |
| | $\rho = 0.05$ | 0 | 12 | 0.08 | 0.82 | 0.81 | 14 | 0.05 | 0.81 | 0.80 |
| | | 0.3 | 12 | 0.08 | 0.79 | 0.81 | 16 | 0.05 | 0.86 | 0.86 |
| | | 0.6 | 14 | 0.08 | 0.83 | 0.85 | 16 | 0.05 | 0.83 | 0.84 |
| | | 0.9 | 14 | 0.08 | 0.79 | 0.83 | 16 | 0.06 | 0.80 | 0.82 |
| | $\rho = 0.10$ | 0 | 22 | 0.06 | 0.81 | 0.81 | 24 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 22 | 0.06 | 0.80 | 0.81 | 24 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 22 | 0.06 | 0.78 | 0.80 | 26 | 0.05 | 0.81 | 0.83 |
| | | 0.9 | 24 | 0.07 | 0.78 | 0.82 | 26 | 0.06 | 0.78 | 0.82 |

**WEB TABLE 19** Required number of clusters $n_{B2}$, empirical type I error $\psi$, empirical power $\phi$, and predicted power $\hat{\phi}$ corresponding to the test for the marginal effect of the individual-level treatment. The unequal cluster sizes approximately follow a normal distribution. Notation: $\delta_Z$ is the marginal effect size of the individual-level treatment, $\bar{m}$ is the mean cluster size, $\rho$ is the ICC, CV is the coefficient of variation of cluster sizes. The results were based on 5,000 simulations.

| | | | $\delta_Z = 0.1$ | | | | $\delta_Z = 0.15$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | CV | $n_{A2}$ | $\psi$ | $\phi$ | $\hat{\phi}$ | $n_{A2}$ | $\psi$ | $\phi$ | $\hat{\phi}$ |
| $\bar{m} = 20$ | $\rho = 0.02$ | 0 | 158 | 0.05 | 0.81 | 0.80 | 70 | 0.05 | 0.80 | 0.80 |
| | | 0.3 | 158 | 0.05 | 0.81 | 0.80 | 70 | 0.04 | 0.80 | 0.80 |
| | | 0.6 | 156 | 0.05 | 0.81 | 0.80 | 70 | 0.05 | 0.82 | 0.80 |
| | | 0.9 | 156 | 0.05 | 0.82 | 0.80 | 70 | 0.05 | 0.83 | 0.80 |
| | $\rho = 0.05$ | 0 | 154 | 0.05 | 0.82 | 0.80 | 70 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 154 | 0.05 | 0.82 | 0.80 | 68 | 0.05 | 0.80 | 0.80 |
| | | 0.6 | 154 | 0.05 | 0.82 | 0.80 | 68 | 0.05 | 0.82 | 0.80 |
| | | 0.9 | 154 | 0.05 | 0.84 | 0.80 | 68 | 0.05 | 0.83 | 0.80 |
| | $\rho = 0.10$ | 0 | 148 | 0.05 | 0.81 | 0.80 | 66 | 0.05 | 0.83 | 0.81 |
| | | 0.3 | 148 | 0.05 | 0.81 | 0.80 | 66 | 0.05 | 0.83 | 0.81 |
| | | 0.6 | 146 | 0.05 | 0.81 | 0.80 | 66 | 0.05 | 0.82 | 0.81 |
| | | 0.9 | 146 | 0.05 | 0.83 | 0.80 | 66 | 0.05 | 0.84 | 0.81 |
| $\bar{m} = 50$ | $\rho = 0.02$ | 0 | 64 | 0.04 | 0.81 | 0.81 | 28 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 64 | 0.05 | 0.83 | 0.81 | 28 | 0.05 | 0.80 | 0.81 |
| | | 0.6 | 64 | 0.05 | 0.81 | 0.81 | 28 | 0.04 | 0.81 | 0.81 |
| | | 0.9 | 64 | 0.05 | 0.84 | 0.81 | 28 | 0.05 | 0.83 | 0.81 |
| | $\rho = 0.05$ | 0 | 62 | 0.05 | 0.82 | 0.81 | 28 | 0.05 | 0.83 | 0.82 |
| | | 0.3 | 62 | 0.05 | 0.81 | 0.81 | 28 | 0.05 | 0.81 | 0.82 |
| | | 0.6 | 62 | 0.05 | 0.81 | 0.81 | 28 | 0.04 | 0.83 | 0.82 |
| | | 0.9 | 62 | 0.05 | 0.84 | 0.81 | 28 | 0.05 | 0.84 | 0.82 |
| | $\rho = 0.10$ | 0 | 58 | 0.05 | 0.81 | 0.80 | 26 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 58 | 0.05 | 0.81 | 0.80 | 26 | 0.05 | 0.82 | 0.81 |
| | | 0.6 | 58 | 0.05 | 0.80 | 0.80 | 26 | 0.05 | 0.82 | 0.81 |
| | | 0.9 | 58 | 0.05 | 0.83 | 0.80 | 26 | 0.05 | 0.82 | 0.81 |
| $\bar{m} = 100$ | $\rho = 0.02$ | 0 | 32 | 0.05 | 0.81 | 0.81 | 14 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 32 | 0.05 | 0.81 | 0.81 | 14 | 0.05 | 0.80 | 0.81 |
| | | 0.6 | 32 | 0.05 | 0.82 | 0.81 | 14 | 0.05 | 0.80 | 0.81 |
| | | 0.9 | 32 | 0.05 | 0.82 | 0.81 | 14 | 0.06 | 0.80 | 0.81 |
| | $\rho = 0.05$ | 0 | 32 | 0.05 | 0.83 | 0.82 | 14 | 0.05 | 0.83 | 0.82 |
| | | 0.3 | 32 | 0.05 | 0.83 | 0.82 | 14 | 0.05 | 0.81 | 0.82 |
| | | 0.6 | 32 | 0.05 | 0.83 | 0.82 | 14 | 0.05 | 0.82 | 0.82 |
| | | 0.9 | 32 | 0.05 | 0.83 | 0.82 | 14 | 0.06 | 0.81 | 0.82 |
| | $\rho = 0.10$ | 0 | 30 | 0.05 | 0.82 | 0.82 | 14 | 0.05 | 0.85 | 0.84 |
| | | 0.3 | 30 | 0.05 | 0.82 | 0.82 | 14 | 0.05 | 0.84 | 0.84 |
| | | 0.6 | 30 | 0.05 | 0.82 | 0.82 | 14 | 0.05 | 0.83 | 0.84 |
| | | 0.9 | 30 | 0.05 | 0.84 | 0.82 | 14 | 0.06 | 0.83 | 0.84 |

**WEB TABLE 20** Required number of clusters $n_C$, empirical type I error $\psi$, empirical power $\phi$, and predicted power $\hat{\phi}$ corresponding to the interaction test. The unequal cluster sizes approximately follow a normal distribution. Notation: $\delta_4$ is the interaction effect size, $\bar{m}$ is the mean cluster size, $\rho$ is the ICC, CV is the coefficient of variation of cluster sizes. The results were based on 5,000 simulations.

| | | | $\delta_4 = 0.2$ | | | | $\delta_4 = 0.3$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | CV | $n_C$ | $\psi$ | $\phi$ | $\hat{\phi}$ | $n_C$ | $\psi$ | $\phi$ | $\hat{\phi}$ |
| $\bar{m} = 20$ | $\rho = 0.02$ | 0 | 158 | 0.05 | 0.81 | 0.80 | 70 | 0.05 | 0.81 | 0.80 |
| | | 0.3 | 158 | 0.05 | 0.80 | 0.80 | 70 | 0.05 | 0.80 | 0.80 |
| | | 0.6 | 156 | 0.05 | 0.82 | 0.80 | 70 | 0.05 | 0.81 | 0.80 |
| | | 0.9 | 156 | 0.05 | 0.83 | 0.80 | 70 | 0.05 | 0.83 | 0.80 |
| | $\rho = 0.05$ | 0 | 154 | 0.05 | 0.81 | 0.80 | 70 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 154 | 0.05 | 0.81 | 0.80 | 68 | 0.05 | 0.81 | 0.80 |
| | | 0.6 | 154 | 0.05 | 0.81 | 0.80 | 68 | 0.05 | 0.81 | 0.80 |
| | | 0.9 | 154 | 0.05 | 0.84 | 0.80 | 68 | 0.04 | 0.83 | 0.80 |
| | $\rho = 0.10$ | 0 | 148 | 0.05 | 0.83 | 0.80 | 66 | 0.05 | 0.83 | 0.81 |
| | | 0.3 | 148 | 0.05 | 0.81 | 0.80 | 66 | 0.05 | 0.82 | 0.81 |
| | | 0.6 | 146 | 0.05 | 0.82 | 0.80 | 66 | 0.05 | 0.82 | 0.81 |
| | | 0.9 | 146 | 0.05 | 0.83 | 0.80 | 66 | 0.05 | 0.84 | 0.81 |
| $\bar{m} = 50$ | $\rho = 0.02$ | 0 | 64 | 0.05 | 0.81 | 0.81 | 28 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 64 | 0.05 | 0.82 | 0.81 | 28 | 0.05 | 0.80 | 0.81 |
| | | 0.6 | 64 | 0.05 | 0.81 | 0.81 | 28 | 0.05 | 0.80 | 0.81 |
| | | 0.9 | 64 | 0.05 | 0.83 | 0.81 | 28 | 0.05 | 0.82 | 0.81 |
| | $\rho = 0.05$ | 0 | 62 | 0.05 | 0.82 | 0.81 | 28 | 0.05 | 0.82 | 0.82 |
| | | 0.3 | 62 | 0.05 | 0.81 | 0.81 | 28 | 0.05 | 0.82 | 0.82 |
| | | 0.6 | 62 | 0.05 | 0.82 | 0.81 | 28 | 0.05 | 0.81 | 0.82 |
| | | 0.9 | 62 | 0.05 | 0.83 | 0.81 | 28 | 0.05 | 0.83 | 0.82 |
| | $\rho = 0.10$ | 0 | 58 | 0.05 | 0.81 | 0.80 | 26 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 58 | 0.05 | 0.81 | 0.80 | 26 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 58 | 0.05 | 0.81 | 0.80 | 26 | 0.05 | 0.81 | 0.81 |
| | | 0.9 | 58 | 0.05 | 0.83 | 0.80 | 26 | 0.05 | 0.83 | 0.81 |
| $\bar{m} = 100$ | $\rho = 0.02$ | 0 | 32 | 0.05 | 0.81 | 0.81 | 14 | 0.05 | 0.80 | 0.81 |
| | | 0.3 | 32 | 0.05 | 0.81 | 0.81 | 14 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 32 | 0.05 | 0.81 | 0.81 | 14 | 0.05 | 0.80 | 0.81 |
| | | 0.9 | 32 | 0.05 | 0.82 | 0.81 | 14 | 0.05 | 0.79 | 0.81 |
| | $\rho = 0.05$ | 0 | 32 | 0.05 | 0.83 | 0.82 | 14 | 0.05 | 0.81 | 0.82 |
| | | 0.3 | 32 | 0.05 | 0.83 | 0.82 | 14 | 0.05 | 0.82 | 0.82 |
| | | 0.6 | 32 | 0.05 | 0.82 | 0.82 | 14 | 0.05 | 0.82 | 0.82 |
| | | 0.9 | 32 | 0.05 | 0.83 | 0.82 | 14 | 0.05 | 0.81 | 0.82 |
| | $\rho = 0.10$ | 0 | 30 | 0.05 | 0.82 | 0.82 | 14 | 0.05 | 0.83 | 0.84 |
| | | 0.3 | 30 | 0.05 | 0.83 | 0.82 | 14 | 0.05 | 0.84 | 0.84 |
| | | 0.6 | 30 | 0.05 | 0.83 | 0.82 | 14 | 0.05 | 0.84 | 0.84 |
| | | 0.9 | 30 | 0.05 | 0.83 | 0.82 | 14 | 0.05 | 0.82 | 0.84 |

**WEB TABLE 21** Required number of clusters $n_{A1}$, empirical type I error $\psi$, empirical power $\phi$, and predicted power $\hat{\phi}$ corresponding to the test for the controlled effect of the cluster-level treatment with and without finite-sample correction. The controlled effect size of the cluster-level treatment is $\delta_2 = 0.2$. The unequal cluster sizes approximately follow a uniform distribution. Notation: $\bar{m}$ is the mean cluster size, $\rho$ is the ICC, CV is the coefficient of variation of cluster sizes. The results were based on 5,000 simulations.

| | | | z-test | | | | t-test | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | CV | $n_{A1}$ | $\psi$ | $\phi$ | $\hat{\phi}$ | $n_{A1}$ | $\psi$ | $\phi$ | $\hat{\phi}$ |
| $\bar{m} = 20$ | $\rho = 0.02$ | 0 | 94 | 0.05 | 0.81 | 0.80 | 96 | 0.05 | 0.80 | 0.80 |
| | | 0.3 | 96 | 0.06 | 0.81 | 0.81 | 98 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 98 | 0.05 | 0.81 | 0.80 | 100 | 0.05 | 0.81 | 0.80 |
| | | 0.9 | 104 | 0.06 | 0.84 | 0.80 | 106 | 0.05 | 0.83 | 0.80 |
| | $\rho = 0.05$ | 0 | 116 | 0.05 | 0.81 | 0.80 | 118 | 0.05 | 0.80 | 0.80 |
| | | 0.3 | 118 | 0.05 | 0.81 | 0.80 | 120 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 124 | 0.05 | 0.81 | 0.81 | 126 | 0.05 | 0.80 | 0.81 |
| | | 0.9 | 136 | 0.05 | 0.83 | 0.81 | 136 | 0.05 | 0.83 | 0.80 |
| | $\rho = 0.10$ | 0 | 152 | 0.05 | 0.81 | 0.80 | 154 | 0.05 | 0.81 | 0.80 |
| | | 0.3 | 154 | 0.05 | 0.81 | 0.80 | 156 | 0.05 | 0.81 | 0.80 |
| | | 0.6 | 160 | 0.05 | 0.79 | 0.80 | 162 | 0.05 | 0.79 | 0.80 |
| | | 0.9 | 176 | 0.05 | 0.82 | 0.80 | 176 | 0.05 | 0.81 | 0.80 |
| $\bar{m} = 50$ | $\rho = 0.02$ | 0 | 48 | 0.06 | 0.81 | 0.81 | 50 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 48 | 0.05 | 0.80 | 0.81 | 50 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 50 | 0.06 | 0.79 | 0.80 | 52 | 0.05 | 0.78 | 0.80 |
| | | 0.9 | 56 | 0.06 | 0.83 | 0.81 | 58 | 0.05 | 0.82 | 0.81 |
| | $\rho = 0.05$ | 0 | 70 | 0.05 | 0.81 | 0.80 | 72 | 0.05 | 0.81 | 0.80 |
| | | 0.3 | 72 | 0.05 | 0.80 | 0.81 | 74 | 0.05 | 0.82 | 0.81 |
| | | 0.6 | 74 | 0.05 | 0.79 | 0.80 | 76 | 0.05 | 0.80 | 0.80 |
| | | 0.9 | 80 | 0.05 | 0.79 | 0.80 | 82 | 0.05 | 0.80 | 0.80 |
| | $\rho = 0.10$ | 0 | 108 | 0.05 | 0.80 | 0.80 | 110 | 0.04 | 0.82 | 0.80 |
| | | 0.3 | 110 | 0.05 | 0.80 | 0.81 | 112 | 0.05 | 0.80 | 0.81 |
| | | 0.6 | 112 | 0.05 | 0.79 | 0.80 | 114 | 0.05 | 0.79 | 0.80 |
| | | 0.9 | 118 | 0.05 | 0.77 | 0.80 | 120 | 0.05 | 0.77 | 0.80 |
| $\bar{m} = 100$ | $\rho = 0.02$ | 0 | 32 | 0.05 | 0.80 | 0.81 | 34 | 0.04 | 0.82 | 0.81 |
| | | 0.3 | 32 | 0.06 | 0.80 | 0.80 | 34 | 0.05 | 0.80 | 0.80 |
| | | 0.6 | 34 | 0.06 | 0.79 | 0.81 | 36 | 0.05 | 0.79 | 0.81 |
| | | 0.9 | 38 | 0.06 | 0.81 | 0.82 | 40 | 0.05 | 0.81 | 0.82 |
| | $\rho = 0.05$ | 0 | 56 | 0.05 | 0.81 | 0.81 | 58 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 56 | 0.05 | 0.80 | 0.81 | 58 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 58 | 0.05 | 0.79 | 0.81 | 60 | 0.05 | 0.79 | 0.81 |
| | | 0.9 | 60 | 0.06 | 0.77 | 0.80 | 62 | 0.05 | 0.76 | 0.80 |
| | $\rho = 0.10$ | 0 | 94 | 0.05 | 0.81 | 0.81 | 96 | 0.06 | 0.80 | 0.81 |
| | | 0.3 | 94 | 0.06 | 0.81 | 0.80 | 96 | 0.05 | 0.80 | 0.80 |
| | | 0.6 | 96 | 0.05 | 0.78 | 0.80 | 98 | 0.05 | 0.79 | 0.80 |
| | | 0.9 | 100 | 0.05 | 0.76 | 0.81 | 102 | 0.05 | 0.75 | 0.81 |

**WEB TABLE 22** Required number of clusters $n_{A1}$, empirical type I error $\psi$, empirical power $\phi$, and predicted power $\hat{\phi}$ corresponding to the test for the controlled effect of the cluster-level treatment with and without finite-sample correction. The controlled effect size of the cluster-level treatment is $\delta_2 = 0.4$. The unequal cluster sizes approximately follow a uniform distribution. Notation: $\bar{m}$ is the mean cluster size, $\rho$ is the ICC, CV is the coefficient of variation of cluster sizes. The results were based on 5,000 simulations.

| | | | $z$-test | | | | $t$-test | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | CV | $n_{A1}$ | $\psi$ | $\phi$ | $\hat{\phi}$ | $n_{A1}$ | $\psi$ | $\phi$ | $\hat{\phi}$ |
| $\bar{m} = 20$ | $\rho = 0.02$ | 0 | 24 | 0.05 | 0.81 | 0.81 | 26 | 0.04 | 0.82 | 0.81 |
| | | 0.3 | 24 | 0.05 | 0.81 | 0.81 | 26 | 0.04 | 0.80 | 0.81 |
| | | 0.6 | 26 | 0.05 | 0.82 | 0.82 | 28 | 0.04 | 0.83 | 0.82 |
| | | 0.9 | 26 | 0.06 | 0.81 | 0.80 | 30 | 0.05 | 0.85 | 0.83 |
| | $\rho = 0.05$ | 0 | 30 | 0.05 | 0.83 | 0.82 | 32 | 0.04 | 0.83 | 0.82 |
| | | 0.3 | 30 | 0.06 | 0.82 | 0.81 | 32 | 0.04 | 0.83 | 0.81 |
| | | 0.6 | 32 | 0.06 | 0.81 | 0.82 | 34 | 0.05 | 0.81 | 0.82 |
| | | 0.9 | 34 | 0.05 | 0.83 | 0.81 | 36 | 0.05 | 0.83 | 0.81 |
| | $\rho = 0.10$ | 0 | 38 | 0.05 | 0.81 | 0.80 | 40 | 0.05 | 0.80 | 0.80 |
| | | 0.3 | 40 | 0.05 | 0.82 | 0.82 | 42 | 0.05 | 0.82 | 0.82 |
| | | 0.6 | 40 | 0.06 | 0.80 | 0.80 | 42 | 0.05 | 0.79 | 0.80 |
| | | 0.9 | 44 | 0.06 | 0.82 | 0.80 | 46 | 0.05 | 0.81 | 0.80 |
| $\bar{m} = 50$ | $\rho = 0.02$ | 0 | 12 | 0.06 | 0.80 | 0.81 | 14 | 0.04 | 0.81 | 0.80 |
| | | 0.3 | 12 | 0.06 | 0.80 | 0.81 | 16 | 0.04 | 0.86 | 0.86 |
| | | 0.6 | 14 | 0.06 | 0.81 | 0.84 | 16 | 0.04 | 0.83 | 0.84 |
| | | 0.9 | 14 | 0.07 | 0.81 | 0.81 | 16 | 0.05 | 0.81 | 0.81 |
| | $\rho = 0.05$ | 0 | 18 | 0.06 | 0.82 | 0.81 | 20 | 0.04 | 0.81 | 0.81 |
| | | 0.3 | 18 | 0.06 | 0.80 | 0.81 | 20 | 0.04 | 0.81 | 0.81 |
| | | 0.6 | 20 | 0.06 | 0.81 | 0.83 | 22 | 0.05 | 0.81 | 0.83 |
| | | 0.9 | 20 | 0.05 | 0.78 | 0.80 | 24 | 0.05 | 0.82 | 0.84 |
| | $\rho = 0.10$ | 0 | 28 | 0.06 | 0.82 | 0.82 | 30 | 0.06 | 0.81 | 0.82 |
| | | 0.3 | 28 | 0.06 | 0.81 | 0.81 | 30 | 0.05 | 0.80 | 0.81 |
| | | 0.6 | 28 | 0.05 | 0.78 | 0.80 | 30 | 0.05 | 0.80 | 0.80 |
| | | 0.9 | 30 | 0.06 | 0.78 | 0.81 | 32 | 0.05 | 0.78 | 0.81 |
| $\bar{m} = 100$ | $\rho = 0.02$ | 0 | 8 | 0.07 | 0.80 | 0.81 | 12 | 0.04 | 0.90 | 0.88 |
| | | 0.3 | 8 | 0.07 | 0.80 | 0.80 | 12 | 0.04 | 0.87 | 0.87 |
| | | 0.6 | 10 | 0.08 | 0.83 | 0.87 | 12 | 0.04 | 0.85 | 0.86 |
| | | 0.9 | 10 | 0.07 | 0.80 | 0.84 | 12 | 0.04 | 0.81 | 0.83 |
| | $\rho = 0.05$ | 0 | 14 | 0.07 | 0.81 | 0.81 | 16 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 14 | 0.07 | 0.81 | 0.81 | 16 | 0.04 | 0.80 | 0.80 |
| | | 0.6 | 16 | 0.07 | 0.83 | 0.85 | 18 | 0.05 | 0.83 | 0.84 |
| | | 0.9 | 16 | 0.08 | 0.77 | 0.83 | 18 | 0.05 | 0.78 | 0.82 |
| | $\rho = 0.10$ | 0 | 24 | 0.07 | 0.82 | 0.81 | 26 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 24 | 0.07 | 0.80 | 0.81 | 26 | 0.06 | 0.81 | 0.81 |
| | | 0.6 | 24 | 0.07 | 0.77 | 0.80 | 26 | 0.05 | 0.77 | 0.80 |
| | | 0.9 | 26 | 0.06 | 0.77 | 0.82 | 28 | 0.05 | 0.77 | 0.82 |

**WEB TABLE 23** Required number of clusters $n_{A2}$, empirical type I error $\psi$, empirical power $\phi$, and predicted power $\hat{\phi}$ corresponding to the test for the controlled effect of the individual-level treatment. The unequal cluster sizes approximately follow a uniform distribution. Notation: $\delta_3$ is the controlled effect size of the individual-level treatment, $\bar{m}$ is the mean cluster size, $\rho$ is the ICC, CV is the coefficient of variation of cluster sizes. The results were based on 5,000 simulations.

| | | | $\delta_3 = 0.15$ | | | | $\delta_3 = 0.3$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | CV | $n_{A2}$ | $\psi$ | $\phi$ | $\hat{\phi}$ | $n_{A2}$ | $\psi$ | $\phi$ | $\hat{\phi}$ |
| $\bar{m} = 20$ | $\rho = 0.02$ | 0 | 140 | 0.05 | 0.81 | 0.80 | 36 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 140 | 0.05 | 0.80 | 0.80 | 36 | 0.06 | 0.82 | 0.81 |
| | | 0.6 | 140 | 0.05 | 0.81 | 0.80 | 36 | 0.05 | 0.82 | 0.81 |
| | | 0.9 | 140 | 0.05 | 0.83 | 0.80 | 36 | 0.05 | 0.83 | 0.82 |
| | $\rho = 0.05$ | 0 | 138 | 0.06 | 0.81 | 0.81 | 36 | 0.05 | 0.84 | 0.82 |
| | | 0.3 | 136 | 0.05 | 0.81 | 0.80 | 34 | 0.05 | 0.81 | 0.80 |
| | | 0.6 | 136 | 0.05 | 0.81 | 0.80 | 34 | 0.05 | 0.80 | 0.80 |
| | | 0.9 | 136 | 0.05 | 0.83 | 0.80 | 34 | 0.05 | 0.83 | 0.80 |
| | $\rho = 0.10$ | 0 | 132 | 0.05 | 0.81 | 0.81 | 34 | 0.05 | 0.83 | 0.82 |
| | | 0.3 | 130 | 0.05 | 0.82 | 0.80 | 34 | 0.05 | 0.83 | 0.82 |
| | | 0.6 | 130 | 0.05 | 0.82 | 0.80 | 34 | 0.05 | 0.82 | 0.82 |
| | | 0.9 | 130 | 0.05 | 0.84 | 0.80 | 34 | 0.05 | 0.84 | 0.82 |
| $\bar{m} = 50$ | $\rho = 0.02$ | 0 | 56 | 0.05 | 0.81 | 0.81 | 14 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 56 | 0.05 | 0.82 | 0.81 | 14 | 0.05 | 0.80 | 0.81 |
| | | 0.6 | 56 | 0.05 | 0.81 | 0.81 | 14 | 0.04 | 0.79 | 0.81 |
| | | 0.9 | 56 | 0.05 | 0.83 | 0.81 | 14 | 0.05 | 0.80 | 0.81 |
| | $\rho = 0.05$ | 0 | 54 | 0.05 | 0.82 | 0.80 | 14 | 0.05 | 0.83 | 0.82 |
| | | 0.3 | 54 | 0.05 | 0.80 | 0.80 | 14 | 0.05 | 0.81 | 0.82 |
| | | 0.6 | 54 | 0.06 | 0.80 | 0.80 | 14 | 0.04 | 0.80 | 0.82 |
| | | 0.9 | 54 | 0.05 | 0.82 | 0.80 | 14 | 0.05 | 0.81 | 0.82 |
| | $\rho = 0.10$ | 0 | 52 | 0.05 | 0.82 | 0.81 | 14 | 0.05 | 0.85 | 0.83 |
| | | 0.3 | 52 | 0.05 | 0.82 | 0.81 | 14 | 0.05 | 0.83 | 0.83 |
| | | 0.6 | 52 | 0.05 | 0.80 | 0.81 | 14 | 0.04 | 0.82 | 0.83 |
| | | 0.9 | 52 | 0.05 | 0.83 | 0.81 | 14 | 0.05 | 0.83 | 0.84 |
| $\bar{m} = 100$ | $\rho = 0.02$ | 0 | 28 | 0.06 | 0.81 | 0.81 | 8 | 0.05 | 0.85 | 0.86 |
| | | 0.3 | 28 | 0.05 | 0.81 | 0.81 | 8 | 0.06 | 0.86 | 0.86 |
| | | 0.6 | 28 | 0.05 | 0.81 | 0.81 | 8 | 0.05 | 0.82 | 0.86 |
| | | 0.9 | 28 | 0.05 | 0.82 | 0.81 | 8 | 0.05 | 0.81 | 0.86 |
| | $\rho = 0.05$ | 0 | 28 | 0.06 | 0.82 | 0.82 | 8 | 0.05 | 0.86 | 0.87 |
| | | 0.3 | 28 | 0.05 | 0.82 | 0.82 | 8 | 0.06 | 0.87 | 0.87 |
| | | 0.6 | 28 | 0.05 | 0.82 | 0.82 | 8 | 0.05 | 0.83 | 0.87 |
| | | 0.9 | 28 | 0.05 | 0.83 | 0.82 | 8 | 0.06 | 0.83 | 0.87 |
| | $\rho = 0.10$ | 0 | 26 | 0.05 | 0.81 | 0.81 | 8 | 0.05 | 0.88 | 0.88 |
| | | 0.3 | 26 | 0.05 | 0.82 | 0.81 | 8 | 0.06 | 0.88 | 0.88 |
| | | 0.6 | 26 | 0.05 | 0.80 | 0.81 | 8 | 0.05 | 0.85 | 0.88 |
| | | 0.9 | 26 | 0.05 | 0.82 | 0.81 | 8 | 0.05 | 0.84 | 0.88 |

**WEB TABLE 24** Required number of clusters $n_{B1}$, empirical type I error $\psi$, empirical power $\phi$, and predicted power $\hat{\phi}$ corresponding to the test for the marginal effect of the cluster-level treatment with and without finite-sample correction. The marginal effect size of the cluster-level treatment is $\delta_X = 0.2$. The unequal cluster sizes approximately follow a uniform distribution. Notation: $\bar{m}$ is the mean cluster size, $\rho$ is the ICC, CV is the coefficient of variation of cluster sizes. The results were based on 5,000 simulations.

| | | | $z$-test | | | | $t$-test | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | CV | $n_{B1}$ | $\psi$ | $\phi$ | $\hat{\phi}$ | $n_{B1}$ | $\psi$ | $\phi$ | $\hat{\phi}$ |
| $\bar{m} = 20$ | $\rho = 0.02$ | 0 | 56 | 0.05 | 0.81 | 0.81 | 58 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 56 | 0.05 | 0.80 | 0.81 | 58 | 0.05 | 0.80 | 0.81 |
| | | 0.6 | 60 | 0.06 | 0.80 | 0.81 | 62 | 0.05 | 0.81 | 0.81 |
| | | 0.9 | 66 | 0.06 | 0.84 | 0.81 | 68 | 0.05 | 0.84 | 0.81 |
| | $\rho = 0.05$ | 0 | 78 | 0.05 | 0.80 | 0.81 | 80 | 0.05 | 0.80 | 0.81 |
| | | 0.3 | 80 | 0.05 | 0.81 | 0.81 | 82 | 0.05 | 0.80 | 0.81 |
| | | 0.6 | 86 | 0.05 | 0.81 | 0.81 | 88 | 0.05 | 0.81 | 0.81 |
| | | 0.9 | 96 | 0.05 | 0.83 | 0.80 | 98 | 0.05 | 0.82 | 0.80 |
| | $\rho = 0.10$ | 0 | 114 | 0.06 | 0.79 | 0.80 | 116 | 0.05 | 0.81 | 0.80 |
| | | 0.3 | 118 | 0.05 | 0.81 | 0.81 | 118 | 0.05 | 0.80 | 0.80 |
| | | 0.6 | 124 | 0.05 | 0.80 | 0.80 | 126 | 0.05 | 0.79 | 0.80 |
| | | 0.9 | 138 | 0.06 | 0.81 | 0.80 | 140 | 0.05 | 0.79 | 0.80 |
| $\bar{m} = 50$ | $\rho = 0.02$ | 0 | 32 | 0.06 | 0.80 | 0.81 | 34 | 0.04 | 0.82 | 0.81 |
| | | 0.3 | 32 | 0.06 | 0.80 | 0.80 | 34 | 0.04 | 0.80 | 0.80 |
| | | 0.6 | 36 | 0.06 | 0.80 | 0.82 | 38 | 0.05 | 0.80 | 0.82 |
| | | 0.9 | 40 | 0.06 | 0.82 | 0.81 | 42 | 0.06 | 0.82 | 0.81 |
| | $\rho = 0.05$ | 0 | 56 | 0.05 | 0.83 | 0.81 | 58 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 56 | 0.06 | 0.81 | 0.81 | 58 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 60 | 0.05 | 0.79 | 0.81 | 62 | 0.05 | 0.80 | 0.81 |
| | | 0.9 | 66 | 0.06 | 0.79 | 0.81 | 68 | 0.05 | 0.79 | 0.81 |
| | $\rho = 0.10$ | 0 | 94 | 0.05 | 0.79 | 0.81 | 96 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 94 | 0.05 | 0.79 | 0.80 | 96 | 0.05 | 0.80 | 0.80 |
| | | 0.6 | 98 | 0.05 | 0.78 | 0.80 | 100 | 0.04 | 0.79 | 0.80 |
| | | 0.9 | 104 | 0.05 | 0.77 | 0.80 | 106 | 0.05 | 0.77 | 0.80 |
| $\bar{m} = 100$ | $\rho = 0.02$ | 0 | 24 | 0.06 | 0.81 | 0.81 | 26 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 24 | 0.07 | 0.79 | 0.80 | 26 | 0.06 | 0.79 | 0.80 |
| | | 0.6 | 26 | 0.06 | 0.78 | 0.81 | 28 | 0.05 | 0.79 | 0.81 |
| | | 0.9 | 30 | 0.06 | 0.80 | 0.82 | 32 | 0.05 | 0.80 | 0.82 |
| | $\rho = 0.05$ | 0 | 48 | 0.06 | 0.81 | 0.81 | 50 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 48 | 0.06 | 0.81 | 0.81 | 50 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 50 | 0.05 | 0.78 | 0.81 | 52 | 0.05 | 0.78 | 0.81 |
| | | 0.9 | 54 | 0.05 | 0.77 | 0.81 | 56 | 0.05 | 0.77 | 0.81 |
| | $\rho = 0.10$ | 0 | 86 | 0.06 | 0.79 | 0.80 | 88 | 0.05 | 0.80 | 0.80 |
| | | 0.3 | 88 | 0.05 | 0.80 | 0.81 | 90 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 88 | 0.05 | 0.78 | 0.80 | 90 | 0.06 | 0.77 | 0.80 |
| | | 0.9 | 92 | 0.05 | 0.75 | 0.80 | 94 | 0.05 | 0.74 | 0.80 |

**WEB TABLE 25** Required number of clusters $n_{B1}$, empirical type I error $\psi$, empirical power $\phi$, and predicted power $\hat{\phi}$ corresponding to the test for the marginal effect of the cluster-level treatment with and without finite-sample correction. The marginal effect size of the cluster-level treatment is $\delta_X = 0.4$. The unequal cluster sizes approximately follow a uniform distribution. Notation: $\bar{m}$ is the mean cluster size, $\rho$ is the ICC, CV is the coefficient of variation of cluster sizes. The results were based on 5,000 simulations.

| | | | $z$-test | | | | $t$-test | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | CV | $n_{B1}$ | $\psi$ | $\phi$ | $\hat{\phi}$ | $n_{B1}$ | $\psi$ | $\phi$ | $\hat{\phi}$ |
| $\bar{m} = 20$ | $\rho = 0.02$ | 0 | 14 | 0.05 | 0.80 | 0.81 | 16 | 0.04 | 0.81 | 0.81 |
| | | 0.3 | 14 | 0.06 | 0.79 | 0.81 | 16 | 0.04 | 0.79 | 0.80 |
| | | 0.6 | 16 | 0.07 | 0.80 | 0.83 | 18 | 0.05 | 0.82 | 0.83 |
| | | 0.9 | 18 | 0.07 | 0.84 | 0.84 | 20 | 0.05 | 0.85 | 0.84 |
| | $\rho = 0.05$ | 0 | 20 | 0.07 | 0.82 | 0.82 | 22 | 0.04 | 0.82 | 0.82 |
| | | 0.3 | 20 | 0.07 | 0.79 | 0.81 | 22 | 0.05 | 0.80 | 0.81 |
| | | 0.6 | 22 | 0.07 | 0.79 | 0.82 | 24 | 0.05 | 0.80 | 0.82 |
| | | 0.9 | 24 | 0.06 | 0.82 | 0.80 | 28 | 0.06 | 0.85 | 0.83 |
| | $\rho = 0.10$ | 0 | 30 | 0.05 | 0.82 | 0.82 | 32 | 0.05 | 0.82 | 0.82 |
| | | 0.3 | 30 | 0.06 | 0.81 | 0.81 | 32 | 0.05 | 0.82 | 0.81 |
| | | 0.6 | 32 | 0.06 | 0.79 | 0.81 | 34 | 0.06 | 0.81 | 0.81 |
| | | 0.9 | 36 | 0.06 | 0.82 | 0.82 | 38 | 0.05 | 0.82 | 0.82 |
| $\bar{m} = 50$ | $\rho = 0.02$ | 0 | 8 | 0.08 | 0.80 | 0.81 | 12 | 0.04 | 0.88 | 0.88 |
| | | 0.3 | 8 | 0.08 | 0.79 | 0.80 | 12 | 0.04 | 0.87 | 0.87 |
| | | 0.6 | 10 | 0.08 | 0.82 | 0.86 | 12 | 0.05 | 0.83 | 0.85 |
| | | 0.9 | 10 | 0.09 | 0.80 | 0.81 | 14 | 0.06 | 0.87 | 0.87 |
| | $\rho = 0.05$ | 0 | 14 | 0.07 | 0.80 | 0.81 | 16 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 14 | 0.08 | 0.80 | 0.81 | 16 | 0.05 | 0.80 | 0.80 |
| | | 0.6 | 16 | 0.07 | 0.81 | 0.84 | 18 | 0.05 | 0.81 | 0.83 |
| | | 0.9 | 18 | 0.08 | 0.81 | 0.84 | 20 | 0.05 | 0.82 | 0.84 |
| | $\rho = 0.10$ | 0 | 24 | 0.06 | 0.82 | 0.81 | 26 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 24 | 0.07 | 0.80 | 0.81 | 26 | 0.05 | 0.80 | 0.81 |
| | | 0.6 | 26 | 0.06 | 0.80 | 0.83 | 28 | 0.05 | 0.81 | 0.83 |
| | | 0.9 | 26 | 0.06 | 0.76 | 0.80 | 28 | 0.05 | 0.77 | 0.80 |
| $\bar{m} = 100$ | $\rho = 0.02$ | 0 | 6 | 0.11 | 0.79 | 0.81 | 10 | 0.05 | 0.88 | 0.89 |
| | | 0.3 | 6 | 0.11 | 0.79 | 0.80 | 10 | 0.05 | 0.89 | 0.89 |
| | | 0.6 | 8 | 0.09 | 0.84 | 0.88 | 10 | 0.06 | 0.84 | 0.87 |
| | | 0.9 | 8 | 0.10 | 0.81 | 0.84 | 10 | 0.06 | 0.79 | 0.83 |
| | $\rho = 0.05$ | 0 | 12 | 0.08 | 0.82 | 0.81 | 14 | 0.05 | 0.81 | 0.80 |
| | | 0.3 | 12 | 0.08 | 0.80 | 0.81 | 16 | 0.05 | 0.86 | 0.86 |
| | | 0.6 | 14 | 0.08 | 0.83 | 0.85 | 16 | 0.05 | 0.83 | 0.84 |
| | | 0.9 | 14 | 0.09 | 0.76 | 0.83 | 16 | 0.06 | 0.76 | 0.82 |
| | $\rho = 0.10$ | 0 | 22 | 0.06 | 0.81 | 0.81 | 24 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 22 | 0.06 | 0.80 | 0.81 | 24 | 0.06 | 0.80 | 0.81 |
| | | 0.6 | 22 | 0.07 | 0.77 | 0.80 | 26 | 0.05 | 0.81 | 0.83 |
| | | 0.9 | 24 | 0.06 | 0.76 | 0.82 | 26 | 0.05 | 0.77 | 0.82 |

**WEB TABLE 26** Required number of clusters $n_{B2}$, empirical type I error $\psi$, empirical power $\phi$, and predicted power $\hat{\phi}$ corresponding to the test for the marginal effect of the individual-level treatment. The unequal cluster sizes approximately follow a uniform distribution. Notation: $\delta_Z$ is the marginal effect size of the individual-level treatment, $\bar{m}$ is the mean cluster size, $\rho$ is the ICC, CV is the coefficient of variation of cluster sizes. The results were based on 5,000 simulations.

| | | | $\delta_Z = 0.1$ | | | | $\delta_Z = 0.15$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | CV | $n_{A2}$ | $\psi$ | $\phi$ | $\hat{\phi}$ | $n_{A2}$ | $\psi$ | $\phi$ | $\hat{\phi}$ |
| $\bar{m} = 20$ | $\rho = 0.02$ | 0 | 158 | 0.05 | 0.81 | 0.80 | 70 | 0.05 | 0.80 | 0.80 |
| | | 0.3 | 158 | 0.05 | 0.81 | 0.80 | 70 | 0.05 | 0.81 | 0.80 |
| | | 0.6 | 156 | 0.05 | 0.80 | 0.80 | 70 | 0.05 | 0.82 | 0.80 |
| | | 0.9 | 156 | 0.05 | 0.83 | 0.80 | 70 | 0.05 | 0.82 | 0.80 |
| | $\rho = 0.05$ | 0 | 154 | 0.05 | 0.82 | 0.80 | 70 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 154 | 0.05 | 0.81 | 0.80 | 68 | 0.05 | 0.82 | 0.80 |
| | | 0.6 | 154 | 0.05 | 0.81 | 0.80 | 68 | 0.05 | 0.81 | 0.80 |
| | | 0.9 | 154 | 0.04 | 0.83 | 0.80 | 68 | 0.05 | 0.83 | 0.80 |
| | $\rho = 0.10$ | 0 | 148 | 0.05 | 0.81 | 0.80 | 66 | 0.05 | 0.83 | 0.81 |
| | | 0.3 | 148 | 0.05 | 0.83 | 0.80 | 66 | 0.05 | 0.83 | 0.81 |
| | | 0.6 | 146 | 0.05 | 0.80 | 0.80 | 66 | 0.05 | 0.80 | 0.81 |
| | | 0.9 | 146 | 0.04 | 0.84 | 0.80 | 66 | 0.06 | 0.84 | 0.81 |
| $\bar{m} = 50$ | $\rho = 0.02$ | 0 | 64 | 0.04 | 0.81 | 0.81 | 28 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 64 | 0.06 | 0.82 | 0.81 | 28 | 0.05 | 0.80 | 0.81 |
| | | 0.6 | 64 | 0.05 | 0.81 | 0.81 | 28 | 0.05 | 0.81 | 0.81 |
| | | 0.9 | 64 | 0.05 | 0.83 | 0.81 | 28 | 0.06 | 0.82 | 0.81 |
| | $\rho = 0.05$ | 0 | 62 | 0.05 | 0.82 | 0.81 | 28 | 0.05 | 0.83 | 0.82 |
| | | 0.3 | 62 | 0.05 | 0.81 | 0.81 | 28 | 0.05 | 0.81 | 0.82 |
| | | 0.6 | 62 | 0.05 | 0.82 | 0.81 | 28 | 0.05 | 0.82 | 0.82 |
| | | 0.9 | 62 | 0.05 | 0.83 | 0.81 | 28 | 0.06 | 0.83 | 0.82 |
| | $\rho = 0.10$ | 0 | 58 | 0.05 | 0.81 | 0.80 | 26 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 58 | 0.05 | 0.80 | 0.80 | 26 | 0.04 | 0.81 | 0.81 |
| | | 0.6 | 58 | 0.05 | 0.82 | 0.80 | 26 | 0.05 | 0.81 | 0.81 |
| | | 0.9 | 58 | 0.04 | 0.83 | 0.80 | 26 | 0.05 | 0.81 | 0.81 |
| $\bar{m} = 100$ | $\rho = 0.02$ | 0 | 32 | 0.05 | 0.81 | 0.81 | 14 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 32 | 0.05 | 0.82 | 0.81 | 14 | 0.05 | 0.80 | 0.81 |
| | | 0.6 | 32 | 0.05 | 0.80 | 0.81 | 14 | 0.05 | 0.79 | 0.81 |
| | | 0.9 | 32 | 0.05 | 0.83 | 0.81 | 14 | 0.05 | 0.78 | 0.81 |
| | $\rho = 0.05$ | 0 | 32 | 0.05 | 0.83 | 0.82 | 14 | 0.05 | 0.83 | 0.82 |
| | | 0.3 | 32 | 0.05 | 0.83 | 0.82 | 14 | 0.05 | 0.81 | 0.82 |
| | | 0.6 | 32 | 0.05 | 0.81 | 0.82 | 14 | 0.05 | 0.80 | 0.82 |
| | | 0.9 | 32 | 0.05 | 0.84 | 0.82 | 14 | 0.05 | 0.80 | 0.82 |
| | $\rho = 0.10$ | 0 | 30 | 0.05 | 0.82 | 0.82 | 14 | 0.05 | 0.85 | 0.84 |
| | | 0.3 | 30 | 0.04 | 0.82 | 0.82 | 14 | 0.05 | 0.83 | 0.84 |
| | | 0.6 | 30 | 0.05 | 0.82 | 0.82 | 14 | 0.05 | 0.82 | 0.84 |
| | | 0.9 | 30 | 0.05 | 0.83 | 0.82 | 14 | 0.05 | 0.82 | 0.84 |

**WEB TABLE 27** Required number of clusters $n_C$, empirical type I error $\psi$, empirical power $\phi$, and predicted power $\hat{\phi}$ corresponding to the interaction test. The unequal cluster sizes approximately follow a uniform distribution. Notation: $\delta_4$ is the interaction effect size, $\bar{m}$ is the mean cluster size, $\rho$ is the ICC, CV is the coefficient of variation of cluster sizes. The results were based on 5,000 simulations.

| | | | $\delta_4 = 0.2$ | | | | $\delta_4 = 0.3$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | CV | $n_C$ | $\psi$ | $\phi$ | $\hat{\phi}$ | $n_C$ | $\psi$ | $\phi$ | $\hat{\phi}$ |
| $\bar{m} = 20$ | $\rho = 0.02$ | 0 | 158 | 0.05 | 0.81 | 0.80 | 70 | 0.05 | 0.81 | 0.80 |
| | | 0.3 | 158 | 0.05 | 0.82 | 0.80 | 70 | 0.05 | 0.81 | 0.80 |
| | | 0.6 | 156 | 0.05 | 0.81 | 0.80 | 70 | 0.05 | 0.81 | 0.80 |
| | | 0.9 | 156 | 0.05 | 0.84 | 0.80 | 70 | 0.05 | 0.84 | 0.80 |
| | $\rho = 0.05$ | 0 | 154 | 0.05 | 0.81 | 0.80 | 70 | 0.05 | 0.82 | 0.81 |
| | | 0.3 | 154 | 0.05 | 0.80 | 0.80 | 68 | 0.05 | 0.81 | 0.80 |
| | | 0.6 | 154 | 0.05 | 0.81 | 0.80 | 68 | 0.05 | 0.80 | 0.80 |
| | | 0.9 | 154 | 0.05 | 0.83 | 0.80 | 68 | 0.05 | 0.82 | 0.80 |
| | $\rho = 0.10$ | 0 | 148 | 0.05 | 0.83 | 0.80 | 66 | 0.05 | 0.83 | 0.81 |
| | | 0.3 | 148 | 0.05 | 0.81 | 0.80 | 66 | 0.05 | 0.82 | 0.81 |
| | | 0.6 | 146 | 0.05 | 0.80 | 0.80 | 66 | 0.06 | 0.81 | 0.81 |
| | | 0.9 | 146 | 0.05 | 0.84 | 0.80 | 66 | 0.05 | 0.83 | 0.81 |
| $\bar{m} = 50$ | $\rho = 0.02$ | 0 | 64 | 0.05 | 0.81 | 0.81 | 28 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 64 | 0.05 | 0.83 | 0.81 | 28 | 0.05 | 0.80 | 0.81 |
| | | 0.6 | 64 | 0.05 | 0.81 | 0.81 | 28 | 0.05 | 0.81 | 0.81 |
| | | 0.9 | 64 | 0.05 | 0.83 | 0.81 | 28 | 0.05 | 0.81 | 0.81 |
| | $\rho = 0.05$ | 0 | 62 | 0.05 | 0.82 | 0.81 | 28 | 0.05 | 0.82 | 0.82 |
| | | 0.3 | 62 | 0.05 | 0.83 | 0.81 | 28 | 0.05 | 0.81 | 0.82 |
| | | 0.6 | 62 | 0.05 | 0.81 | 0.81 | 28 | 0.05 | 0.82 | 0.82 |
| | | 0.9 | 62 | 0.05 | 0.83 | 0.81 | 28 | 0.05 | 0.81 | 0.82 |
| | $\rho = 0.10$ | 0 | 58 | 0.05 | 0.81 | 0.80 | 26 | 0.05 | 0.81 | 0.81 |
| | | 0.3 | 58 | 0.05 | 0.80 | 0.80 | 26 | 0.05 | 0.81 | 0.81 |
| | | 0.6 | 58 | 0.05 | 0.81 | 0.80 | 26 | 0.05 | 0.81 | 0.81 |
| | | 0.9 | 58 | 0.05 | 0.83 | 0.80 | 26 | 0.04 | 0.82 | 0.81 |
| $\bar{m} = 100$ | $\rho = 0.02$ | 0 | 32 | 0.05 | 0.81 | 0.81 | 14 | 0.05 | 0.80 | 0.81 |
| | | 0.3 | 32 | 0.05 | 0.81 | 0.81 | 14 | 0.05 | 0.80 | 0.81 |
| | | 0.6 | 32 | 0.05 | 0.81 | 0.81 | 14 | 0.05 | 0.80 | 0.81 |
| | | 0.9 | 32 | 0.05 | 0.81 | 0.81 | 14 | 0.05 | 0.80 | 0.81 |
| | $\rho = 0.05$ | 0 | 32 | 0.05 | 0.83 | 0.82 | 14 | 0.05 | 0.81 | 0.82 |
| | | 0.3 | 32 | 0.05 | 0.82 | 0.82 | 14 | 0.05 | 0.81 | 0.82 |
| | | 0.6 | 32 | 0.05 | 0.82 | 0.82 | 14 | 0.05 | 0.81 | 0.82 |
| | | 0.9 | 32 | 0.05 | 0.82 | 0.82 | 14 | 0.05 | 0.81 | 0.82 |
| | $\rho = 0.10$ | 0 | 30 | 0.05 | 0.82 | 0.82 | 14 | 0.05 | 0.83 | 0.84 |
| | | 0.3 | 30 | 0.05 | 0.82 | 0.82 | 14 | 0.05 | 0.83 | 0.84 |
| | | 0.6 | 30 | 0.05 | 0.81 | 0.82 | 14 | 0.05 | 0.83 | 0.84 |
| | | 0.9 | 30 | 0.05 | 0.83 | 0.82 | 14 | 0.05 | 0.83 | 0.84 |